

JAIR JOSÉ FERRONATO

**PM4SOS: Um *Framework* para suporte
à tomada de decisão operacional**

Tese apresentada ao Programa de Pós-Graduação
em Informática da Pontifícia Universidade Católica
do Paraná – PUC-PR como requisito parcial para
obtenção de título de Doutor em Informática.

Curitiba

2022

JAIR JOSÉ FERRONATO

**PM4SOS: um *Framework* para suporte
à tomada de decisão operacional**

Tese apresentada ao Programa de Pós-Graduação em Informática da Pontifícia Universidade Católica do Paraná como requisito parcial para obtenção de título de Doutor em Informática.

Área de concentração: Ciência da Computação

Orientador: Prof. Dr. Edson Emílio Scalabrin

Curitiba

2022

FICHA CATALOGRÁFICA

Ferronato, Jair José

PM4SOS: Um Framework para suporte à tomada de decisão operacional

J.J. Ferronato. -- Curitiba, 2022.

Número de páginas p. 157

Tese (Doutorado) – Pontifícia Universidade Católica do Paraná. Curitiba. Programa de Pós-Graduação em Informática.

1. Tomada de decisão 2. Mineração de processos 3. Simulação
4. Agendamento 5. Raciocínio Baseado em Casos
Pontifícia Universidade Católica do Paraná. Programa de Pós-Graduação em Informática.

DEDICATÓRIAS

Dedico esta tese à minha família.

AGRADECIMENTOS

Primeiramente, a Deus, pela oportunidade em evoluir pessoal e profissionalmente, por me dar forças na caminhada.

À minha família, em especial à minha esposa Sibeli, meus filhos Angelo Luigi e Maria Chiara, ao apoio incondicional e apoio na busca constante dos nossos sonhos.

Especialmente ao meu orientador, prof. Edson Emílio Scalabrin, que não mediu esforços para proporcionar um ambiente motivador e inspirador, pelo seu exemplo como pesquisador e por conduzir a orientação para o conhecimento científico no campo computacional e de pesquisa. Os seus ensinamentos são diferenciais nesta etapa da vida.

A todos os professores do PUCPR, dos diversos programas que contribuíram para a minha formação acadêmica durante o desenvolvimento do doutorado.

Ao Instituto Federal de Santa Catarina (IFSC), pela concessão ao desenvolvimento de minha pesquisa de forma integral.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela concessão de bolsa de estudos durante a realização do doutorado.

Start by doing what is necessary, then what is possible, and suddenly you are doing the impossible.

RESUMO

A gestão de processo de negócio é uma área do conhecimento em evolução. Novas técnicas e métodos são criados para gerar informações de apoio à tomada de decisão. Nesse sentido, a mineração de processos constitui uma disciplina que permite obter informações relevantes sobre os comportamentos de um processo de negócio, tomando como ponto de partida *logs* de eventos prontamente disponíveis em sistemas de informação. Tais comportamentos refletem situações do passado, referentes ao processo de negócio em questão. Porém, quando se precisa de um “olhar à frente”, a simulação computacional fornece suporte à tomada de decisão. A simulação computacional imita situações por meio da replicação de cenários/comportamentos reais em um ambiente virtual. E, na medida em que indicadores de melhoria podem ser visualizados e avaliados sem o ônus de uma operação real, a simulação, com baixo custo operacional, torna-se de fato uma vantagem competitiva. Porém, a construção de um modelo de simulação que represente a realidade pode ser onerosa em termos de esforço humano. Adicionalmente, há limitações quanto à integração completa de ferramentas, à criação automática do modelo de simulação e ao suporte à tomada de decisão operacional. No entanto, para tomar decisões assertivas e que agreguem valor ao negócio, é preciso dotar o sistema de inteligência, que, a partir de problemas e soluções passadas, por exemplo, possam reutilizar ou adaptar soluções para novos problemas operacionais. Neste trabalho busca-se preencher tais lacunas com a criação do framework PM4SOS, que, com base em uma fonte de dados consistente e completa, permite criar um modelo de simulação de forma automatizada e com baixo esforço cognitivo. Demonstra-se a atividade de simulação e resolução de problema, baseada em CBR, aplicada ao estudo de caso de um centro cirúrgico com foco no ambiente de filas, com base na criação de um protótipo computacional, incluindo métodos multicritério e otimização multiobjetivo. Nesse sentido, este trabalho utiliza *Design Science Research*. A aplicação do framework PM4SOS está fortemente estruturada na complementariedade da relação entre mineração de processos, simulação computacional, CBR e tomada de decisão.

Palavras-chaves: mineração de processos, simulação, otimização, agendamento, suporte à decisão.

ABSTRACT

Business process management is an evolving area of knowledge. New techniques and methods are created to generate information to support decision making. In this line, process mining is a discipline that allows obtaining relevant information about the behavior of a business process, taking as a starting point event logs readily available in information systems. Such behaviors reflect past situations regarding the business process in question. However, when a “look ahead” is needed, computer simulation provides support for decision making. Computer simulation mimics situations by replicating real scenarios/behaviors in a virtual environment. To the extent that improvement indicators can be visualized and evaluated without the burden of an operation, the simulation, with low operational cost, becomes a competitive advantage. However, to create a simulation model that represents reality can be costly in terms of human effort. Additionally, there are limitations regarding the complete integration of tools, the automatic creation of the simulation model and the support for operational decision making. However, to make assertive decisions that add value to the business, it is necessary to provide the intelligence system, which, based on past problems and solutions, for example, can reuse or adapt solutions to new operational problems. This research aims to fill these gaps with the creation of the PM4SOS framework, which from a consistent and complete data source, allows the creation of a simulation model in an automated way and with low cognitive effort. The activity of simulation and problem solving based on CBR, applied to the case study of a surgical center focusing on the queue environment, are demonstrated from the creation of a computational prototype, including multicriteria methods and multi objective optimization. This research uses Design Science Research. The application of the PM4SOS framework is strongly structured in the complementarity of the relationship between process mining, computer simulation, CBR and decision making.

Keywords: *Process Mining, Simulation, Optimization, Scheduling, Decision-Making.*

LISTA DE FIGURAS

Figura 1 – metodologia da pesquisa – adaptada de (PEFFERS et al., 2007)	22
Figura 2 – PM4SOS e seus componentes em IDEF0 – o autor 2022	23
Figura 3 – ciclo de vida BPM – adaptada de (DUMAS et al., 2013).....	26
Figura 4 – sistema de filas – adaptada de (VAN DER AALST et al., 2010)	30
Figura 5 – mineração de processos – adaptada de (VAN DER AALST, 2016).....	32
Figura 6 – qualidade do log – adaptada de (RAMOS-GUTIÉRREZ et al., 2021).	34
Figura 7 – árvore de processos – adaptada de (VAN DER AALST et al., 2019).....	36
Figura 8 – estrutura da simulação e otimização – adaptada de (FU, 2002).....	42
Figura 9 – otimização unidimensional – (KOCHENDERFER; WHEELER, 2019)	44
Figura 10 – obtenção das soluções de Pareto – (DEB et al., 2002).....	45
Figura 11 – ciclo de vida do CBR – adaptada de (RIESBECK; SCHANK, 1989).....	52
Figura 12 – diagrama das técnicas de busca – (LINDEN, 2008).....	55
Figura 13 – fluxo da seleção de estudos para a revisão – o autor 2022.	72
Figura 14 – técnicas de simulação versus suporte operacional – o autor 2022	84
Figura 15 – DSRM – modelo de processo – adaptada de (PEFFERS et al., 2007) ..	92
Figura 16 – <i>framework</i> PM4SOS – o autor 2022.....	97
Figura 17 – modelo de processo de consultas médicas – o autor 2022.....	100
Figura 18 – modelo de processo de um centro cirúrgico - o autor 2022	101
Figura 19 – fase 1: fonte de dados – o autor 2022.....	103
Figura 20 – fase 2: mineração de processos – o autor 2022	105
Figura 21 – modelo de processo em DFG – o autor 2022.	107
Figura 22 – modelo de processo em árvore – o autor 2022.....	107
Figura 23 – distribuição dos cases diários e semanais – o autor 2022.	111
Figura 24 – fase 3: simulação computacional – o autor 2022	114
Figura 25 – fase 4: otimização de processos – o autor 2022	118
Figura 26 – etapa 4-1: raciocínio baseado em casos – o autor 2022.....	128
Figura 27 – fase 5: visualização de dados – o autor 2022	128
Figura 28 – análise da qualidade dos logs – o autor 2022	131
Figura 29 – estatísticas e modelo organizacional – o autor 2022.....	132
Figura 30 – modelo de simulação – o autor 2022.	133
Figura 31 – validação do modelo de simulação – o autor 2022.	134
Figura 32 – relatório dos indicadores da simulação – o autor 2022.	135

Figura 33 – recuperação de casos por similaridade – o autor 2022.....	136
Figura 34 – agendamento completo das salas de cirurgia – o autor 2022.	137
Figura 35 - evolução do tempo de processamento em relação o número de salas e cirurgias.....	138
Figura 36 - desempenho do Framework PM4SOS.....	140
Figura 37 – fluxo de atividades em IDEF0 do PM4SOS – o autor 2022.....	142

LISTA DE TABELAS

Tabela 1 - exemplo de log de eventos na área da saúde.....	16
Tabela 2 – tabela resumo da fundamentação teórica	64
Tabela 3 – resultados obtidos para cada fonte de dados.....	70
Tabela 4 – estudos selecionados em ordem cronológica.....	74
Tabela 5 – técnicas de simulação – trabalhos e respectivos autores.....	75
Tabela 6 – atividades de apoio operacional – trabalhos e respectivos autores.....	75
Tabela 7 – formalismos de modelagem, trabalhos e respectivos autores	76
Tabela 8 – dados da mineração de processo.....	77
Tabela 9 – dados do estado corrente.....	78
Tabela 10 – contribuições informadas pelos autores em suas publicações	81
Tabela 11 – limitações informadas pelos autores em suas publicações.....	82
Tabela 12 – oportunidades informadas pelos autores em suas publicações	83
Tabela 13 – exemplo de log de eventos.....	102
Tabela 14 – matriz de agendamento cirúrgico.	119
Tabela 15 – exemplo de slots necessários para agendamento.....	120
Tabela 16 – representação da base de casos.....	120
Tabela 17 – segmentos da demonstração.	129
Tabela 18 - taxa de ocupação por sala de cirurgia.....	139

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Contextualização	14
1.2	Problema de pesquisa	17
1.3	Objetivos.....	18
1.4	Justificativa e originalidade.....	19
1.5	Contribuições científicas e tecnológicas	21
1.6	Abordagem metodológica.....	22
1.7	Estrutura do documento de tese.....	24
2	FUNDAMENTAÇÃO TEÓRICA	25
2.1	Gestão de Processos de Negócios	25
2.2	Problema do planejamento e agendamento cirúrgico	26
2.3	Teoria das filas	29
2.4	Mineração de processos.....	31
2.5	Simulação computacional.....	38
2.6	Otimização de processos	41
2.6.1	Otimização multiobjetivo	44
2.6.2	Otimização com base em restrições	46
2.6.3	Otimização de múltiplos critérios	47
2.6.4	Raciocínio Baseado em Casos (CBR).....	50
2.7	Relacionamento dos conhecimentos.....	61
2.8	Resumo do Capítulo	62
3	REVISÃO SISTEMÁTICA DE LITERATURA.....	65
3.1	Introdução.....	66
3.2	Objetivos da pesquisa	69
3.3	Protocolo de pesquisa	70
3.4	Descobertas e perspectivas: Suporte à tomada de decisão operacional	74
3.4.1	Modelo de simulação	74
3.4.2	Dados da mineração de processos.....	76
3.4.3	Dados do estado corrente.....	77
3.4.4	Gerenciamento de recursos.....	79
3.4.5	Suporte à tomada de decisão operacional.....	80
3.4.6	Contribuições, lacunas e oportunidades	81
3.5	Discussão	84
3.6	Conclusão.....	88
3.7	Resumo do capítulo.....	90

4	ESTRUTURAÇÃO DA PESQUISA	91
4.1	Caracterização da pesquisa	91
4.2	Design Science Research Methodology	91
4.3	Identificação do problema e motivação	92
4.4	Definição dos objetivos para a solução	93
4.5	Projeto e desenvolvimento	94
4.6	Demonstração	95
4.7	Avaliação	95
4.8	Comunicação.....	96
5	DESENVOLVIMENTO DA PESQUISA	97
5.1	<i>Framework</i> PM4SOS.....	97
5.2	Projeto e desenvolvimento	100
5.2.1	Domínio do problema de pesquisa	100
5.2.2	Fase 1 – preparar dados.....	103
5.2.3	Fase 2 - mineração de processos.....	105
5.2.4	Fase 3 - simulação computacional.....	114
5.2.5	Fase 4 - otimização de processos	118
5.2.6	Fase 5 - visualização de dados.....	128
5.3	Resultados.....	129
5.4	Avaliação	139
5.5	Resumo do capítulo.....	142
6	CONCLUSÕES	144
7	CONSIDERAÇÕES FINAIS	147
7.1	Relevância do estudo	147
7.2	Contribuições da pesquisa	148
7.3	Limitações da pesquisa	149
7.4	Trabalhos futuros.....	149
	REFERÊNCIAS	150

1 INTRODUÇÃO

“Pesquisar é acordar para o mundo.”
(*Marcelo Lamy*).

1.1 Contextualização

Na vida diária, sempre que as pessoas tentam encontrar uma solução para um problema, há a necessidade de decisão. Essa necessidade está presente nas escolhas pessoais relativas ao que e como comer; ao que, como e com quem falar, dentre outras. Em nível empresarial, as decisões são tomadas com o intuito de reduzir custos e desperdícios, aumentar a produtividade ou obter maior eficiência operacional para aumentar os lucros.

No entanto, qual metodologia deve ser utilizada em relação à tomada de decisão? Qual é o esforço humano cognitivo necessário para decidir e concretizar as escolhas? As decisões precisam ser tomadas para solucionar problemas. O mundo corporativo é repleto de desafios, e cada desafio pode ser desdobrado e requer outras decisões paralelas, as quais influenciam os resultados da decisão principal. É necessário, nesse sentido, questionar: em quanto tempo, é possível operacionalizar um plano de ação? Qual é o esforço gerencial necessário para estruturar e executar ações emergenciais?

As decisões são tomadas por todos os níveis hierárquicos das empresas. De forma geral, utiliza-se a estrutura organizacional com os níveis: a) estratégico: decide o que fazer; b) tático ou gerencial: decide como fazer e c) nível operacional: executa as ações planejadas. Esses níveis devem funcionar de forma integrada e harmônica para alcançar os objetivos estabelecidos.

Em nível operacional, o papel de um gestor é estruturar e implementar planos de ação, identificar gargalos, fazer controle de qualidade, garantir as entregas no prazo estipulado e alocar recursos conforme a necessidade. Esse profissional é responsável, também, por otimizar os processos para mitigar problemas e prever falhas graves antes que elas aconteçam. A gestão operacional constitui a interface entre o problema ou necessidade e o serviço prestado.

Porém, deve-se refletir como as decisões são tomadas. Elas encerram base emocional ou valores pessoais? É uma decisão de cunho racional ou subjetiva?

Borges (2019) realizou uma análise referente às crenças dos decisores, que influenciam o uso de informação para o processo decisório em contextos organizacionais, utilizando de métodos e técnicas das ciências cognitivas e psicologia cognitiva. Essa pesquisa ressalta que as pessoas são dotadas de valores, sentimentos e crenças que determinam seu comportamento. No contexto organizacional, os gestores usam esses mecanismos para apoiar situações de decisão, omitindo a correlação entre informação e decisão (BORGES, 2019).

Quando decisões são tomadas meramente com base em valores, crenças e sentimentos pessoais dos decisores, não há padronização e as soluções contêm o viés do tomador de decisão. Essa perspectiva pode ser superada com processos de tomada de decisão baseados na racionalidade.

Os aspectos cognitivos envolvidos no processo decisório racional utilizam os seguintes componentes propostos por (BAZERMAN; MOORE, 2012): a) definição do problema; b) identificação das opções existentes; c) ponderação e peso das opções; d) geração de alternativas; e) ponderação das opções com as alternativas; e f) computação da decisão ótima. Deverá ser escolhida a solução com maior valor. Esse modelo assume que os passos para a tomada de decisão são executados de forma totalmente racional.

Por outro lado, observa-se que, em relação à contemporaneidade, há inúmeros recursos de armazenamento de dados no formato digital para conteúdo como textos, vídeos, imagens, registros e outros subsídios tecnológicos. As pessoas geram e consomem informações todos os dias, interagindo com diversas aplicações de sistemas de informação. Diante disso, é importante administrar as informações e utilizá-las a favor dos processos de negócio. Com informações prontamente disponíveis em sistemas de informação, é possível tomar decisões mais assertivas em vários setores empresariais.

Uma das tecnologias que lida com conjuntos volumosos de dados é “big data”. Esse termo é usado para representar um conjunto de dados massivos com estruturas variadas e complexas (SAGIROGLU; SINANC, 2013). O processo de pesquisa em *big data* permite revelar padrões ocultos e correlações (SAGIROGLU; SINANC, 2013). Esse cenário motiva o questionamento sobre o uso desses dados para apoiar a tomada de decisão organizacional.

O fator motivador desta pesquisa é a resolução de problemas com base em dados disponíveis em sistemas de informação. O tratamento e análise de dados

armazenados em sistemas de informação permite a automação de procedimentos manuais. A mineração de processos consiste em analisar dados obtidos a partir de sistemas de informação. A Tabela 1 mostra um exemplo de log de eventos de um sistema de consultas médicas. As colunas representam os dados dos pacientes, as ações ou atividades, os recursos que executam as ações, o espaço de tempo da realização da ação e dados extras.

Tabela 1 - exemplo de log de eventos na área da saúde

Para quem?	O quê?	Por quem?	Quando?	Dados extras
Paciente	Ação	Recurso	Data e Hora	Anotações
Bianca	Primeira consulta	Dr. Anna	2021-01-05 11:15	
Pedro	Primeira consulta	Dr. Anna	2021-01-05 11:30	
Bianca	Teste de sangue	Laboratório	2021-01-09 15:30	2 tubos
Bianca	Teste físico	Dr. Bem	2021-01-09 16:30	Teste OK
Bianca	Raio-X	Equipe 1	2021-01-11 9:30	3 imagens
Pedro	Raio-X	Equipe 1	2021-01-11 10:30	2 imagens
Bianca	Segunda consulta	Dr. Anna	2021-01-12 12:45	Cirurgia necessária
Bianca	Cirurgia	Dr. Charlie	2021-01-24 13:00	Agendada
Jamile	Primeira consulta	Dr. Ben	2021-01-24 13:40	
Bianca	Consulta Final	Dr. Anna	2021-01-27 10:20	
Pedro	Teste físico	Dr. Anna	2021-01-30 8:30	

Fonte: o autor 2022

A mineração de processos permite estabelecer uma sequência cronológica dos fatos. A cada linha registra um evento e o conjunto de eventos denotam o que aconteceu em um processo. O log de eventos é dado como entrada para as ferramentas de análise, que implementam técnicas de mineração de processos, tais como, descoberta do modelo de processos, verificação de conformidade e aprimoramento do processo. O conjunto de dados para o experimento desta pesquisa foram obtidos de um centro cirúrgico. Os dados fazem parte do projeto de pesquisa intitulado: “HUC¹ 4.0 - otimização de recursos do centro cirúrgico de um hospital universitário” e aprovado em comitê científico da instituição hospitalar.

A integração desses sistemas permite usufruir dos benefícios tecnológicos, proporcionados aos tomadores de decisões, como a redução do tempo de planejamento e esforço necessário ao nível gerencial para organizar o setor operacional das organizações.

¹ Projeto da Plataforma Brasil (CAAE: 61432722.4.0000.0020)

1.2 Problema de pesquisa

Mesmo tendo acesso a recursos tecnológicos avançados, incluindo os dados presentes em sistemas de informação, o uso prático desses dados pode não ocorrer de forma eficiente e os problemas podem persistir. Os tomadores de decisões podem, ainda, empregar alto esforço cognitivo e tempo para organizar controles manuais, esquemas gráficos, rascunhos e, dessa forma, a decisão é tomada de forma intuitiva e com base em valores pessoais.

Outra situação possível e desejada é a tomada de decisão racional com base em dados presentes em sistemas de informação. Essa perspectiva merece atenção, pois utiliza dados disponíveis gerados pela própria organização. Esses dados podem representar processos produtivos, dados funcionais, dados de vendas ou campanhas de *marketing*, dados hospitalares, dentre outros.

Os dados históricos são elementares para a análise de situações passadas para orientar decisões futuras. Os dados de um sistema em execução também são importantes para comparar os dados históricos com os dados atuais e, dessa forma, detectar desvios na execução de um processo. Outra perspectiva é um “olhar à frente” para prever como as ações executadas no passado e aquelas identificadas em sistemas em operação interferirão, negativa ou positivamente, no futuro.

Se, de um lado, há problemas a serem solucionados, processos a serem melhorados e decisões a serem tomadas por gestores; de outro lado, há dados prontamente disponíveis em sistemas de informação. Assim, lança-se o seguinte questionamento para a pesquisa apresentada nesta tese: como os dados presentes em sistemas de informação podem dar suporte para o processo decisório de gestão operacional com baixo esforço cognitivo e tempo reduzido?

Na literatura, não estão claras as limitações para colocar em prática a operacionalização e a estruturação de aplicações em relação à simulação e otimização com a ajuda da mineração de processos. Todavia, um requisito está claramente definido, todo o log de eventos extraído de um sistema de informação deve ser consistente e refletir o comportamento do processo real em análise.

O contexto do problema desta tese, em termos do domínio de aplicação, refere-se à busca de solução para o problema clássico de planejamento e agendamento cirúrgico. É um cenário em que vários atores participam, como gestores de hospitais, cirurgiões e pacientes. Um dos focos da pesquisa consiste em

avaliar a utilização de recursos, como salas de cirurgias com base em indicadores de filas de atendimento. O objeto final é a obtenção do agendamento de cirurgias demandando a menor quantidade de informação possível ao ator do processo.

No contexto, a fonte principal de dados deve ser um log de eventos bem-formado e no formato XES². O log deve conter dados sobre instâncias ou execuções das atividades ligadas às cirurgias, assim como os recursos alocados para as suas execuções. Ainda, na fase de preparação do log, devem-se filtrar as ocorrências de variantes de processo pouco frequentes. Atendidos esses requisitos de qualidade da informação de partida, advoga-se que a interação do usuário deve-se restringir à validação do modelo de processo descoberto e estabelecimento dos limites dos indicadores de gestão de filas. Outra interação do analista/usuário do processo com o sistema se refere à definição de preferências.

Salienta-se que a abordagem, a respeito do *framework* denominado PM4SOS, é experimental e a resolução de cada desafio, face a sua aplicação, deve permitir o refinamento sucessivo do PM4SOS na direção da sua evolução e correto funcionamento.

1.3 Objetivos

O objetivo geral desta pesquisa consiste no projeto de um *framework* que coloca em prática um conjunto de procedimentos de apoio ao suporte operacional com baixo esforço cognitivo e tempo reduzido. Os pilares da proposta envolvem a mineração de processos com base em logs de eventos, a simulação computacional e a otimização com o reuso de experiências.

Os objetivos específicos para a consecução do objetivo geral são:

- O1.** definir um ambiente computacional integrando a mineração de processos, simulação computacional e otimização para suporte à tomada de decisão com baixo esforço cognitivo e tempo reduzido;
- O2.** implementar método de otimização de agendamento baseado no reuso de experiências; e

² <https://www.xes-standard.org/>

O3. aplicar o método, com o suporte do ambiente integrado, para resolver o problema de planejamento e agendamento cirúrgico.

Nesse contexto de integração, exibem-se os resultados obtidos com as execuções dos mecanismos do PM4SOS em uma interface visual. Essa interface deve facilitar o usuário visualizar *insights* e demais dados da aplicação. Essa última se refere ao agendamento de recursos de um centro cirúrgico.

1.4 Justificativa e originalidade

Esta pesquisa justifica-se pela proposta e apresentação de um *framework* para suporte à tomada de decisão operacional. O uso combinado da mineração de processos e simulação computacional para otimizar processos deve resultar em um ambiente tecnológico, possível de ser executado em tempo reduzido e com baixo esforço humano.

A pesquisa bibliográfica permitiu investigar outros trabalhos sobre suporte à decisão e identificar lacunas e oportunidades. O avanço no campo teórico oferece subsídios para a resolução de problemas com suporte de dados disponíveis em sistemas de informação. O projeto, desenvolvimento e avaliação do PM4SOS deve mostrar que a integração de dados, mineração de processos, simulação computacional, otimização de planejamento e agendamento cirúrgico são viáveis e permitem redução de esforços humanos, de custo e tempo.

Em geral, para que o usuário possa obter *insights* e propor ações de melhoria a nível operacional, ele necessita de informações assertivas; a assertividade se eleva em função do uso de dados reais para compor as estimativas e extrair informações dos processos de negócio e estimativas para a simulação e otimização. A extração de parâmetros de logs de eventos para criar o modelo de simulação, de forma automatizada e usada em conjunto com a otimização de processo, é a base para o desenvolvimento desta pesquisa.

Ainda na linha de automatização, dada a crescente demanda para a tomada de decisão no curto prazo, a nossa ênfase no suporte operacional está concentrada na perspectiva de análise dos casos em execução, como a detecção de desvios para acionar gatilho o mecanismo de otimização. O módulo de otimização proposto

inclui um método de decisão multicritério para mitigar critérios de decisão conflitantes.

O diferencial desta pesquisa é utilizar dados provenientes da mineração de processos para criar um modelo de simulação de forma automatizada. Ao informar um log de eventos como a fonte de dados, a extração de parâmetros/estimativas para o modelo de simulação não exige esforços humanos adicionais. O suporte à tomada de decisão operacional inclui a perspectiva de recomendação de ações. Esta pesquisa propõe a recomendação de agendamento cirúrgico e inclui suporte do método multicritério. A simulação permite evidenciar o tempo de atendimento, tempo em fila e uso de recursos, no contexto da aplicação construída, a partir de um log de eventos de um centro cirúrgico.

A essência do suporte operacional é a utilização de eventos incompletos, os quais representam a oportunidade de interagir com as instâncias em execução antes de seu desfecho. A partir desta perspectiva têm-se as informações de curto prazo, as quais são inseridas como estado inicial da simulação. Os dados de curto prazo podem ser obtidos, se disponíveis, a partir de sistemas de workflow ou log de eventos ou quando não disponíveis estes dados podem ser gerados artificialmente.

Deve-se notar que as principais lacunas existentes na literatura decorrem, principalmente, da falta de integração efetiva dos mecanismos de mineração de processos, ferramentas de simulação, visualização e análise (ROZINAT et al., 2008b). A mineração de processos frequentemente é usada de forma *offline* e raramente é usada na visão *online* para o suporte operacional (DER AALST; ARCIERI; ADRIANSYAH, 2012).

Estes dois aspectos fundamentam a abordagem desta proposta com ganhos esperados sobre a redução dos esforços na criação de um ambiente de gestão operacional proativa e na geração de modelo de simulação de forma automatizada. Também são evidenciados ganhos com o mecanismo de otimização de planejamento e agendamento cirúrgico, evitando erros, parâmetros inconsistentes e decisões equivocadas.

Os resultados desta pesquisa devem contribuir com *insights* e tomada de decisões mais assertivas face a problemas na gestão de processos de negócio.

1.5 Contribuições científicas e tecnológicas

A principal contribuição desta pesquisa é o PM4SOS. Este último pode ser visto como um ambiente operacional para execução de suporte à tomada de decisão operacional. O PM4SOS recebe como *input* um log de eventos e as demais análises são realizadas de forma simplificada *vis-à-vis* a interação com o usuário. Dentre os recursos estão: o gerenciamento dos recursos usados nos processos e a gestão de tempo de espera e de execução das atividades.

A identificação de novas técnicas para apoio à tomada de decisão com o aporte da mineração de processos e da simulação permite avanço no conhecimento teórico e prático referente ao tema de pesquisa.

O presente trabalho apresenta os seguintes resultados **científicos**:

- a) Criação de um método para modelagem e simulação para servir de referência e aplicação em diversos domínios de conhecimento.
- b) Uma discussão sobre o uso combinado da mineração de processos, com otimização e simulação de indicadores para apoiar a tomada de decisão.
- c) Uso da mineração de processos para descoberta do modelo de processos a partir de um log de eventos e com obtenção de parâmetros úteis para a simulação e otimização de processos.
- d) Uma revisão sistemática de literatura sobre mineração de processos e simulação para atualização do estado da arte deste tema de pesquisa.
- e) Uso da abordagem de raciocínio, baseada em casos para reutilizar experiências passadas face a resolução de outros problemas.

O presente trabalho inclui os seguintes resultados **tecnológicos**:

- a) Desenvolvimento de uma ferramenta geradora de log de eventos sintéticos Sim2Log³.
- b) Implementação de um protótipo de software para aplicar os conceitos do PM4SOS proposto e discutido nesta tese.

³ <https://github.com/ferronato/Sim2Log>

1.6 Abordagem metodológica

A abordagem metodológica utiliza como base a estratégia DSRM (Design Science Research Methodology). Para (JACOBSON, 1999), o *design* significa comunicação por palavras, imagens, tabelas, gráficos, mapas e desenhos, por meios convencionais ou digitais. O DSRM caracteriza-se pela organização sistemática para projetar artefatos e facilitar o processo de aquisição de informação e conhecimento.

O método DSRM consiste nas seguintes etapas, conforme (PEFFERS et al., 2007): 1: identificação do problema e sua motivação; 2: definição dos objetivos para a solução; 3: projeto e desenvolvimento; 4: demonstração; 5: avaliação e 6: comunicação. As etapas metodológicas são detalhadas no Capítulo 3. A Figura 1 mostra as etapas do método DSRM utilizadas no desenvolvimento do PM4SOS.

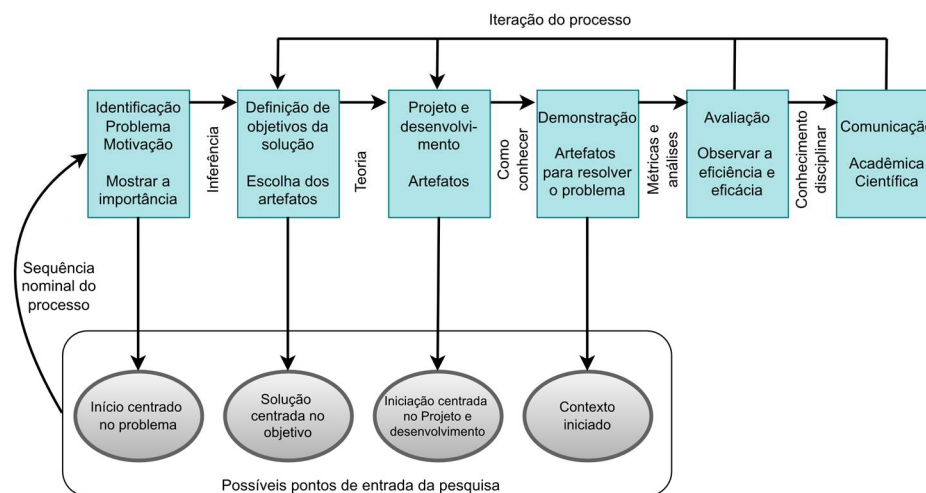


Figura 1 – metodologia da pesquisa – adaptada de (PEFFERS et al., 2007)

A pesquisa iniciou com a Etapa 1 e produziu uma revisão sistemática de literatura sobre o binômio mineração de processos e simulação. Os resultados da Etapa 1 juntamente com as características do problema de pesquisa permitiram definir os objetivos da Etapa 2.

A Etapa 3 consiste no projeto e desenvolvimento do PM4SOS, tomando como ponto de partida a descoberta do modelo de processo, a criação automatizada do modelo de simulação e otimização para apoio à tomada de decisão.

A Etapa 4 consiste na demonstração dos algoritmos e das ferramentas utilizadas no projeto e sua utilização. Nesta etapa foi instanciado um log de eventos de um centro cirúrgico para resolver o problema do agendamento de salas de

cirurgia. A demonstração do PM4SOS consiste em analisar os esforços demandados para colocar em prática a aplicação.

A Etapa 5 avalia o PM4SOS mediante os objetivos estabelecidos.

Por fim, a Etapa 6, da metodologia DSRM, consiste em comunicar os aprendizados com a utilização do PM4SOS. Essa comunicação envolve a publicação de artigos científicos.

O resultado das etapas 1 e 2 resultaram na elaboração de um capítulo desta tese conduzido na forma de revisão sistemática de literatura. Esse capítulo foi submetido ao “Jornal da Simulação Computacional ACM” e está em fase de revisão.

O resultado das etapas de projeto, desenvolvimento, demonstração e avaliação, respectivamente 2, 3 e 4 resultam em publicações em congressos internacionais, exceto a Publicação 4 que é em evento nacional.

- Publicação 2: Austrália 2021: IEEE - (FERRONATO; SCALABRIN, 2021).
- Publicação 3: Praga 2022: IEEE – (FERRONATO; SCALABRIN; CARVALHO, 2022).
- Publicação 4: SICTSUL 2022 – (FERRONATO; SCALABRIN, 2022).

A Figura 2 mostra o ciclo completo do PM4SOS, o qual encerra em cinco fases: a) fonte de dados, b) mineração de processos, c) simulação computacional, d) otimização, e) visualização de dados.

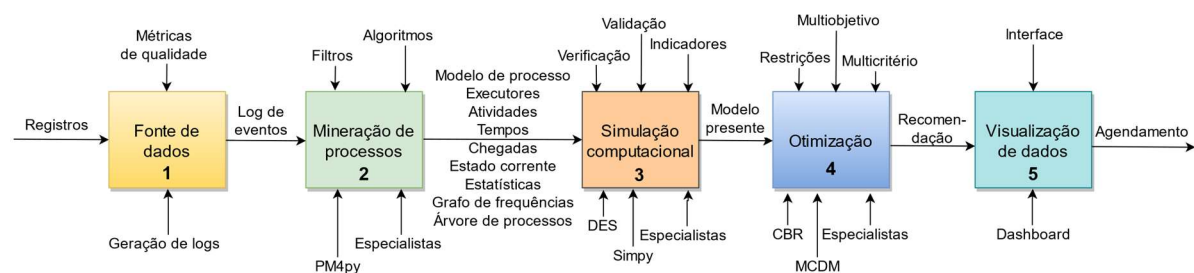


Figura 2 – PM4SOS e seus componentes em IDEF0 – o autor 2022

A modelagem de uma organização em IDEF0 é estruturada em ações, atividades e decisões (PRESLEY; LILES, 1998). O modelo IDEF0 da Figura 2 descreve um modelo de processo em que as setas que chegam nas *caixas* no lado esquerdo representam as entradas (Inputs), as setas que saem das *caixas* são as

saídas (Outputs). As setas da parte superior das caixas representam os controles (métodos e procedimentos) e as setas na parte inferior das caixas representam os mecanismos (humanos e tecnológicos como ferramentas computacionais).

A seguir é apresentada a estrutura deste documento de tese.

1.7 Estrutura do documento de tese

Este documento foi estruturado com as etapas de concepção e experimentação. O Capítulo 1 descreve a contextualização, o problema de pesquisa, os objetivos, a justificativa, as contribuições científicas e tecnológicas e a abordagem metodológica. O Capítulo 2 apresenta uma fundamentação teórica dos elementos abordados na pesquisa. O Capítulo 3 apresenta a estrutura metodológica. O Capítulo 4 apresenta a revisão sistemática de literatura. O Capítulo 5 detalha o desenvolvimento do PM4SOS. O Capítulo 6 apresenta as conclusões. O Capítulo 7 descreve as considerações finais, a relevância do estudo, as limitações e trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Os livros não são feitos para que alguém acredite neles, mas para serem submetidos à investigação. (Humberto Eco)

Neste capítulo, para facilitar o entendimento subsequente, os fundamentos desta tese serão descritos de forma conceitual. Esta pesquisa aborda diferentes perspectivas e áreas do conhecimento na resolução da classe de problemas de planejamento e agendamento, gestão de processos de negócio, indicadores de desempenho, teoria das filas, mineração de processos, simulação computacional, otimização de processos.

Os leitores já familiarizados com estes conceitos, poderão iniciar a leitura a partir da estruturação da pesquisa no Capítulo 3.

2.1 Gestão de Processos de Negócios

A essência de um processo de negócio está no seu gerenciamento e na sua governança. E a implementação de tal gerenciamento deve estar concentrada prioritariamente nos processos essenciais do negócio principal (JESTON; NELIS, 2008). *Business Process Management* (BPM) é uma disciplina de cunho operacional que fornece vários métodos para modelar, analisar, mensurar, otimizar, automatizar e melhorar processos de negócio. Sabe-se que um processo de negócio pode somente ser melhorado com base em dados e fatos. Neste sentido, a mineração de processos é uma disciplina que facilita a análise dos dados e fatos, e coloca em evidência as possibilidades de melhoria do processo de negócio em questão.

BPM compreende estratégias, metas, cultura, estruturas organizacionais, funções, políticas, metodologias e ferramentas de controle para: a) projetar, implementar, analisar, controlar e melhorar continuamente processos fim a fim e b) estabelecer processos de governança (SATALOFF; JOHNS; KOST, 2009).

A eficácia e eficiência dos processos de negócios nas organizações são vitais para a sobrevivência no mundo competitivo. Um processo de negócio mal estruturado pode levar a longos tempos de resposta, baixos níveis de serviço, utilização desequilibrada de recursos, clientes insatisfeitos, atrasos e perdas financeiras. É importante analisar os processos antes de colocá-los em produção

(falhas de projeto), mas também se faz necessária a análise dos processos em execução (diagnóstico e apoio à tomada de decisão) (VAN DER AALST, 2015).

A simulação de processos de negócio (BPS—*Business Process Simulation*) é uma disciplina amplamente utilizada para a análise quantitativa de processos de negócios (DUMAS et al., 2013). Na Figura 3, apresentam-se o ciclo de vida e a arquitetura de um processo de negócio.

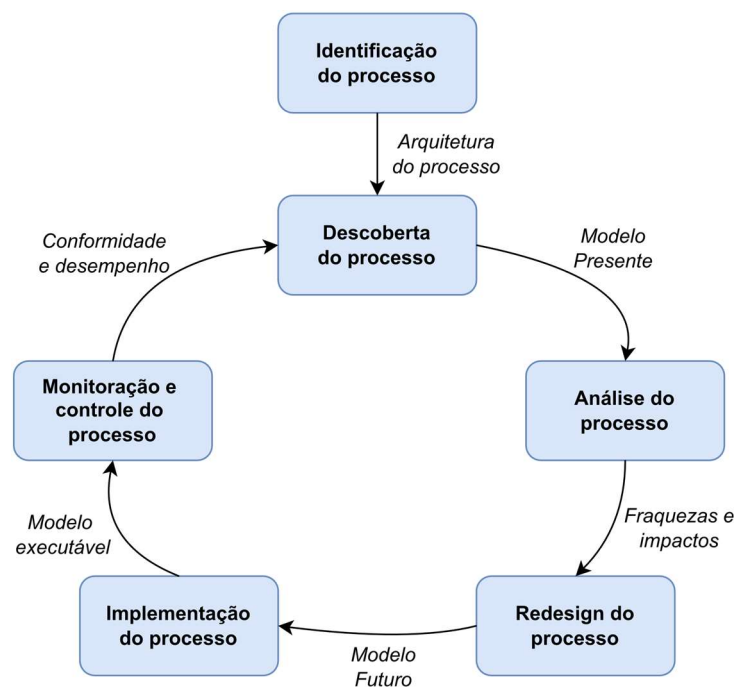


Figura 3 – ciclo de vida BPM – adaptada de (DUMAS et al., 2013)

A ideia principal do BPS é gerar um conjunto de possíveis traços de execução de um processo a partir de um modelo de processo de negócio anotado com, por exemplo, taxa de chegada de novas instâncias de processo, tempo de processamento de cada atividade. Os traços de execução resultantes são usados para calcular as medidas de desempenho do processo, como por exemplo, tempo de ciclo, utilização de recursos e tempos de espera para cada tarefa no processo (CAMARGO; DUMAS; GONZÁLEZ-ROJAS, 2020).

2.2 Problema do planejamento e agendamento cirúrgico

Para Albert Einstein “Insanidade é fazer sempre a mesma coisa, mas esperar resultados diferentes”. A ideia expressa a necessidade de melhoria contínua nos

processos organizacionais. O avanço tecnológico permite armazenar e extrair dados de forma rápida e automatizada em sistemas de informação.

A gerência de recursos em ambientes hospitalares é fundamental para a obtenção de lucratividade/sustentabilidade, redução de custos e satisfação dos pacientes e funcionários envolvidos nos processos. O setor de cirurgias atende procedimentos eletivos e não eletivos consideradas atividades de urgência e emergência. O cenário de alta complexidade de um centro cirúrgico exige procedimentos adequados de planejamento e agendamento de cirurgias.

O problema de planejamento e agendamento está situado no campo da pesquisa operacional. O **planejamento** é descrito em (SLACK; LEWIS, 2005) como “o processo de conciliar oferta e demanda e lidar com decisões de capacidade”. O **agendamento** é descrito como “a definição da sequência e tempo alocado para cada atividade. É a construção de um cronograma detalhado que mostra em que horário ou data os procedimentos cirúrgicos devem começar e quando devem acabar” (SLACK; LEWIS, 2005).

A abordagem dos temas de planejamento e agendamento é feita conforme os níveis organizacionais estratégico, tático e operacional. Ao tratar de um problema específico, no caso em questão centros cirúrgicos, (CARDOEN; DEMEULEMEESTER; BELIËN, 2010) realizaram uma revisão de literatura e classificaram seus estudos sob outras perspectivas: a) **características do paciente**: categorizados em cirurgias eletivas e não eletivas; b) **medidas de desempenho**: tempo de espera, recursos utilizados, valores financeiros, preferências ou tempo para percorrer o sistema; c) **delineamento da decisão por paciente ou cirurgião**: tipo de decisão (data, hora, sala ou capacidade); d) **características de incerteza**: tempo de chegada e duração de cada procedimento; e) **aplicabilidade da pesquisa**: implementação prática e testes *in loco*. Esta tese considera estes cinco aspectos para delineamento da proposta de solução do problema em foco.

Problemas de atribuição e sequenciamento de tarefas são considerados NP-completos (GAREY; JOHNSON, 1979). Esse tipo de problema é de difícil resolução com métodos convencionais, e a quantidade de computação necessária para encontrar a solução ótima aumenta exponencialmente com o tamanho do problema (GAREY; JOHNSON, 1979).

O agendamento de salas cirúrgicas pode ser dividido em duas versões: o atendimento emergencial e imediato, ou a programação antecipada a qual

estabelece uma data para a cirurgia (eletiva) acontecer, e seu objetivo é minimizar tempo de espera do paciente. O agendamento de alocação antecipado envolve a montagem da sala de cirurgia e uma data e horário de início do processo cirúrgico (ABDALKAREEM et al., 2021).

Um cronograma deve se adequar a várias restrições. Geralmente, para lidar com problemas de horários emprega-se a resolução de restrições (AMPONSAH; AGYEMAN; OKRAH, 2011). O agendamento de cirurgias envolve a satisfação de condicionantes categorizadas em restrições *rígidas* e *suaves* — *leves ou flexíveis*. Restrições *rígidas* inviabilizam o agendamento se não forem atendidas. Restrições *suaves* podem reduzir ou aumentar a qualidade da solução em inviabilizar o agendamento. O objetivo é satisfazer todas as restrições *rígidas* e minimizar a violação de restrições *suaves*.

Justificativa: A otimização do uso dos espaços ou salas cirúrgicas em ambientes hospitalares é um desafio a ser superados pelos gestores (NEPOTE; ILZA URBANO MONTEIRO; ELLEN HARDY, 2009). A taxa de ocupação das salas cirúrgicas no Brasil é de 53% e nos países desenvolvidos 85% conforme tese de (GATTO, 1995) e corroborando com a temática, (STROPARO; BICHINHO; PROTIL, 2005) criou um ambiente de modelagem e simulação para fornecer subsídios para melhorar a utilização de profissionais envolvidos e aumentar a taxa de ocupação das salas cirúrgicas. Conforme (CGU, 2021), a impossibilidade de efetivo agendamento da cirurgia no Mapa Cirúrgico Diário pode resultar em falta de comparecimento do paciente e conseqüentemente ocasionar ociosidade do centro cirúrgico, por falta de tempo hábil para encaixe de outra cirurgia. A baixa taxa de utilização das salas de cirurgia e a necessidade de otimização dos espaços para redução de custos e maximização das receitas embasam a escolha do nicho de atuação da pesquisa desta tese. O problema de agendamento envolve uma complexa estrutura e detalhes antes de obter o agendamento cirúrgico. As implicações podem estar na autorização dos procedimentos, nos materiais e suprimentos necessários, farmácia, Unidade de Terapia Intensiva, estoque de sangue, diagnósticos e exames, laboratórios, leitos disponíveis, profissionais para execução dos procedimentos e outros fatores. Nesse sentido, a abordagem central desta tese contempla apenas o processo de agendamento em específico, não sendo objeto de estudo a complexidade de um centro cirúrgico. Esta pesquisa está concentrada em obter a automatização dos agendamentos cirúrgicos com base em dados presentes em

sistemas de informação e avaliar indicadores de ocupação de salas de cirurgia, bem como o ciclo de atendimento e a entrega dos serviços dentro do prazo estabelecido.

Um número crescente de sistemas de informação usa ferramentas de software, como gerenciamento de fluxo de trabalho e sistemas de planejamento de recursos organizacionais para apoiar o controle, a coordenação, o monitoramento e a execução de processos de negócios. Como resultado, estes sistemas de informação empresarial têm informações dos processos de negócios, do estado atual de cada processo, e de dados históricos (REIJERS; AALST, 1999).

A mensuração de um processo de um negócio pode ser realizada com a análise do comportamento de filas.

2.3 Teoria das filas

O comportamento das filas compreende uma forma de alterar sistemas com gargalos de estrangulamentos. A ideia central desta teoria é prestar um melhor atendimento e proporcionar redução de custos, na medida que as filas consideram o uso de recursos e tempos de espera. Existem alguns fatores que podem influenciar na operação de um sistema: forma de atendimento, disciplina da fila, taxa de chegada e estrutura do sistema de atendimento (PRADO, 2009).

As filas estão presentes em nosso dia a dia, seja em bancos, em lojas, no trânsito ou qualquer local que seja necessário aguardar para obter um serviço. As filas também estão presentes em ambientes de produção na espera de uma peça na linha de montagem. Na área da saúde, as filas representam a espera pelo atendimento médico especializado. O objetivo da teoria de filas é avaliar o comportamento de um sistema e seus parâmetros. E nesta linha de raciocínio pode-se citar alguns exemplos de parâmetros: a) tempo médio de espera; b) probabilidade de formação de fila; c) probabilidade de um cliente esperar mais do que certo tempo; d) número médio de clientes em fila.

Justificativa: A teoria de filas consiste em método analítico que utiliza fórmulas matemáticas e a simulação é uma técnica apoiada ao uso de computador, procuram gerar um modelo que represente o Sistema em estudo (PRADO, 2009). A simulação permite replicar, por exemplo, o funcionamento de uma fábrica antes de efetuar alterações na fábrica real, ou seja, ela permite interagir com uma “fábrica virtual” (PRADO, 2009).

As filas ocorrem devido ao desequilíbrio entre a demanda e a execução de um serviço. (PRADO, 2009) classifica uma fila nos seguintes aspectos: a) **população**: indivíduos que usam filas como clientes ou pacientes; b) **taxa de chegada**: valores médios do intervalo entre cada chegada dos indivíduos no Sistema. Neste aspecto são usadas as distribuições de frequência, pois este processo sofre variações no decorrer da sua execução; c) **taxa de atendimento**: caracteriza-se pelo tempo médio de atendimento de um servidor a um indivíduo. O cálculo é realizado para obter o ritmo de atendimento, por exemplo, 5 clientes por hora e o tempo médio de atendimento por cliente é de 12 minutos; d) **número de servidores**: os atendimentos são realizados de forma individualizadas por servidores. Cada servidor tem uma capacidade de atendimento e a qualidade no atendimento pode ser influenciada. Existem vários sistemas de filas como MM1 (vários indivíduos aguardando e 1 único servidor atendendo) ou MMS (vários indivíduos aguardando e vários servidores atendendo); e) **disciplina da fila**: é a regra que define quem será o próximo indivíduo a ser atendido. Podem ser: *First In First Out (FIFO)*, *Last In First Out (LIFO)*, serviço por ordem de prioridade ou serviço.

O tratamento de filas em processos de negócio apresenta características específicas para os recursos como taxa de chegada dos clientes em fila, o tempo de preparação ou ativação do serviço, a disponibilidade, horizonte de tempo da atividade, tempo de serviço e liberação do recurso cf. Figura 4.

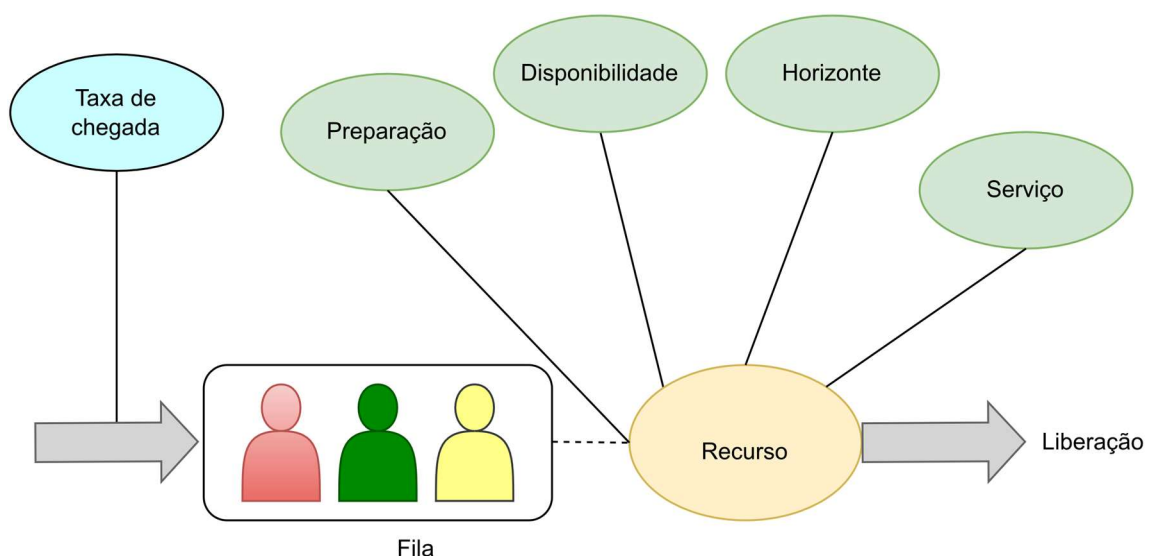


Figura 4 – sistema de filas – adaptada de (VAN DER AALST et al., 2010)

Um recurso pode apresentar três estados: a) **inativo**: recurso não alocado no processo porque não está em operação; b) **pronto**: está alocado no processo, mas não está pronto e c) **ocupado**: recurso está alocado no processo e trabalhando em alguma atividade. Quando uma nova instância de simulação é criada, o recurso está inativo, este *case* permanece uma parcela de tempo em um estado chamado *preparação*. Os parâmetros de um modelo de filas em um processo (AALST et al., 2010) são: a) **taxa de chegada (λ)**: número médio de instâncias chegando por unidade de tempo. O intervalo entre duas chegadas subsequentes é descrito pela distribuição exponencial negativa $\frac{1}{\lambda}$, sendo $\lambda > 0$; b) **serviço (μ)**: número médio de instâncias atendidas por unidade de tempo. O tempo de processamento é descrito pela distribuição exponencial negativa $\frac{1}{\mu}$, sendo $\mu > 0$; c) **utilização (ρ)**: é a fração de tempo que o recurso está ocupado, descrito por $\rho = \frac{\lambda}{\mu}$; d) **preparação (c)**: pequena fração de tempo em que o recurso é alocado ao processo. O período ativo deve ser múltiplo de *c*; e) **horizonte (h)**: período considerado, sendo $h > 0$; f) **disponibilidade (a)**: fração de tempo que o recurso está disponível para o processo.

Os indicadores de filas permitem detectar anormalidades e necessidade de melhorias. Para a adequada análise dos indicadores de filas, é necessário analisar dados presentes em sistemas de informação, tais como eventos. A mineração de processos executa esta tarefa de processamento/análise de eventos.

2.4 Mineração de processos

A mineração de processos encerra (cf. Figura 5) um cenário com diferentes sistemas de informação que controlam operações do mundo real e armazenam registros de eventos e transações (VAN DER AALST, 2016).

Justificativa: A mineração de processos foi escolhida, nesta pesquisa, para obter, a partir de um log de eventos, informações para o suporte à tomada de decisão. As estatísticas obtidas permitem analisar como um processo se comporta e quais são as suas deficiências e anomalias. As informações do modelo de processo descoberto favorecem a criação de um modelo de simulação.

A mineração de processos contempla três perspectivas: descoberta, verificação de conformidade e aprimoramento. E para a mineração de processos ser

realizada é necessário um log de eventos bem estruturado para permitir a aplicação das técnicas de mineração.

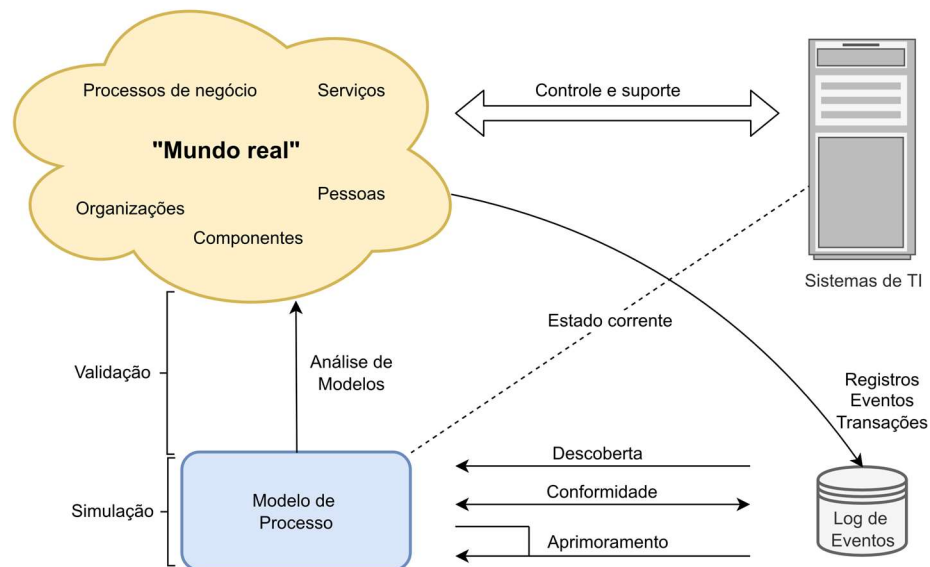


Figura 5 – mineração de processos – adaptada de (VAN DER AALST, 2016)

O log de eventos representa uma coleção de registros históricos de instâncias de um processo. Cada instância contém um conjunto de eventos gerados durante a execução de um processo.

Definição 1 – Casos e traços (VAN DER AALST, 2016):

Seja \mathcal{C} o universo de casos, ou seja, o conjunto de todos os possíveis identificadores de casos. Casos e eventos possuem atributos. Para qualquer caso $c \in \mathcal{C}$ e nome $n \in AN$: $\#_n(c)$ é o valor do atributo n para o caso c ($\#_n(c) = \perp$ se caso c não contém atributo n). Cada caso tem um traço com um atributo obrigatório especial $\#_{trace}(c) \in \mathcal{E}^*$. E para se referir ao traço de um caso usa-se $\hat{c} = \#_{trace}(c)$. Um traço é uma sequência finita de eventos $\sigma \in \mathcal{E}^*$ tal que cada evento aparece apenas uma vez, ou seja, para $1 \leq i < j \leq |\sigma|$: $\sigma(i) \neq \sigma(j)$.

Definição 2 – Log de eventos (VAN DER AALST, 2016):

Um log de eventos é um conjunto de casos $L \subseteq \mathcal{C}$ tal que cada evento c aparece no máximo uma vez em todo o log, ou seja, para qualquer $c1, c2 \in L$ tal que $c1 \neq c2$: $\partial_{set}(\hat{c1}) \cap \partial_{set}(\hat{c2}) = \emptyset$.

Se um log de eventos contém carimbo de tempo, então a ordenação dos *eventos* deve respeitar esses carimbos de tempo, ou seja, para qualquer $c \in L$, i e j tal que $1 \leq i < j \leq |\hat{c}|$: $\#_{time}(\hat{c}(i)) \leq \#_{time}(\hat{c}(j))$.

O conjunto de atividades que aparecem no log L são representadas por $\{\#_{activity}(c) \mid c \in L \wedge c \in \hat{c}\}$.

Seja \mathcal{A} o conjunto de nomes de atividades. Um traço σ é uma sequência de atividades, ou seja, $\sigma \in \mathcal{A}^*$. Um log de eventos L é um múltiplo conjunto de traços sobre \mathcal{A} , ou seja, $L \in \mathbb{B}(\mathcal{A}^*)$.

Os logs de eventos são essenciais para a execução da mineração de processos. A avaliação da qualidade dos logs permite compreender se os seus dados—como entrada—são adequados e suficientes para que a mineração de processos ocorra com eficiência.

De fato, conforme (VAN DER AALST, 2016), é necessário revisar a qualidade dos logs de eventos antes de aplicar a mineração de processos. Essa sugestão também foi descrita no manifesto da mineração de processos definindo cinco níveis de maturidade dos logs de eventos, atribuindo a denotação de estrelas, sendo que uma estrela representa um log com qualidade pobre e cinco estrelas representa um log com qualidade excelente (VAN DER AALST; ARCIERI; ADRIANSYAH, 2012).

Um log de eventos com baixa qualidade contém dados errôneos, valores duplicados ou dados “perdidos” representam modelos desestruturados do tipo “*spagueti*”. É difícil interpretar os modelos de negócio, bem como modelos obtidos com a mineração de processos se os dados não refletem a realidade (MARIN-CASTRO; TELLO-LEAL, 2021).

A avaliação dos logs permite identificar diversas características da qualidade como descritos em (FISCHER et al., 2020), a saber: a) **acurácia**: descreve o desvio do valor registrado do valor real representado pelos dados; b) **integridade**: descreve o registro de todos os valores para uma variável específica. Os registros de tempo são verificados para cada evento de um log; c) **consistência**: representação igualitária e granular de valores para todos os logs de eventos; e d) **singularidade**: garante a não existência de dados duplicados.

Apresenta-se, na Figura 6, a avaliação da qualidade de um log de eventos a partir de um log de eventos. Cada dimensão de qualidade utiliza medidas e avaliações específicas de cada dimensão. O resultado é um escore de qualidade

para análise do log de eventos. E de acordo com (VAN DER AALST, 2016), um log de eventos contém casos. Os casos contêm eventos, de forma que cada evento está precisamente relacionado a um caso. Cada evento que pertence a um caso está ordenado. Os eventos podem conter atributos básicos como ID, nome da atividade, carimbo de tempo e atributos opcionais como custos e recursos originadores. E depois de obter a avaliação da qualidade do log de eventos, a mineração de processos permite obter informações relevantes dos processos de negócio.

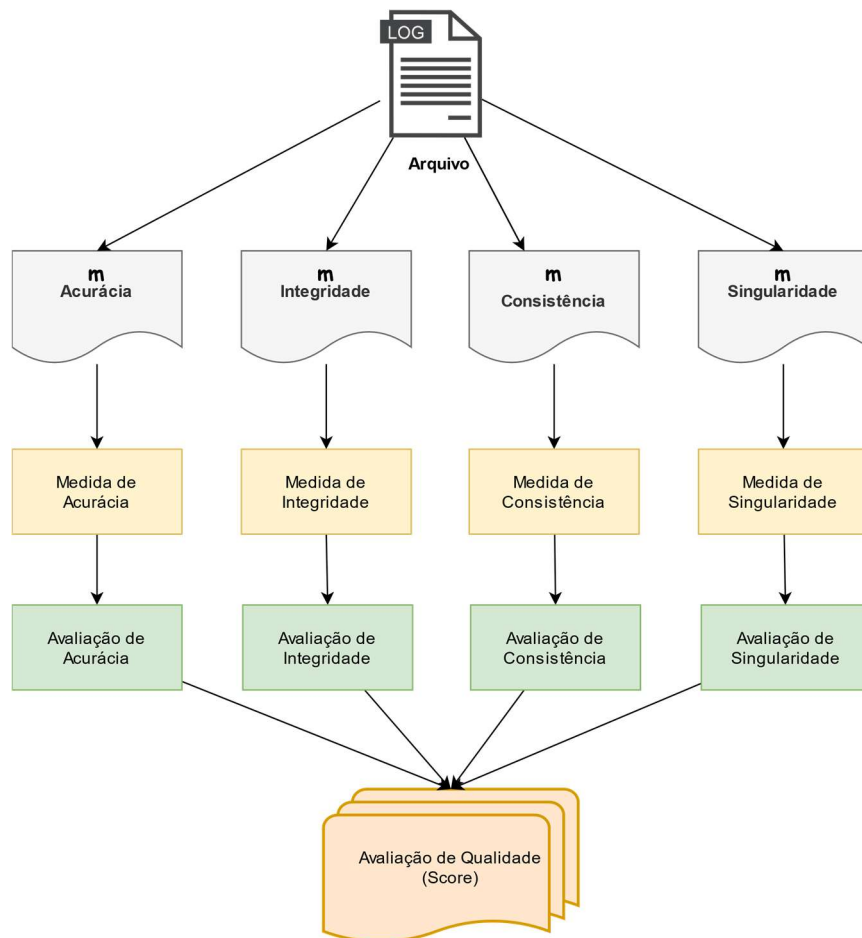


Figura 6 – qualidade do log – adaptada de (RAMOS-GUTIÉRREZ et al., 2021).

Uma abordagem qualitativa para avaliação dos logs de eventos deve ser realizada como etapa anterior à mineração de processos. A avaliação de qualidade de um log de eventos permite identificar se os dados estão adequados para prosseguir com a mineração de processos. Caso o log de eventos não atinja o escore de avaliação desejado faz-se necessário aplicar o pré-processamento, conforme realizado por (KHERBOUCHE; LAGA; MASSE, 2016).

A perspectiva descoberta da mineração de processos permite, a partir de logs de eventos, obter modelos de instâncias de processos. Dentre os formalismos para representar modelos de processos têm-se as redes de Petri, a notação de modelagem de processos de negócios (BPMN), grafos direcionados — ou *DFG*, árvores de processos dentre outros (VAN DER AALST, 2016).

O algoritmo *Alpha-Miner* (VAN DER AALST; WEIJTERS; MARUSTER, 2004) foi o primeiro algoritmo implementado para a tarefa de descoberta de modelo de instâncias processo. *Alpha-Miner* toma como entrada um log de eventos e transforma-o em um modelo de processo de negócio no formato de uma Rede de Petri. Este formalismo permite analisar os comportamentos registrados no log de eventos. Quando o log de eventos contém informações relativas a recursos, pode-se vinculá-los ao modelo de processo, assim como, as tarefas pessoais executadas e registradas em um sistema de informação juntamente com o modelo do ambiente (VAN DER AALST, 2016).

O algoritmo *Alpha-Miner* não lida corretamente com *outliers*, dados faltantes e não reconhece processos de iteração sequencial (VAN DER AALST; WEIJTERS; MARUSTER, 2004). O algoritmo *Heuristic Miner* (WEIJTERS; VAN DER AALST; ALVES DE MEDEIROS, 2006) não garante a consistência do modelo.

As redes de Petri foram as precursoras dos formalismos para a descoberta de modelos de processos. Todos os *traces* aceitáveis correspondem ao caminho que inicia em M_0 e termina em M_f . Pode não haver nenhum traço alcançável a partir da origem, isto é, M_f não é alcançável a partir de M_0 . O conceito de solidez dos modelos de processos foi introduzido no contexto dos *workflows* por (VAN DER AALST et al., 2011).

Uma forma de garantir a solidez de um modelo de processo é utilizar o algoritmo *Inductive Miner*. Este algoritmo garante a consistência do modelo descoberto e executa a divisão do log de eventos como a sequencialidade, paralelização, concorrência e laços. A descoberta do modelo de processo com *Inductive Miner* foi apresentada por (LEEMANS; FAHLAND; VAN DER AALST, 2013), (LEEMANS; POPPE; WYNN, 2019) e utiliza árvore de processos para obtenção de modelos sólidos. Na Figura 7, mostra-se uma árvore de processos e seus operadores para a descoberta do modelo de processos. A ordem de execução das atividades do processo ocorre da esquerda para a direita a partir do nodo raiz \rightarrow .

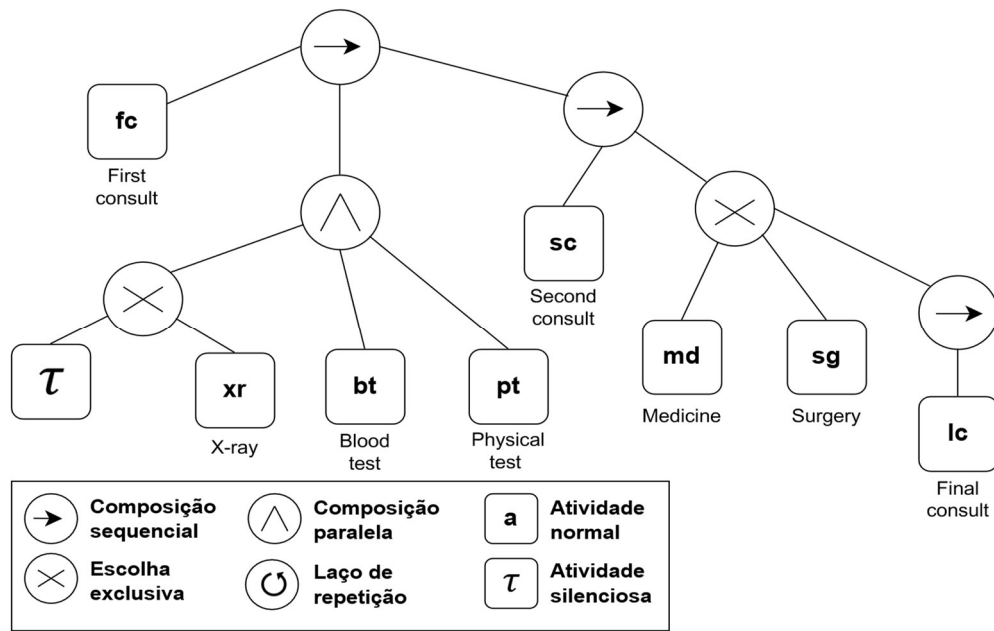


Figura 7 – árvore de processos – adaptada de (VAN DER AALST et al., 2019)

Uma árvore de processos é definida de forma gráfica conforme Figura 7 e textual de acordo com a expressão: $\rightarrow (fc, \wedge (\times (\tau, xr), bt, pt), (sc), (\times (md, sg)), (lc))$.

Definição 3 – Árvore de processos (VAN DER AALST et al., 2019):

Seja $A \subseteq \mathcal{A}$ um conjunto finito de atividades com $\tau \notin \mathcal{A}$. O conjunto dos operadores de uma árvore de processos é definido por $\oplus = \{\rightarrow, \times, \wedge, \cup\}$. A definição indutiva de árvores de processos (VAN DER AALST et al., 2019):

- Se $a \in A \cup \{\tau\}$, então $Q = a$ é uma árvore de processo,
- Se $n \geq 1, Q_1, Q_2, \dots, Q_n$ são árvores de processos e $\oplus \in \{\rightarrow, \times, \wedge\}$, então $Q = \oplus (Q_1, Q_2, \dots, Q_n)$ é uma árvore de processos, e
- Se $n \geq 2$ e Q_1, Q_2, \dots, Q_n são árvores de processos, então $Q = \cup (Q_1, Q_2, \dots, Q_n)$ é uma árvore de processos.

O resultado da etapa de descoberta do modelo de processo pode ser representado visualmente como um grafo direcionado. De forma semântica, a árvore de processos corresponde ao conjunto de traços obtidos a partir do nó raiz, percorrendo os arcos direcionais das conexões entre as atividades do modelo até o nó final. O grafo direcionado utiliza duas métricas: a) **frequência**: número de vezes que o evento de origem se conecta ao evento de destino e b) **execução**: agregação do tempo utilizado entre dois eventos (LEEMANS; POPPE; WYNN, 2019).

Definição 4 – Grafos diretamente conectados (VAN DER AALST et al., 2019):

Seja L um log de eventos, tal que $L \in \mathbb{B}(\Sigma^*)$. O grafo diretamente conectado de L é $G(L) = (A_L, \mapsto_L, A_L^{start}, A_L^{end})$ com:

- $A_L = \{a \in \sigma \mid \sigma \in L\}$ é o conjunto de atividades que pertencem a L ,
- $\mapsto_L = \{(a, b) \in A \times A \mid a >_L b\}$ é a relação do grafo direcionado,
- $A_L^{start} = \{a \in A \mid \exists \sigma \in L, a = first(\sigma)\}$ representa a atividade inicial, e
- $A_L^{end} = \{a \in A \mid \exists \sigma \in L, a = last(\sigma)\}$ representa a atividade final.

A ideia geral da abordagem do algoritmo *inductive miner* é construir grafos diretamente conectados para representar o modelo presente no log de eventos. O log é decomposto em sub logs criados de forma que cada evento aparece precisamente em um sub log (VAN DER AALST et al., 2019).

A perspectiva de verificação de conformidade permite que um modelo de processo seja comparado com o log de eventos do mesmo processo. Tal verificação pode ser usada para validar se os fatos registrados no log de eventos pertencem ao modelo de processo existente e se eles encerram alinhamentos de conformidade com o modelo de processo existente (ou de referência). A verificação de conformidade pode também ser usada para detectar, localizar, explicar e medir a gravidade de desvios em um processo de negócio (VAN DER AALST, 2016).

A perspectiva de aprimoramento permite confrontar um modelo de processo a um log de eventos e gerar um novo modelo de processo. A ideia central é estender e melhorar um modelo de processo existente com o uso de informações registradas em logs de eventos (VAN DER AALST, 2016).

A mineração de processos define um conjunto de tarefas de suporte operacional (VAN DER AALST, 2016). Essas tarefas são especializadas na detecção de desvios, predição de desfechos e recomendação de ações, tomando como parâmetro de entrada dados históricos de eventos. Neste mesmo ambiente operacional, a simulação permite responder questões “*What if*” ligadas às tarefas de diagnóstico e predição com base nos modelos de processos *de facto* e *de jure*. Ambos os modelos *de facto* e *de jure* encerram perspectivas de controle de fluxo, dados/regras e recursos organizacionais. O modelo de simulação pode aprender

continuamente com dados históricos factuais e artificiais oriundos de um ambiente de simulação (VAN DER AALST, 2016).

Embora a ênfase da análise repouse sobre os dados históricos, diversos repositórios de dados são atualizados em tempo real e requerem análises instantâneas (DER AALST; ARCIERI; ADRIANSYAH, 2012). A mineração de processos fornece suporte operacional às tarefas de detecção, predição e recomendação. Por exemplo, quando um caso desvia do modelo pré-definido no processo *de jure*, um alerta pode ser gerado. Dados históricos e de processos em execução podem ser usados para construir modelos de simulação, por exemplo, para predizer o tempo restante de execução de uma instância de processo (CHOUËIRI et al., 2020), para recomendar ações com vistas à redução de custo operacional (THABET; GHANNOUCHI; BEN GHÉZALA, 2014) ou controle do tempo de execução do processo (ROZINAT et al., 2008a). A capacidade de predição da análise operacional permite fornecer subsídios para a verificação de conformidade e melhorar o suporte à tomada de decisão para gestores de processos de negócio (PRAVILOVIC; APPICE; MALERBA, 2014).

Para prover suporte operacional adequado—no tempo correto e com informações que espelham a realidade próxima dos fatos—passa pela conjugação do binômio mineração de processo e simulação com vista a construção de um ambiente integrado e *online* de simulação de curto prazo. Considerando a simulação de curto prazo, quanto mais longo for o período de simulação, menor será o efeito do estado inicial. Quando o estado corrente é transformado em dado histórico, a tomada de decisão se afasta do nível operacional (REIJERS; AALST, 1999).

2.5 Simulação computacional

A simulação e a análise de um processo de negócio são reconhecidas como partes integrantes da otimização de desempenho. Uma organização pode recorrer à simulação de processos quando enfrenta níveis crescentes de pressão para entregar eficiência operacional. A simulação de um processo de negócio é efetivamente relevante quando o teste de um novo processo de negócio no mundo real demanda: perturbação nas operações correntes, rapidez na avaliação de cenários potenciais e custo relativamente baixo de exploração/experimentação da transformação do negócio.

A simulação de um processo de negócio utiliza três perspectivas: (a) **controle de fluxo**, que consiste no ordenamento de atividades, no estabelecimento de seqüências de atividades, controle de laços e roteamento de decisões; (b) **dados e regras**, que permitem armazenar valores durante a execução de cada instância de processo; e c) **dados de recursos**, que representam a disponibilidade dos recursos humanos e não humanos. A duração de cada atividade e a probabilidade de seguir um caminho específico são importantes para permitir à cada gestor controlar os recursos na tomada de decisão (VAN DER AALST et al., 2010).

O binômio simulação e mineração de processos habilita ter um “olhar à frente” e explorar eficientemente cenários envolvendo processos de negócio. Esse binômio, como já dito, é considerado por van der Aalst (VAN DER AALST, 2018), como “*A combinação perfeita!*”. Essa contribuição mútua entre mineração de processos e simulação facilita a construção de um suporte operacional efetivo.

A simulação computacional é uma ferramenta de análise de cenários que pode ser combinada com técnicas de otimização para obter soluções ótimas de um problema, i.e., obter a melhor solução dentre todas aquelas viáveis. Existem várias técnicas de simulação computacional e elas podem ser categorizadas (KATSALIAKI; MUSTAFEE, 2011), como segue: simulação baseada em agentes (ABS), simulação de eventos discretos (DES) e simulação de sistemas dinâmicos (SDS). Essas técnicas de simulação consideram o fator tempo.

A ABS reúne vários agentes autônomos que atuam continuamente em um ambiente. Eles interagem por meio da troca de mensagens ou indiretamente por meio de estímulos recebidos do ambiente. A abordagem é interessante na medida que se pode modificar as regras comportamentais no nível individual e adicionar novos agentes com modelos comportamentais personalizados. Essas características permitem modelar um sistema com comportamentos complexos e emergentes, resultantes da interação dos seus indivíduos com o ambiente (FERBER, 1999).

A DES permite modelar cada operação do Sistema como uma seqüência de eventos discretos e cada evento ocorre em um determinado instante de tempo e denota uma mudança de estado no Sistema (BECK, 2008). Os elementos da simulação DES são: a) **estado**: define o comportamento do Sistema ao longo do tempo; b) **relógio**: controla o tempo da simulação; c) **lista de eventos**: representa os eventos que ainda podem ser executados; d) **gerador de números aleatórios**: cria valores de variáveis aleatórias para a execução da simulação; e) **estatística**:

estima as medidas do comportamento do Sistema; f) **condição de finalização da execução**: determina a condição de parada da simulação (BECK, 2008).

A SD permite analisar sistemas baseados em relações de causa e efeito entre os componentes. A abordagem dinâmica é relevante para executar simulações de longo prazo de nível estratégico. Ela permite simular comportamentos complexos e utiliza dois conceitos básicos: a) **estoque**: representa os elementos que são acumulados ou se esgotam e b) **fluxos**: taxa de variação de um estoque. São exemplos de sistemas dinâmicos: população (estoque), taxa de fertilidade (fluxo positivo) ou taxa de mortalidade (fluxo negativo) (CHWIF; MEDINA, 2015).

A simulação como abordagem metodológica encerra fases, tarefas e métodos específicos. Em algumas situações, as técnicas são usadas de forma associada, i.e., combinando as principais características de cada técnica de simulação. A seguir são descritas as abordagens de simulação para processo de negócio.

De forma geral, as tarefas de mineração de processos ligadas às perspectivas de descoberta, verificação de conformidade e aprimoramento de processos permitem executar diversas análises, como *replay* da execução de um processo, verificação de conformidade e *performance* de casos. As fontes de dados primárias para a extração, por exemplo, do estado corrente ou situação atual do processo de negócio em questão são obtidas por meio da análise de casos. A obtenção do estado corrente de um processo tem como propósito alimentar a criação de um modelo de simulação em seu estado inicial.

A simulação de curto prazo demanda informações prontamente disponíveis nos sistemas de informação das organizações. Ela fornece suporte às decisões que afetam o processo de negócio no curto espaço de tempo e podem ser avaliadas sem a necessidade de modelagem extra (REIJERS; AALST, 1999).

Há três elementos principais para colocar em prática a simulação de curto prazo: (a) a parametrização do modelo de processo; (b) a disponibilidade *online* de dados históricos para estimar os valores dos parâmetros da simulação; e (c) a disponibilidade *online* do estado atual do processo. O estado atual representa a essência para a consecução da simulação de curto prazo; ele reflete o estado corrente do processo em execução. Ele dá o suporte necessário para a implementação da funcionalidade “*avanço rápido*”. Com a simulação pode-se explorar diferentes cenários relacionados aos seus efeitos no futuro próximo (REIJERS; AALST, 1999). Para ir na direção da automatização da simulação de

curto prazo, requer-se que a coleta de dados dos modelos *de jure* e *de facto* sejam sistemáticas e *online*. Nesse contexto, (ROZINAT et al., 2008a) combinou a especificação do modelo de processo de negócio, dados históricos e estado corrente do sistema de processo para a consecução de um modelo de simulação.

A simulação de curto prazo antecipa o efeito de certas decisões, como por exemplo, a contratação de novos funcionários ou a renúncia ou a aceitação de novos pedidos. A simulação de curto prazo pode ser usada para detectar desequilíbrios de recursos em um sistema, examinar efeitos de indicadores de desempenho operacional, gerar e validar modelos de simulação, dado que o resultado da simulação deve refletir a real execução de um processo (REIJERS; AALST, 1999).

Em (SCHOLL et al., 2012), discute-se os requisitos de um modelo de simulação de curto prazo. Destaca-se a exigência de alta precisão da tarefa de predição, o que demanda elevado nível de detalhes e qualidade dos dados coletados e uma inicialização do modelo de simulação com o estado corrente do processo e automatização da geração do modelo de simulação, i.e., qualquer ajuste manual, na parametrização do modelo de simulação de curto prazo, o inviabilizaria.

2.6 Otimização de processos

Otimizar significa buscar a melhor solução para um problema. A intenção é maximizar ou minimizar um ou mais objetivos, como por exemplo, o tempo e/ou o custo. Para alcançar um objetivo de negócio é preciso respeitar restrições, como orçamentos, especificações, leis. A otimização é realizada colocando-se em prática ajustes de características — ou parâmetros — que compõem uma possível solução, como por exemplo, dimensões ou tipo de materiais (ÁVILA, 2020).

Em projetos de simulação pode-se ter um conjunto aleatório de variáveis de entrada. É comum o processamento de diversos cenários em um simulador com o objetivo de obter informações sobre o comportamento do sistema. Porém, parte das simulações realizadas tendem a serem desprezadas, na medida em que os resultados não apresentam impacto significativo de melhoria do modelo simulado em questão. O desenho de um experimento é um método que permite estimar como as variáveis de entrada afetam a resposta do modelo de simulação. Para um ambiente

simulado, pode-se planejar cenários de forma racional, evitando o método de tentativa e erro, comum em projetos de simulação (CHWIF; MEDINA, 2015).

A simulação permite responder questões como: “O que aconteceria se?”. Por exemplo: O que aconteceria com as entregas ao se aumentar o número de entregadores de 6 para 7? A simulação é considerada uma técnica avaliadora e não geradora de soluções (CHWIF; MEDINA, 2015). Para mudar essa condição, utilizam-se técnicas de otimização em conjunto com a simulação computacional.

A otimização pode ser vista como complemento à simulação, pois permite obter possíveis soluções candidatas para a solução de um problema. Apresenta-se o ciclo de otimização na Figura 8 (FU, 2002). A partir de funções de otimização, soluções candidatas podem ser testadas com a simulação. O resultado obtido com a simulação permite efetuar análise estatística. O ciclo é repetido até atingir um critério de parada. Se o critério de parada não for satisfeito, o ciclo de otimização reinicia e gera novas soluções, com técnicas específicas, como por exemplo, algoritmos genéticos e um novo ciclo de otimização é gerado e testado com a simulação e caso o critério de parada for satisfeito, os resultados são apresentados.

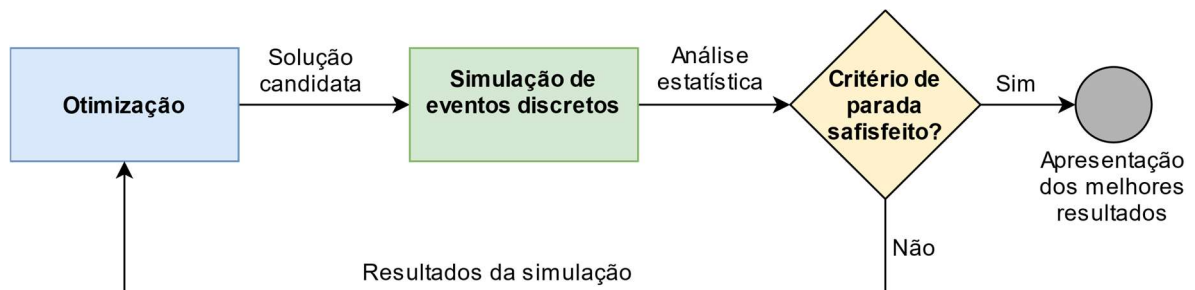


Figura 8 – estrutura da simulação e otimização – adaptada de (FU, 2002)

A otimização consiste em tentar obter diferentes combinações de valores de variáveis de entrada. Tais combinações de valores das variáveis provê uma saída viável. No entanto, este processo que usa diferentes combinações para as variáveis de entrada para gerar uma solução ótima, torna-se inviável de serem colocadas em prática em sistemas reais e por isso realiza-se tal processo por meio de modelos (HARREL et al., 2000). Um problema de otimização pode ser formulado a partir de uma função objetivo (ou medida de desempenho) que se deseja maximizar ou minimizar, sujeito a um conjunto de restrições que delimitam os valores que as variáveis podem assumir (CHWIF; MEDINA, 2015). Uma das limitações para uso da

otimização em simulação é a quantidade de variáveis. O desempenho computacional pode ser reduzido diante de um elevado número de variáveis a serem manipuladas pela otimização (FU, 2002).

A busca pela solução ótima pode ser feita de forma manual ou automatizada. Essa última é colocada em prática por meio de algoritmos específicos para obtenção da solução ótima sem a necessidade de avaliação de todas as possíveis soluções. A automatização da otimização, por meio de algoritmos, permite, segundo (HARREL et al., 2000):

- Economizar tempo do analista para automatizar o processo de análise;
- Explorar eficientemente as melhores combinações de soluções possíveis usando um método lógico;
- Mostrar apenas as “melhores” soluções para o analista. Tal visualização possibilita analisar soluções desconhecidas pelo analista.

A otimização pode ser usada para encontrar uma solução “próxima ao ótimo”. Dada uma função objetivo incorpora-se mecanismos de compensação em custo e benefício. Os primeiros métodos que surgiram tinham enfoque na natureza linear, como por exemplo, o Simplex. Os métodos não lineares evoluíram com o uso de meta heurísticas com base em inteligência artificial, redes neurais, algoritmos genéticos (BANKS et al., 2013).

Um problema básico de otimização pode ser representado como um problema de minimização (KOCHENDERFER; WHEELER, 2019):

$$\text{Minimizar } f(x) = c_1x_1 + c_2x_2 + \dots + c_nx_n \quad \text{Função objetivo (Min)} \quad \text{Eq. 1}$$

Sujeito a: $x \in X$ **Conjunto de restrições**

O conjunto de variáveis x pode ser representado no formato de um vetor de valores unidimensionais:

$$[x_1, x_2, x_3 \dots x_n] \quad \text{Variáveis} \quad \text{Eq. 2}$$

Os elementos do vetor são usados para minimizar a função objetivo. Qualquer valor de x dentre todas as variáveis no conjunto viável x , que minimiza a função objetivo, é chamado de solução.

De forma geral, para qualquer problema de otimização, a Eq. 1 pode ser reescrita para maximizar a função objetivo:

$$\text{Maximizar } f(x) = c_1x_1 + c_2x_2 + \dots + c_nx_n \quad \text{Função objetivo (Max)} \quad \text{Eq. 3}$$

Sujeito a: $x \in X$ **Conjunto de restrições**

A solução particular da Eq. 1 está descrita como x^* . O gráfico da Figura 9 mostra um exemplo de problema de otimização unidimensional.

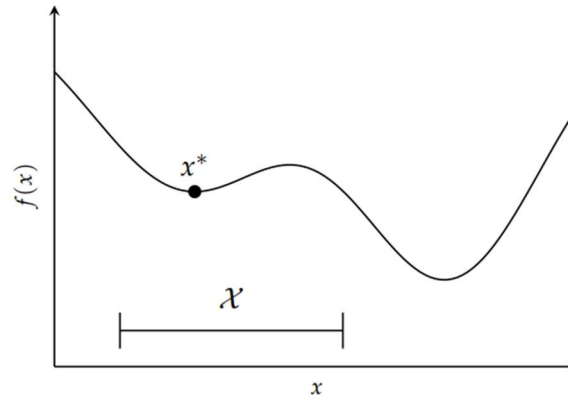


Figura 9 – otimização unidimensional – (KOCHENDERFER; WHEELER, 2019)

Os métodos de otimização com simulação têm sido aplicados, normalmente, com um único objetivo. Porém, as aplicações destes métodos em processos de negócios demandam a otimização de múltiplos objetivos, como por exemplo, reduzir o tempo de espera e aumentar a quantidade de produtos expedidos.

2.6.1 Otimização multiobjetivo

O problema de otimização multiobjetivo lida com mais de uma função objetivo. Na prática, são evidenciados de forma recorrente problemas de tomada de decisão com múltiplos objetivos. Para problemas de otimização com um único objetivo, encontra-se apenas uma solução que otimiza a função objetivo. No entanto, para problemas de otimização com múltiplos objetivos, encontra-se um conjunto de soluções viáveis (DEB, 2001).

Uma solução ótima para um problema multiobjetivo deve apresentar equilíbrio entre as diversas funções de interesse, pois se tratadas isoladamente, a melhora em uma função poderia implicar na piora em outras funções. Esta abordagem foi denominada como “Ótimo de Pareto”. A seguir são descritas de forma geral um problema de otimização multiobjetivo (DEB, 2001):

Minimizar/Maximizar	$f_m(x),$	$m = 1, 2, \dots, M;$	}	Eq. 4
Restrições	$g_j(x) \geq 0,$	$j = 1, 2, \dots, J;$		
	$h_k(x) = 0,$	$k = 1, 2, \dots, K;$		
	$x_i^{(L)} \leq x_i \leq x_i^{(U)},$	$i = 1, 2, \dots, n.$		

A solução x é um vetor com n variáveis de decisão: $x = (x_1, x_2, \dots, x_n)^T$. Os três conjuntos de restrições são: a) desigualdade $g_j(x)$; b) igualdade $h_k(x)$ e c) limite de variáveis pois a restrição de cada variável x_i para obter um valor entre o mínimo $x_i^{(L)}$ e o máximo $x_i^{(U)}$. Estes limites constituem o espaço de decisão D . Assim, toda solução x que satisfaz todas as restrições e limites é conhecida como solução viável.

O conjunto das soluções obtidas com a otimização das funções multiobjetivo são consideradas soluções viáveis conforme ilustração da Figura 10. As soluções obtidas podem ser caracterizadas em dois tipos: a) **dominadas**: estão na área interna e estas não são as melhores soluções, pois há outras melhores que elas e b) **não dominadas**: estas soluções estão na fronteira do Pareto ótimo. Para obter o *ranking* das soluções não dominadas usa-se a medida de distância Euclidiana entre duas soluções (DEB et al., 2002).

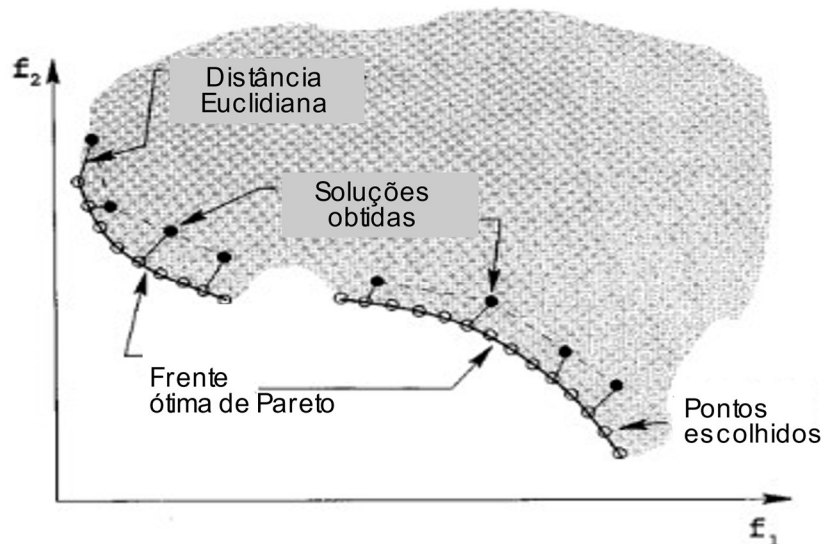


Figura 10 – obtenção das soluções de Pareto – (DEB et al., 2002)

A otimização multiobjetivo é caracterizada em dois objetivos para o conjunto de Pareto-ótimo: a) a convergência para o conjunto de Pareto-ótimo; e b) manutenção da diversidade das soluções. A primeira métrica de convergência

consiste, para cada solução obtida, computar a distância Euclidiana mínima. A média das distâncias é a métrica de convergência. A avaliação da diversidade é obtida com a mensuração da distância de aglomeração. A solução escolhida é aquela que não estiver nos extremos ou fronteiras do Pareto e aquelas que estão localizadas em regiões com menor aglomeração de pontos de solução (DEB, 2001).

Para a resolução de problemas computacionais de forma rápida quando os métodos clássicos são lentos para encontrar uma solução exata, faz-se necessário o uso de heurísticas. Funções heurísticas permitem classificar alternativas em algoritmos de pesquisa para encontrar uma solução aproximada para determinado objetivo com base nas informações disponíveis. As heurísticas existentes exploram resultados em busca de solução global. A otimização com base em restrições pertence a esta classe de heurísticas e são apresentadas a seguir.

2.6.2 Otimização com base em restrições

A otimização com base em restrições consiste em otimizar uma função objetivo com relação a algumas variáveis na presença de restrições. A função objetivo é uma função de custo ou função de energia, que deve ser minimizada, ou uma função de recompensa ou função de utilidade, que deve ser maximizada (MARTINS; NING, 2021). As restrições podem ser: (a) rígidas que definem condições para as variáveis que devem ser satisfeitas; ou (b) flexíveis que possuem alguns valores de variáveis que são penalizados na função objetivo se as condições sobre as variáveis não estão satisfeitas (MARTINS; NING, 2021).

Justificativa: O problema investigado visa alocar procedimentos cirúrgicos em salas com recursos limitados. As restrições referem-se ao tempo de execução da cirurgia, cirurgiões que atendem em turnos específicos, uso de equipamentos específicos (HSU et al., 2021).

Um problema de otimização baseado em restrições pode ser representado por (MARTINS; NING, 2021):

$$\begin{array}{ll} \min & f(x) \\ \text{sujeito a:} & g_i(x) = c_i \quad \text{para } i = 1, \dots, n \quad \text{restrições de igualdade} \\ & h_j(x) \geq d_j \quad \text{para } j = 1, \dots, m \quad \text{restrições de desigualdade} \end{array}$$

Onde $g_i(x) = c_i$ para $i = 1, \dots, n$ e $h_j(x) \geq d_j$ para $j = 1, \dots, m$ são restrições que devem ser satisfeitas—também chamadas de restrições *rígidas*—e $f(x)$ é a função objetivo que precisa ser minimizada sujeita às restrições.

Em alguns problemas com base em restrições, a função objetivo é a soma das funções de restrição, e cada uma penaliza a extensão (se houver) em que uma restrição *flexível* (uma restrição de preferência, mas não obrigatória) é violada.

Para a otimização baseada em restrições, as restrições *rígidas* devem ser integralmente satisfeitas e as restrições *flexíveis* não precisam necessariamente serem satisfeitas por completo. Contudo, as restrições *flexíveis* agregam qualidade à solução. Quanto maior o grau de resolução das restrições *flexíveis*, maior será a qualidade da solução, pois este tipo de restrição leva em consideração preferências dos usuários e executores de atividades.

2.6.3 Otimização de múltiplos critérios

Em geral, em um processo de tomada de decisão é necessário atingir múltiplos objetivos, como por exemplo, aumentar a produtividade e reduzir custos. Porém, como é possível atingir múltiplos objetivos? Deve-se organizar os critérios por meio de pesos conforme o domínio da aplicação.

Os problemas de decisão com múltiplos e conflitantes objetivos ou critérios é conhecido como Método de tomada de decisão multicritério (MCDM). Os métodos multicritério abrangem problemas discretos (conjunto finito de alternativas, também chamados de ações ou soluções) e problemas contínuos (otimização multiobjetivo). Problemas multiobjetivo e multicritério são definidos para dar suporte à tomada de decisão quanto ao processo colocado em prática para encontrar as soluções que melhor se ajustam às preferências dos decisores (FONSECA; FLEMING, 2004).

Justificativa: o método MCDM é utilizado, nesta pesquisa, para apoiar e qualificar a seleção de alternativas. A decisão de cada decisor influencia o processo de escolha de alternativas conforme as preferências pré-definidas. Dessa forma é possível privilegiar, por exemplo, um critério de utilização de espaços e tempo de espera em fila de atendimento de pacientes. Pesos são atribuídos a cada critério para contribuir com o mecanismo de seleção de alternativas.

Os métodos MCDM são classificados como um conjunto de ferramentas para padronizar o processo de tomada de decisão com base na modelagem matemática. Eles ajudam nas tomadas de decisões *vis-à-vis* eleição de alternativas e critérios conflitantes (BRANS; MARESCHAL, 2005). A organização e ordenação das

alternativas são definidas por um especialista do domínio de conhecimento. E os métodos MCDM permitem levar em conta as preferências do decisor.

As alternativas não dominadas também são chamadas de soluções eficientes, ou seja, quando uma alternativa é melhor em um critério, e nenhuma outra é melhor do que ela, considera-se que a solução não é dominada por nenhuma outra e outras informações podem auxiliar no processo decisório (BRANS; MARESCHAL, 2005), a saber: a) decisão que consiste na escolha de uma opção em detrimento de outra; b) uma função de valor que agrega todos os critérios em uma única função, obtendo uma solução para um problema monocritério para o qual há uma solução ótima; c) pesos relativos à importância de cada critério; d) preferências associadas a avaliação por pares entre critérios; e e) fixação de limites de preferências.

Apesar de existirem diversos métodos MCDM na literatura, como por exemplo *AHP (Analytic Hierarchy Process)*, *Promethee (Preference Ranking Organization Method for Enrichment Evaluations)*, *ELECTRE (Elimination and Choice Translating algorithm)*, dentre outros, não é possível comprovar qual é o melhor método a ser utilizado. É necessário compreender o contexto do processo de tomada de decisão a fim de escolher o método que responda satisfatoriamente a tomada de decisão. Dentre os métodos multicritério, o TOPSIS considera que a solução ideal positiva é a solução que maximiza os critérios de benefício e minimiza os critérios de custo. Esta característica foi determinante na escolha do método TOPSIS utilizado nesta pesquisa.

O método multicritério TOPSIS consiste na ordenação das preferências por similaridade e tem como princípio de que a alternativa mais adequada é aquela que mais se aproxima da solução ideal e que mais se afasta da solução não desejada. A métrica utilizada para a ordenação é o coeficiente de similaridade para mensurar a semelhança entre as alternativas (HWANG; YOON, 1981).

Justificativa: O TOPSIS empregado, nesta pesquisa, utiliza a aplicação simplificada da ordenação advinda das medidas de desempenho considerando cada solução candidata juntamente com os pesos e escalas fornecidas pelo tomador de decisões, não sendo necessária a comparação par a par (BOTTANI; RIZZI, 2006).

As etapas do TOPSIS descritas conforme (JUNIOR; CARPINETTI, 2015), são:

Etapas 1: Definir a matriz de decisão: Dada uma matriz de decisão D , a qual corresponde o desempenho das alternativas e um vetor de pesos W . A matriz D

descreve as alternativas avaliadas A_i , os critérios C_j , os números das alternativas $i = 1, 2, \dots, n$, o número de critérios de decisão $j = 1, 2, \dots, m$. O vetor de pesos w_j de cada critério C_j compõem W de modo a satisfazer $\sum_{i=1}^m w_j = 1$.

$$D = \begin{matrix} & C_1 & C_2 & \dots & C_j & \dots & C_m \\ A_1 & d_{11} & d_{12} & \dots & d_{1j} & \dots & d_{1m} \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ A_i & d_{i1} & d_{i2} & \dots & d_{ij} & \dots & d_{im} \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ A_n & d_{n1} & d_{n2} & \dots & d_{nj} & \dots & d_{nm} \end{matrix}$$

$$W = [w_1, w_2, \dots, w_n]$$

Etapa 2: Normalizar e ponderar a matriz de decisão D : Os elementos da matriz normalizada $N = [n_{ij}]$ são definidos cf. Eq. 5.

$$n_{ij} = w_j d_{ij} / \sqrt{\sum_{j=1}^n w_j d_{ij}^2} \quad \text{Eq. 5}$$

Etapa 3: Determinar as soluções: Determinar a solução ideal positiva (PIS, A^+) e a solução ideal negativa (NIS, A^-) cf. as equações Eq. 6 e Eq. 7.

$$A^+ = \{MAX_j n_{ij} \mid j = 1, 2, \dots, m\} = \{n_1^+, \dots, n_j^+, \dots, n_m^+\} \quad \text{Eq. 6}$$

$$A^- = \{MIN_j n_{ij} \mid j = 1, 2, \dots, m\} = \{n_1^-, \dots, n_j^-, \dots, n_m^-\} \quad \text{Eq. 7}$$

Etapa 4: Calcular as distâncias entre soluções: Para cada alternativa avaliada, calcula-se a distância D_i^+ entre os valores de desempenho normalizados e ponderados da matriz N e os valores da solução ideal positiva cf. Eq. 8. O cálculo da distância D_i^- com as pontuações das alternativas da matriz N e os valores da solução ideal negativa cf. Eq. 9.

$$D_i^+ = \sqrt{\sum_{i=1}^n (n_{ij} - n_j^+)^2} \quad \text{Eq. 8}$$

$$D_i^- = \sqrt{\sum_{i=1}^n (n_{ij} - n_j^-)^2} \quad \text{Eq. 9}$$

Etapa 5: Calcular o coeficiente de aproximação: A pontuação final ou desempenho global da alternativa i é definida pela Eq. 10.

$$CC_i = \frac{D_i^-}{(D_i^+ + D_i^-)} \quad \text{Eq. 10}$$

Etapa 6: Classificar alternativas: as alternativas são classificadas em ordem decrescente conforme os valores do coeficiente de aproximação, definido no intervalo $[0, 1]$. As alternativas em que o desempenho global é próximo de 1 são consideradas as melhores.

Outra abordagem empregada neste projeto é CBR (**C**ase **B**ased **R**easoning), cuja principal iniciativa é manter e reusar experiências passadas para resolver problemas similares a tais experiências.

2.6.4 Raciocínio Baseado em Casos (CBR)

O CBR encerra a ideia da resolução de problemas se utilizando de experiências passadas, denominadas de casos. Em outras palavras, a abordagem CBR tem por objetivo fomentar a resolução de um novo problema a partir da recuperação de casos similares, de uma memória de casos, e adaptação desses casos recuperados para compor a solução do problema em questão (RIESBECK; SCHANK, 1989). Essa abordagem é a mesma que as pessoas utilizam, tanto para resolver problemas quanto para tomar decisões, em particular, na medida que as pessoas têm facilidade de raciocinar por meio de analogias (KOLODNER, 1993).

Justificativa: O raciocínio baseado em casos permite que a etapa de otimização do framework PM4SOS utilize uma base de conhecimento existente para reduzir o espaço de busca de soluções. Os atributos/valores armazenados na base de casos são comparados com o desempenho de indicadores obtidos em conjunto com a simulação do estado presente e somados às preferências estabelecidas previamente pelos gestores. Essa estratégia permite estabelecer um refinamento mais preciso sobre os casos recuperados/selecionados que vão, na próxima etapa,

fazer parte do processo de adaptação. Os mecanismos da abordagem CBR permitem complementar o mecanismo de otimização de agendamento cirúrgico; a formalização de cada tarefa da abordagem CBR—recuperação, reutilização, revisão e retenção—pode ser implementada por meio de regras.

O mecanismo CBR pode ser classificado em duas categorias: a) **estático**: há dependência do *designer* do processo e este deve ter antecipadamente informações que descrevem os casos. Estes casos são estruturados segundo um modelo e representados no formato de uma lista de atributos e valores. b) **dinâmico**: lidam com casos temporais, em que a análise de casos semelhantes é em tempo real. Os casos evoluem dinamicamente ao longo do tempo e são tratados como situações dinâmicas (EL MOKHTAR; ZOUHAIR, 2016).

Na Figura 11, apresenta-se o ciclo de vida do CBR e suas principais tarefas, que são: recuperação, reutilização, revisão e retenção. A partir da descrição de um dado problema, as experiências passadas são recuperadas a partir de uma base de casos. O caso recuperado é reusado depois de adaptado segundo as regras de adaptação de casos visando compor uma solução ao problema atual—ou informado como entrada. A solução proposta é revisada: caso ocorra uma falha durante a revisão, gera-se uma explicação da falha, gera-se uma reparação para a falha e reavalia-se se a reparação estabelecida foi efetiva; esse processo de revisão, explicação, reparação e reavaliação se executa até se obter uma solução aceitável. Em outras palavras, se o teste de aceitação da solução proposta for positivo *vis-à-vis* as métricas estabelecidas, a solução proposta é aplicada e armazenada na base de casos—ou memória de casos.

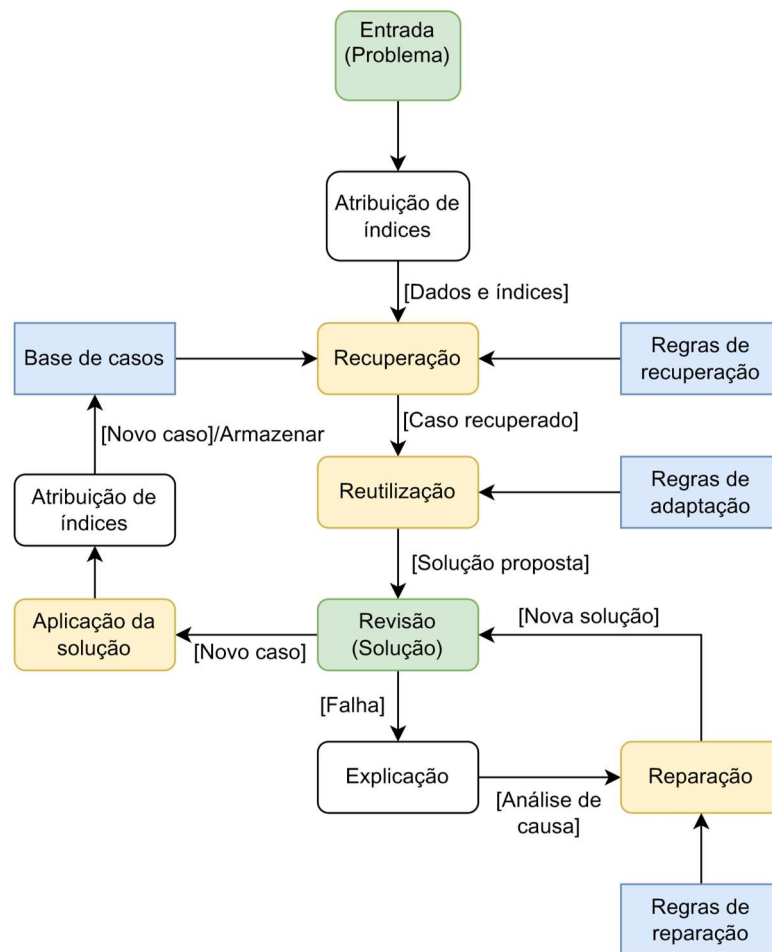


Figura 11 – ciclo de vida do CBR – adaptada de (RIESBECK; SCHANK, 1989)

O raciocínio baseado em caso está associado ao apoio à tomada de decisão e a gestão da memória de experiências passadas; essas últimas são utilizadas como uma base de cenários para possíveis soluções futuras.

2.6.4.1 Representação dos casos

A intuição do raciocínio baseado em casos é que as situações se repetem com regularidade e o que foi feito em uma situação passada provavelmente será feito em uma situação semelhante no futuro (KOLODNER, 1993).

A aquisição do conhecimento por meio de casos consiste em indexar os atributos relevantes da descrição do problema presentes na base de casos. Esta tarefa pode ser automatizada com uso de aprendizagem de máquina para inferir regras com dados minerados em sistemas de informação (GUO; HU; PENG, 2011).

Justificativa: Esta pesquisa utiliza casos representados e armazenados em uma memória de casos, onde é incorporado um conjunto de características para cada caso armazenado para permitir a sua indexação. Seguindo (SÀNCHEZ-

MARRÈ, 1994), as características em questão são: a) identificador; b) descrição; c) diagnóstico; d) solução; e) derivação; e f) resultado da solução.

2.6.4.2 *Recuperação de casos*

A recuperação consiste em encontrar na base de casos registros similares ao novo problema que se deseja resolver, cuja descrição é fornecida como entrada do mecanismo de recuperação. A comparação de atributos/índices da descrição do novo problema com os índices dos casos registrados na base de casos é feita se utilizando de regras específicas de recuperação os casos com vistas a recuperar os casos mais semelhantes à descrição do problema em questão. Pode-se também recuperar casos por meio da aplicação de métricas/cálculos de similaridade.

Pode-se usar, por exemplo, o cálculo do vizinho mais próximo (ou KNN) com a ponderação de atributos ou distância Euclidiana para recuperar os casos mais semelhantes. O procedimento de análise de similaridade pode considerar, a partir dos atributos de um dado caso, os seguintes passos: a) verificar a existência da característica de interesse na base de casos; b) comparar os valores de cada característica de interesse e calcular o grau de similaridade; c) somar o coeficiente da importância do atributo para a similaridade (KOLODNER, 1993).

A tarefa de indexação de casos é um elemento crítico, na medida que ela permite que, a partir de características que representem os casos, sejam estabelecidos os índices para melhorar a precisão da seleção/recuperação de casos por meio do cálculo de similaridade. Em outras palavras, os índices facilitam a identificação dos casos e reúnem as características/atributos mais relevantes. Esses atributos formam as informações indexadas. Porém, os atributos que não foram usados para recuperar casos são chamados de informações não indexadas (BARBALETTA; MARK, 1988).

Justificativa: Esta pesquisa emprega o algoritmo do vizinho mais próximo e a distância Euclidiana para o cálculo de similaridade entre casos. A indexação de casos leva em conta os atributos específicos das descrições de casos e as preferências dos especialistas, previamente definidas para refinar a recuperação de casos; as preferências são também levadas em conta nas decisões multicritério.

Seja a descrição de um caso com t atributos $c[a_1...a_t]$. Para calcular a similaridade global entre dois casos, primeiro são especificados os atributos das

similaridades locais. Normalmente, a similaridade global pode ser calculada como a agregação das similaridades locais (MELVILLE; SINDHWANI, 2010).

$$Sim(c^I, c^R) = \frac{\sum_{i=1}^n w_i \times sim(a_i^I, a_i^R)}{\sum_{i=1}^n w_i}, \quad \text{Eq. 11}$$

Na Eq. 11, a_i^I, a_i^R são valores dos atributos de entrada c^I e atributos recuperados c^R respectivamente. A similaridade local é dada por $sim(a_i^I, a_i^R)$ usada para computar a similaridade entre o novo caso e armazenados na base de casos. O fator de ponderação é dado peso $w_i \in [0, 1]$. Os fatores de ponderação podem ser especificados por expressões de preferência (MELVILLE; SINDHWANI, 2010).

2.6.4.3 Reuso de casos

O reuso de casos concerne a tarefa mais desafiadora, na medida que se requer a adaptação, por exemplo, do caso recuperado mais similar ao problema que se deseja resolver. Há várias iniciativas de adaptação que decorrem da característica do problema em questão. Pode-se ter a necessidade da adaptação da estrutura e/ou dos valores dos atributos do caso. A adaptação de valores dos atributos é normalmente executada com vistas a propor uma nova solução e que tal solução seja ótima. Aqui, frequentemente se utiliza derivar uma adaptação ótima de um caso com a aplicação de técnicas evolucionárias, por exemplo, algoritmo genético (BORGES, 2015) e evolução diferencial (SOUZA, 2018).

Justificativa: A abordagem genética foi usada nesta pesquisa para adaptar casos recuperados da base de casos. Usou-se essa abordagem em particular para adaptar casos cuja descrição envolve restrições que precisam ser satisfeitas. Pode-se citar como exemplo a otimização das restrições para estabelecer o agendamento de recursos em torno de um centro cirúrgico; nesse exemplo, a estratégia consiste na busca de soluções sub ótimas para organizar o agendamento de cirurgias envolvendo diferentes recursos como salas, procedimentos, datas, horários, cirurgiões etc. A estratégia de adaptação pode usar algoritmo genético para encontrar combinações de agendamentos que atendam às restrições *rígidas*, ou seja, que precisam ser satisfeitas peremptoriamente.

Os algoritmos evolucionários usam modelos computacionais dos processos naturais de evolução como ferramenta para resolver problemas. Embora exista uma

variedade de modelos computacionais propostos, eles têm em comum o conceito de simulação da evolução das espécies por meio de processos de seleção, mutação e reprodução. Estes processos que dependem do desempenho dos indivíduos da espécie em estudo dentro do ambiente (LINDEN, 2008).

Na Figura 12, mostra-se uma classificação dos algoritmos evolucionários como técnicas de busca. Eles fazem parte de uma categoria denominada *Técnicas aleatórias-guiadas*, ou seja, contém componentes aleatórios, mas usam as informações do estado corrente para determinar o próximo estado. O algoritmo de resfriamento simulado também pertence à categoria *aleatória-guiada* (LINDEN, 2008). A seguir descrevem-se os algoritmos genéticos com vistas a adaptação e reuso de casos.

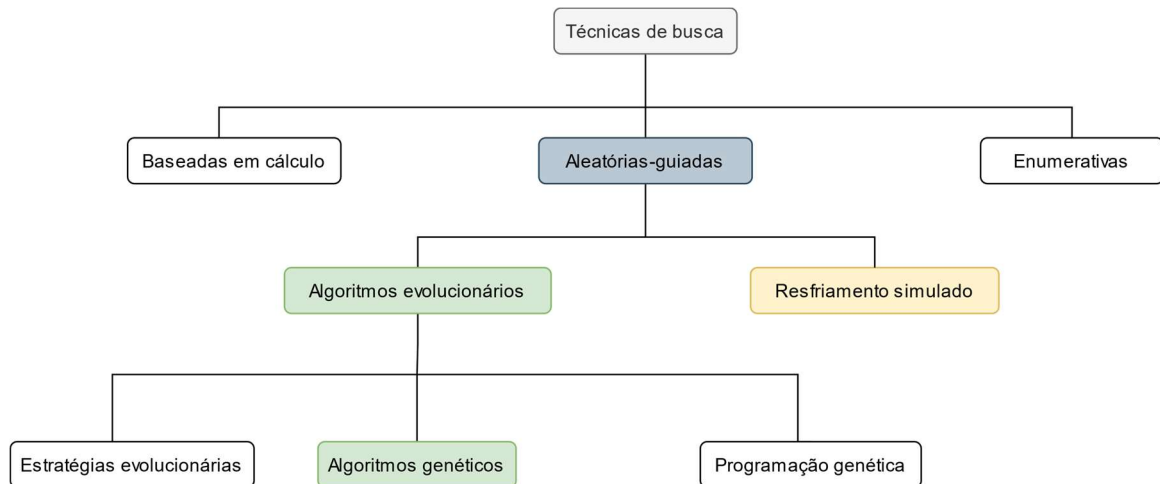


Figura 12 – diagrama das técnicas de busca – (LINDEN, 2008)

Algoritmos genéticos

Os algoritmos genéticos usam estratégias de busca que empregam escolhas randômicas guiadas para explorar o espaço de busca visando obter a otimização global para um problema. Esta estratégia é análoga à evolução biológica, onde os organismos com maior habilidade sobrevivem no ambiente e são determinados pelo seu DNA. Os operadores utilizados em algoritmos genéticos são seleção, cruzamento e mutação em sistemas artificiais. Esses operadores contém estratégias para solução de problemas de otimização (YANG, 2014).

O algoritmo genético é representado por (GEN; CHENG, 1997):

Algoritmo 1 – Algoritmo genético

```

Begin
  t ← 0;           {contador de gerações}
  inicializar P(t); {população inicial}
  avaliar P(t);   {função de avaliação da aptidão}
  while (condição de parada não encontrada) do:
    selecionar indivíduos da população P(t) para reproduzir
    cruzar P(t) para produzir C(t) com probabilidade pc;
    mutar P(t) para produzir C(t), probabilidade pm;
    avaliar C(t);
    t ← t + 1;     {próxima geração}
  endwhile
end

```

O Algoritmo 1 mantém a população de indivíduos $P(t)$ por t gerações. Cada indivíduo representa uma solução potencial para o problema. Cada indivíduo é avaliado para calcular a sua aptidão. Alguns indivíduos passam por transformações estocásticas com operadores genéticos. As duas transformações são: recombinação que cria indivíduos combinando partes de dois cromossomos e mutação que cria indivíduos trocando partes do cromossomo de cada indivíduo. Os novos indivíduos $C(t)$ são avaliados. Uma nova população é formada a partir da seleção dos indivíduos, da população anterior, com as maiores aptidões. Depois de algumas gerações, o algoritmo converge para o ótimo individual, o qual representa uma solução ótima ou sub ótima para o problema (GEN; CHENG, 1997).

Os algoritmos genéticos são utilizados para compor os algoritmos de otimização em problemas não lineares (YANG, 2014).

O uso de algoritmos genéticos pode ser aplicado a uma série de problemas computacionais como a otimização multiobjetivo. Para executar a otimização, vários objetivos devem ser analisados, pode-se aplicar algoritmos genéticos que lidem com a questão de múltiplos objetivos. O *Archive-Based Micro-Genetic Algorithm* (AMGA2) proposto por (TIWARI; FADEL; DEB, 2011) lida com esta questão.

O algoritmo AMGA2 pertence à classe de micro algoritmos genéticos proposto por (COELLO; PULIDO, 2001). A ideia desta abordagem é a segmentação em duas memórias: a) **Memória populacional**: usada como entrada para manter a diversidade. Uma parte da memória populacional é substituível e a outra não substituível, e b) **Memória externa**: usada em conjunto com o conjunto de Pareto ótimo. A intenção da abordagem é reduzir o número de comparações executadas

por métodos tradicionais de algoritmos evolucionários multiobjetivo para manter a diversidade e reduzir o custo computacional (COELLO; PULIDO, 2001).

O algoritmo AMGA2 é uma evolução do algoritmo AMGA proposto por (TIWARI et al., 2008) e melhora vários conceitos de algoritmos de otimização multiobjetivo existentes. As principais modificações consideram as técnicas de avaliação de diversidade existente e operadores de variação genética. O algoritmo é considerado de estado estacionário porque sua principal frente de Pareto tem um número pequeno de soluções, embora outras boas soluções sejam mantidas armazenadas em um arquivo. Para produzir a próxima geração de uma população, o algoritmo usa todas as soluções da frente de Pareto principal combinadas com soluções do arquivo (TIWARI; FADEL; DEB, 2011).

Para decidir quais soluções incluir na nova frente de Pareto, usa-se dois critérios: a) o grau de **dominância** da solução; e b) a **diversidade** de soluções. Assim, dois objetivos são alcançados, ou seja, um pequeno número de avaliações de funções e alta diversidade. As boas soluções que não são selecionadas para a nova frente de Pareto são incluídas no arquivo. Para manter o arquivo, as soluções que lotam uma região específica do espaço de soluções são eliminadas usando a estratégia de busca do vizinho mais próximo (TIWARI; FADEL; DEB, 2011).

Os recursos mais importantes do AMGA2 são (TIWARI; FADEL; DEB, 2011): a) permitir melhoria independente de cada solução; b) utilizar população dinâmica e pequena; c) utilizar arquivo externo para armazenar as melhores soluções; d) definir número de soluções desejadas antes de iniciar execução do algoritmo de otimização; e e) usar seleção dinâmica adaptável do operador de cruzamento.

Justificativa: O algoritmo AMGA2 aborda alguns dos pontos fracos de outros algoritmos evolucionários multiobjetivo. Por exemplo, para cada novo indivíduo, as restrições são avaliadas antes de fazer uma avaliação desnecessária da função objetivo. AMGA2 foi utilizado neste trabalho porque o seu desempenho no estudo feito por (MARQUEZ-CALVO, 2020), mostrou-se ser um dos algoritmos mais eficientes. Outro estudo comparativo entre algoritmos genéticos híbridos foi realizado por (BADALLÓ et al., 2013), em que se mostrou que o algoritmo AMGA híbrido permitiu encontrar soluções definidas próximo da frente de Pareto ótimo para a maioria dos problemas examinados. Neste contexto, o algoritmo AMGA foi

escolhido por tratar adequadamente da dominância e diversidade das soluções, bem como preservar um arquivo de soluções.

Deve-se notar que adaptação de abordagens híbridas que combinam diferentes conceitos de várias técnicas em aplicações de inteligência artificial e técnicas de gestão operacional está presente desde muito tempo (KOBACZY; VADERA; RASMY, 2007). Esses autores relataram a tendência de se usar algoritmos genéticos para realizar busca eficientes visando solucionar problemas de agendamento e, complementarmente, outras técnicas como redes neurais, lógica difusa ou *simulated annealing*. A seguir será descrito o funcionamento do algoritmo *simulated annealing*.

2.6.4.4 Revisão de casos

A revisão de um caso consiste em duas etapas: a) **avaliação**, colocada em prática para avaliar o resultado da etapa de adaptação e verificar se os resultados são satisfatórios conforme os índices das preferências pré-definidas pelos gestores; b) **reparação**, que coloca em prática a tarefa de reparação da solução quando não se atinge um valor determinado de qualidade ou eficiência; essa etapa é usada para encontrar uma solução que agregue qualidade na solução inicialmente adaptada.

A reparação de solução, em nossa proposta, é colocada em prática pela abordagem a seguir.

Simulated annealing

O mecanismo de otimização com base no recozimento simulado ou “*simulated annealing*” (SA) foi proposto por (KIRKPATRICK; GELATT; VECCHI, 1983). SA consiste em combinar o comportamento de um sistema térmico com temperatura finita e a otimização multivariada ou combinatória, tendo como objetivo, a partir de diversos parâmetros, encontrar o mínimo de uma determinada função. Este algoritmo de otimização explora a analogia entre o resfriamento de um metal até o congelamento em uma estrutura de energia mínima. O algoritmo SA foi formalizado por (AARTS; KORST; MICHIELS, 2005), sua descrição se encontra no Algoritmo 2.

No Algoritmo 2, a busca começa com uma solução inicial aleatória e uma dada temperatura inicial ($T = T_0$). Para cada interação, uma solução na vizinhança é obtida. A função objetivo da nova solução é comparada com a solução atual. Se a solução obtida é melhor, ela substitui a solução atual, e se for pior, a solução atual é

substituída por uma probabilidade. A função de probabilidade é $e^{-\Delta/T}$, onde Δ é a diferença entre as soluções e T um parâmetro de controle de temperatura. No início, quando a temperatura é alta, mais movimentos são aceitos e quando a temperatura se aproxima de zero, os movimentos que aumentam a função custo são rejeitados.

Algoritmo 2 – Algoritmo simulated annealing

```

procedure SA ( $f(\cdot), N(\cdot), S_{Max}, T_0, s$ )
   $s^* \leftarrow s$ ;           {Melhor solução obtida até então}
   $IterT \leftarrow 0$       {Número de iterações na Temperatura T}
   $T \leftarrow T_0$        {Temperatura corrente}
  while ( $T > 0.0001$ )
    while ( $IterT < S_{Max}$ ) do
       $IterT \leftarrow IterT + 1$ ;
      Gerar um vizinho ( $s'$ ) aleatoriamente na vizinhança  $N^k(s)$ 
       $\Delta = f(s') - f(s)$ 
      if ( $\Delta < 0$ ) then
         $s \leftarrow s'$ 
        if ( $f(s') < f(s^*)$ ) then
           $s^* \leftarrow s'$ 
        else
          Para  $x \in [0,1]$ 
          if ( $x < e^{-\Delta/T}$ ) then
             $s = s'$ 
          endif
        endif
      endwhile
       $T \leftarrow T x$ ;
       $IterT = 0$ 
    endwhile
  return  $s^*$ 
endprocedure

```

A busca por soluções nos vizinhos mais próximos permite encontrar e aceitar soluções “melhores”, porém também se aceita soluções piores. A ideia é utilizar uma função de probabilidade que reduza a ocorrência de soluções piores enquanto a temperatura vai decrescendo. A medida quantitativa que representa quão boa é uma solução é uma função custo (KIRKPATRICK; GELATT; VECCHI, 1983).

De forma diversa a outros algoritmos como “subida da encosta”, no qual a solução fica presa a mínimos locais, o SA a partir de uma temperatura alta T evita que a solução fique vinculada a um mínimo local.

Justificativa: A heurística do algoritmo de SA permite escapar dos ótimos locais. Os algoritmos genéticos e SA têm diferentes pontos fortes e fracos: Algoritmos genéticos podem percorrer no espaço de busca e encontrar uma solução razoável, mas muitas vezes não consegue produzir uma solução ótima em tempo

razoável. SA trabalha continuamente em uma única solução e não pesquisa o suficiente no espaço de busca para obter uma visão geral de todas as áreas de solução promissoras. Quando usados juntos, eles podem produzir resultados promissores (ELHADDAD; SALLABI, 2010). O framework proposto, nesta pesquisa, utiliza algoritmos genéticos e SA em conjunto para usufruir do benefício de cada algoritmo e mitigar suas deficiências.

Algumas pesquisas apontam o uso integrado de heurísticas para otimização de sistemas como descrito em (ŁAWRYNOWICZ, 2011). Neste estudo, várias aplicações de heurísticas são apresentadas para agendamento de tarefas industriais e aplicadas ao problema de planejamento e agendamento. O sistema híbrido proposto por (ŁAWRYNOWICZ, 2011) não oferecem apenas planejamento e programação da produção de curto prazo para atender às mudanças nos requisitos de mercado, que podem utilizar de forma mais adequada a capacidade disponível de um sistema de fabricação, mas também fornece suporte para controle operacional.

O algoritmo híbrido definido por (JARDZIOCH; SKOBIEJ, 2016), aborda o escalonamento de tarefas em máquinas de capacidade limitada. O estudo atende ao planejamento e agendamento de tarefas e considera dois critérios de otimização: tempo máximo de execução das tarefas e produtividade de cada processo.

As heurísticas de forma geral facilitam o processo de tomada de decisão pois possibilitam a estruturação de mecanismos para reduzir o tempo de planejamento e execução de tarefas árduas. A automatização de tarefas rotineiras é planejada e programada com apoio de heurísticas e dados presentes em sistemas de informação.

A seguir apresenta-se a retenção de casos na base de casos.

2.6.4.5 Retenção de casos

Um novo caso precisa ser indexado e armazenado apropriadamente na memória para uso futuro. A descrição de um caso compreende a descrição do problema, sua solução e outras características necessárias para o sistema de raciocínio. Um dos pontos a ser considerado no armazenamento do caso é a indexação para permitir acesso às estruturas de dados posteriormente (KOLODNER, 1993). Para (AAMODT; PLAZA, 2001), a retenção de um caso encerra o produto do episódio de resolução do problema mais recente e incorporado à memória de casos.

A visão de retenção acomoda uma perspectiva mais ampla do que os meios para armazenamento de um sistema CBR aprender com sua experiência de resolução de problemas. A retenção de casos está fortemente ligada a questões relacionadas ao gerenciamento de múltiplas formas de representar e recuperar o conhecimento em sistemas CBR (MÁNTARAS et al., 2005).

2.7 Relacionamento dos conhecimentos

Nesta seção será detalhado o relacionamento entre as áreas de mineração de processos, simulação computacional, otimização e tomada de decisão aplicadas ao problema de gerenciamento de filas para o planejamento e agendamento.

De forma geral, a mineração de processos fornece os meios para descobrir, a partir de um log de eventos, modelos de processos reais, detectar desvios de processos normativos e analisar gargalos. A mineração de processos permite analisar eventos passados, ou seja, dados históricos. Complementarmente, a simulação pode ser usada para explorar diferentes alternativas de *projeto* e antecipar futuros problemas de desempenho. Como já dito, Van der Aalst qualifica a relação entre a mineração de processos e a simulação como “*A combinação perfeita feita nos céus!*”.

A mineração de processos pode ser usada para facilitar a criação de modelos de simulação e comparar execuções de comportamentos simulados com comportamentos reais. A simulação, aliada a mineração de processos, possibilita explorar diferentes mudanças de um processo (VAN DER AALST, 2018).

A mineração de processos permite criar modelos de simulação mais confiáveis com base em dados de eventos reais. Além disso, a simulação pode ser transformada em uma ferramenta ponderosa para apoiar a tomada de decisão operacional, usando dados de processos em tempo real (VAN DER AALST, 2015).

A combinação de simulação de eventos discretos e algoritmo genético foi usada para definir a melhor política de estoque em sistemas de multiproduto. A hibridização de meta heurística e técnica de simulação é promissora, e apresenta soluções eficientes a nível de otimização de inventário, levando em conta componentes estocásticos, funções objetivo e restrições a nível de otimização (JACKSON; TOLUJEVS; REGGELIN, 2018).

A relação entre otimização e configuração de filas considerou períodos de pico para a chegada de pacientes em filas. Dois modelos foram criados: o primeiro com filas comuns e o segundo com filas rápidas. Cálculos do tempo de espera entre esses modelos foram realizados para a análise de desempenho de serviço. O método em questão indicou valores de referência à prática gerencial (LI; PAN; CHENG, 2018).

A relação entre otimização e MCDM foi colocada em prática como um modelo gerado por algoritmo genético para obter informações referentes à tomada de decisão. O modelo pode ser usado para ajudar tomadores de decisões, com diferentes interesses e de forma iterativa, a alcançarem um acordo sobre como classificar critérios de forma individual, refletindo as preferências a nível individual e coletivo (ÁLVAREZ et al., 2015).

Deve-se notar que o uso de técnicas de otimização com algoritmos evolucionários segue o paradigma de “aprender enquanto otimiza”. Elas permitem a criação de modelos robustos para identificar uma solução satisfatória (BATTITI; PASSERINI, 2010); esse modo de aprender permite reduzir o esforço cognitivo necessário pelo tomador de decisão para buscar uma solução.

A solução de problemas de agendamento pode utilizar uma abordagem em duas etapas. Na primeira etapa, a solução é interativamente construída com uso de heurísticas sequenciais e posteriormente a solução é gradualmente aprimorada com uso de meta-heurísticas para melhorar a solução inicial (PETROVIC; YANG; DROR, 2004). Esta abordagem também foi implementada por (THOMPSON; DOWSLAND, 1996) com uso de heurísticas para encontrar uma solução viável e posteriormente aplicar a estratégia SA no segundo estágio para lidar com as restrições *suaves*. Deve-se notar que essas abordagens antigas não são o foco da nossa proposta, elas representam apenas complementos a aplicação da mineração de processo e simulação para compor uma solução de agendamento.

2.8 Resumo do Capítulo

Esse capítulo apresentou uma breve descrição das técnicas empregadas, em diferentes profundidades, na construção do objeto principal desta tese, que pode ser resumido por meio da seguinte pergunta: Como os dados de eventos presentes em sistemas de informação podem dar suporte ao processo decisório de gestão

operacional com baixo esforço cognitivo e tempo reduzido? Dentre as técnicas, duas delas assumem maior destaque: simulação computacional e mineração de processo.

Em nosso contexto, a simulação é a técnica que permite imitar um processo de negócio real. E cada modelo de simulação pode ser construído sobre um dado paradigma de simulação, a saber: simulação baseada em agentes, simulação de eventos discretos e simulação de sistemas dinâmicos. O ciclo de vida da simulação compreende a concepção e a implementação de diferentes subprocessos: formulação e implementação de modelo, verificação e validação, experimentação e análise, bem como a redefinição do processo.

A construção de um modelo simulação requer a obtenção das distribuições estatísticas que descrevem o comportamento do sistema alvo. Classicamente, a etapa de construção do modelo de simulação envolve esforços humanos e de tempo. No entanto, este processo pode ser automatizado evitando erros operacionais e reduzindo esforços na medida que a extração de parâmetros pode ser feita diretamente dos eventos de logs provenientes de sistemas de informação. Em outras palavras, a partir de um log de eventos, a mineração de processos permite obter, de forma facilitada, os parâmetros necessários para criar um modelo de simulação. Tal facilidade requer que seja colocado em prática um processo de extração de dados.

A mineração de processos pode ser *off-line* para análise de dados históricos e *online* para análise de instâncias em execução. Destas análises podem ser colocadas em prática tarefas de suporte operacional voltadas à detecção de desvios, previsão de tempos e recomendação de ações. Essa última requer, em geral, a priorização das ações recomendadas em função de critérios definidos pelo gestor do processo analisado. E tal diversidade de critérios pode ser tratada por métodos multicritério, o que permite usufruir de base matemática para estabelecer concordância entre as partes envolvidas com vistas a ação de maior peso/relevância na tomada de decisão.

O uso integrado da mineração de processos e simulação permite criar aplicações para apoiar tomadas de decisão mais assertivas. A mineração de processos é a fonte principal de parâmetros/estimativas para criar um modelo de simulação relativo ao processo de negócio descoberto. Em resumo, o binômio mineração de processos e simulação permite ampliar a análise de um processo com um “*olhar à frente*”.

Na Tabela 2, mostra-se uma síntese dos assuntos abordados neste capítulo e a aplicação no Framework proposto.

Tabela 2 – tabela resumo da fundamentação teórica

Assunto	Framework proposto	Descrição
Problema de agendamento	Agendamento de cirurgias	Recomendação de ações operacionais.
Indicador de desempenho	Gatilho de otimização	Detecção de desvios para executar a otimização
Teoria das Filas	Simulação de Eventos Discretos	Execução da simulação para um modelo de processo de negócio.
Fonte de dados	Log de eventos	Extração de parâmetros/estimativas e modelo de processos.
Mineração de processos	Modelo de processo	Extração de modelo de processos e obtenção de parâmetros para criar o modelo de simulação.
Simulação computacional	Modelo de simulação e experimentação.	Criação do modelo de simulação com dados extraídos da mineração de processos.
Método multicritério	Tomada de decisão	Definição de pesos para cada critério e geração de <i>ranking</i> de critérios para recuperação de casos.
Raciocínio Baseado em Casos.	Otimização de processo.	Reutilização do conhecimento relativo a um modelo de processo.

Fonte: o autor 2022

No próximo capítulo, apresentaremos uma revisão sistemática da literatura com foco na interseção entre mineração de processos e simulação de curto prazo, visando fornecer suporte à tomada de decisões operacionais. Esta revisão abrangerá estudos recentes que investigam a utilização de técnicas de mineração de processos para extrair informações relevantes de dados históricos, bem como as abordagens que utilizam a simulação para antecipar possíveis cenários futuros e avaliar diferentes estratégias operacionais. Além disso, serão discutidos os principais desafios e oportunidades dessa abordagem combinada, bem como as lacunas existentes na literatura que merecem atenção futura.

3 REVISÃO SISTEMÁTICA DE LITERATURA

As palavras só têm sentido
se nos ajudam a ver o mundo melhor
(Rubem Alves)

Este capítulo é dedicado à revisão sistemática da literatura com o foco no binômio mineração de processos e simulação de curto prazo para suporte à tomada de decisão operacional.

Analisando a mineração de processos para a simulação de curto prazo na tomada de decisão operacional: uma revisão sistemática de literatura

RESUMO

A mineração de processos pode ser combinada com a simulação computacional. A mineração de processos tem foco na análise de logs de eventos históricos ou passados, e a simulação permite obter uma visão à frente, por exemplo, para recomendar cenários e antecipar comportamentos de instâncias de processos. Uma contribuição deste artigo remete à compreensão de como os modelos de simulação são gerados para executar tarefas de suporte operacional (e.g., predição e recomendação), e mais particularmente envolvendo a simulação de curto prazo com suporte da mineração de processos. Assim, este artigo apresenta uma revisão sistemática com foco na simulação de curto prazo combinada com a mineração de processos. Os critérios metodológicos colocados em prática exploraram essencialmente fontes de dados digitais. As fases empreendidas para a seleção dos artigos foram: identificação, triagem, elegibilidade e inclusão dos estudos analisados. Dentre o conjunto de 223 artigos recuperados a partir da aplicação de um conjunto de critérios de busca estabelecidos nesta revisão sistemática, 10 deles tratam especificamente de simulação de curto prazo e mineração de processos. A análise de cada um destes artigos permitiu identificar lacunas, oportunidades e tendências envolvendo o binômio simulação de curto prazo e mineração de processos. Identificou-se uma tendência quanto a utilização em conjunto *machine learning* e *process mining* para apoiar a criação de modelo de simulação. Constatou-se também que o paradigma de simulação mais frequentemente colocado em prática

foi a simulação de eventos discretos. Percebeu-se um crescimento das arquiteturas híbridas visando equilibrar benefícios e desvantagens de diferentes técnicas (e.g., *Agent-Based Simulation, Discrete-Event Simulation, System Dynamic Simulation*).

3.1 Introdução

Em ambientes corporativos, o processo de tomada de decisão envolve a avaliação de alternativas com intuito de compreender uma situação problema para uma tomada de decisão mais assertiva. Tal processo de tomada de decisão é contínuo e pode envolver diversos setores de uma organização. E cada tomada de decisão deve contribuir para a manutenção da sustentabilidade do negócio e do adequado funcionamento tanto a nível administrativo quanto a nível operacional. Neste contexto dinâmico é imprescindível que o gestor seja munido de subsídios eficientes quanto ao suporte à tomada de decisão. Sabe-se que as decisões de curto prazo podem impactar, significativamente, nas decisões estratégicas de longo prazo e a previsibilidade de desfechos permite antecipar possíveis problemas operacionais evitando desperdícios, por exemplo, de tempo e de recursos.

Os sistemas computacionais permitem obter informações rapidamente sobre uma situação operacional de uma empresa. Em outras palavras, os sistemas de informação associados aos processos de negócio de uma empresa permitem obter informações sobre o estado corrente e dados históricos de cada processo, que em conjunto permitem colocar em prática o suporte operacional para as tomadas de decisões. A simulação é a ferramenta que contribui para a criação de cenários de situações futuras com base em dados atuais e históricos. E em sua especificidade, a simulação de curto prazo explora informações que estão prontamente disponíveis, e preferencialmente, de forma automatizada. Assim, as decisões que afetam os processos de negócio em futuro próximo podem ser avaliadas sem a necessidade de grandes esforços adicionais de modelagem. O armazenamento estruturado de informações em sistemas de informação empresariais permite a criação relativamente simples de modelos de simulação (REIJERS; AALST, 1999).

A mineração de processos é o ponto de partida para a construção do suporte operacional de curto prazo, pois viabiliza a simulação de curto prazo com a extração de parâmetros de dados históricos de um processo de negócio. A ideia básica da mineração de processos é descobrir, monitorar e melhorar processos de negócio

reais mediante a extração de padrões de comportamento a partir de logs de eventos prontamente disponíveis em sistemas de informação (VAN DER AALST, 2005).

A relevância da mineração de processos decorre, por exemplo, do fato que o *design* do modelo de processo nem sempre retrata o processo real, pois cada instância em execução está sujeita a influência de fatores tanto previsíveis (e.g., ajustes da capacidade produtiva) quanto imprevisíveis (e.g., falhas de equipamentos). Adicionalmente, em um ambiente flexível, em que, pelo menos, parte dos recursos decidem o próximo passo no processo, a execução desse processo também está sujeita à influência de fatores parcialmente previsíveis, como as preferências exageradas por certos recursos em detrimentos de outros. Tal flexibilidade pode gerar muitas variantes de um processo de negócio, dificultando tanto a descoberta de um modelo compreensível—intuitivo e correto—de instâncias de processo quanto a sua reprodução por meio de um modelo de simulação. Embora não seja objeto de pesquisa desta tese, há várias abordagens que se utilizam, em particular, de técnicas de agrupamento de traços de processos como etapa preparatória à tarefa de descoberta de modelo de instâncias de processo mais compreensível (e.g., *k-means*, *Quality Threshold*, *Agglomerative hierarchical*, *Self-organizing map*) (SONG; GÜNTHER; VAN DER AALST, 2009).

A simulação é uma ferramenta em potencial para a melhoria da tomada de decisão. Nessa linha, o controle de processo, o suporte à tomada de decisão e o planejamento proativo são sustentados em (ABUKHOUSA et al., 2014). Este coloca em evidência os esforços necessários para a criação de modelos de simulação para o suporte à tomada de decisão na área da saúde e com enfoque no gerenciamento da cadeia de suprimentos. Em geral, sem um suporte adequado, a construção de bons modelos de simulação e otimização requerem mais esforços e experiência do modelador—ou do especialista—que ele pode realizar ou maior investimento que uma empresa tem disponível para colocar em prática. Assim, como também a construção de modelos de simulação de curto prazo demanda esforços significativos, torna-se necessário avaliar a efetividade dessa técnica para dar suporte à tomada de decisão de curto prazo; dado que o modelo de simulação pode rapidamente tornar-se obsoleto e não refletir o estado atual do processo.

O foco desta revisão da literatura está centrado no binômio mineração de processo e simulação computacional, que, segundo van der Aalst (VAN DER AALST, 2018), representa “*A Match Made in Heaven!*”. Em outras palavras, a

descoberta de um modelo de instâncias de processo, em relação a dados históricos, revela padrões comportamentais ocorridos no passado e permite ter uma melhor compreensão do presente, e a simulação, se utilizando dessa informação de situação presente, viabiliza a geração de cenários, bem como a construção de uma base informacional para planejar e prever o futuro (LIU, 2015). Tal sinergia se dá da seguinte forma: informações são extraídas do modelo de instâncias de processo minerado e usadas para calibrar modelos de simulação e potenciais comportamentos, provendo uma visão “*fast forward*”. Essa visão consiste em prever comportamentos futuros de um sistema e explorar de forma antecipada vários cenários (WYNN et al., 2010).

A simulação computacional aplicada a um processo de negócio, mesmo tendo vários benefícios, apresenta uma lacuna, na medida em que seu foco está geralmente voltado ao *design* de processo (ROZINAT et al., 2008b). Esse foco não está alinhado com as necessidades do gestor de negócio no tocante à geração de informações relativas ao apoio à tomada de decisão operacional, que naturalmente é de curto prazo, e conseqüentemente, requer que a simulação de cenários do negócio em questão seja também de curto prazo. A natureza dessa simulação de curto prazo não permite, de um lado, que sua preparação seja longa e, de outro lado, que sua consecução demande um esforço demasiado. Ainda, tal simulação de curto prazo deve gerar resultados altamente aderentes e compreensíveis *vis-à-vis* ao domínio de aplicação.

Nessa linha, uma forma de operacionalizar a simulação de curto prazo é aproveitando-se do resultado da tarefa de descoberta de modelo de instâncias de processo e do modelo de *workflow* em operação como fontes de informações para alimentar em tempo real um sistema de simulação de curto prazo. Taxa de chegada de cada traço, tempo de execução de cada caso e custo de cada instância de processo são exemplos de informações extraídas de um processo descoberto. O modelo organizacional, modelo de referência, recursos—pessoas e máquinas—e estado corrente constituem exemplos de informações extraídas do ambiente de *workflow*. Essas informações contribuem para que a simulação não se inicie no estágio zero, i.e., sem atividades em execução (ROZINAT et al., 2008a), e conseqüentemente, seja mais eficiente, i.e., a consecução do suporte à tomada de decisão de nível operacional deve ocorrer em um tempo próximo dos fatos.

3.2 Objetivos da pesquisa

Esta pesquisa tem como objetivo compreender como a mineração de processos tem sido aplicada para apoiar a simulação de curto prazo. Nesse contexto, seis questões de pesquisa foram respondidas por meio de uma pesquisa exploratória:

QP1: como são criados os modelos de simulação a partir da mineração de processos? Essa questão visa identificar de que forma são obtidos os modelos de simulação tendo como entrada logs de eventos.

QP2: quais dados da mineração de processos são usados para executar a simulação de curto prazo? Esta questão visa identificar os dados obtidos a partir da mineração de processos para a simulação de curto prazo.

QP3: como o estado corrente é extraído e quais dados são usados na simulação de curto prazo? Esta questão visa identificar o mecanismo de extração e os dados do estado corrente usados na simulação de curto prazo. Deve-se também salientar que um olhar cuidadoso foi estabelecido com relação ao uso de *workflow* e de logs de eventos na composição do ambiente de simulação de curto prazo para fornecer o estado corrente de um processo.

QP4: como são tratados os recursos na simulação de curto prazo? Esta questão visa identificar como a modelagem é realizada e quais dados da estrutura organizacional e recursos são utilizados na simulação de curto prazo.

QP5: como a mineração de processos fornece subsídios para o suporte à tomada de decisão operacional? Esta questão visa identificar qual é o suporte efetivo que a mineração de processos fornece à simulação de curto prazo com vista à tomada de decisão operacional.

QP6: quais contribuições, limitações e estudos futuros foram delineados em cada pesquisa? Essa questão visa identificar como as pesquisas estão sendo feitas para fomentar o binômio mineração de processos e simulação de curto prazo.

O restante do texto apresenta, na Seção 4.3, os procedimentos metodológicos e a estrutura detalhada da pesquisa realizada. Na Seção 4.4, apresentam-se, de forma estruturada, os resultados e as principais descobertas da pesquisa bibliográfica. Na Seção 4.5, discutem-se os resultados obtidos, bem como enunciadas algumas ponderações. Na Seção 4.6, apresentam-se as conclusões, limitações e oportunidades.

A seguir será descrita a metodologia e procedimento de análise empregado nesta pesquisa, bem como os critérios definidos com base em perguntas de pesquisa, as quais visam compreender o efeito da relação entre simulação e mineração de processos, em especial, a respeito da simulação de curto prazo.

3.3 Protocolo de pesquisa

O caráter da pesquisa é bibliográfico e visa selecionar, mapear e discutir trabalhos e conceitos que perfazem a intersecção entre as áreas de mineração de processos e simulação, especificamente simulação de curto prazo. Este trabalho é descritivo e visa também apontar lacunas e encontrar oportunidades de investigação. Primeiramente, realizou-se uma pesquisa exploratória para delimitar a abrangência do tema. Os artigos selecionados para esta etapa foram: (REIJERS; AALST, 1999), (WYNN et al., 2008), (VAN DER AALST et al., 2010), (VAN DER AALST, 2015), (VAN DER AALST, 2018), (ROZINAT et al., 2008a), (ROZINAT et al., 2008b) e (ROZINAT et al., 2008c).

Selecionaram-se as fontes de artigos para compor esta revisão sistemática com base nos seguintes critérios: a) permitir pesquisa por título, abstract e palavras-chave; b) possibilitar a aplicação de filtros: por período, idioma, tipo de publicação e c) permitir exportar resultados para formatos RIS ou *BibTex*.

Com base nos critérios acima, selecionaram-se as bases de dados digitais: ACM Digital, IEEE Xplore, Science Direct, Scopus *data base* e Web of Science. A *string* de busca aplicada foi: ((*"process* mining"* OR *"workflow* mining"* OR *"mining workflow*"*) AND *"simulation*"* AND *"decision"*). Na Tabela 3, apresentam-se as quantidades de trabalhos obtidos para cada fonte de pesquisa selecionada após a aplicação da *string* de busca. A técnica de *snowballing*, também, foi aplicada com a mesma *string* de busca inicial para pesquisa no Google Scholar e sete artigos foram obtidos por *snowballing*. A pesquisa nas bases de dados está atualizada até 2022.

Tabela 3 – resultados obtidos para cada fonte de dados

Fonte de busca	Total de resultados
ACM Digital	2
IEEE Xplore	12
Science Direct	7
Scopus database	67
Web of Science	29
Total de referências	117

Fonte: o autor 2022

Na Figura 13, apresentam-se as fases dessa revisão sistemática: identificação, triagem, elegibilidade e inclusão dos estudos analisados. Cada etapa é descrita a seguir: a) **identificação**: nessa etapa, executaram-se as buscas nas fontes digitais para identificar pesquisas. Removeram-se os trabalhos repetidos. b) **seleção**: nessa etapa realizou-se a triagem dos trabalhos por título, abstract e palavras-chave. c) **elegibilidade**: nesta etapa realizou-se a triagem por meio da leitura completa de cada artigo e aplicação dos critérios de exclusão. d) **inclusão**: o resultado dessa etapa consistiu na obtenção de 10 artigos para responder à pergunta principal de pesquisa e realização da avaliação qualitativa dos artigos.

Lançaram-se as perguntas para a etapa de extração de dados dos artigos:

Q1. **Modelo de simulação** – quais foram as técnicas de simulação, atividades de suporte operacional, formalismos e aspectos de arquitetura usadas para a criação do modelo de simulação?

A resposta de Q1 possibilita relatar como os modelos de simulação foram criados. A identificação das técnicas de simulação que podem ser simples ou híbridas, os formalismos aplicados, os artefatos gerados e as arquiteturas propostas nos trabalhos analisados visam sistematizar a geração dos modelos de simulação. A automatização da geração dos modelos de simulação também é analisada como manual, semiautomática ou automática.

Q2. **Mineração de processos** – quais dados históricos foram considerados (e.g., taxa de chegada, tempo de execução dos processos) e como eles foram usados para a criação do modelo de simulação?

A resposta de Q2 descreve como a mineração de processos é executada e quais dados presentes nos *logs* de eventos são necessários para a criação dos modelos de simulação. Dentre as perspectivas da mineração de processos, a descoberta do modelo de processo é essencial para a identificação das etapas que o modelo de simulação percorrerá.

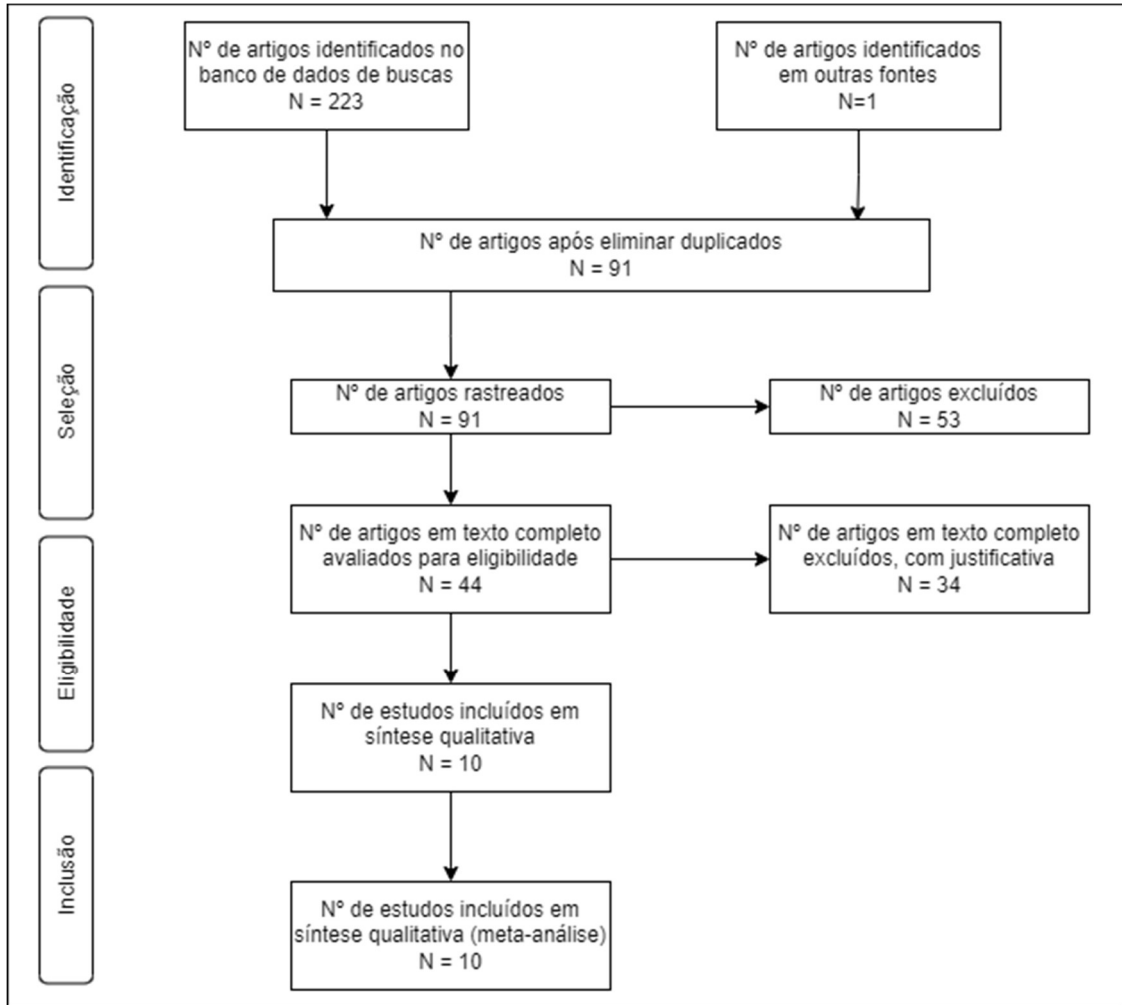


Figura 13 – fluxo da seleção de estudos para a revisão – o autor 2022.

Q3. **Estado corrente** – quais dados foram considerados do estado corrente (e.g., casos em execução, progresso dos casos) e como eles foram usados na criação do modelo de simulação?

O estado corrente dos processos é definido como o estado inicial do modelo de simulação. A resposta de Q3 permite identificar os dados necessários presentes nos logs de eventos para serem inseridos no início da simulação. Os dados são vitais, pois a simulação de curto prazo necessita lidar com processos em execução.

Q4. **Uso de recursos** – como foi abordado o uso de recursos e quais foram os dados obtidos?

A utilização dos recursos (e.g., recursos ocupados), durante a simulação, permite identificar lacunas envolvendo o estado presente e seu impacto futuro. A resposta de Q4 permite analisar como os recursos foram usados em ambientes que utilizem a combinação da mineração de processos e da simulação.

Q5. Suporte à tomada de decisão operacional – como o binômio mineração de processos e simulação propiciou a tomada de decisão operacional?

A resposta de QP5 descreve a obtenção de suporte à tomada de decisão com a aplicação conjunta de mineração de processos e simulação. As análises devem refletir o uso de KPIs e os mecanismos necessários para provimento de apoio à tomada de decisão com base em dados históricos e com ênfase em curto prazo.

Q6. Contribuições, limitações e trabalhos futuros – quais foram as principais contribuições, limitações e trabalhos futuros informados pelos autores?

A resposta de QP6 visa evidenciar as principais contribuições e limitações das abordagens estudadas nesta revisão sistemática. Também se destacam as oportunidades vislumbradas pelos autores das pesquisas. Essa questão projeta desenvolvimento de arquiteturas e aplicações futuras com a combinação da mineração de processos e simulação.

Os critérios de inclusão estabelecidos foram os seguintes:

- IC1. pesquisas avaliadas por pares;
- IC2. pesquisas de estudos empíricos;
- IC3. pesquisas publicadas em língua inglesa;
- IC4. pesquisas publicadas entre 2002 e junho de 2022. Em 2002 foi publicado o primeiro algoritmo de mineração de processos, Alpha Miner;
- IC5. pesquisas que consideram as abordagens mineração de processos e simulação voltadas à tomada de decisão operacional.

Os critérios de exclusão estabelecidos foram os seguintes:

- EC1. estejam duplicadas;
- EC2. não estejam em versão final de publicação;
- EC3. secundárias como *surveys*, mapeamentos, teses, dissertações;
- EC4. publicadas em meios diferentes de *Journals or Proceedings*;
- EC5. com enfoque introdutório, teórico ou tendência;
- EC6. com enfoque de mineração diferentes de processos de negócio;
- EC7. pesquisas com enfoque de simulação espacial, baseada em jogos e abordagens diferentes de processos de negócio;
- EC8. sem estabelecer relação entre mineração de processo e simulação;
- EC9. pesquisas que não elucidem o uso da mineração de processos e simulação de curto prazo aplicadas à tomada de decisão operacional.

A seguir são examinados os principais resultados dessa revisão sistemática.

3.4 Descobertas e perspectivas: Suporte à tomada de decisão operacional

Na Tabela 4, apresenta-se iniciativas emergentes, referente a artigos selecionados na fase de elegibilidade, e organizados em ordem cronológica.

Tabela 4 – estudos selecionados em ordem cronológica

Estudo	Título	Autores
S1	Workflow simulation for operational decision support using design, historic and state information	(ROZINAT et al., 2008a)
S2	Workflow simulation for operational decision support using event graph through process mining	(LIU, 2015)
S3	Discrete Modeling and Simulation of Business Processes Using Event Logs	(KHODYREV; POPOVA, 2014)
S4	Evaluation of discovered clinical pathways using process mining and joint agent-based discrete-event simulation	(AUGUSTO et al., 2016)
S5	Holistic Modeling of Chronic Diseases for Recommendation Elaboration and Decision Making	(BALAKHONTCEVA et al., 2018)
S6	Is there a Need for Agent-based Modeling and Simulation in Business Process Management?	(HALAŠKA; ŠPERKA, 2018)
S7	A Hybrid Process Mining Framework for Automated Simulation Modeling for Healthcare	(MESABBAH; MCKEEVER, 2018)
S8	An Evidence-Based Decision Support Framework for Clinician Medical Scheduling	(CHO et al., 2019)
S9	Scenario-based prediction of business processes using system dynamics	(POURBAFRANI; VAN ZELST; VAN DER AALST, 2019)
S10	Automated discovery of business process simulation models from event logs	(CAMARGO; DUMAS; GONZÁLEZ-ROJAS, 2020)

Fonte: o autor 2022

Nas próximas seções, analisam-se os elementos de interesse extraídos de cada artigo selecionado. Cada questão de pesquisa é respondida, de maneira que se apresenta uma visão do contexto de uso do estado atual do fluxo de trabalho e da mineração de processos em relação à criação de um modelo de simulação de curto prazo para o gerenciamento de recursos.

3.4.1 Modelo de simulação

QP1 – Quais as técnicas de simulação, atividades de suporte operacional, formalismos e aspectos estruturais foram usadas para criar o modelo de simulação?

A simulação atende a vários objetivos e aplicações em conjunto com a mineração de processos. Predominantemente, o uso da simulação está vinculado à criação do modelo de simulação e ao suporte à tomada de decisão, assim como a predição de tempos e comportamentos. Na Tabela 5, reúnem-se as técnicas de simulação empregadas para a criação de modelos de simulação.

Tabela 5 – técnicas de simulação – trabalhos e respectivos autores

Técnica de simulação	N.º trabalhos	Percentual	Autores
Eventos discretos	5	50%	S1, S2, S3, S8, S10
Simulação híbrida	3	30%	S4, S5, S7
Baseada em agentes	1	10%	S6
Sistemas dinâmicos	1	10%	S9

Fonte: o autor 2022

Observa-se que há uma preferência pela simulação discreta de eventos (MIELCZAREK, 2016). Em seguida, a hibridização entre técnicas ABS e DES é evidenciada como uma tendência, visto que ao combinar as técnicas consideram-se os benefícios de cada uma delas. A ABS foi reportada, apenas, na pesquisa S6. Ao utilizar ABS, é possível avaliar a simulação em nível operacional, essa esta técnica baseou-se em agentes e abrange o nível estratégico, tático e operacional (BORSHCHEV; FILIPPOV, 2004). Por outro lado, a simulação SD, também, foi reportada na pesquisa S9, sendo utilizada em nível estratégico, o que permite uma análise sistêmica de um processo.

A seguir são descritas as tarefas de apoio operacional (Tabela 6). Computaram-se mais de uma tarefa por trabalho, por exemplo, detecção e predição.

Tabela 6 – atividades de apoio operacional – trabalhos e respectivos autores

Atividade operacional	N.º trabalhos	Percentual	Autores
Predição	8	80%	S1,S2,S3,S4,S5,S6,S7,S8,S9,S10
Detecção	4	40%	S6,S8,S10
Recomendação	1	10%	S5

Fonte: o autor 2022

Os dados apresentados na Tabela 6 evidenciam que a combinação de mineração de processos e simulação é usada em 80% das pesquisas analisadas para a predição de comportamentos. O resultado reforça o papel da simulação, que é prover um “olhar à frente”. A atividade de detecção de desvios é identificada em 40% das pesquisas, evidenciando a importância em verificar anormalidades que ocorrem na execução dos processos. A recomendação é encontrada em, apenas, uma pesquisa. Para a recomendação assertiva de cenários, é importante acoplar um método de tomada de decisão com o intuito de validar os dados inseridos no modelo de simulação com formalismos matemáticos de suporte à decisão.

Para a QP1, identificaram-se vários formalismos para a criação de modelos de simulação. Na sequência, descrevem-se os principais elementos das arquiteturas e dos experimentos colocados em prática nos artigos analisados.

Tabela 7 – formalismos de modelagem, trabalhos e respectivos autores

Formalismo	Nº trabalhos	Percentual	Autores
Petri Nets	2	20%	S1, S3
Event Graphs	2	20%	S2, S5
Data-Based	2	20%	S7, S8
State Charts	1	10%	S4
BPMN	1	10%	S10
System Dynamics	1	10%	S9
Agent-Based	1	10%	S6

Fonte: o autor 2022

Os formalismos identificados na Tabela 7 respondem à questão de pesquisa QP1. Podem-se destacar os trabalhos baseado em: (a) Petri Nets que permite transpor o modelo matemático para o modelo de simulação; (b) *Event Graphs* que permite operacionalizar o modelo de simulação por meio de um conjunto de *threads* aplicadas a um modelo de filas; e (c) *Data-Based* que permite descobrir modelos de processos e aplicar aprendizagem de máquina para potencializar os modelos de simulação. Além disso, identificaram-se, com menor frequência, outros formalismos utilizados para a criação de modelos de simulação: *State Charts*, *BPMN*, *System Dynamics* e *Agent-Based*.

3.4.2 Dados da mineração de processos

A questão QP2 busca saber quais dados históricos foram considerados e como eles foram usados na criação do modelo de simulação. A análise para responder a essa questão colocou em evidência, em particular, os casos e as atividades de um processo, bem como a taxa de chegada de casos, o tempo médio de duração das atividades e os recursos que executam as atividades. Os dados históricos são usados para: a) descobrir modelos de processos; b) estabelecer as funções de transição das atividades; c) obter distribuições de probabilidade e modelos preditivos; e d) identificar o comportamento dos recursos em um processo.

Na Tabela 8, resume-se a tarefa realizada sobre os dados históricos e as respectivas informações obtidas como resultado.

Tabela 8 – dados da mineração de processo

Autores	Informação obtida	Tarefa Executada
S1	Distribuição de frequências, tempos de execução; Taxa de chegada de <i>cases</i> ; Disponibilidade de recursos.	Extração de log de eventos do workflow (MXML) para efetuar <i>merge</i> entre <i>design</i> , dados de eventos e estado corrente do <i>workflow</i> .
S2	Número de <i>cases</i> ; Taxa de chegada de <i>cases</i> ; Número de gerentes e funcionários.	Construção do modelo de processo; Mineração de condições de limite; Estabelecimento de funções de transição;
S3	Nome do evento; Identificador da instância do processo; Tipo de evento; Originadores do evento. Valores de variáveis de contexto; <i>Timestamp</i> de eventos;	Extração de logs de eventos (XES); Extração de modelo de instâncias de processo (Rede de Petri).
S4	Causal Nets.	Filragem das principais classes de estados; Descoberta de modelo de instâncias de processo (rede causal).
S5	Sequência de eventos; Atributos e restrições; Relacionamentos de causa e efeito; Restrições e condições;	Extração de logs de eventos. Cada paciente é modelado como um agente.
S6	Dados de pacientes e processos. CaseID, activity, timestamp e resource	Extração de logs de eventos (ERP to XES).
S7	ID do objeto, nome da atividade, ID da instância, estado da atividade, ID do recurso, <i>Timestamp do evento</i> , idade do paciente; sexo do paciente, histórico de doenças do paciente.	Extração do comportamento de recursos; Extração de dados de treinamento para modelo preditivo.
S8	Taxa de chegada de pacientes; Tempo de serviço (somente tarefa completa); Capacidade de atendimento.	Extração de dados para criação de modelo de simulação, usando algoritmo de <i>frequency mining</i> ; Extração de probabilidade de transição.
S9	Taxa de chegada de <i>cases</i> ; Uso de recursos; Tempo ocupado e livre de recursos; Tempo de execução por <i>case</i> e por evento; Duração dos <i>cases</i> e tempos de espera;	Transformação de logs de eventos em logs de sistemas dinâmicos (SD); Identificação de padrões que alimentam os parâmetros de <i>stock</i> e <i>flow</i> de um SD, usados como parâmetros de simulação.
S10	Identificador de <i>case</i> ; Recurso que executou a atividade; Tempo de início e término de atividade.	Extração individualizada de parâmetros, dado que o esforço de descoberta deve estar na precisão do parâmetro.

Fonte: o autor 2022

3.4.3 Dados do estado corrente

A questão QP3 busca identificar quais dados do estado corrente (e.g., casos em execução, progresso dos casos) foram considerados e como eles foram usados na criação do modelo de simulação. Na Tabela 9, sintetizam-se os contextos de uso do estado corrente, mapeando fonte de dados, tarefa executada e informação obtida.

Tabela 9 – dados do estado corrente

Autores	Fonte de dados	Informação obtida	Tarefa Executada
S1	Workflow <i>online</i>	Estado de progresso dos casos (marcadores de estado); casos em execução; tempos de execução dos casos; data e hora do estado corrente; dados associados a cada caso; informações de recursos;	Extração dos dados do estado corrente dos <i>cases</i> ; criação do modelo de simulação a partir do estado corrente para o modelo CPN Tools; inserção do estado inicial ao modelo de simulação;
S2	Workflow <i>online</i>	<i>tags</i> attribute e originator;	extração do estado de cada variável
S3	Log de eventos	<i>cases</i> em execução (<i>status</i> dos últimos registros do log de eventos).	Iniciação da simulação com os Últimos <i>traces</i> do log de eventos;
S4	Fluxo de mensagens	estados (hospitalizado ou aguardando transferência); Custos; <i>Wait state</i> (i.e., estado do <i>pathway</i> do paciente);	extração do estado atual de cada agente/paciente, usando troca de mensagens ponto-a-ponto;
S5	Log de eventos	transições de estado e estado corrente de cada processo;	extração do log de eventos: estado corrente e modelo de simulação com base em cenários de tratamento de pacientes; aplicação de métodos evolucionários para recomendação; análise de instâncias de processo para recomendação; caracterização do estado corrente permite considerar irregularidades na execução de eventos e dados faltantes;
S6	Fluxo de mensagens	estado do agente;	extração do estado corrente de cada agente/vendedor do ERP;
S7	Fluxo de mensagens	estado do agente (com contexto); recursos usados;	extração de dados de atualização, em dados intervalos de tempos, para o modelo DES (contexto de execução <i>live-stream</i>); adição de agentes em tempo de execução da simulação.
S8	Log de eventos	estimativa de tempo de trabalho;	extração de estimativas de <i>logs</i> de eventos;
S9	Log de eventos	número de <i>cases</i> e de eventos ativos;	extração dos estados dos <i>cases</i> e dos recursos de <i>logs</i> de eventos;
S10	Fluxo de mensagens	identificador do caso; recurso que ativou o caso; nome da atividade.	mapeamento de fluxos de sequência com valores booleanos (se há um <i>token</i> no fluxo de sequência no estado atual, ele é considerado estado corrente).

Fonte: o autor 2022.

A análise para responder QP3 permitiu colocar em evidência que o estado corrente foi extraído em 40% a partir de *logs* de eventos, 40% a partir de fluxos de mensagens e 20% a partir de *workflows online*. As informações obtidas, relativas ao estado corrente de um processo, foram, de forma resumida, a identificação do *case* e da atividade, o estado atual do processo, o recurso que originou o *case*, bem como o estado do agente quando a abordagem usada era ABS. A principal tarefa executada usando o estado corrente diz respeito à inserção do estado corrente ao início da simulação. Nas pesquisas que utilizaram os *logs* de eventos como fonte de

dados para a obtenção do estado corrente, utilizou-se o estado incompleto dos últimos casos dos traços como ponto de partida da simulação.

3.4.4 Gerenciamento de recursos

A questão QP4 permitiu colocar em destaque como abordou-se o uso dos recursos, e quais foram os dados obtidos. (e.g., recursos ocupados).

Em S1, utilizou-se o modelo organizacional para enriquecer o modelo de simulação. Os dados de interesse estão ligados aos processos em execução e recursos em uso. E, no contexto da simulação, enfocou-se a geração de diferentes cenários, variando as quantidades de recursos e analisando os seus efeitos exibidos nos *logs* de eventos.

Em S2, o interesse centrou-se na contagem de eventos originados por cada recurso; nesta tese, destaca-se que a alocação desses recursos seguiu a ordem de chegada de cada caso. As variáveis do sistema foram categorizadas em variáveis do tipo recurso e não recurso. Para as variáveis do tipo não recurso, realizou-se análise estatística e com base nos resultados obtidos desta análise, determinaram-se as variáveis que compõem o estado corrente do processo. Para as variáveis do tipo recurso, extraíram-se os originadores de eventos e estes permitiram calcular o número total de recursos usados em diferentes eventos. Com base nesses dados, os recursos são mapeados de acordo com os seus diferentes papéis.

Em S6, utilizaram-se os recursos para analisar os efeitos —distribuídos aleatoriamente— do comportamento individual/pessoal nos resultados da simulação. As mudanças no comportamento dos recursos, com base em estatísticas, permitiram identificar efeitos significativos tanto em relação ao desempenho dos processos como também às decisões tomadas pelos indivíduos.

Em S7, o interesse centrou-se nos dados relativos às disponibilidades de recursos. Obtiveram-se as características e os comportamentos dos recursos com base nos *logs* de eventos. O *framework* proposto (ASMB-H) apresenta uma abordagem de predição de desempenho, baseada em técnicas de aprendizagem de máquina. O ASMB-H combina simulação por blocos de atividades com geração automática do modelo de simulação e atualização do *stream* de dados de pacientes (MESABBAH; MCKEEVER, 2018).

Em S9, o interesse diz respeito aos dados do tempo necessário e do tempo livre de cada recurso no processo. Com tais dados, definiram-se diferentes políticas

para melhorar o uso dos recursos ou das horas de trabalho. O uso mais eficiente dos recursos — com controle de produtividade — permitiu observar um aumento de produtividade.

3.4.5 Suporte à tomada de decisão operacional

A questão QP5 buscou capturar como o binômio mineração de processos e simulação propiciam a tomada de decisão operacional.

Tanto em S1 como em S2 foram realizados estudos comparativos, confrontando dados históricos reais com dados gerados por meio da simulação. Nesse sentido, pode-se observar a importância de saber qual é a amplitude do "curto prazo", pois o "*delay*" impacta a tomada de decisão e os valores estimados para um dado intervalo de predição. Em S2, avaliou-se a *performance* de um grupo de recursos, observando a chegada de cada caso. E a predição de *performance* foi feita de forma individual por recurso.

Em S4, a tomada de decisão está centrada em um processo de decisão médica, na área de diagnóstico (estado de saúde do paciente). As decisões baseiam-se nas probabilidades extraídas dos dados históricos, relativos à taxa de chegada de pacientes a ser executado no modelo de simulação. Em S5, objetiva-se a obtenção de um modelo preditivo para o estabelecimento de recomendações. Realizou-se o ajustamento da execução do modelo de simulação e de predição com base em um processo de calibragem de modelo. Nesse sentido, estado final do processo e da análise operacional orientam a condução dos modelos preditivos para avaliar o estado real do paciente. A abordagem auxilia a predição do desenvolvimento de doenças e a geração de recomendações ao tomador de decisões.

Em S6, a proposta encerra uma modelagem baseada em agentes. Tal modelagem permite observar que a tomada de decisão, com base na dinâmica comportamental dos agentes, é mais flexível em face ao modelo de eventos discretos. Em S7, a proposta está centrada na investigação do impacto das políticas de ações quanto à *performance* geral de uma unidade de atendimento ao paciente. O modelo construído permite prever o tempo médio de atendimento de cada alta de paciente, bem como o tamanho da fila de espera para um determinado serviço. A alocação de recursos alimenta o cálculo de um conjunto de KPIs que permite o

tomador de decisões identificar em quais dias os recursos serão subutilizados ou escassos. Em S9, a proposta assume a forma de um sistema dinâmico.

3.4.6 Contribuições, lacunas e oportunidades

A questão QP6 buscou identificar quais foram as principais contribuições, limitações e trabalhos futuros informados pelos autores. As principais contribuições encontradas foram: (a) extração de dados de logs de eventos para criar modelos de simulação; (b) criação automática do modelo de simulação; e (c) uso da simulação em conjunto com *machine learning*. As limitações encontradas mais severas foram: (a) generalização das abordagens; (b) dificuldade em lidar com recursos e (c) falta de ferramental para análise da simulação. E, finalmente, os trabalhos futuros mais frequentemente enunciados foram: (a) automatizar a extração de características de logs de eventos; (b) automatizar o processo de tomada de decisão; e (c) melhorar a validação do sistema de simulação. Na Tabela 10, mostram-se as contribuições dos estudos analisados.

Tabela 10 – contribuições informadas pelos autores em suas publicações

Contribuições identificadas	Estudos	Contextos
• Extração de parâmetros de log de eventos para compor modelos de simulação.	S1, S2, S3, S5, S9	O log de eventos permite obter um conjunto de informações tais como: parâmetros de desempenho.
• Criação automática de modelo de simulação.	S1, S2, S7, S10	O modelo de simulação deve representar objetos reais de sistemas. A automatização do processo de criação do modelo de simulação permite reduzir esforços humanos e vieses.
• Simulação para avaliar comportamentos de recursos.	S6, S7, S9	O comportamento do ser humano é difícil de simular dado o seu aspecto subjetivo que as pessoas podem expressar na execução de atividades não mensuráveis, como por exemplo, o ritmo de execução de cada tarefa e a execução simultânea de atividades.
• Inserção do Estado Corrente em modelos de simulação.	S1, S2	O Estado Corrente é essencial para a simulação de curto prazo, na medida que o estado de cada atividade em andamento no processo é tomado como ponto partida do modelo de simulação.
• Utilização de variantes de processos na simulação.	S4	A decisão quanto à utilização de variantes de um processo na composição de modelo de simulação é relevante.
• Utilização da simulação com <i>machine learning</i> para predição.	S5, S7	A experimentação baseada em simulação requer a implementação de um conjunto de ciclos de simulação para a obtenção de dados estatísticos. O uso do aprendizado de máquina permite que os modelos de simulação sejam executados com maior rapidez, pois, a cada ciclo de execução da simulação, os resultados são ajustados com os novos dados gerados.
• Utilização de <i>machine learning</i> para análise de simulação em tempo real.	S7	A simulação possibilita, a partir de entradas bem-definidas, testar cenários. A alimentação de entradas dinâmicas em tempo real permite obter <i>insights</i> ininterruptos para apoiar a tomada de decisão operacional.
• Simulação como suporte à otimização e tomada de decisão.	S8	A combinação de técnicas de simulação e otimização permite o fornecimento um ambiente mais favorável para a utilização dos recursos, obtendo parâmetros de entrada que produzem os resultados esperados para o processo em questão; e (b) apoiar a geração de soluções customizadas para a tomada de decisão.

Fonte: o autor 2022.

As principais contribuições identificadas nos estudos analisados dizem respeito à extração de parâmetros de registros de eventos para alimentar modelos de simulação. Além disso, destacam-se a criação automática de modelos de simulação, a análise do comportamento dos recursos e o uso de aprendizado de máquina para prever eventos de ação futura.

Dentre as limitações identificadas, expõem-se as dificuldades quanto à obtenção do estado corrente de um processo, a possibilidade de viés de especialistas, a falta de ferramentas para integrar ambientes e automatizar tarefas, por exemplo, de descoberta automática de modelo de simulação. O grau de complexidade dos modelos de processos e os mecanismos para tratamento de dados que envolvem recursos, também, são destacados como lacunas nas pesquisas analisadas. Na Tabela 11, apresentam-se as principais lacunas encontradas.

Tabela 11 – limitações informadas pelos autores em suas publicações.

Limitações indicadas	Estudos	Contextos
Falta de generalização da abordagem.	S1, S4, S6, S10	A generalização ocorre pelo caráter geral da abordagem de simulação adotada, com o suporte da mineração de processos, ser aplicada a domínios variados.
Dificuldade em lidar com recursos.	S3, S6, S7, S10	O controle dos comportamentos dos recursos é um desafio devido aos comportamentos subjetivos destes em um ambiente de processos.
Falta de ferramenta de análise da simulação.	S2, S8, S10	O uso de ferramenta adequada facilita o mecanismo de análise da simulação, dado que tal uso pode reduzir esforços e vieses.
Viés do modelo de simulação.	S4, S6, S9	A formalização e a consistência do modelo de simulação são requisitos peremptórios para que a execução da simulação seja fidedigna à realidade.
Dificuldade na representação do espaço de estados e da realidade.	S5, S8, S10	A aleatoriedade característica, referente à representação do espaço de estados, viabiliza o controle de acontecimentos em intervalos de tempo não fixos.
Falta de integração de ferramentas.	S1, S9	A integração de ferramentas é necessária para reduzir os esforços humanos face à execução da simulação.
Grau de complexidade do sistema simulado.	S7	O estabelecimento de níveis de controle da complexidade é necessário para colocar em prática a simulação de processos complexos de forma viável.
Viés do especialista.	S3	O uso de dados reais, obtidos de forma automatizada das instâncias de modelos de processos, evita viés do especialista.
Falta de mecanismo para obtenção do estado corrente	S2	A extração automatizada do estado corrente é essencial na simulação de curto prazo, pois sua correta identificação permite que ele seja utilizado, estrategicamente, como estado inicial da simulação.
Falta de automatização da validação do modelo de simulação.	S9	A tarefa de validação do modelo de simulação é necessária para comparar, objetivamente, modelos reais e simulados com base em medidas estatísticas.

Fonte: o autor 2022

Dentre as oportunidades apontadas pelos autores presentes nesta revisão sistemática, merecem destaque os mecanismos para melhorar a extração de

parâmetros dos logs de eventos. Nesta tese, o requisito é que tal tarefa de extração seja automatizada, bem como a tarefa de construção do modelo de simulação em conjunto com o estado corrente. Essa automatização é essencial para a simulação de curto prazo. Também se destaca a generalização de soluções e a integração de ferramentas para automatizar o processo de tomada de decisão. Na Tabela 12, apresentam-se as principais oportunidades dos estudos analisados.

Tabela 12 – oportunidades informadas pelos autores em suas publicações

Oportunidades identificadas	Estudos	Contextos
Melhorar e automatizar os mecanismos de extração de características de logs de eventos.	S1, S3, S4, S9	Os mecanismos de extração de características de <i>logs</i> de eventos são os pontos de partida essenciais para colocar em prática um processo automatizado de simulação.
Automatizar processo de tomada de decisão.	S4, S8, S10	O processo decisório é complexo por sua natureza, consequentemente, deve-se oferecer subsídios para que se tome decisões mais assertivas, criando diferenciais competitivos para uma organização, como a redução do atraso de uma tomada de decisão.
Melhorar a validação do sistema.	S7, S8, S9	A validação do sistema simulado é condição obrigatória, e sua operacionalização deve colocar em relação dados reais e dados simulados, e automatização dessa validação reduz o tempo e o esforço de ciclo de novas tomadas de decisões.
Criar soluções customizáveis.	S4, S10	A customização de uma solução permite ampliar a aplicação de uma abordagem a diversos domínios.
Enriquecer a obtenção do estado corrente.	S2, S5	O uso de uma estrutura de dados em grafos permite estabelecer os relacionamentos entre objetos de um sistema, onde os estados são os objetos e as arestas as transições. A instância de um processo em um dado tempo, na forma de um grafo, denota o estado corrente.
Melhorar a integração de sistemas baseado em dados.	S5, S7	A departamentalização da informação reduz a complexidade de concepção e construção de sistemas de informação, contudo a obtenção de uma visão que vai além de um departamento requer a integração da informação de vários departamentos. Tal integração é essencial para facilitar a criação de modelos de simulação com esforços reduzidos.
Automatizar a criação do modelo de simulação.	S5, S9	A automatização pode ser parcial ou completa, e ela produz a redução da iteração/dependência do especialista na criação do modelo de simulação, diminuindo o risco de viés.
Criar métodos inteligentes de observação do comportamento dos recursos.	S3, S6	A observação do comportamento de cada recurso constitui um desafio e é essencial para a criação de modelo de simulação inteligente—ou automatizado; nesse sentido, a mineração de processos dispõe de meios para extrair, por exemplo, do <i>log</i> de eventos com informações relativas ao modelo organizacional associado a um processo de negócio.

Fonte: o autor 2022

Posteriormente à descrição das questões de pesquisa, apresentam-se as discussões e relacionamento dos trabalhos analisados. Discutem e posicionam-se questões definidas neste trabalho em relação aos trabalhos analisados. Destacam-se, para a discussão, as técnicas de simulação, a extração de parâmetros e do estado corrente dos *logs* de eventos, a abordagem de uso dos recursos, a

combinação do binômio mineração de processos e simulação, e as contribuições, limitações e oportunidades dos estudos analisados nesta revisão sistemática.

3.5 Discussão

A discussão desta pesquisa está organizada conforme a ordem das questões de pesquisa apresentadas anteriormente. E, em termos de avanço da área de processos de negócios, entendemos que as reflexões mais importantes se referem aos desafios em relação à tomada de decisão operacional, com base na mineração de processo e simulação de curto prazo.

QP1 – quais foram as técnicas de simulação, atividades de suporte operacional, formalismos e aspectos de arquitetura usadas para a criação do modelo de simulação?

As técnicas de simulação e tarefas de suporte operacional possuem relacionamentos significativos entre si (cf. Figura 14). A simulação DES se encontra associada frequentemente a detecção e predição. Associa-se a simulação híbrida, mais frequentemente, à predição de cenários do que à recomendação ações, e observou-se apenas na pesquisa S5. Associa-se a simulação baseada em agentes à predição de desfechos e à detecção de desvios. E os modelos de simulação estão mais presentes nos seguintes domínios de aplicação: saúde (S4, S5, S7 e S8) e comercial (S1, S2 e S6).

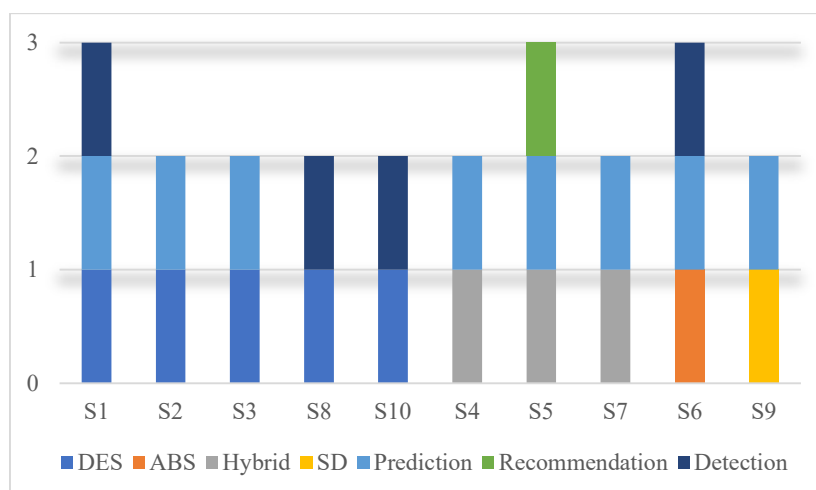


Figura 14 – técnicas de simulação versus suporte operacional – o autor 2022

Nesta revisão observou-se que 30% dos trabalhos utilizaram técnicas de simulação híbridas (S4, S5, S7). Outra constatação refere-se à utilização em conjunto da mineração de processo e da mineração de dados (S4 e S7). Em S4,

utilizou-se a mineração de processos para a obtenção de redes causais e a mineração de dados para os parâmetros das distribuições de probabilidades. Em S7, utilizaram-se os dados históricos para o modelo de um processo de negócio e fundamentação da análise estatística para a detecção de comportamentos dos eventos e seus tempos de processamento. Em S5, elaborou-se um sistema de recomendação para pacientes com doenças crônicas. Dado o exposto — números reduzidos de trabalhos — na Figura 14, novas técnicas e ferramentas integradas são necessárias para facilitar a confecção de modelos de simulação com esforços humanos e de tempo reduzidos.

QP2 – quais dados históricos foram considerados (e.g., taxa de chegada, tempo de execução dos processos) e como eles foram usados para a criação do modelo de simulação?

Partindo dos dados históricos à simulação, a simulação deparou-se com algumas dificuldades adicionais, como a indisponibilidade de procedimentos claros que permitam realizar a configuração e parametrização de um modelo de simulação com esforços reduzidos. A preparação do ambiente para a execução da simulação envolve a definição clara de um objetivo, de um método específico de simulação e de um conjunto de recursos tecnológicos; estes requerem experiência e a ajuda de especialista. Outra dificuldade da simulação refere-se ao fato de que, de forma geral, ela é usada para a análise de cenários de maneira estática, ou seja, os parâmetros iniciais da simulação são inseridos pelo analista. Em outras palavras, o suporte operacional está mais alinhado à simulação online, que inclui algumas particularidades, como a inserção contínua de parâmetros, e a extração de dados de histórico de *logs*.

QP3 – quais dados foram considerados do estado corrente (e.g., casos em execução) e como foram usados para a criação do modelo de simulação?

Assume-se que a simulação de curto prazo é fundamental para o gerenciamento tático de um processo de negócio. E tal suporte à tomada de decisão exige interação contínua com a simulação — como fonte de informação — em relação a cada processo de negócio.

As abordagens baseadas no estado corrente de um processo para alimentar a simulação de curto prazo foram observadas em, apenas, S1 e S2. Nesse sentido, obteve-se o estado corrente diretamente do *workflow* e enriquecido com valores de parâmetros obtidos de registros históricos de eventos. Em S3, S5, S8 e S9, o estado

corrente foi obtido com base no último estado observado no log de eventos. Em S4, S6 e S7, o estado corrente do processo foi obtido com base no estado de cada agente do Sistema, a partir das mensagens trocadas. Especificamente em S7, em intervalos determinados de tempo, os dados do estado corrente, com os atributos de cada agente, são atualizados com vista a alimentar o modelo de simulação em execução. Tal abordagem permite inserir novos agentes durante a execução da simulação. Em S10, o estado corrente é obtido por meio da identificação de *tokens*, observando a sequência do fluxo do processo. Iniciou-se o estado corrente do processo a partir da ocorrência de *tokens* no fluxo de saída das atividades.

Colocou-se a obtenção do estado corrente em prática usando diferentes fontes de dados, que são: logs de eventos (S3, S5, S8 e S9), sequenciamento de *tokens* e mensagens (S4, S6, S7 e S10), em um ambiente de *workflow* em operação. Porém, o estado corrente é inserido uma única vez no modelo de simulação (S1 e S2). Os dados obtidos sobre os recursos associados a um processo limitaram-se à ocupação e flutuação do uso de cada recurso (S1 e S6). Realizou-se a extração desses dados com a ajuda de tarefas específicas de mineração de dados (S2 e S7), guiada pelo estado dos casos e do estado de cada recurso, ambos presentes no *log* de eventos (S9). Em S6, a utilização do estado corrente foi conjugada com o estado de cada agente do Sistema (com contexto), e, em S7, o estado corrente foi usado com objetivo de monitorar KPIs e gerenciar filas de atendimento.

QP4 – como abordou-se o uso de recursos e quais foram os dados obtidos? (e.g., recursos ocupados).

O suporte ao gerenciamento de recursos, com ajuda da simulação e da mineração de processos, está presente em S1, S2 e S3. Tal suporte assenta-se na captura do estado corrente de cada recurso em uso e de todos os eventos gerados com base na taxa de chegada dos casos, assim como na contagem do número de eventos e de alterações na quantidade de recursos presentes na simulação. Adotou-se essa abordagem de gerenciamento (NAKATUMBA; VAN DER AALST, 2010) para a alocação de recursos no serviço público. Os autores destacam a importância da equalização da simulação de curto prazo e da tomada de decisão, visto que ambas influenciam o desempenho do modelo de simulação.

Abordou-se a perspectiva de modelagem de recursos com mais profundidade em S6, cuja abordagem baseou-se na simulação de agentes. Nessa linha, em S7, a

obtenção do comportamento dos agentes baseou-se na aplicação de tarefas de *machine learning*. Após execução da simulação, analisaram-se os recursos por meio de KPIs. Em S2 e S6, realizaram-se estatísticas *f* para obtenção de informações quanto à utilização dos recursos. As distribuições estatísticas permitiram a observação de efeitos positivos no desempenho dos processos e tomada de decisão operacional com base no estado dos agentes.

Observou-se o foco no uso de recursos, também, em S9, em que se definiram dois critérios associados a cada recurso no processo: tempo livre e tempo ocupado. Em S1 e S10, o modelo de simulação foi criado de forma que cada recurso execute, apenas, uma tarefa por vez. Adicionalmente, em S1 argumentou-se que o modelo de simulação, criado com a ferramenta CPN Tools, necessita evoluir para permitir tratar, adequadamente, o comportamento dos recursos. O gerenciamento de recursos, em ambientes de simulação, necessita de mecanismos e formalismos para extrair informações fundamentais para a simulação, como o relacionamento entre recursos disponíveis, a definição de parâmetros de custos e de *performance* associados aos recursos, a fim de mensurar e prever situações que envolvem remanejamento ou eventos esporádicos — catástrofes ou acidentes.

QP5 – como o binômio mineração de processos e simulação propiciaram a tomada de decisão operacional?

Sabe-se que a simulação de curto prazo tem sua base nas atividades de suporte operacional. No entanto, é importante notar que a simulação de curto prazo requer um intervalo de tempo específico para acontecer, conforme as exigências de cada domínio de aplicação; i.e., a recomendação de uma decisão operacional deve ser feita em tempo hábil para a observação do seu efeito. Em S1 e S2, destaca-se a necessidade de integração das ferramentas existentes à medida que se torna complexo executar várias ferramentas em plataformas diferentes para obter resultados operacionais. Essa argumentação é relevante dado que o gestor, de forma geral, não tem conhecimento técnico específico de simulação e seu foco encontra-se na produtividade.

QP6 – quais foram as principais contribuições, limitações e trabalhos futuros informados pelos autores?

Identificaram-se as seguintes limitações: modelagem de recursos (S1), generalização do modelo de simulação (S1 e S4), integração de ferramentas (S1, S2 e S5), dificuldade na alocação de tarefas (S2 e S10), dificuldades na análise e

visualização de resultados obtidos nas simulações (S1, S2 e S3), complexidade elevada dos modelos (S6 e S7), e necessidade de estabelecer padrões de comportamento (S6). As limitações identificadas indicam desafios constantes em criar modelos de simulação baseados em fontes de dados dinâmicas. A integração de sistemas e ferramentas favorece a tomada de decisão operacional. Porém, os trabalhos analisados não contemplam a integração completa dos mecanismos de alimentação de dados de entrada da simulação, da utilização do estado corrente como um fator relevante para a composição da parte preditiva dos processos e dos sistemas de recomendação de ações a serem realizadas com base em critérios específicos para cada domínio de aplicação.

3.6 Conclusão

As decisões operacionais estão relacionadas às operações do cotidiano da empresa. Elas têm um horizonte de curto prazo à medida que as execuções são ações recorrentes. As decisões são baseadas em eventos factuais e não requerem um julgamento significativo. E o controle operacional determina como realizar as tarefas específicas desenhadas pelos tomadores de decisões de nível tático. O desenho do encaminhamento operacional determina as unidades da organização que realizam as tarefas, estabelecem os critérios para conclusão de cada tarefa, utilizam recursos e avaliam as saídas. Deve-se notar que, em geral, todas as unidades demandam decisões de controle operacional.

Os dados gerados com base no modelo de processo são úteis para a criação de modelos de simulação, especialmente simulação de curto prazo, calcular tempos de execução e de espera, bem como gerir a distribuição de casos. É possível também operar cenários “*What if*” e alocar recursos em curto prazo com base nos modelos de simulação. Por outro lado, a mineração de processo pode beneficiar-se do resultado da simulação como um meio de geração de *logs* de eventos confiáveis e reduzido índice de *outliers*.

As deficiências e limitações detectadas com base nessa revisão sistemática estão centradas na modelagem de recursos, na extração dinâmica de dados históricos para a obtenção dos parâmetros para a composição do modelo de simulação e na extração de dados do estado corrente de um processo em execução. Há também um apontamento que encerra demandas relativas à criação de

ferramentas eficientes para apoiar os tomadores de decisão quanto à integração de sistemas.

Outra lacuna identificada nos trabalhos analisados refere-se à utilização de padrões para processos de simulação. Nesse sentido, o *BPSim*⁴ tem por objetivo apoiar a otimização da pré-execução e pós-execução dos referidos modelos de simulação de processos. Uma observação importante refere-se ao fato de que nenhum trabalho analisado usou a padronização *BPSim*. Tal especificação define a parametrização para a troca de dados para a análise da simulação de processos, permitindo análise estrutural e de capacidade dos modelos de simulação baseados em processos.

Deve-se observar que as tarefas de suporte operacional como predição e recomendação encerram valores agregados, e são aderentes às demandas centradas na gestão operacional. A reestruturação de um processo de negócio é uma oportunidade de melhoria nos cenários em que se utiliza a alocação dinâmica de recursos humanos, apoiada pela simulação.

A simulação é um pilar importante da Indústria 4.0, à medida que a economia está sendo digitalizada e logs de eventos estão sendo prontamente disponibilizados pelos sistemas de informações, ou seja, novas demandas devem emergir. Em outras palavras, a simulação e mineração de processo, aliadas à Indústria 4.0, ampliam as possibilidades de geração de informações úteis e rápidas em um ambiente com intensas mudanças e necessidades de entregar resultados.

Finalmente, o uso de indicadores de desempenho de tempo de execução e de conclusão de um processo, análise em tempo real de um processo, e criação de modelo de simulação, baseado em dados sem viés do especialista, devem permitir projetar soluções eficientes, de baixo custo e de alta aderência a realidade. A simulação é forte aliada à criação de *logs* artificiais e à geração de evidências para a recomendação de cenários e situações que envolvem comportamentos de processos em situações realistas.

⁴ <https://www.bpsim.org/>

3.7 Resumo do capítulo

A principal contribuição desta revisão sistemática de literatura diz respeito à organização de forma sistemática dos conceitos relacionados à simulação aplicada em conjunto com a mineração de processos, com ênfase na simulação de curto prazo e na geração de modelos de simulação.

Essa revisão sistemática permitiu analisar, detalhadamente, a relação entre a mineração de processos e a simulação. Nesse sentido, destacam-se cinco oportunidades:

1. criar ferramentas para operacionalizar a simulação de curto prazo, em especial, ferramentas *open source* para fomentar novas pesquisas;
2. simplificar arquiteturas existentes com a integração de vários componentes de software para executar todo o processo, iniciando com mineração e fechando com análise e operacionalização de um modelo de simulação de curto prazo;
3. capacitar gestores e pessoas quanto ao uso de ferramentas de análises de cenários *vis-à-vis* ao aproveitamento de recursos, gerenciamento de pessoas, bem como melhorar a tomada de decisão operacional e criar um diferencial competitiva baseado na eficiência operacional;
4. viabilizar a aplicação da tomada de decisão operacional e da simulação em tempo real;
5. explorar e criar técnicas de mineração de processos que possam ser utilizadas em conjunto com técnicas de simulação, com baixos esforços humanos.

A seguir, apresentaremos a estruturação detalhada da pesquisa, seguindo cada etapa da metodologia previamente definida. Descreveremos cada uma das etapas, incluindo sua justificativa, objetivo e atividades específicas a serem realizadas. Dessa forma, os leitores terão uma visão clara e abrangente do processo de pesquisa, compreendendo como cada parte se encaixa no todo e qual é a finalidade de cada uma delas. Com esta estruturação, busca-se garantir uma abordagem sistemática e rigorosa da pesquisa, assegurando que os resultados obtidos sejam confiáveis e possam ser validados por outros pesquisadores.

4 ESTRUTURAÇÃO DA PESQUISA

Quanto mais se racionaliza, mais se cria.

(Raymond Chandler)

Este capítulo apresenta os procedimentos metodológicos utilizados e tem o objetivo de gerar conhecimentos para a aplicação prática, e de caráter exploratório da problemática estabelecida para o tema de planejamento e agendamento.

4.1 Caracterização da pesquisa

A nossa pesquisa foi colocada em prática por meio da abordagem exploratória e experimental. A pesquisa exploratória abrange os procedimentos técnicos executados por meio da abordagem dos conceitos essenciais para o suporte ao tema de estudo. A consulta em bases de conhecimentos de materiais científicos permite fundamentar conceitos e embasar o método proposto. A pesquisa experimental abrange essencialmente a criação, experimentação e validação de modelos. A experimentação foi realizada a partir da concepção e criação de modelos de simulação, onde tais modelos encerram a lógica e o comportamento de um processo de negócio. A pesquisa é também de natureza tecnológica e aplicada à geração de valor econômico e social.

Dada a natureza tecnológica da pesquisa para gerar novos conhecimentos, a metodologia aplicada para o desenvolvimento do projeto foi “Design Science Research Methodology” (DSRM).

4.2 Design Science Research Methodology

A DSRM tem sua origem em ambientes naturais e artificiais, e foi proposta por Herbert Simon (SIMON, 1996). A metodologia baseia-se na “ciência do artificial”, na medida que ela se ocupa da concepção de artefatos que permitem realizar objetivos.

Os artefatos podem ser classificados em constructos, modelos, métodos e instanciações. Os **constructos** encerram conceitos que vão formar o vocabulário de domínio com termos usados sobre as tarefas. Os **modelos** definem sequências de passos para expressar as ações entre vários conceitos de um domínio. Os **métodos** reúnem o conjunto de passos necessários para a execução das tarefas. As

instanciações definem concretamente cada artefato concebido em seu ambiente, mostrando a viabilidade e eficácia dos modelos e métodos (MARCH; SMITH, 1995).

A DSRM oferece uma abordagem que permite criar artefatos para resolver problemas em contexto científico e gerar novos conhecimentos técnicos e científicos. Ela encerra seis etapas. E para cada etapa define-se procedimentos específicos, que são ordenados de forma adequada à necessidade do projeto (PEFFERS et al., 2007).

Na Figura 15 mostra-se a sequência da metodologia DSRM utilizada.

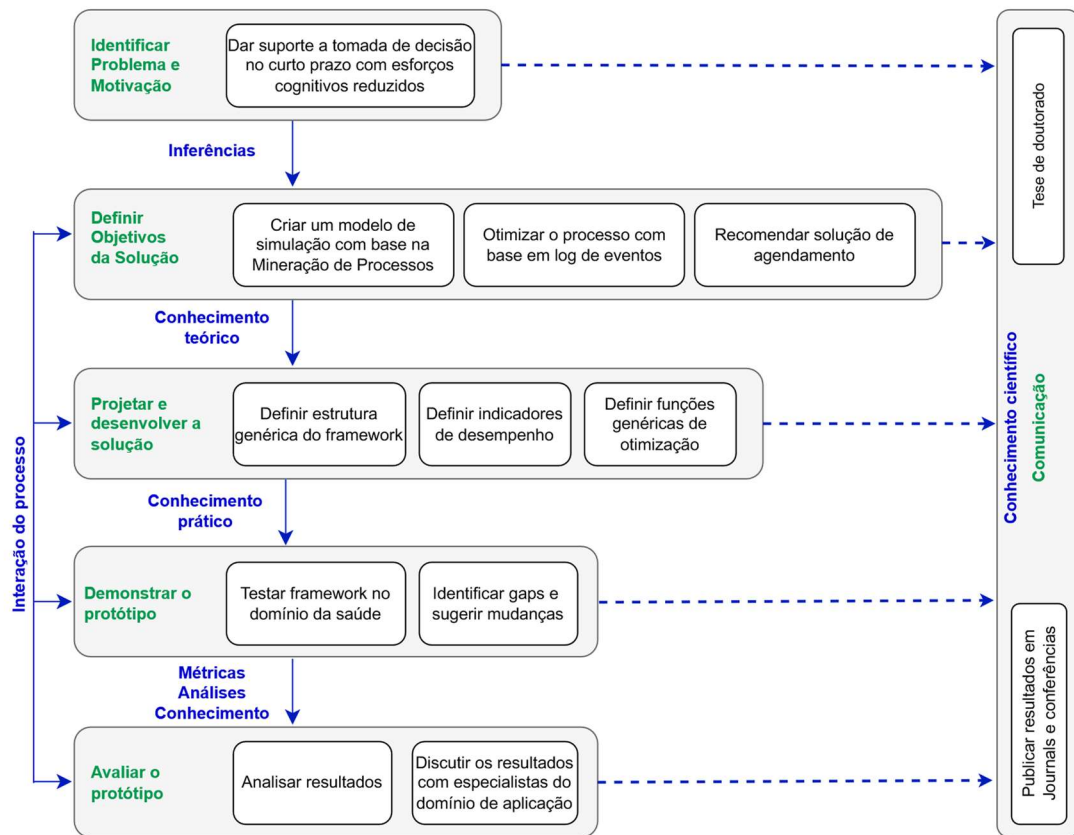


Figura 15 – DSRM – modelo de processo – adaptada de (PEFFERS et al., 2007)

A seguir, cada etapa da metodologia DSRM é detalhada.

4.3 Identificação do problema e motivação

Na primeira etapa da pesquisa foi realizado um estudo da literatura, conduzido na forma de uma revisão sistemática de literatura e apresentada no Capítulo 4. Nesta etapa, enumerou-se diversas questões com vista a colocar em evidência os elementos da pesquisa: 1) Como são criados os modelos de simulação a partir da mineração de processos? Essa questão visa identificar como são obtidos os

modelos de simulação tendo como *input* os logs de eventos; 2) Quais dados da mineração de processos são usados para executar uma simulação? Esta questão visa identificar os dados dos processos em execução necessários para a simulação; 3) Como o estado corrente de um processo é extraído e quais dados são usados na simulação? Esta questão permite identificar o mecanismo de extração e os dados do estado corrente para uso na simulação de curto prazo; 4) Como são tratados os recursos na simulação? Esta questão visa identificar como são modelados os recursos e quais dados da estrutura organizacional são utilizados na simulação; 5) Como a mineração de processos fornece subsídios para a tomada de decisão operacional? Esta questão visa identificar qual é o suporte efetivo que a mineração de processos fornece à simulação com vista à tomada de decisão operacional; 6) Quais contribuições, limitações e oportunidades podem ser enumeradas? Esta questão visa identificar as tendências das pesquisas para fomentar o binômio mineração de processos e simulação.

Parte das respostas às questões supranumeradas foram formuladas a partir da análise exploratória e inferencial da literatura, tomando como ponto de partida fontes de artigos as seguintes bases de dados: *Science Direct*, *IEEE*, *ACM* e *Springer*. E a outra parte das respostas foram geradas a partir dos resultados da modelagem, instanciação e avaliação do objeto de pesquisa.

4.4 Definição dos objetivos para a solução

Tomando como ponto de partida as respostas às questões apresentadas na seção anterior, definiu-se como meta resolver o problema do agendamento e alocação de recursos. Trata-se de um problema clássico, entretanto, ele permite colocar em prática um conjunto de procedimentos de automatização, que vão da coleta de dados, passando pela geração e avaliação do modelo de otimização, até a priorização das ações; tem-se como, um dos resultados esperados, a redução do esforço humano na implementação de um método de agendamento e alocação de recursos inteligente. Aqui, a inteligência se encerra na atividade de que o método pode operar de forma autônoma *vis-à-vis* a detecção, adaptação e recomendação de ações para voltar-se ao estado de homeostase do sistema. A ideia central é que o sistema resultante tenha habilidade de manter o meio interno—insumos e recursos—em um equilíbrio quase constante, independentemente das demandas no

ambiente externo. O termômetro para a indicação de equilíbrio se faz por meio de um conjunto mínimo de indicadores de desempenho.

De forma pragmática, a concepção geral do projeto passa, essencialmente, pela consecução das seguintes metas: a) automatização da criação do modelo de simulação—para um dado processo de negócio—com base na extração de dados/estimativas do modelo de processo descoberto a partir de um log de eventos; b) otimização do desempenho de indicadores com base na mineração de processos e na definição das preferências do especialista de processo negócio analisado. Dada a multiplicidade de indicadores e preferências, o método de otimização colocado em prática considera a avaliação de funções multiobjetivo. Deve-se notar que essencialmente os indicadores de desempenho—que forma o conjunto mínimo supracitado—estão relacionados à teoria de filas com vistas ao gerenciamento de uso de recursos e redução do tempo de espera; e c) recomendação de ações operacionais com vistas a otimização do agendamento e alocação de recursos.

4.5 Projeto e desenvolvimento

Consideraram-se, em relação à modelagem conceitual e arquitetural do *framework*, diversas áreas do conhecimento, a saber: mineração de processos, simulação computacional, otimização multiobjetivo e tomada de decisão.

O ambiente de desenvolvimento reúne um conjunto de bibliotecas escritas em *Python*⁵, a saber: *PM4py*⁶ para a tarefa de descoberto de modelos de instâncias de processo; *Simpy*⁷ para a simulação computacional de eventos discretos; e *Pycbr*⁸ para a adaptação de casos com otimização baseada em algoritmos evolucionários. Complementar, em Python, implementou-se também um método de otimização de agendamentos. Em termos de algoritmos de otimização, as principais estratégias colocadas em prática foram: evolucionário — *genética* — e *Simulated Annealing*.

A parte experimental da pesquisa fez uso, essencialmente, de modelos de simulação e dados reais anonimizados. A parte de otimização de indicadores de desempenho baseou-se essencialmente na resolução de restrições *hard* — ou

⁵ <https://www.python.org/>

⁶ <https://pm4py.fit.fraunhofer.de/>

⁷ <https://simpy.readthedocs.io/en/latest/>

⁸ <https://github.com/Dih5/pycbr/>

rígidas — e *soft* — ou *suáveis*; lembrando, as restrições *hard* precisam obrigatoriamente serem satisfeitas na composição da solução final, e *soft* não precisam necessariamente serem satisfeitas. O projeto e desenvolvimento do *framework PM4SOS* foi realizado em quatro etapas: a) definição da estrutura genérica do *framework*; b) definição do conjunto mínimo de indicadores de desempenho; c) definição das funções genéricas de otimização de agendamento cirúrgico, e d) projeto do *framework*. O Capítulo 5 apresenta em detalhes a concepção e consecução de cada artefato do *framework*.

4.6 Demonstração

A demonstração do *framework PM4SOS*⁹ considera, como exemplo de aplicação, a área da saúde, em específico a alocação de recursos e agendamento de cirurgias eletivas em um pronto atendimento de um hospital universitário. Deve-se notar, no entanto, que esta abordagem não se restringe apenas a essa área do conhecimento. A generalização do *framework* permite aplicação a outros domínios, observando as orientações e restrições.

O ponto de partida são logs de eventos extraídos dos sistemas de informação do Hospital Universitário, em particular, descendo as atividades, recursos e insumos do centro cirúrgico. Por meio da instanciação do agendamento de recursos e de classes de cirurgias eletivas do centro cirúrgico em questão, apresentou-se o funcionamento da arquitetura do *framework*.

A demonstração do *framework PM4SOS* encerra essencialmente dois quesitos: a) a execução do *framework* com dados do domínio da saúde, e b) identificação *gaps* e sugestão de melhorias no processo de agendamento cirúrgico. O conjunto completo dessas etapas estão descritas no Capítulo 5.

4.7 Avaliação

O ponto de partida para a construção do modelo de simulação é a obtenção de um log de eventos do processo de negócio em estudo. Assume-se, nesta tese, que o *log* de eventos atende o conjunto mínimo de qualidade—acurácia, integridade,

⁹ O *framework* proposto não está disponível em plataformas públicas de compartilhamento de código fonte devido à restrição de registro de patente de software da solução desenvolvida.

consistência e singularidade. Assim, a primeira avaliação consiste em comparar o log de eventos original—extraído do sistema de informação do processo de negócio em estudo—usado como entrada e o log de eventos obtido pela simulação. Deve-se notar que o modelo de simulação foi gerado a partir dos dados do log eventos originais ou do processo de negócio real. Quanto mais similares forem os logs de eventos real e simulado melhor, ou seja, mais aderente é o modelo de simulação ao processo real. A segunda parte da avaliação consiste em medir o quanto o modelo de simulação contribuiu para a melhoria do processo em estudo face aos indicadores de desempenho. A terceira parte da avaliação encerra o esforço de adaptação de uma experiência passada na composição da solução de um problema; lembramos que o problema concerne o redimensionamento de um processo de negócio em termos de alocação de recursos e agendamento de atividades. E a última avaliação centra-se em medir o quanto o método de otimização é capaz de compor uma solução que atenda todas as restrições *hard* e *soft*, bem como a redução do esforço humano na reconfiguração do processo de negócio, mantendo os indicadores de desempenho dentro de limites de satisfação previamente definidos para o processo.

4.8 Comunicação

A comunicação dos resultados encerra três momentos específicos: (a) submissão de artigo de revisão sistemática de literatura sobre o estudo do estado-da-arte de simulação de curto prazo e mineração de processos—revista científica *ACM Transactions on Modeling and Computer Simulation*; (b) submissão de três artigos em conferências relativos às etapas de projeto e desenvolvimento, demonstração e avaliação; e (c) elaboração do texto da tese—publicado no site da Universidade. O conjunto de análise e discussão dos resultados são parte integrante da estratégia de comunicação desta tese.

No próximo capítulo, apresenta-se o desenvolvimento da pesquisa conforme as etapas do método DSRM. Inicialmente será apresentado o projeto e desenvolvimento do framework, em seguida, serão apresentados os experimentos e os respectivos resultados obtidos com a aplicação do framework para agendamento cirúrgico e posteriormente a avaliação do framework proposto.

5 DESENVOLVIMENTO DA PESQUISA

*Conhecimento não tem valor,
a menos que você o coloque em prática. (Anton Chekhov)*

Neste capítulo descrevem-se o projeto, desenvolvimento, demonstração e avaliação do *framework PM4SOS*, bem como os resultados obtidos com a sua aplicação.

5.1 Framework PM4SOS

O framework PM4SOS compreende uma estrutura integrada de elementos interligados para o suporte à tomada de decisão. Na Figura 16, apresenta-se o ponto de partida para a execução do fluxo de trabalho do *framework* referente à extração e validação de um log de eventos com base em uma dada fonte de dados — que pode ser formada pelos sistemas de informação de uma organização.

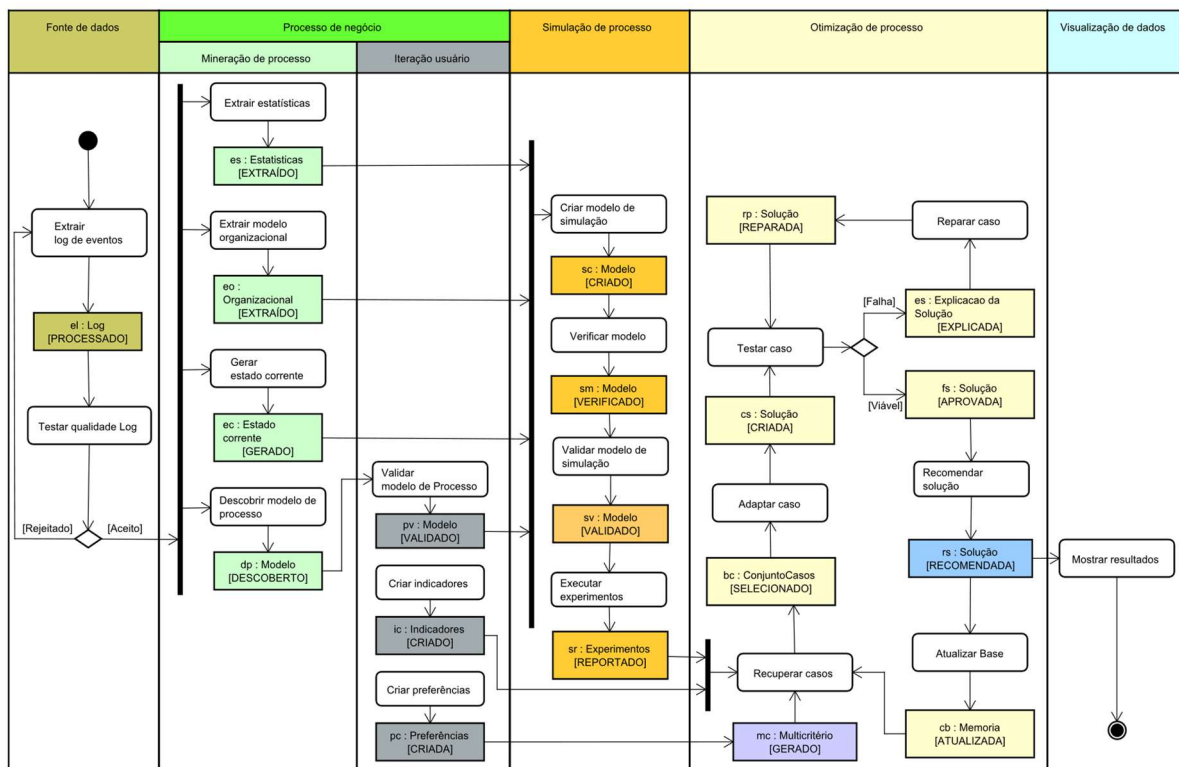


Figura 16 – *framework PM4SOS* – o autor 2022

Considerando-se *el* um log de eventos válido de um processo de negócio *pn*, *es* um conjunto de estimativas estatísticas extraídas de *el*, *eo* o modelo organizacional extraído de *el*, *ec* o estado corrente do processo *pn* gerado a partir de

el , dp o modelo de processo descoberto a partir de el e pv um modelo de processo definido e validado a partir de dp .

Seja sc um modelo de simulação criado a partir de es , eo , ec e pv . Seja sm um modelo de simulação definido e verificado a partir de sc e sv um modelo de simulação validado a partir de sm . Seja sr o resultado obtido a partir da execução/experimentação do modelo sv . Ainda, seja $ic = \{k1, k2, \dots, kn\}$ um conjunto de indicadores de desempenho pré-definido para o processo pn , $pc = \{f1, f2, \dots, fm\}$ o conjunto de preferências pré-definidas para o processo pn e mc um conjunto de critérios normalizados e gerados a partir das preferências pc . Deve-se notar participação do especialista no processo se dá essencialmente em dois momentos: (a) na validação do modelo de processo antes da simulação e b) na definição dos conjuntos indicadores ic e preferências pc .

A terceira etapa concerne à otimização do processo pn . Esse método de otimização assume o ciclo canônico da abordagem raciocínio baseado em casos, em que um caso c_i é dado por um par $\langle p_i, s_i \rangle$, p_i é a descrição do problema e s_i é a solução de p_i . E a base de casos ou memória $cb = \{c1, c2, \dots, cn\}$ é formada por um conjunto de c_i . O processo de otimização toma como entrada: sr, ic, pc, cb e executa a tarefa de recuperação de casos sobre cb . Seja $cb.tf$ um método de cb definido para a recuperação de casos, onde sua instanciação é dada por $cb.tf(sr, ic, pc)$. Os parâmetros de $cb.tf$ resultam o conjunto índices $i_x = \{i_1, i_2, \dots, i_q\}$ para o cálculo de similaridade entre o processo que se deseja otimizar e cada elemento da base cb . Seja $bc = \{c1, c2, \dots, cp\}$ o subconjunto de casos de cb mais similares aos índices definidos por i_x .

A próxima etapa consiste em construir uma solução para o processo que se deseja otimizar adaptando um ou mais casos para comportar a solução em questão. Essa adaptação de caso, em nosso projeto, foi implementada fazendo uso de uma abordagem evolucionária — algoritmo genético —, em que toda solução válida cs é dada por um conjunto de valores que satisfazem todas as restrições *hard* definidas para o problema. A solução cs é submetida a um conjunto de regras de teste ou de viabilidade, e, se o teste falhar: (a) gera-se uma explicação es da falha, (b) repara-se a solução cs — solução reparada rp — por meio da aplicação de um conjunto de regras de reparação. Se a solução cs é testada, e o teste indicar viabilidade, então o fluxo de trabalho segue indicando a solução aprovada fs . A última etapa envolve a

recomendação da solução *rs*, seguida pela inclusão de *fs* na base *cb* e sua aplicação, representada pela ação *mostrar resultados*.

Deve-se salientar que o modelo de simulação *sc* é gerado, automaticamente, a partir de *es*, *eo*, *ec* e *pv*, respectivamente, estimativas estatísticas, modelo organizacional, estado corrente e modelo de processo validado. O modelo gerado *sc* é validado a partir da aplicação de métricas de cálculo de similaridade entre valores obtidos de desempenho da execução da simulação e aqueles obtidos diretamente do log de eventos *el*. O que é esperado com o resultado dessa comparação? Espera-se que o processo de simulação imite o processo real *pn* o mais fidedignamente possível. Em geral, espera-se, também, que resultado da simulação gere valores para os indicadores pré-estabelecidos *ic* iguais ou melhores daqueles obtidos para o processo real *pn*. Caso ocorra desvios nos indicadores e estes estejam fora das faixas preestabelecidas, recomenda-se a ativação do gatilho de otimização do processo.

Em resumo, a otimização de processo proposta utiliza o ciclo canônico do raciocínio baseado em casos para reusar experiências passadas na expectativa de contribuir para a geração de boas soluções com um custo computacional reduzido. Obviamente, a estrutura de memória e raciocínio dá suporte a um ciclo de aprendizagem virtuoso, em que a base de conhecimento é enriquecida na medida em que novas soluções são produzidas e aceitas.

Utilizam-se, também, as preferências do tomador de decisão quanto à composição multicritério para a tomada de decisão. A análise multicritério, envolvendo objetivos conflitantes, permite melhorar a precisão da tarefa de recuperação de casos similares ao problema estabelecido. Tal precisão contribui, significativamente, para a redução de esforço do mecanismo de adaptação de caso/solução, em que se pode utilizar algoritmo genético para adaptar a solução inicial, por exemplo, do agendamento cirúrgico. A solução adaptada é testada para verificar se ela atingiu qualidade mínima *vis-à-vis*, e se os critérios estabelecidos pelo especialista, em que se podem ter duas respostas mutuamente exclusivas, a saber: *falha* ou *sucesso*. Quando a resposta obtida é uma *falha*, disparam-se dois mecanismos em sequencias, o primeiro mecanismo para gerar uma explicação da *falha*, e o segundo para reparar a solução levando à explicação da *falha*. O mecanismo de reparação implementa, no contexto da nossa aplicação, a lógica do

algoritmo *Simulated Annealing*. Executa-se o ciclo teste, explicação e reparação até a obtenção de uma resposta indicando que a solução proposta é aplicável — ou *sucesso*. A solução aplicável segue o fluxo colocando-a em prática; finalmente, a solução e seus efeitos enriquecerão a base de experiências. A recomendação é aplicada, sobre as soluções identificadas, ao final da aplicação do raciocínio, baseado em casos na base de conhecimento. A etapa final encerra a descrição da solução, gerando sua identificação/assinatura — na forma de um conjunto de índices — e recomendação. Tal descrição, centrada no usuário, pode incluir gráficos e textos.

A seguir será detalhado o projeto e desenvolvimento do *framework* PM4SOS.

5.2 Projeto e desenvolvimento

O *framework* PM4SOS encerra uma arquitetura modular e cada módulo contempla, em uma certa medida, uma etapa ou objeto ligado à investigação, a saber: preparação de fonte de dados, mineração de processos, simulação computacional, otimização de processo e visualização de dados. E cada etapa será representada seguindo a notação IDEF0.

5.2.1 Domínio do problema de pesquisa

O domínio de aplicação encerra as atividades ligadas à Saúde, em particular, as consultas médicas que antecedem as cirurgias eletivas de um centro cirúrgico e as atividades inerentes ao processo cirúrgico que inicia com o agendamento e encerra na alta hospitalar. Na Figura 17, representa-se o modelo de processo para consultas médicas.

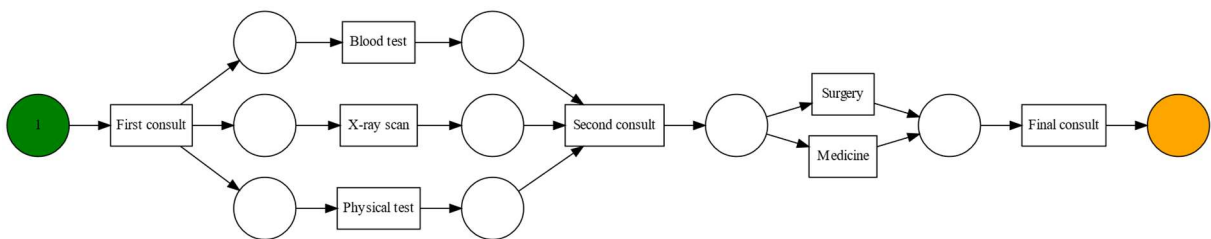


Figura 17 – modelo de processo de consultas médicas – o autor 2022

Nesse contexto, o desafio da pesquisa está na geração de políticas de ações para agendar cirurgias e alocar recursos, incluindo salas de cirurgia, cirurgiões e assistentes às cirurgias, além da sequência de cirurgias em salas cirúrgicas e o

horário de início. Este cenário é também presente em (GHAZALBASH et. al., 2012), ilustrado no Anexo II. Parte da solução do desafio inspirou-se no sequenciamento e definição das atividades de um centro cirúrgico dado por (SOLTI et. al., 2013).

A compreensão do processo de estruturação de um centro cirúrgico conta com diversas fontes de dados, por exemplo, o documento de Autorização para Internação Hospitalar (BRASIL, 2007). Este é composto pelo laudo do solicitante do procedimento, hospital onde vai ser realizada a cirurgia, bem como detalhes do procedimento cirúrgico, diagnóstico do paciente e outras informações relevantes ao procedimento cirúrgico.

Na Figura 18 representa-se o modelo de processo identificado, contemplando a demanda cirúrgica e inserção na fila de espera para o processo de agendamento.

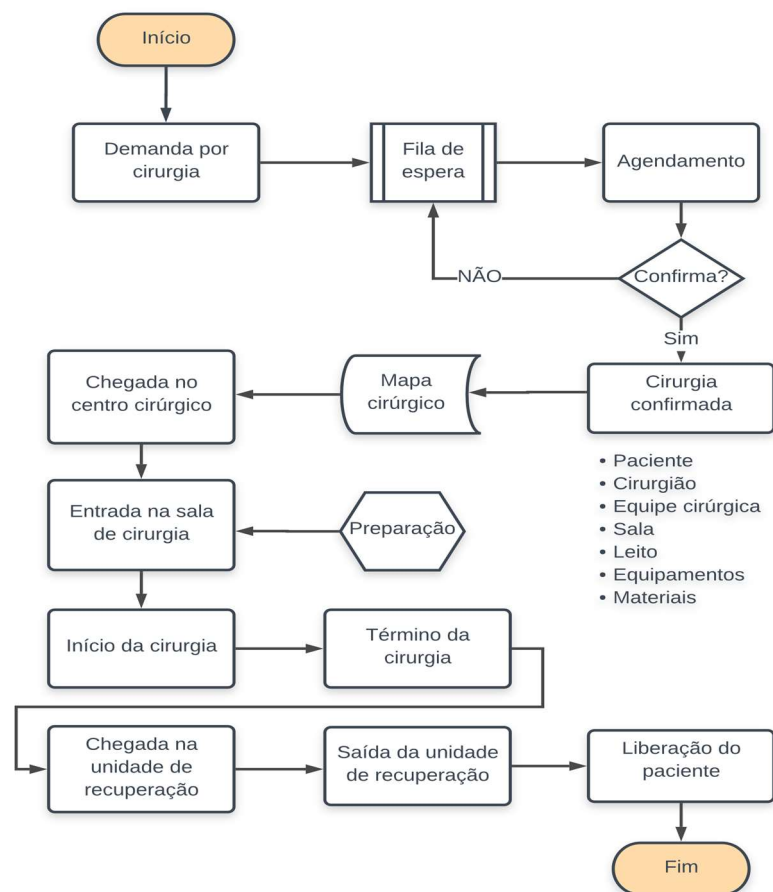


Figura 18 – modelo de processo de um centro cirúrgico - o autor 2022

A demanda cirúrgica permanece na fila de espera até que se tenham disponível os recursos para a realização da demanda, incluindo local e equipamentos necessários. Ao obter a confirmação de disponibilidade, gera-se o

mapa cirúrgico para o agendamento semanal do centro cirúrgico. Na data do agendamento, iniciam-se os procedimentos de preparação do ato cirúrgico. Tal inicialização desdobra-se em preparação do ambiente, limpeza e condicionamento da sala, verificação da disposição de equipamentos e materiais até a sedação do paciente para a realização da cirurgia. As atividades seguintes contemplam o início e término da cirurgia, a recuperação do paciente na unidade de recuperação. A etapa final consiste na liberação do paciente.

Esse modelo genérico contempla a rotina de um centro cirúrgico convencional. Nessa abordagem, não fazem parte as cirurgias de emergência, ou seja, apenas cirurgias eletivas estão sendo admitidas devido às suas uniformidades em termos comportamentais do processo; deve-se, também, notar que não foram considerados atrasos *vis-à-vis* a preparação do paciente e do ambiente, execução do procedimento cirúrgico, assim como a reabilitação e alta do paciente.

A definição do modelo de processo permitiu colocar em prática a preparação de dados, iniciando pela criação de logs de eventos artificiais. Deve-se salientar que se utilizou da informação de tempo médio para a realização de cada tarefa de um centro cirúrgico definido em (STROPARO; BICHINHO; PROTIL, 2005). Tal informação foi usada para replicar as atividades do processo real em execução e o resultado registrado na forma de um log de eventos.

Na Tabela 13, informa-se parte do *log* de eventos criado de forma sintética, seguindo modelo de processo da Figura 18.

Tabela 13 – exemplo de log de eventos.

Case	Atividade	Data-Hora	Usuário	Cirurgião	Sala	Proc.	Duração
1	Solicitação de cirurgia	04/05/2022 14:52	Marcia				
1	Agendamento de cirurgia	10/05/2022 08:30	Gabriel				
1	Chegada Centro Cirúrgico	10/05/2022 08:00	Marcos				
1	Entrada Sala Cirurgia	10/05/2022 08:25	Lóris				
1	Início da Cirurgia	10/05/2022 08:30	Eduardo	345	2	65	2
1	Final da Cirurgia	10/05/2022 10:00	Eduardo				
1	Entrada Recuperação	10/05/2022 10:10	Antônio				
1	Saída Recuperação	10/05/2022 12:30	Antônio				
1	Saída Centro Cirúrgico	10/05/2022 16:15	Eduardo				

Fonte: o autor 2022

A preparação de dados é uma seção importante em relação à proposta apresentada nesta tese. Em outras palavras, como a proposta passa pela redução do esforço humano no que tange ao esforço e conhecimento para colocar em prática um cenário de simulação de curto prazo, fazer com que o processo seja independente do especialista é uma tarefa desejada, e não trivial.

5.2.2 Fase 1 – preparar dados

A primeira etapa do fluxo de trabalho consiste na preparação dos dados. A representação IDEF0 (cf. Figura 19) mapeia como entrada os registros de dados, como controle das métricas de qualidade, como mecanismo a geração de *logs* e como saída os *logs* de eventos.

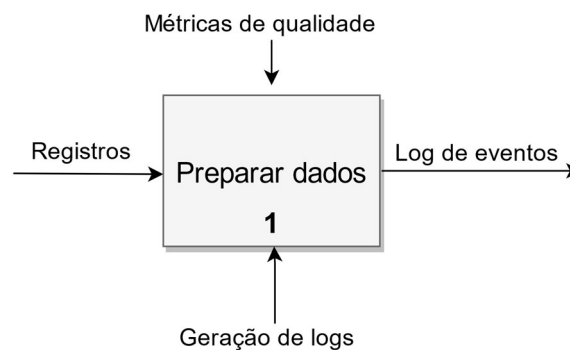


Figura 19 – fase 1: fonte de dados – o autor 2022

Depois de criar uma base de dados para o log de eventos, criou-se uma ferramenta computacional para gerar *logs* artificiais chamado Sim2Log¹⁰. Essa ferramenta é utilizada quando não se tem acesso diretamente a um conjunto de dados reais. Os parâmetros utilizados como entrada de dados para a geração sintética dos *logs* de eventos foram: a) tempo médio de execução das atividades; b) distribuições estatísticas para os intervalos de chegadas de cada caso; c) recursos como cirurgiões e salas e atributos adicionais d) identificação do procedimento cirúrgico e sua duração. Deve-se informar, também, a quantidade de casos a serem gerados, por exemplo, 1.000.

¹⁰ <https://github.com/ferronato/Sim2Log>

A ferramenta Sim2Log toma como entrada um log de eventos no formato XES e gera dados estatísticos referentes ao processo identificado/descoberto — as distribuições estatísticas são obtidas de forma automática —, bem como saída um novo *log* populado com a quantidade estabelecida de casos, por exemplo, 1.000.

5.2.2.1 Avaliação de qualidade do log de eventos

Os logs gerados podem ser avaliados com base na comparação do comportamento, obtido nos *logs* de eventos e os comportamentos presentes no modelo de processo. A avaliação dos logs de eventos, segundo Van der Aalst (2016), pode ser feita sob quatro aspectos: a) **aptidão**, que avalia se pode ser feito *replay* dos logs; b) **precisão**, que avalia se há desvios em relação ao modelo; c) **generalização**, que avalia amplitude dos comportamentos dos logs, e d) **simplicidade**, que avalia a estrutura de um modelo com vista a se ter o mais simples possível. Ainda segundo Van der Aalst (2016), não se pode ter todos os aspectos de avaliação dos *logs* atendidos, pois as métricas de qualidade possuem objetivos conflitantes. É importante, intuitivamente, ter um equilíbrio sobre os critérios de avaliação dos logs.

O PM4SOS permite, ao fazer a leitura de um log de eventos, identificar o modelo de processo e mensurar a qualidade dos logs. A validação do modelo de processo é uma tarefa na qual o especialista decide em continuar a execução do *framework* ou decide ajustar ou alterar a fonte de dados.

A utilização adequada de uma fonte de dados para a mineração de processos compreende algumas premissas, como: a) uma fonte de dados bem-formada, sem *outliers* e sem a necessidade de correção manual dos dados; b) cada evento do log de eventos deve ter uma estrutura mínima — identificador de caso, nome de atividade e registro de tempo do evento. Outros dados podem presentes no *log* de eventos, como identificação do paciente, cirurgião, sala, procedimento e duração do procedimento.

A utilização de um mecanismo de avaliação do *log* de eventos de um processo habilita que, a partir de uma entrada estabelecida com nível mínimo de qualidade, seja possível executar as demais etapas do *framework*. E, para garantir que a fonte de dados tenha um nível de qualidade adequado, utilizou-se a abordagem do *framework* proposto por (VERHULST, 2016), o qual definiu quatro dimensões de qualidade para um *log* de eventos. Ele estabeleceu, para cada dimensão da qualidade, uma métrica e para cada métrica uma avaliação. Em outras

palavras, um escore de qualidade é obtido como resultado da aplicação de um conjunto de critérios avaliativos. O *pipeline* para avaliação da qualidade dos logs de eventos está disponível no PROM¹¹.

Descreveram-se, os critérios de qualidade desta tese, na Seção 2.5, e sintetizaram-se as métricas de acurácia, integridade, consistência e singularidade de um *log* de eventos. O cálculo de cada métrica e seus respectivos pesos produz um valor geral estabelecido como escore de qualidade. O usuário pode definir o peso para cada critério. Nesse sentido, um *log* de eventos é válido, somente, se cumprir os requisitos mínimos estabelecidos de qualidade.

Após estabelecer a fonte de dados e os critérios de qualidade de um *log* de eventos, a próxima à etapa é de mineração de processos.

5.2.3 Fase 2 - mineração de processos

Na Fase 2 do PM4SOS, considerando a presença de um log de eventos, positivamente avaliado em termos dos critérios de qualidade, executa-se a tarefa de mineração de processos para obtenção de informações relevantes para a composição da simulação.

Na Figura 20, representa-se o subprocesso da Fase 2 do PM4SOS. A entrada desse subprocesso é um log de eventos bem-formatado e já validado por critérios objetivos de qualidade. Então, aplicam-se filtros para a obtenção das variantes de processo informando um nível de interesse entre 0% e 100%. O nível informado é 100%; filtra-se o número total de variantes presentes no *log* de eventos, e, se o nível informado for 80% o número de variantes de interesse para estabelecer o modelo de processo, será reduzido em 20%.

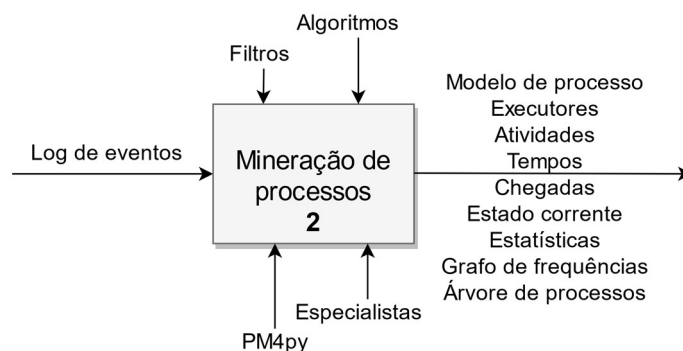


Figura 20 – fase 2: mineração de processos – o autor 2022

¹¹ <https://www.promtools.org>

Utilizaram-se algoritmos para as atividades de descoberta do modelo de processos com *inductive miner*, geração do grafo de frequências com *Directly-Follows Graphs* e árvore de processos com *inductive miner*. A biblioteca base para a mineração de processos utilizado foi o PM4py e realizou-se o controle de validação do modelo descoberto por especialistas. A intervenção do especialista limita-se em avaliar se o processo descoberto representa a realidade do negócio. A seguir são detalhadas as saídas geradas pela fase de mineração de processos. O resultado prático dessa fase é a geração automática do modelo de simulação.

5.2.3.1 Descoberta do modelo de processo

Na fase de mineração de processos, aplicou-se o algoritmo de mineração de processos *inductive miner* para a descoberta do modelo de processo. A escolha desse algoritmo justifica-se, pois ele garante a solidez do modelo descoberto e implementa estratégias que permitem descobrir estruturas, como sequencialidade, paralelização, concorrência e laços, conforme (LEEMANS et. al., 2013).

Complementarmente à descoberta do modelo de processo com o algoritmo *inductive miner*, extraiu-se o grafo direcionado, derivado do *log* de eventos. A título de exemplo, criaram-se dois grafos direcionados. O grafo direcionado extraído do *log* de eventos contém dois formatos: (a) baseado na frequência da execução das atividades Figura 21a; e (b) baseado na performance de execução, ou seja, os tempos médios detectados para a execução de cada atividade do processo em análise Figura 21b.

Os grafos direcionados revelam o relacionamento direto entre as atividades e onde o traço começa e termina. Na Figura 21, observa-se que a atividade “*First consult*” está conectada diretamente com um arco à atividade “*Blood Test*”. Pode-se também dizer que a atividade “*First consult*” é seguida pela atividade “*Blood Test*” (LEEMANS; FAHLAND; VAN DER AALST, 2018).

O peso de cada arco denota a frequência da ocorrência da conexão das atividades. Por exemplo, o grafo da Figura 21a denota que a frequência do caminho identificado pelos arcos “*Second consult*” até “*Final consult*” tem uma frequência 72 para a atividade “*Medicine*” e frequência 8 para “*Surgery*”. Para obter a frequência do grafo direcionado, conta-se o número de vezes que a atividade de origem é seguida pela atividade de destino. Esse mecanismo é usado para atribuir regras de decisão em certos pontos do modelo de simulação em que é necessário estabelecer

probabilidades de execução das atividades para representação adequada do modelo.

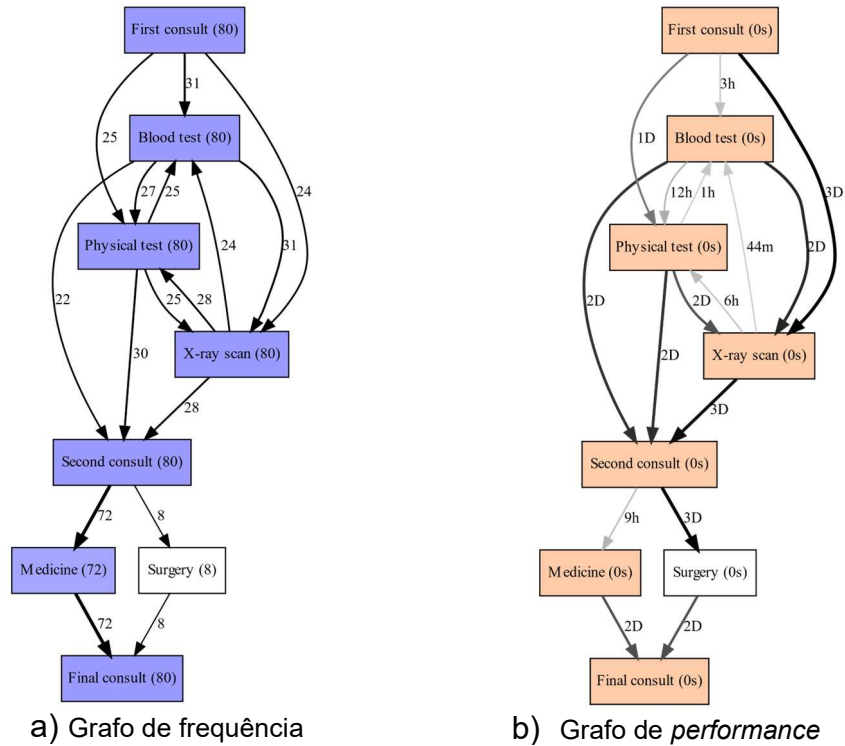


Figura 21 – modelo de processo em DFG – o autor 2022.

Outra etapa associada à descoberta do modelo de processos e utilizada para a criação do modelo de simulação é a extração da árvore de processos. Com a aplicação do algoritmo *inductive miner*, é possível obter árvore de processos, conforme Figura 22. A árvore de processos permite identificar as atividades do processo, sua sequência de execução, as execuções em paralelo e estruturas de decisão. A abordagem do *inductive miner* permite a descoberta de modelos de processos sólidos e adequados em tempo finito (LEEMANS et. al., 2013).

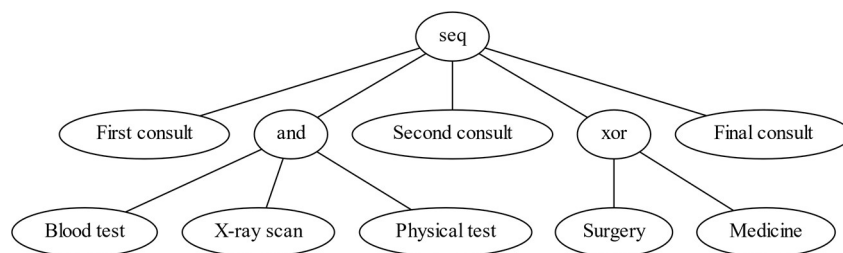


Figura 22 – modelo de processo em árvore – o autor 2022

A seguir é descrito-se a expressão regular extraída do *log* de eventos, correspondente à árvore de processo da Figura 22:

seq ('*First consult*', ***and*** ('*Blood test*', '*Xrayscan*', '*Physical test*'), '*Second consult*', ***xor*** ('*Medicine*', '*Surgery*'), '*Final consult*'

Os símbolos extraídos e suas representações: “***seq***” representam as sequências das atividades. No exemplo acima, pode-se perceber que a sequência inicia na atividade “*First consult*” e, posteriormente, o elemento “***and***” é utilizado para executar uma das atividades interna ao parêntesis de forma paralela, ou seja, a ordem de execução pode variar de acordo com cada paciente. A expressão “***xor***” representa uma decisão exclusiva com uso dos elementos entre parêntesis. Nesse exemplo a decisão é tomada para as atividades “*Medicine*” ou “*Surgery*”.

A estrutura de árvore de decisão é complementada pelas informações obtidas no grafo direcionado — ou DFG. Assim, é possível utilizar a frequência do DFG para determinar a probabilidade de escolha entre duas opções decisórias para definir o modelo de simulação. A partir da estrutura da árvore de processo, pode-se inferir o tamanho mínimo do traço permitido no modelo de processo descoberto.

A seguir será descrita a descoberta do modelo organizacional.

5.2.3.2 *Descoberta do modelo organizacional*

Outro elemento extraído do *log* de eventos é o modelo organizacional. Esse modelo determina quem são os executores das atividades e como eles participam e interagem na execução do processo. A extração do modelo organizacional é feita seguindo a lógica do Algoritmo 3. Esse último recebe como entrada um *log* de eventos e extrai-se o modelo organizacional orientando-se pelas atividades do processo presentes no *log* em questão. Deve-se, também, notar que, a partir de cada atividade se identifica e extrai-se o(s) executor(es) ou recurso(s) associado(s).

Algoritmo 3 - descoberta do modelo organizacional

```

Input : Log de eventos (L)
Output: Modelo Organizacional (EO) {<Atividade, Recurso>}


---


program discoveryOrganizationalModel
  EO =  $\emptyset$ ; {EO-conjunto de pares  $\langle a_i, r_i \rangle$ ,  $a_i$  atividade e  $r_i$  recurso}
begin
  A  $\leftarrow$  getActivity(L, "concept:name"); {A é um vetor de atividades}
  for k = 1 to A.length do:
    T  $\leftarrow$  getTraces(L, k);
    R  $\leftarrow$  getResources(T, "org:resource"); {R é um vetor de recursos}
    EO = EO  $\cup$  {T,R}
  endfor
  return EO;
end

```

A tarefa de extração recurso é executada com base no atributo "*org:resource*" presente no *log* de eventos. A presença desse atributo denota um requisito de log de eventos bem-formado. A presença desse atributo, no *log* de eventos, é verificada na etapa de avaliação do *log* de eventos na Fase 1 do fluxo de trabalho do PM4SOS. A seguir será descrita a etapa de extração de parâmetros e estatísticas a partir do log de eventos.

5.2.3.3 Extração de parâmetros e estatísticas

A extração de parâmetros e estatísticas constantes no *log* de eventos sobre um processo é uma tarefa importante para a direção da automatização da criação do modelo de simulação. Esse mecanismo viabiliza que o modelo de simulação seja criado com base nos valores observados em termos de parâmetros e estatísticas presentes no log de eventos do processo de interesse. Nesse sentido, deve-se observar que o que se busca é a redução dos esforços humanos em relação à criação do modelo de simulação, pois os dados para a criação do modelo de simulação podem ser obtidos de forma computacional, com base no resultado da aplicação da mineração de processos.

Outros benefícios também podem ser citados como isenção de viés do analista de processo ou qualquer interferência humana. Como o *log* de eventos é atualizado constantemente ou de forma ininterrupta, o modelo de simulação também pode ser atualizado rapidamente por meio de novas obtenções de parâmetros, recursos e estruturas. Os seguintes parâmetros foram extraídos dos *logs* de eventos para criar o modelo de simulação:

Taxa de chegada: ela corresponde à distância média entre a chegada de dois casos consecutivos presentes no log de eventos. Esse parâmetro permite inserir no modelo de simulação as chegadas de novas instâncias para serem processadas pelo simulador.

A distribuição de probabilidade utilizada para a geração da taxa de chegada das instâncias consiste na **distribuição de Poisson**. Esse mecanismo descreve o número de vezes que um evento ocorre em dado intervalo de tempo. As principais características dessa distribuição são: a) o número de ocorrências possíveis em qualquer intervalo de tempo é ilimitado; b) as ocorrências são independentes e não afetam outras ocorrências; e c) o número médio de ocorrências permanece o mesmo de intervalo a intervalo (MUN, 2008).

Descreve-se, na Eq. 12 a representação da distribuição de Poisson.

$$P(x) = \frac{e^{-\lambda} \lambda^x}{x!}, \text{ para } x \text{ e } \lambda > 0 \quad \text{Eq. 12}$$

em que

- e é base do logaritmo natural
- $x!$ é fatorial de x ,
- λ é um número real que denota o número esperado de ocorrências que ocorrem em um dado intervalo de tempo.

Duração dos casos: ela representa o tempo médio que um caso necessita para ser executado. Dessa forma, é possível utilizar esse parâmetro como base para comparar o tempo de execução de um caso simulado. A duração média dos casos é obtida com base nos casos presentes no *log* de eventos.

Taxa de saída: ela representa a distância média entre a conclusão de dois casos consecutivos. Esse parâmetro permite, também, ajustar o modelo de simulação para que ele produza resultados similares àqueles observados no *log* de eventos. Além disso, ele pode ser usado na fase de validação do modelo de simulação, em que os dados gerados pelo modelo de simulação podem ser comparados ao log de eventos dado como entrada à mineração de processos. A taxa de dispersão dos casos é obtida com base nos casos presentes no *log* de eventos.

Taxa de execução das atividades: ele permite identificar a distribuição estatística das atividades do processo. As atividades são executadas de forma

diferente a cada caso. Então, deve-se obter a distribuição estatística de cada atividade. Esta taxa é obtida a partir das atividades presentes no log de eventos

A distribuição de probabilidade para obtenção dos tempos das atividades utilizada foi a **distribuição Exponencial**. Essa distribuição é amplamente usada para descrever a geração randômica de eventos em pontos de tempo. Ela está relacionada à distribuição de Poisson, a qual descreve o número de ocorrências de um evento em um intervalo de tempo. Uma característica da distribuição exponencial é “*memoryless*”, que significa que um objeto tem a mesma distribuição futura independente do tempo em que existiu (MUN, 2008).

A Eq. 13 representa a distribuição exponencial.

$$f(x) = \lambda e^{-\lambda x}, \text{ para } x \geq 0; \lambda > 0 \quad \text{Eq. 13}$$

em que:

- e é base do logaritmo natural;
- λ é um número real que denota o número esperado de ocorrências que ocorrem em um dado intervalo de tempo.

Depois de obter os parâmetros para o modelo de simulação, completa-se o sequenciamento do modelo de simulação com base na árvore de processos obtida com a descoberta do modelo de processo.

De forma geral, a distribuição de probabilidade segue a distribuição de Poisson (tempo entre duas chegadas segue uma distribuição exponencial negativa) com média de $\frac{1}{\lambda}$ [(HILLIER, 1967; ROZINAT et al., 2009), (DUMAS et al., 2013), (ZHOU; WANG; LI, 2014)].

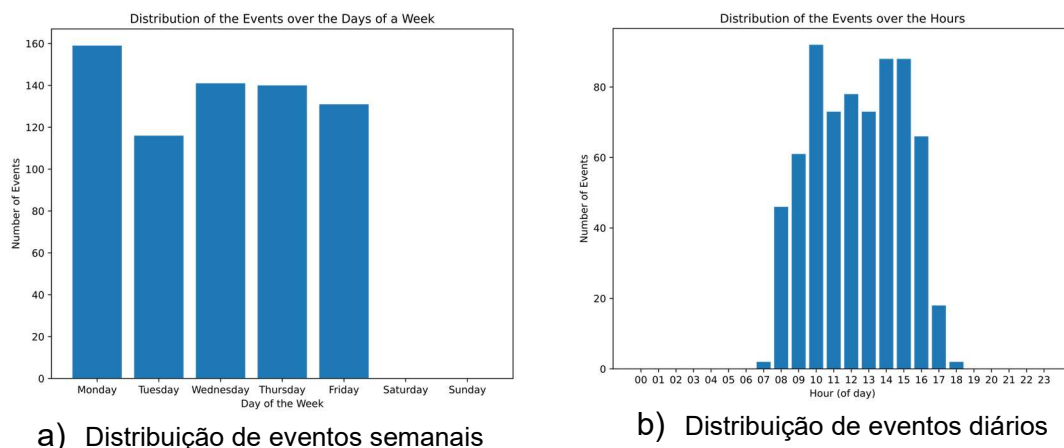


Figura 23 – distribuição dos cases diários e semanais – o autor 2022.

A distribuição de probabilidade relacionadas às atividades do processo seguem a frequência detectada em cada transição no grafo DFG. Outra distribuição de probabilidade utilizada é a distribuição dos casos em dias da semana, conforme Figura 23a; e horas do dia, Figura 23b. Esse mecanismo permite que o modelo de simulação imite o comportamento da distribuição da execução de eventos.

Deve-se destacar que os dados disponíveis em log de eventos são ricos em estatísticas e parâmetros úteis para a criação de um modelo de simulação de processo. Essa disponibilidade de dados permite reduzir o tempo de criação de um modelo de simulação, bem como evitar a inserção de parâmetros incorretos pelos usuários.

A seguir, será descrita a extração do estado corrente do log de eventos.

5.2.3.4 Extração do estado corrente dos processos

De uma forma geral, os *logs* de eventos contêm casos em execução que são eliminados das análises, pois são considerados incompletos. No entanto, esses casos podem ser úteis para a criação de modelos de simulação mais realísticos. Os modelos de simulação usados comercialmente iniciam a simulação em estado estático, sem nenhum caso em execução. A simulação é executada por longos períodos, partindo do tempo estático zero. Porém, conforme defendido por Rozinat et. Al (2008a), esse tempo não existe, pois, a execução real de um processo está em contínua mudança.

Um modelo de simulação real não pode iniciar no tempo zero ou vazio. O estado corrente do processo deve ser inserido como ponto inicial da simulação. Essas informações são obtidas do *log* de eventos conforme o Algoritmo 4.

 Algoritmo 4 - extração do estado corrente dos processos

```

Input : Log de eventos (L)
Output: Lista de casos em execução (EC) {<atividade, recurso>}
1  begin
2    PN, IM ← inductive_miner(L) {PN: petrinet e IM: marcação inicial}
3    A ← getActivities(PN, IM) {A: Conjunto de atividades identificadas}
4    FM ← getFinalActivities(L, A) {FM: marcação final}
5    EC = ∅; {CE : conjunto de pares <ai, ri>, ai atividade e ri recurso}
6    AC = ∅; {AC : Atividades correntes ou não finalizadas}
7    for (a = 1 TO A.lenght) do:
8      if a NOT in FM then    {Verifica se a não é atividade final}
9        AC = AC ∪ a;
10     endif
11   endfor
12   logEC ← getFinalActivities(L, AC);
13   for case in enumerate(logEC) do:
14     for event in enumerate(case) do: {Enumera os eventos do case}
15       activity ← event["concept:name"];
16       resource ← event["org:resource"];
17       EC = EC ∪ {<activity, resource>}
18     endfor
19   endfor
20   return EC;
21 end

```

No Algoritmo 4, linha 2, aplica-se o *inductive miner* para descobrir o modelo de processo no formato de rede de Petri e marcação de início do processo. As atividades são recuperadas na linha 3 e na linha 4 são recuperadas as atividades finais do processo. Posteriormente, verifica-se se cada atividade do *log* pertence a uma atividade “final” do processo, da linha 8 a linha 12. Então, verifica-se se o caso foi finalizado ou está em execução. Se o caso não estiver sido concluído, ele é inserido no conjunto AC de estado corrente (linha 10). Um objeto *logEC* é criado com um sub *log* dos casos em execução (linha 13). Esse sub *log* é percorrido e selecionam-se informações como nome da atividade e o recurso que estão atribuídos à atividade e recurso, nas linhas 14 a 20. As informações são usadas para contabilizar os casos em execução e as atividades e executores correntes. O retorno desse algoritmo é um conjunto de pares <atividade, recurso> que estão em execução, ou ativos e ainda não foram concluídos e representam o estado corrente do processo.

A tarefa de extração do estado corrente, nesta pesquisa, é usada para alimentar os parâmetros da inicialização do modelo de simulação. No entanto, outras funções podem ser extraídas com o estado corrente do processo, como suporte operacional.

A seguir será descrita a Fase 3 do PM4SOS – a simulação computacional.

5.2.4 Fase 3 - simulação computacional

A terceira fase do PM4SOS compreende a simulação computacional. Descrevem-se, no diagrama da Figura 24, as etapas do subprocesso do PM4SOS.

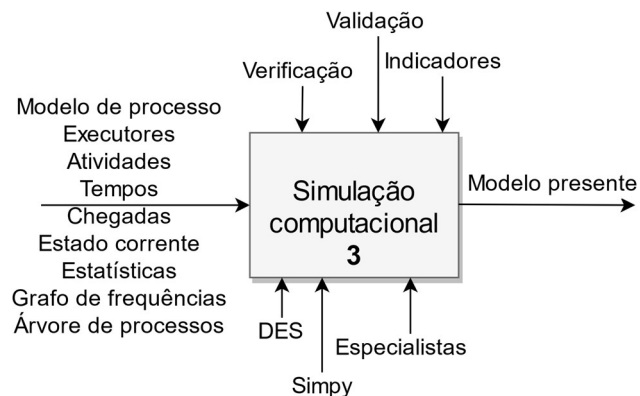


Figura 24 – fase 3: simulação computacional – o autor 2022

Na Figura 24, representa-se o subprocesso simulação computacional. Os parâmetros, estatísticas e objetos obtidos na Fase 2 do PM4SOS definem as entradas para a criação do modelo de simulação. Essa fase usa os mecanismos de controle como a simulação discreta de eventos e biblioteca de simulação *Simpy*. Os especialistas do processo são responsáveis por estabelecer os indicadores do mecanismo de simulação. O resultado dessa fase é o *modelo presente*, ou seja, a representação, por meio do modelo de simulação, do processo descoberto na fase de mineração de processos.

5.2.4.1 Criação do modelo de simulação

A criação do modelo de simulação é realizada de forma automatizada, sem a interação de usuários. A partir dos dados presentes no *log* de eventos, as diferentes extrações de dados são realizadas por meio de *scripts* de programação.

O componente *Simpy* permite executar modelos de simulação usando geradores aleatórios, agindo de forma autônoma com a possibilidade de interagir por meio de interrupções pré-estabelecidas. O Algoritmo 5 mostra os passos necessários para a criação do modelo de simulação.

Algoritmo 5 - geração do modelo de simulação

Input : Árvore de processos{TR}, Grafo Frequência{GF}, Modelo Organizacional (EO), Estado Corrente (CE) e Estatísticas do Log (EL), Número de Instâncias (NI)

Output: Modelo de simulação {SM}

```

1  program createSimulationModel
2  AR ← EL(arrival) {Obtém a taxa de chegada dos cases do Log}
3  SM = ∅;
4  for (elem = 1 TO TR.lenght) do:
5      SR ← EO(activity(elem))           {Taxa de serviço da atividade}
6      RC ← EO(resource(elem))          {Recurso usado na atividade}
7      GF ← GF(TR(activity(elem)))      {Frequência entre atividades}
8      SM ← SM ∪ {SR, RC, GF}
9  endfor
10 SM ← CE; i ← 0
11 while(i < NI) do:
12     for (seq = 1 TO SM.lenght) do:
13         while(timeout random( $f(x) = \lambda e^{-\lambda SM(SR,RC)}$  para  $0 \leq SM(SR,RC) < \infty$ )) do:
14             SVC ← random( $f(x) = \lambda e^{-\lambda S(SR)}$  para  $0 \leq SM(SR) < \infty$ )      {Tx serviço}
15             LOAD ← random( $f(x) = \lambda e^{-\lambda AR/SVC}$  para  $0 \leq AR/SVC < \infty$ )      {Recurso}
16         endwhile
17     endfor
18     i ← i + 1
19     NEXT ← random( $f(x) = \lambda e^{-\lambda AR}$  para  $0 \leq AR < \infty$ )           {Próxima instância}
20 endwhile
21 endprogram

```

O Algoritmo 5 ilustra as principais etapas de geração do modelo de simulação e sua execução. As entradas do algoritmo são respectivamente TR (Árvore de Processos), GF (Grafo de Frequências), MO (Modelo Organizacional), CE (Estado Corrente) e EL (Estatísticas do Log). Inicialmente, a AR mantém a taxa de chegada dos casos obtidos a partir do Log de eventos. O modelo de simulação SM é construído com o mecanismo de sequenciamento de TR desde seu estado inicial até o estado final da árvore de processo. Para cada atividade de TR, são definidos os seguintes parâmetros: SR (Taxa de serviço da atividade), RC (Recurso usado na atividade), GF (Grafo de frequência das atividades) para determinar, por exemplo, a frequência em que as atividades são executadas no modelo. A união dos parâmetros SR, RC e GF geram o modelo de simulação.

A execução do modelo de simulação utiliza um número pré-definido de instâncias para simulação (NI). O modelo de simulação SM inicia com EC (Estado Corrente), na medida que o modelo de simulação deve imitar, mais fidedignamente, a realidade. O estado corrente inclui as atividades em execução e assim o modelo não inicia sem atividades. Essa abordagem evita que a simulação inicie em estado estacionário.

A simulação é executada enquanto houver instâncias a serem executadas. Na linha 12 do Algoritmo 5, as instâncias percorrem o modelo de simulação SM, reservam e alocam os recursos (linha 13), considerando a função exponencial e randômica do tempo de execução das atividades pelos respectivos recursos. Para cada atividade do modelo SM, obtêm-se as taxas de serviço SVC, seguindo uma função exponencial e randômica com base na execução da simulação (linha 14) e a taxa de utilização do recurso LOAD (linha 15). As instâncias entram no sistema de simulação com base na função exponencial e randômica da taxa de chegada dos casos NEXT (linha 19). Os recursos são vinculados a um repositório de recursos para serem utilizados com as atividades. Por exemplo, determinada atividade “Cirurgia” pode conter um repositório com 4 recursos (e.g. Cirurgiões). Para cada instância gerada, recursos são requisitados do repositório. Se todos os recursos estiverem em uso, o sistema gera fila de espera.

Outras funções também são executadas com o modelo de simulação, a saber: a) auditoria a cada intervalo de tempo — enquanto os eventos são gerados na simulação; b) geração de resultados e estatísticas globais para cada instância de simulação; c) visualização de resultados em formato gráfico e texto; e d) registro dos eventos gerados para cada instância da simulação em formato XES.

5.2.4.2 Verificação do modelo de simulação

A verificação do modelo de simulação está relacionada ao modelo computacional (ou modelo implementado em algum software de simulação), e a pergunta que se deve fazer é: “será que se está desenvolvendo corretamente o modelo?” (CHWIF; MEDINA, 2015).

As técnicas utilizadas, nesta tese, para a verificação do modelo de simulação criado foram: a) implementação e verificação modular: a cada módulo, realizam-se testes para verificar o correto funcionamento e corrigir erros detectados ao longo do desenvolvimento do software; b) valores de funções e variáveis: os resultados da simulação são comparados à execução do processo real. Taxas de utilização, taxa de chegada entre instâncias, tempo de serviço para cada atividade são comparados às taxas do processo real; e c) *debugger*: a cada execução, os passos do simulador são acompanhados para verificar o comportamento do simulador e analisar se este está funcionando corretamente.

5.2.4.3 Validação do modelo de simulação

A validação está relacionada ao modelo conceitual. O objetivo da validação do modelo de simulação equivale a responder a seguinte pergunta: “será que se está desenvolvendo o modelo correto?” (CHWIF; MEDINA, 2015).

O mecanismo de validação do modelo de simulação utilizado, nesta pesquisa, é aplicado de duas maneiras:

a) **Comparativo dos logs** (ROZINAT et al., 2009): esta etapa compreende em comparar os *logs* gerados pelo simulador e o *log* inicial inserido para execução da mineração de processos. Dessa forma, é possível observar e comparar os dois comportamentos em nível de resultados estatísticos, tais como uso de recursos, tempo para percorrer o sistema e tempo de espera em fila.

b) **Custo de alinhamento** (VAN ZELST et al., 2020): teste realizado para obtenção do custo do alinhamento entre o *log* simulado e o modelo de processo.

Os parâmetros desejados para validar um modelo de simulação são pré-definidos pelo especialista. Por exemplo, se o limite estabelecido for 75%, as estatísticas da simulação devem atingir esse número e o custo para o alinhamento dos *logs* deve ser inferior a 25% para considerar o modelo de simulação válido. As duas validações descritas acima são realizadas de forma automática, com base nos estabelecidos limites de aceitação.

5.2.4.4 Execução do modelo de simulação

Após o modelo de simulação ser criado de forma automática, verificou e validou-se o correto funcionamento deste com os parâmetros pré-estabelecidos. Em seguida, realizou-se a etapa de execução de experimentos.

O usuário deve definir, previamente, os limites para os indicadores desejados para o sistema de simulação. Esses parâmetros são referentes aos indicadores de filas: a) número médio de instâncias no sistema; b) tempo médio das instâncias no sistema; c) número médio de instâncias em fila; d) tempo médio das instâncias em fila e) utilização dos recursos.

Ao executar o modelo de simulação e obter os dados estatísticos do modelo presente, os valores resultantes são comparados aos parâmetros estabelecidos pelo especialista descritos acima. Se forem detectados desvios, ou seja, se os resultados estiverem fora dos limites estabelecidos, o acionamento do gatilho da otimização é recomendado.

A seguir descreve-se a Fase 4 do PM4SOS - otimização.

5.2.5 Fase 4 - otimização de processos

A quarta fase do PM4SOS, ilustrada na Figura 25, representa, de forma abrangente, o mecanismo de otimização. Essa fase de otimização tem como entradas: (a) resultado obtido com a execução da simulação – modelo presente; e (b) o conjunto de preferências, relativas ao processo em análise. Esse modelo é avaliado e os valores dos indicadores de filas são mensurados a fim de analisar a necessidade de otimização. Quando a otimização se faz necessária, lança-se o mecanismo a fim de se buscarem soluções que contribuam para melhorar os indicadores defasados. Salienta-se que as preferências dos especialistas são consideradas para a definição dos pesos para recuperar soluções similares da base de casos.

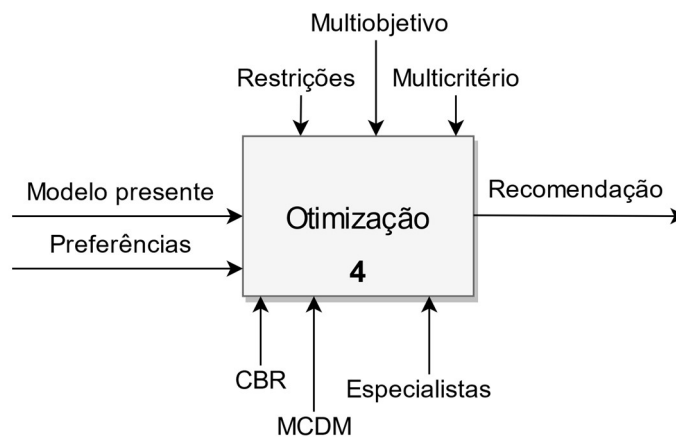


Figura 25 – fase 4: otimização de processos – o autor 2022

A otimização de agendamentos, usado como caso de estudo neste trabalho de pesquisa, é baseada na abordagem multiobjetivo e envolve a satisfação de restrições *rígidas* e *flexíveis*. O método multicritério é utilizado para atuar como um elemento de qualidade na recuperação de casos, que atendam os critérios definidos pelos especialistas. Deve-se notar que cada agendamento de cirurgias pode ser visto como um caso, ou seja, uma ocorrência na base de casos ou de experiências passadas que podem ser reaproveitadas em outros momentos.

De acordo com (SOMAYEH et. al., 2012), o modelo de escalonamento, para determinar a alocação de recursos, envolve salas de cirurgia, cirurgiões, e auxiliares de cirurgias, a sequência de cirurgias dentro salas de cirurgia e seus horários de

início. Os intervalos de tempo para alocação de cirurgias podem ser representados por uma matriz bidimensional.

Considerando-se C , S , P , R , D e T os conjuntos de valores para o problema de agendamento cirúrgico, em que $C=\{c_1, c_2 \dots c_n\}$ é conjunto de cirurgias, $S=\{s_1, s_2 \dots s_n\}$ é o conjunto de cirurgiões, $P=\{p_1, p_2 \dots p_n\}$ é o conjunto de procedimentos, $R=\{r_1, r_2 \dots r_n\}$ é o conjunto de salas de cirurgia, $D=\{d_1, d_2 \dots d_n\}$ é o conjunto de dias prioritários e $T=\{t_1, t_2 \dots t_n\}$ é o conjunto de tempos definidos, há relações fortes – ou *hards* – entre as cirurgias planejadas, cirurgiões e procedimentos cirúrgicos. Supondo que c_1 , s_1 , p_1 seja uma das relações, em que o cirurgião s_1 realiza a cirurgia c_1 e o procedimento p_1 , adicionalmente, s_1 , d_1 , t_1 também definem outra relação, em que uma cirurgia é realizada na sala s_1 , no dia d_1 e horário t_1 . Se há a ligação de c_1 , s_1 , p_1 com r_1 , d_1 , t_1 , o cirurgião s_1 realiza o procedimento p_1 para a cirurgia c_1 , agendada para o dia d_1 , no horário t_1 e na sala r_1 . Dispõem-se essas conexões na Tabela 14.

Tabela 14 – matriz de agendamento cirúrgico.

	c1, s1, p1	c2, s2, p2	c3, s3, p3	...
r1, d1, t1	1	-1	0	...
r1, d1, t2	1	0	1	...
r1, d1, t2	-1	0	1	...
r1, d2, t2	-1	1	0	...
r2, d3, t1	0	0	1	...
...
#Agendados	2	1	3	...

Fonte: o autor 2022

Na Tabela 14, apresenta-se uma matriz M e mlc é uma célula de M , em que l e c representam, respectivamente, suas linhas e colunas. M é a matriz de agendamentos, que pode conter os seguintes valores:

- $mlc = 0$ espaço disponível.
- $mlc = 1$ *slot* ocupado por um horário.
- $mlc = -1$ *slot* não disponível.

O número de *slots* disponíveis N_{slots} é igual a multiplicação do número de dias da semana disponíveis para agendamento D e pelas unidades de períodos (horas) H possível de agendamento em determinado dia conforme a Eq. 14.

$$N_{slots} = D \times H \quad \text{Eq. 14}$$

Na Tabela 15, mostra-se um exemplo de *slots* 30 disponíveis e foi possível alocar 10 cirurgias. O somatório dos tempos cirúrgicos de todas as cirurgias alocadas é de 24 *slots*. No exemplo dado, a taxa de ocupação do agendamento cirúrgico é de 80%.

Tabela 15 – exemplo de slots necessários para agendamento

	Segunda	Terça	Quarta	Quinta	Sexta
13:00-14:00					
14:00-15:00					
15:00-16:00					
16:00-17:00					
17:00-18:00					
18:00-19:00					

Fonte: o autor 2022

Outro aspecto importante da nossa proposta consiste em reusar experiências passadas, ou seja, descrições de problemas e suas respectivas soluções servem de ponto de partida para compor uma nova solução. Essa nova solução se utiliza de dados do processo em relação ao centro cirúrgico e dos agendamentos propriamente ditos.

5.2.5.1 Representação dos casos

A representação dos casos — problemas e suas soluções — é fundamental para se colocar em prática o conjunto de tarefas da abordagem, raciocínio baseado em casos, ou simplesmente, CBR. Na Tabela 16, expõem-se os atributos de descrição e solução de um problema.

Tabela 16 – representação da base de casos

ID	L	Lq	W	Wq	P	Diagnóstico	Resposta	Falha	Solução
1	45	14	15	19	0,65	[1,4]	Falha	Descrição	[]
2	12	10	20	21	0,91	[2,3]	Sucesso	[]	[Agenda]
3	85	13	10	14	0,73	[3]	Sucesso	[]	[Agenda]

Fonte: o autor 2022

Para cada caso, há os seguintes atributos: **ID**, que é um identificador único para cada caso; **atributos do processo**, L é número médio de instâncias no

sistema, L_q é tempo médio das instâncias no sistema, W é número médio de instâncias em fila, W_q é tempo médio das instâncias em fila e P é utilização dos recursos dos indicadores de filas; **diagnóstico** representa o vetor de indicadores que devem ser otimizados; **resposta** indica o resultado da fase de otimização. Possíveis valores para a resposta são: *falha* ou *sucesso*; **falha** é descrição do motivo da falha para ser utilizado na etapa de recuperação. Se o resultado for *sucesso*, a descrição é vazia; e **solução** representa o vetor com o cromossomo da solução do problema em questão. Para o problema de agendamento cirúrgico, a solução é o calendário de agendamentos otimizado.

Uma vez definido o formato de representação de cada caso e sua persistência em uma base de casos, o próximo passo consiste em dispor de um mecanismo para recuperar casos.

5.2.5.2 Recuperação dos casos

A recuperação de casos é fundamental. Ela requer um mecanismo eficiente para encontrar, na base de casos, aqueles mais similares ao problema corrente. A tarefa de recuperação pode ser definida por um conjunto de funções de similaridade para cada valor da base de casos. Utilizou-se a similaridade linear para cada valor.

Para auxiliar a tarefa de recuperação, usamos o método multicritério para estabelecer os pesos entre os atributos da base de casos. Por exemplo, se o gestor tiver como critérios aumentar a taxa de ocupação dos espaços (peso 7) e reduzir a quantidade de pessoas em fila (peso 3), a tarefa de recuperação considera essas preferências ao recuperar os casos da base de casos.

Utilizou-se o método multicritério para elevar a qualidade da obtenção dos casos mais similares. Quanto ao método multicritério, pode-se utilizar aquele que fornece medidas de pesos para cada critério, sendo visto como uma caixa preta. Nesta tese elencou-se o método TOPSIS, pois ele permite ordenar preferências do usuário por meio da avaliação de desempenho das alternativas por meio da similaridade com a solução ideal (HWANG; YOON, 1981).

Para desenvolvimento da tarefa de recuperação da base de casos, utilizou-se *Pycbr*¹². O *Pycbr* é integrado ao PM4SOS como um serviço e ele gerencia a base de

¹² <https://github.com/Dih5/pycbr>

casos. A base de casos é um arquivo no formato “csv”. O algoritmo de recuperação dos casos é baseado na estratégia dos vizinhos mais próximos. Deve-se notar que a base de dados tem um esquema padrão de uma matriz. E, para encontrar os casos mais similares, utiliza-se um vetor de valores para representar a descrição do problema corrente como parte da entrada do algoritmo de busca e a quantidade de casos similares que se deseja obter, por exemplo 5, o que significa que serão retornados os 5 casos mais similares. Em outras palavras, tem-se como retorno para o algoritmo de busca uma lista dos casos similares com seus respectivos índices e escore.

Sobre o caso recuperado, a próxima fase é a sua adaptação. O resultado dessa fase é uma solução que, antes de ser aplicada, é avaliada segundo um conjunto de condições do domínio do problema.

5.2.5.3 Adaptação do caso

Criar um cronograma de cirurgia é um problema que pode ser resolvido usando uma busca heurística. Algoritmos nessa categoria tentam encontrar uma solução ótima, o que pode se tornar ineficaz à medida que a complexidade aumenta. Encontrar uma solução consideravelmente boa pode demorar ou até ser impossível para mais requisitos complexos (GOLDBERG, 1989). A adaptação de casos pode ser *online* ou *offline*. O formato *offline* pode ser demorado. A aquisição *online* leva vantagem de um episódio de adaptação para solicitar pontualmente o especialista. Essa abordagem é motivada pelo objetivo de reduzir o esforço cognitivo (ROLDÁN, et al., 2015). Uma possível solução para isso são os algoritmos genéticos. Esses últimos encerram uma propriedade evolutiva inspirada na biologia, como herança, transformação, seleção natural e recombinação.

A primeira reflexão que se deve fazer ao lidar com um algoritmo genético é como representar nossa solução para adequá-la às operações genéticas como cruzamento e mutação. Adicionalmente, deve-se saber especificar quão adequada é a solução. Para isso, devemos ser capazes de calcular o valor de *fitness* ou aptidão da nossa solução (GOLDBERG, 1989).

A representação de um cromossomo para o problema de agendamento cirúrgico deve gerar *slots* de tempo para alocar procedimentos cirúrgicos. Segmentamos cada *slot* em intervalos de uma hora. Estabelecemos que a primeira cirurgia começará às 6h, a última deve começar às 18h (total de 12 horas), e os dias

úteis são de segunda-feira a sexta-feira (total de 5 dias). Também estabelecemos que o tipo de cirurgia será, apenas, do tipo eletiva, não incluindo urgência e casos de emergência. A escolha do tipo de cirurgia eletiva decorre da forma de execução dos procedimentos e a linearidade encontrada nesse comportamento.

O cálculo de *fitness* de um agendamento cirúrgico é baseado em conjunto de restrições *rígidas* e *flexíveis*. E o resultado da função *fitness* é comparado com um valor pré-estabelecido pelo usuário, por exemplo, 50% de aptidão. E aciona-se o mecanismo de reparação de solução de forma automática se o resultado é inferior ao nível pré-estabelecido.

A função *fitness* retorna uma pontuação com valores entre 0 e 5. A pontuação é incrementada em situações específicas: a) se uma cirurgia usa uma sala de cirurgia livre; b) se uma cirurgia requer equipamento e a sala agendada possui o equipamento; e c) se um cirurgião não estiver sobreposto em outras cirurgias. Não será incrementada a pontuação para o cromossomo que infringir as regras em qualquer intervalo de tempo. A pontuação total para um agendamento cirúrgico consiste na soma dos pontos de todas as cirurgias agendadas. Representamos valores da função *fitness* com valores contínuos entre 0 e 1. Então, calculamos o valor do *fitness* por meio do score gerado para o agendamento dividido pela pontuação máxima. A pontuação máxima consiste no número de cirurgias multiplicado pelo número de dias da semana. A Eq. 15 descreve que a função *fitness* é calculada com base nas penalidades, ou seja, na violação das restrições rígidas. Caso não seja violada nenhuma restrição, a função *fitness* é 1.

$$Fitness = \frac{100}{100 + \sum(\text{penalidades})} \quad \text{Eq. 15}$$

Quando a operação de mutação, a estratégia consiste em selecionar uma cirurgia aleatoriamente e mover para outro *slot* também escolhido aleatoriamente. O tamanho da mutação define o número de cirurgias que são movidas em uma única operação.

No contexto específico desta tese, utilizou-se algoritmo AMGA2 (Micro Algoritmo Genético Baseado em Arquivo). Esse algoritmo opera com tamanho mínimo de população e mantém um arquivo externo com as soluções razoáveis

obtidas e fornece informações sobre o espaço de busca, bem como tende a gerar diversos pontos de Pareto no final da simulação (TIWARI; FADEL; DEB, 2011).

Para a etapa de seleção de cromossomos, o AMGA usa um mecanismo de aptidão com dois níveis: a) aptidão primária: seleção baseada no nível de dominação; e b) aptidão secundária: seleção com base na diversidade das soluções em toda a população. Isso contrasta com o NSGA-II, em que a diversidade é computada, apenas, entre as soluções pertencentes ao mesmo nível (TIWARI et al., 2009).

Ao final de cada ciclo do algoritmo, realiza-se um teste para verificar a adequação da solução encontrada, linha 11 do Algoritmo 6. Esse último apresenta o pseudocódigo das operações realizadas com o algoritmo AMGA2 (TIWARI; FADEL; DEB, 2011).

Algoritmo 6 – pseudocódigo do algoritmo genético AMGA2

```

1  begin
2    Gerar população inicial
3    Avaliar a população inicial
4    Arquivar população inicial
5    repeat
6      Criar população pai a partir do arquivo
7      Criar pool de acasalamento com população-ancestral e arquivar
8      Criar população de descendentes com o cruzamento e mutação
9      Avaliar a população de descendentes
10   Arquivar população de descendentes
11   until(condicao_termino)
12   Mostrar o número desejado de soluções do arquivo
13  end

```

O pseudocódigo do Algoritmo 6 separa, claramente, todas as etapas conceituais do algoritmo AMGA2. O AMGA2 difere de um algoritmo genético convencional quanto aos procedimentos de atualização do arquivo externo, linhas 4 e 10. O diferencial dessa abordagem refere-se ao fato de que as soluções são armazenadas no arquivo somente após serem avaliadas e essas soluções não participam da evolução da população.

A população inicial utilizada neste projeto foi de 100 cromossomos. A taxa de cruzamento foi de 80% e a taxa de mutação de 3% e 2 mutações por cromossomo.

No Algoritmo 7, descreve-se, apenas, o procedimento de atualização do arquivo. As operações envolvidas na atualização do arquivo são comparadas, por meio de um procedimento de verificação de dominância, e as variáveis de decisão e

valores da função objetivo, quando se trata de uma nova solução, é armazenada em arquivo externo (PATIL, 2018).

Algoritmo 7 – atualização do arquivo AMGA2

```

1  begin
2    for  $i_{pop} = 1$  a  $N$  do
3      for  $i_{cell} = 1$  a  $N_{cells}$  do
4        for  $i_{sol}$  in  $i_{cell}$  do
5          Comparar  $i_{pop}$  e  $i_{sol}$  para dominância
6          if  $i_{pop}$  and  $i_{sol}$  são não dominadas then
7            if  $i_{pop}$  and  $i_{sol}$  são idênticas then
8              Prox  $i_{pop}$ 
9            end
10           else if  $i_{sol}$  domina then
11             Prox  $i_{pop}$ 
12           end
13           Encontrar as soluções no arquivo dominadas por  $i_{pop}$  e remover
14           Encontrar  $j_{cell}$  correspondentes a  $i_{pop}$ 
15           if  $j_{cell}$  existe then
16             if  $j_{cell}$  estiver preenchido then
17               Remover randomicamente uma solução de  $j_{cell}$ 
18             end
19             Adicionar  $i_{pop}$  em  $j_{cell}$ 
20           else
21             if  $N_{cells} = N_{cells}^{max}$  then
22               if há células livres then
23                 Comprimir arquivo (remover células livres)
24                 Criar nova célula e adicionar  $i_{pop}$ 
25               else
26                 Declarar "arquivo cheio" e pare
27               endif
28             endif
29           endif
30         endfor
31       endif
32     endfor
33   endfor
34 endfor
35 end

```

O Algoritmo 7 executa a interação sobre a população estabelecida e arquiva as soluções. Os indivíduos da população são avaliados e comparados em termos de dominância de soluções e os valores são armazenados em células (linhas 4 a 13). Caso a célula do arquivo de soluções contenha um valor pior do que aquele da população atual, o valor é removido (linha 14). De outra forma, se o valor de aptidão do cromossomo atual for melhor que o valor aptidão armazenado no arquivo, o valor da solução do arquivo será substituído pela solução correspondente (linhas 17 a 21).

O AMGA gera um número pequeno de novas soluções a cada iteração, e é esse comportamento que o faz ser considerado um micro algoritmo genético. A

geração de um pequeno número de soluções a cada iteração permite reduzir o número de avaliações, minimizando a exploração de regiões e direções de buscas menos promissoras (TIWARI et al., 2008).

5.2.5.4 *Revisão do caso*

Ao concluir a adaptação do caso, se todas as restrições rígidas foram satisfeitas, a etapa de teste é executada e verifica-se, nessa sequência, a porcentagem de restrições flexíveis atendidas. Supondo que a porcentagem de restrições flexíveis atendidas seja menor do que o valor desejado, por exemplo, 50%, então o mecanismo de reparação de solução é executado. Caso a solução encontrada apresente valores superiores ao esperado para as restrições flexíveis, a tarefa de reparação não é executada.

Em nossa proposta, a reparação é implementada usando o algoritmo *Simulated Annealing*. O SA usa um parâmetro de controle de temperatura e um esquema de resfriamento, que evita um ponto ótimo local e encontra uma solução mais próxima de um ótimo global (AARTS; KORST; MICHIELS, 2005).

Para operar com o algoritmo SA, criou-se uma função de custo que leva em conta o esforço para resolver as restrições leves/flexíveis. Assim, o custo total da solução é a soma dos custos para resolver as restrições flexíveis e as restrições rígidas. Os parâmetros da estrutura de resfriamento do SA são: a) temperatura inicial; b) função de queda de temperatura; e c) temperatura final ou condição de parada. Deve-se notar que a probabilidade de aceitar soluções piores que a atual no contexto do SA é maior no início da execução do algoritmo. E à medida que a temperatura cai, o algoritmo encontra um ótimo global com uma probabilidade aproximando-se de 1. O SA foi adaptado de (RUSSEL; NORVIG, 2020) e está descrito no Algoritmo 8.

 Algoritmo 8 - algoritmo adaptado *Simulated Annealing*

```

Input : Solução obtida com o AMGA2
Output: Solução reparada
1  funcaoSA(problema, agenda):
2    solCorrente ← problema.INICIAL; {solução obtida na adaptação}
3    for t=1 a ∞ do:
4      T ← agenda(t)
5      if T = 0 then
6        return solCorrente
7      endif
8      prox ← random(solCorrente)
9      ΔE ← VALOR(solCorrente) - VALOR(prox)
10     if ΔE > 0 then
11       solCorrente ← prox
12     else
13       solCorrente ← prox {somente se probabilidade  $e^{-\Delta E/T}$ }
14     endif
15   endfor
16 endfunction

```

Após a execução da tarefa de reparação da solução, retorna-se à etapa de teste para obter nova avaliação para os valores mínimos desejados. As próximas tarefas, se os valores mínimos forem satisfeitos, consistem na aplicação, solução e registro de uma nova experiência na base de casos.

5.2.5.5 Retenção do caso

Após executar a avaliação e revisão do caso com obtenção dos indicadores desejados, a solução pode ser considerada viável e condizente com os requisitos dados pelas restrições rígidas e flexíveis. Então, gera-se um conjunto de índices para a solução, e ela é armazenada e indexada na base de casos. Essa tarefa de retenção de soluções constitui a base da estratégia de aprendizagem de um sistema de CBR, ou seja, a capacidade do sistema resolver novos problemas aumenta à medida que a base de casos é enriquecida.

Na seção 5.2.5.6 apresenta-se a recomendação da solução obtida na fase de otimização.

5.2.5.6 Recomendação de solução

Além de armazenar a solução na base de casos para futuras consultas e resoluções de problemas similares, aplica-se uma estratégia de recomendação do calendário de agendamento cirúrgico. Essa etapa considera as variáveis de entrada para o

agendamento como: cirurgiões, procedimentos cirúrgicos, salas, restrições rígidas, restrições suaves, datas e horários disponíveis. O relatório gerado é o agendamento que pode ser aceito e operacionalizado pelas equipes responsáveis, ou, no caso de não aceito, reexecutar a fase de otimização buscando novas soluções. A solução não aceita para um problema pode reaparecer para casos com maior similaridade.

De modo sintetizado, ilustra-se a Fase 4 — otimização — na Figura 26.

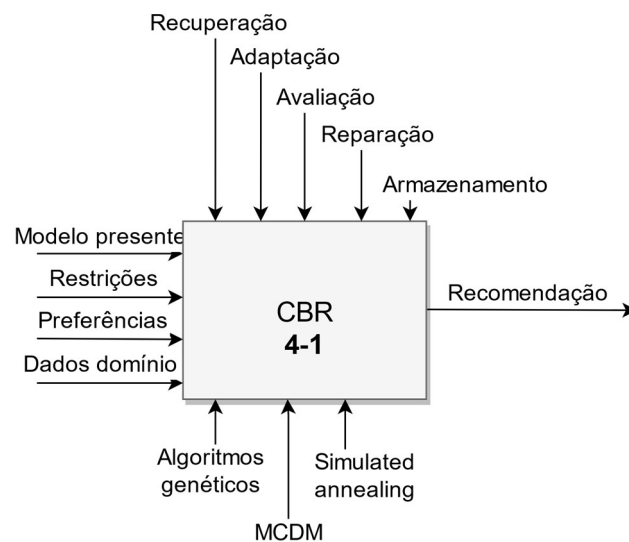


Figura 26 – etapa 4-1: raciocínio baseado em casos – o autor 2022

A Fase 5 do PM4SOS diz respeito à visualização de dados ou de relatórios.

5.2.6 Fase 5 - visualização de dados

A quinta e última fase do PM4SOS compreende a visualização de dados gerados. A Figura 27 mostra o subprocesso, em que se pode destacar a recomendação da solução como entrada e o relatório do agendamento.

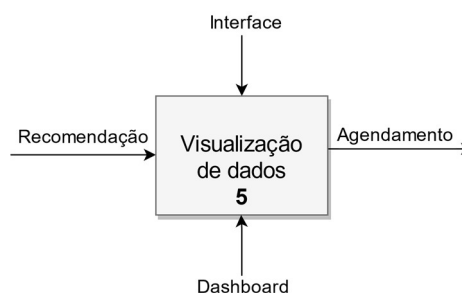


Figura 27 – fase 5: visualização de dados – o autor 2022

A fase de visualização de dados tem como suporte à geração de uma página *web* — *dashboard* —, cuja escolha desse suporte deve facilitar o acesso do usuário e sua experiência com o PM4SOS.

5.3 Resultados

Na etapa de demonstração da aplicação, instanciam-se os artefatos construídos para a resolução de uma ou mais instâncias do problema por meio de experimentos ou simulações, estudos de caso ou provas formais (HEVNER; CHATTERJEE, 2010). Assim, estabeleceu-se uma aplicação com a finalidade de ilustrar o funcionamento operacional do *framework* PM4SOS.

A demonstração do PM4SOS foi dividida em dois segmentos (A e B) para facilitar a exposição e compreensão do tema de forma sistêmica, conforme Tabela 17. Porém a funcionalidade do PM4SOS é completa e inclui os dois segmentos.

Tabela 17 – segmentos da demonstração.

Fase	Descrição	Segmento	Atributos
1	Preparar dados	A,B	Arquivo de <i>log</i> de consultas Arquivo de <i>log</i> de cirurgias
2	Mineração de processos	A,B	Descoberta do modelo de processo, extração de estatísticas, modelo organizacional e estado corrente do processo.
3	Simulação computacional	A,B	Taxa de chegadas, árvore de processos, DFG, executores, atividades, estatísticas da mineração.
4	Otimização de processos	B	Procedimentos, cirurgiões, salas, recursos, restrições, preferências cirurgiões, dias, horários.
5	Visualização de dados	A,B	Mineração de processos, simulação e otimização.

O segmento A compreende o modelo de processo de consulta médica e pode ter dois desfechos distintos (medicação ou cirurgia). O modelo de processo está representado na Figura 17. O segmento B compreende a etapa cirúrgica que inicia no agendamento e encerra na alta hospitalar. Descreve-se, na Figura 18, o modelo de processo, utilizado nesse segmento de demonstração.

Deve-se notar que os executores das tarefas são os cirurgiões que efetuam os procedimentos cirúrgicos, e, para cada um deles, pode-se atribuir um dia preferencial para a consecução das suas cirurgias.

Satisfação de restrições flexíveis: caso seja atribuído ao agendamento cirúrgico o dia de preferência do cirurgião. Esse critério é analisado como uma restrição flexível e o seu tratamento é na etapa de refinamento da solução, utilizando-se da estratégia de reutilização de soluções passadas—CBR. A geração de uma solução que considera, apenas, as restrições flexíveis são sempre possíveis de se obter, pois, mesmo não atendendo tais restrições, a solução final não é comprometida severamente. Contudo, no PM4SOS, pode-se definir um nível aceitável de atendimento das restrições flexíveis; essa definição é parâmetro de configuração do sistema. Por exemplo, se as restrições flexíveis sobre as preferências de horários não atingirem 50% de satisfação, na etapa de otimização — colocada em prática pelo algoritmo AMGA2 —, entra em ação a estratégia de reparação de solução — a etapa CBR —, realizada pelo algoritmo SA (*Simulated Annealing*). Esse último realiza uma busca local nas soluções já encontradas e executa permutações para tentar elevar a quantidade de restrições flexíveis satisfeitas.

Indicadores de filas: a análise dos indicadores de filas é realizada na etapa de simulação do modelo de processo corrente. No entanto, os limites dos indicadores são pré-estabelecidos pelo usuário. Cada indicador de fila encerra os seguintes atributos, a saber: identificador, descrição e tipo. O tipo do indicador define se ele está associado a tempo, pessoa ou recurso. Ainda, cada indicador está associado a um intervalo de valores que define um limite mínimo e um limite máximo. Os limites dos indicadores são importantes, pois, a partir destes se decide pela ativação ou não da otimização do processo em questão.

Fonte de dados: a análise de dados é essencial para indicar a qualidade dos dados dos eventos presentes no log do processo em questão. Essa avaliação gera um conjunto de valores referentes à qualidade dos dados de entrada do PM4SOS. Deve-se notar que, se os dados presentes no *log* de eventos de entrada não possuem qualidade adequada, o restante do processo não terá êxito. No PM4SOS, a etapa de análise da qualidade do *log* de eventos é feita, automaticamente, conforme critérios e faixas de valores pré-estabelecidos. Na Figura 28, mostram-se os critérios de qualidade usados para avaliar o *log* de eventos de entrada, a saber: acurácia, integridade, consistência e singularidade. Deve-se notar, que além dos valores individuais de tais critérios, tem-se a geração de um escore geral da qualidade dos dados de entrada.

De forma concreta, na Figura 28, tem-se a indicação do *log* de eventos em análise, em que “fila_cirurgias.xes” é a identificação do *log* de eventos do processo. A análise desse *log* de eventos resulta no escore geral de qualidade de 7.3. E, para cada atributo de qualidade analisado, há, também, um escore e uma classificação, de notas pelas classes baixo, médio e alto associadas faixas pré-estabelecidas. Na do *log* em questão, detectaram-se 121 eventos duplicados. E a orientação para esse tipo de problema é pré-processar o *log* de eventos, filtrando os eventos duplicados.

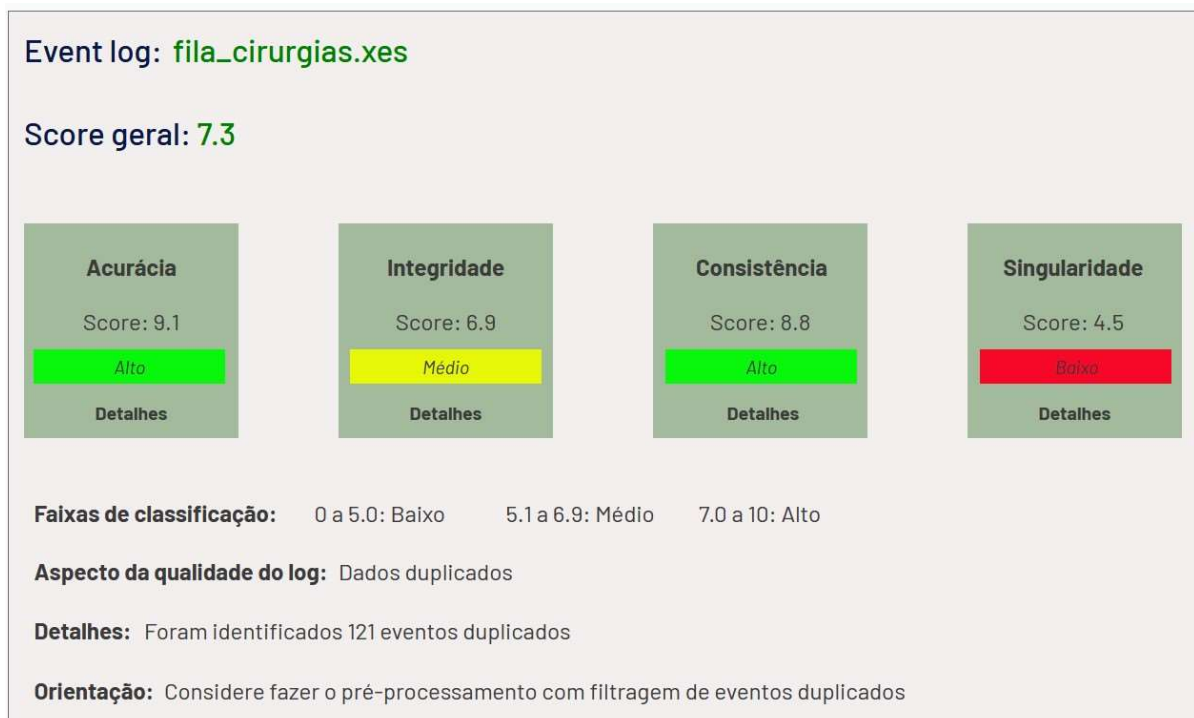


Figura 28 – análise da qualidade dos logs – o autor 2022

Ao avaliar a qualidade do *log* de eventos, deve-se assegurar a existência da qualidade mínima estabelecida, ou seja, um *log* de evento integro, consistente, sem dados duplicados e com alta acurácia para prosseguir para a próxima fase do PM4SOS, a saber: mineração de processos.

Mineração de processos e extração de dados: na Figura 29, mostram-se as informações extraídas do *log* de eventos, seguidas da aplicação da tarefa de descoberta do modelo de processo com base no *log* de eventos em questão. As estatísticas do processo são, a saber: número de casos, variantes utilizadas na descoberta do modelo de processo, ocupação do sistema, taxa de chegada e árvore

de processo. Esses elementos são as estatísticas fundamentais para a criação do modelo de simulação.

O modelo organizacional extraído do *log* de eventos encerra os seguintes elementos básicos, a saber: atividades, recursos. Deve-se notar que há uma relação de execução entre atividade e recurso. Esses elementos e suas relações compõem parte da entrada para a obtenção do modelo de simulação. A obtenção do modelo organizacional a partir do *log* de eventos de um processo permite, intuitivamente, inferir que há uma evidente aproximação do modelo de processo real ao modelo de processo simulado.

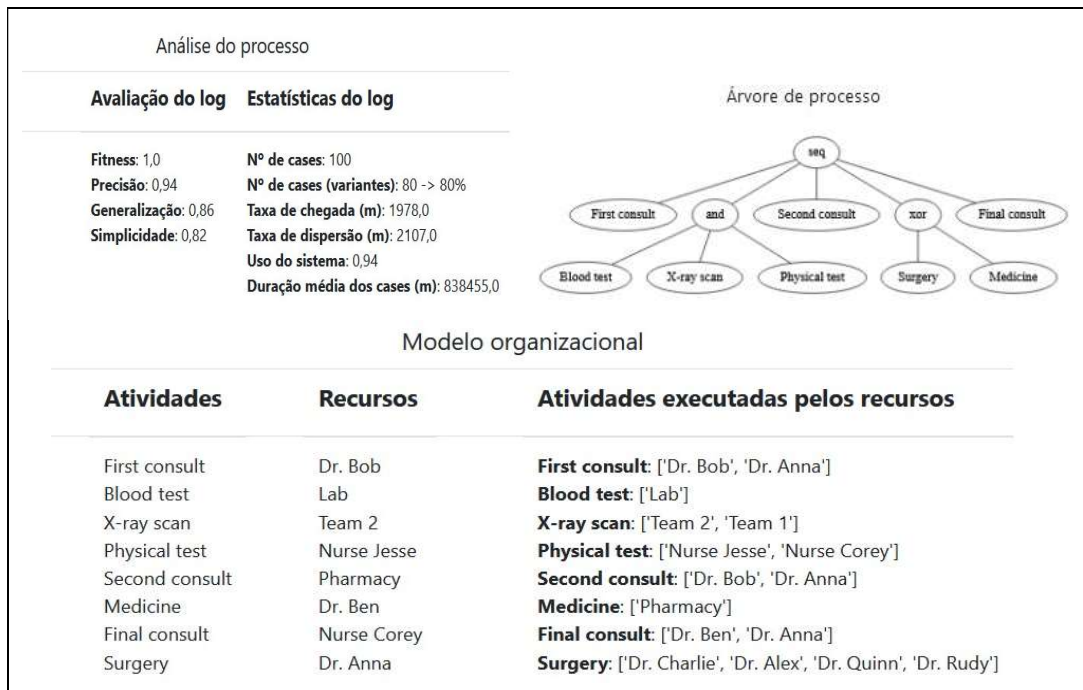


Figura 29 – estatísticas e modelo organizacional – o autor 2022

Modelo de simulação: o modelo de simulação, exibido na Figura 30, mostra os dados necessários para a execução da simulação do processo. A taxa de chegada é utilizada para gerar novas instâncias e controlar as demandas que chegam para ser atendidas no sistema. A árvore de processos guia o modelo de simulação para percorrer as atividades do processo. Isto é feito conforme os operadores de sequência, de paralelismo, ou de decisão para executar o sequenciamento das atividades simuladas com as distribuições de probabilidades.

<p>Modelo de Simulação</p> <p>Taxa de chegada (m): 20</p> <p>Árvore de processos: ->('First consult', +('Blood test', 'X-ray scan', 'Physical test'), 'Second consult', X('Surgery', 'Medicine'), 'Final consult')</p> <p>Ordem de execução das atividades: ['First consult', 'Physical test', 'X-ray scan', 'Blood test', 'Second consult', 'Surgery', 'Final consult']</p> <p>Tempo das atividades (m): {'First consult': 38.0, 'Physical test': 17.0, 'X-ray scan': 68.0, 'Blood test': 2.0, 'Second consult': 70.0, 'Surgery': 85.0, 'Final consult': 52.0}</p> <p>Recursos identificados: ['Dr. Bob', 'Lab', 'Team 2', 'Nurse Jesse', 'Pharmacy', 'Dr. Ben', 'Nurse Corey', 'Dr. Anna', 'Team 1', 'Dr. Charlie', 'Dr. Alex', 'Dr. Quinn', 'Dr. Rudy']</p> <p>Recursos por atividade: {'First consult': 2, 'Physical test': 2, 'X-ray scan': 2, 'Blood test': 1, 'Second consult': 2, 'Surgery': 4, 'Final consult': 2}</p>
--

Figura 30 – modelo de simulação – o autor 2022.

O tempo de execução de cada atividade também é utilizado pelo simulador para informar o tempo de permanecer na atividade. Deve-se observar a presença também dos recursos envolvidos para compor o modelo organizacional em relação à simulação do processo. É importante ressaltar que essa informação deve estar presente no *log* de eventos de entrada no PM4SOS. Se um *log* de eventos não possui as informações relativas aos executores/recursos das atividades, o modelo de simulação será construído sem essa informação. Contudo, o uso do conjunto de recursos, presente no modelo organizacional, eleva a qualidade final do modelo de simulação.

Avaliação do modelo de simulação: além de gerar, automaticamente, o modelo de simulação, prioritariamente com esforços humanos reduzidos, precisa-se avaliar se o modelo gerado representa, coerentemente, um processo real ou de interesse. Na Figura 31, mostra-se a validação do modelo de simulação. A técnica de validação utilizada consiste na comparação do *log* de eventos real — ou de origem — fornecido como entrada no PM4SOS, e do *log* de eventos gerados com a execução do modelo de simulação. A precisão do modelo, no exemplo dado, é de 94% e o custo de alinhamento entre os logs de eventos em questão é de 6.200 unidades, ou seja, significa a quantidade de movimentações necessárias para que o log simulado seja idêntico ao log original. A precisão e o custo são inversamente proporcionais, ou seja, a precisão deve ser máxima possível e o custo de alinhamento mínimo.

Validação do Modelo de Simulação

Alinhamento: Log original x Log Simulado

Fitness: 0,94

Custo : 6200,0

Confirmar a validação do Modelo de Simulação

Modelo validado pelo usuário!

[Analisar Kpis](#)

Figura 31 – validação do modelo de simulação – o autor 2022.

No exemplo em questão (cf. Figura 31), o modelo é indicado como validado pelo usuário. Contudo, o modelo é avaliado pelo PM4SOS, baseando-se nos limites de níveis mínimos exigidos tanto para precisão quanto para o custo de alinhamento.

Na Figura 32, apresenta-se um resumo dos indicadores de fila. É possível visualizar que, tanto o indicador de número médio de instâncias no sistema como o indicador de tempo médio para percorrer o sistema estão com desvios entre os valores medidos e os valores esperados. Mas os valores dos resultados do número médio de instâncias em fila, tempo médio de espera em fila e uso do sistema se encontram dentro dos intervalos esperados.

Análise dos Kpis da simulação com os Indicadores Institucionais

Número médio de instâncias no sistema : 5380 **Atenção** -> Resultado com desvio do Indicador entre [1] e [10]

Número médio de instâncias em fila : 5340 **Atenção** -> Resultado com desvio do Indicador entre [1] e [4000]

Tempo médio de throughput no sistema : 262 **Atenção** -> Resultado com desvio do Indicador entre [1] e [60]

	Variable	N	Mean	SD	SE	95% Conf.	Interval
0	tempo_sistema	47.0	262.1915	157.4624	22.9683	215.9588	308.4242

Tempo médio de espera em fila : 259 **Atenção** -> Resultado com desvio do Indicador entre [1] e [200]

	Variable	N	Mean	SD	SE	95% Conf.	Interval
0	tempo_fila	47.0	259.0638	157.9362	23.0374	212.692	305.4356

Taxa de uso do sistema : 3 **OK!** Resultado entre [1] e [80]

	Variable	N	Mean	SD	SE	95% Conf.	Interval
0	tx_utilizacao	47.0	3.8511	8.4544	1.2332	1.3688	6.3334

Utilize a otimização para melhoramento do processo!

Figura 32 – relatório dos indicadores da simulação – o autor 2022.

O mecanismo de verificação de indicadores de fila é essencial para verificar a necessidade de otimização. Caso os indicadores não apresentem desvios, a otimização na busca de melhoria de indicadores não é lançada. Porém, para efeitos de demonstração, a avaliação dos indicadores encerra a otimização para gerar as soluções de agendamento cirúrgico. Neste trabalho, a estratégia de otimização envolve o suporte de uma base de casos.

Satisfações de restrições flexíveis usando CBR: na Figura 33, mostram-se as instâncias obtidas por meio de cálculo de similaridade entre os atributos que descrevem o problema identificado durante a execução da simulação e os indicadores que apresentam desvios, e os casos já solucionados e armazenados na base de casos. O caso a ser adaptado pode ser selecionado pelo usuário para efeitos didáticos, ou é possível utilizar o caso mais similar dentre os armazenados na base de casos. Se não houver nenhum caso armazenado na base de casos, as informações contidas no resultado da simulação do modelo corrente serão usadas para compor o caso a ser adaptado, testado e reparado (se for necessário).

Casos similares encontrados na base de casos

Case ID	Similaridade	Descrição do Caso
0	0,80373	{'instancias_sistema': 1316.0, 'tempo_sistema': 421.55, 'instancias_fila': 6434.0, 'tempo_fila': 253.04, 'tx_utilizacao': 0.64}
3	0,71066	{'instancias_sistema': 208.0, 'tempo_sistema': 1256.0, 'instancias_fila': 490.58, 'tempo_fila': 311.74, 'tx_utilizacao': 0.3}
1	0,55096	{'instancias_sistema': 8511.0, 'tempo_sistema': 8672.0, 'instancias_fila': 339.27, 'tempo_fila': 274.67, 'tx_utilizacao': 0.55}
2	0,49627	{'instancias_sistema': 7372.0, 'tempo_sistema': 2957.0, 'instancias_fila': 152.28, 'tempo_fila': 364.16, 'tx_utilizacao': 0.15}

Informe o id do caso para adaptar

Adaptar caso CBR

Figura 33 – recuperação de casos por similaridade – o autor 2022.

Ao lançar a opção adaptar caso e considerando os parâmetros registrados do próprio caso selecionado, entra em ação a tarefa de adaptação para o caso em questão. A estratégia de adaptação implementa o algoritmo genético AMGA2. E a primeira ação feita é a verificação se o número de *slots* disponíveis de tempo é maior do que a duração dos procedimentos cirúrgicos. Assim, é possível distribuir as cirurgias para os *slots* observando as restrições.

Deve-se lembrar que a tarefa de otimização requer que todas as restrições rígidas sejam satisfeitas. Já as restrições flexíveis não têm obrigatoriedade de serem satisfeitas para gerar solução completa. No entanto, a solução das restrições flexíveis agrega qualidade e PM4SOS disponibiliza uma que computa o custo para atender as preferências/restrições flexíveis. Por exemplo, para cada cirurgia agendada, verifica-se se a restrição flexível foi satisfeita para o dia de preferência do cirurgião. Caso a preferência não tenha sido satisfeita, a função custo é incrementada.

Em resumo, depois que as restrições rígidas são satisfeitas, a função custo conta quantas restrições flexíveis foram satisfeitas. Se a percentagem de restrições flexíveis satisfeitas for inferior a 50%, por exemplo, acionam-se as regras de reparação de solução. Essa etapa de reparação aplica o algoritmo *Simulated Annealing*, que faz permutas entre cirurgias seguindo as preferências dos cirurgiões. E ao final verifica a efetividade da reparação quanto à redução de custo da solução e à quantidade de restrições flexíveis satisfeitas. Deve-se salientar que as permutas entre procedimentos podem ocorrer somente se os tempos estimados para as execuções das cirurgias forem os mesmos. Essa definição foi adotada para evitar flexibilizar restrições rígidas já atendidas pelo algoritmo genético, como, por

exemplo, a necessidade de equipamentos específicos para os procedimentos cirúrgicos.

Na Figura 34, mostra-se o resultado de um agendamento cirúrgico. A ilustração coloca, lado a lado, para efeito de comparação visual, o agendamento manual, agendamento com algoritmo genético e o com algoritmo *Simulated Annealing*. Duas salas foram consideradas para o agendamento: Sala 2 e Sala 3. Os dias da semana foram de segunda-feira a sexta-feira com horários das 06:00 até as 18:00. E cada célula ocupada apresenta o número da cirurgia agendada. Os slots de tempo sem marcações indicam que não houve agendamentos.

Na Seção 2.3, definiram-se indicadores para analisar mediante os resultados obtidos com o agendamento cirúrgico, a saber: TOC (**T**axa de **O**cupação da **C**apacidade) e TCA (**T**axa de **C**iclo de **A**tendimento). Para a situação em questão, o indicador TOC aumentou em média 17,5% com a otimização e o indicador TCA aumentou em média de 37,5% considerando o número de cirurgias atendidas.

	Agenda Manual					Algoritmo genético AMGA2					Simulated Annealing				
	Seg	Ter	Qua	Qui	Sex	Seg	Ter	Qua	Qui	Sex	Seg	Ter	Qua	Qui	Sex
Sala 2															
06:00 a 07:00	1	4	6	11	14	38		8	14		38		11	8	
07:00 a 08:00	1	4	6	11	14	38		8	14	30	38		11	8	30
08:00 a 09:00		4			14	4	19	12		30	5	19	12		30
09:00 a 10:00	2	4	7	12	14	4	19	12		30	5	19	12		30
10:00 a 11:00	2	4	7	12	14	4		22			5		22		
11:00 a 12:00	2		7				3	22	5	26		3	22	2	26
12:00 a 13:00	3	5	8			2	27	7	5	26	4	27	14	2	26
13:00 a 14:00	3	5	9			2	27	7	5		4	27	14	2	
14:00 a 15:00		5	9	13		2		0	9	39	4		0	9	39
15:00 a 16:00		5	9	13		23		0	9	39	23		0	9	39
16:00 a 17:00		5	9			23	10	0		11	23	10	0		7
17:00 a 18:00			10				10	0		11		10	0		7
Sala 3															
06:00 a 07:00						13	28	34		35	20	28	34		35
07:00 a 08:00	15	19	20	22		13	28	34	24	36	20	28	34	24	36
08:00 a 09:00	15	19	20	22			25	34	24	36		25	34	24	36
09:00 a 10:00	15	19	20	22	24	29	25		18		29	25		18	
10:00 a 11:00		19	20	22		29		16	18	37	29		13	18	37
11:00 a 12:00	16		20	22		20	31	16		37	1	31	13		37
12:00 a 13:00	16		20	22		20	31			37	1	31			37
13:00 a 14:00	17		20			33	31		15	21	33	31		15	21
14:00 a 15:00	17		21	23	25	33	31	32	15	21	33	31	32	15	21
15:00 a 16:00				23	25	33	17	32			33	17	32		
16:00 a 17:00	18			23		33	17	32	6	1	33	17	32	6	16
17:00 a 18:00						33			6	1	33			6	16
	Sala 2	Sala 3				Sala 2	Sala 3				Sala 2	Sala 3			
Slots ocupados	39	32				45	47				45	47			
Ocupação	65%	53%				75%	78%				75%	78%			
Ganhos	-	-				10%	25%				10%	25%			
Slots livres	21	28				15	13				15	13			
Cirurgias	14	11				20	20				20	20			
Restrições leves		Prefer Seg					Prefer Qua					Prefer Sex			

Figura 34 – agendamento completo das salas de cirurgia – o autor 2022.

Para os experimentos com o framework, foi analisada a relação entre o número de salas disponíveis no centro cirúrgico e o número de cirurgias realizadas em cada sala. Para avaliar a correlação entre essas variáveis, foi utilizada uma função linear representada pela Figura 35, que explica a variabilidade dos dados em torno da média. Observou-se que o índice R^2 possui um valor próximo a 1, indicando que o modelo está bem ajustado aos dados e que a correlação entre as variáveis é forte e positiva. Isso significa que à medida que o número de salas disponíveis aumenta, há um aumento proporcional no número de cirurgias realizadas em cada sala. Essa análise é importante para entender a dinâmica do centro cirúrgico e identificar possíveis oportunidades de melhoria no agendamento de cirurgias, que podem ser exploradas pelo framework proposto.

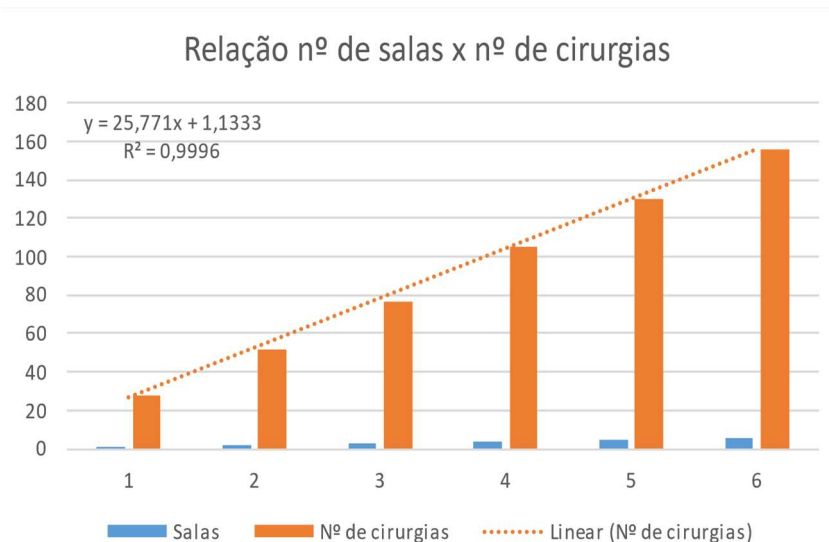


Figura 35 - evolução do tempo de processamento em relação o número de salas e cirurgias

Os experimentos realizados para avaliar a taxa de ocupação das salas cirúrgicas estão apresentados na Tabela 18, na qual é possível observar a variação dessa taxa de acordo com o número de salas disponíveis no centro cirúrgico. A taxa de ocupação média observada foi de 71% dos espaços, o que indica um bom aproveitamento das salas disponíveis para as cirurgias. Além disso, verificou-se que o tempo necessário para executar o framework, principalmente na etapa de otimização, foi significativamente menor do que o tempo necessário para realizar o agendamento manualmente pela equipe de enfermagem. Isso se deve à automatização da tarefa de otimização dos agendamentos, que reduz o tempo necessário para o planejamento e possibilita uma alocação mais eficiente das salas

cirúrgicas. Com essa redução no tempo de planejamento, é possível aumentar a produtividade da equipe de enfermagem e reduzir os atrasos e cancelamentos de cirurgias devido a problemas de agendamento.

Tabela 18 - taxa de ocupação por sala de cirurgia.

	Salas					
	1	2	3	4	5	6
Total slots	60	120	180	240	300	360
Slots ocupados	50	82	140	175	183	226
Taxa ocupação	83%	68%	78%	73%	61%	63%
Nº de cirurgias	28	52	77	105	130	156
Tempo de execução (seg.)	20	30	122	172	189	359
Tempo de execução(min.)	0,3	0,5	2,0	2,9	3,2	6,0
Gerações	681	722	1127	1203	1397	1546

Essa seção mostrou o PM4SOS em ação com uma aplicação de ilustração e reuso de experiências em termos de valores de atributos de casos. Em pesquisas futuras, pretende-se alterar a estrutura do processo para considerar outras atividades e recursos, por exemplo, internamento prévio. Nesse contexto, o reuso de experiências deverá seguir a mesma linha de atuação para permitir a indexação, recuperação e adaptação de casos face a sua estrutura.

5.4 Avaliação

A etapa de avaliação, segundo a metodologia DSRM, orienta encontrar soluções suficientemente adequadas para o contexto em questão. E as soluções devem ser viáveis, não necessariamente ótimas (DRESCH; LACERDA; JÚNIOR, 2020). Nesse contexto, a solução encontrada, com ajuda do PM4SOS, é viável em comparação às suas versões manuais, otimizada, usando a abordagem evolucionária, e aprimorada usando a abordagem *Simulated Annealing*. E, adicionalmente, ela também pode resolver uma classe de problemas semelhantes, especificamente o planejamento e agendamento.

Neste trabalho, o método utilizado para avaliar o artefato produzido foi o método experimental. Esse método é baseado em experimentos controlados em nível de laboratório, nos quais se utiliza modelos que representam um ambiente realista em termos do comportamento do artefato avaliado. O método experimental

permite a realização de testes sistemáticos e rigorosos do artefato, possibilitando a coleta de dados quantitativos e a análise estatística dos resultados. Além disso, esse método possibilita a validação do artefato em diferentes condições e contextos de uso, bem como a comparação com outras soluções existentes.

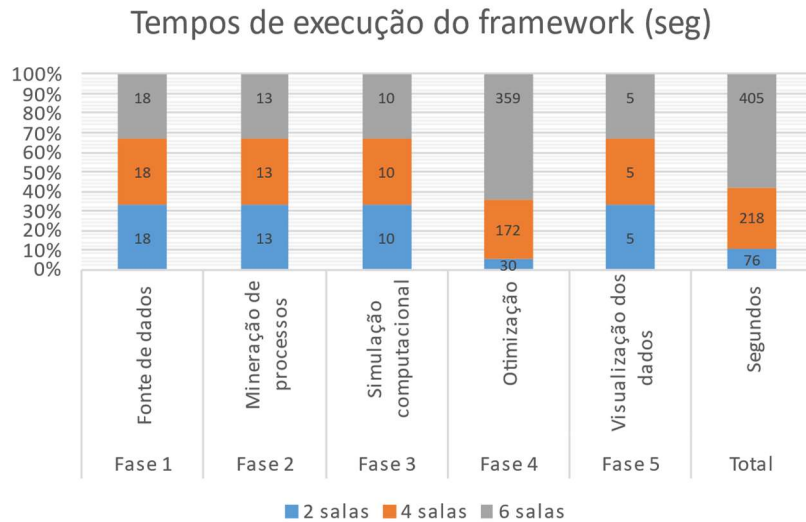


Figura 36 - desempenho do Framework PM4SOS.

A Figura 36 apresenta o desempenho do framework para gerar o agendamento de 2, 4 e 6 salas cirúrgicas, respectivamente. Os resultados indicam que a execução completa de todo o processo do framework para cada uma dessas situações é de 76s, 218s e 405s. É importante destacar que esse tempo cronometrado é significativamente menor em comparação ao esforço manual feito pela equipe de enfermagem para gerar os agendamentos cirúrgicos diários. Com a automatização desse processo por meio do framework desenvolvido, é possível reduzir consideravelmente o tempo necessário para gerar os agendamentos cirúrgicos, aumentando a eficiência do processo e permitindo que a equipe de enfermagem possa se dedicar a outras tarefas relevantes em seu trabalho.

Relembrando o problema de pesquisa, descrito na Seção 1.2, que investiga como os sistemas de informação, em específico *logs* de eventos, podem dar suporte para o processo decisório de gestão operacional com baixo esforço cognitivo e tempo reduzido. Salienta-se que a necessidade de interação do usuário com o protótipo — PM4SOS — é realizada em momentos específicos e dividida em dois momentos: a) antes da execução do fluxo de trabalho do PM4SOS, em que acontece a etapa de parametrização e definição dos limites para os indicadores de desempenho para as filas, e a definição de preferências para a tomada de decisão

multicritério; e b) na validação do modelo de processo, em confirmar ou não o modelo obtido, com base no *log* de eventos bem-formado, representa o processo de forma realista. Essas interações denotam os esforços reduzidos e necessários para interagir/usar o conjunto de conhecimentos presentes no artefato PM4SOS.

Na Figura 34, apresentada na seção anterior, expõe-se a taxa de ocupação das salas de cirurgias, para o cenário de agendamento manual, em comparação ao otimizado, teve aumento de 10% para a Sala 2 e 25% para a Sala 3. Adicionalmente, deve-se notar que, ao utilizar as regras de reparação da solução proposta e otimizada, a taxa de ocupação não foi alterada. Isto aconteceu porque a estratégia de reparação coloca em prática apenas permutas entre cirurgias, sem fazer ajustes de recursos, os quais demandam operações envolvendo restrições rígidas. A tarefa de reparação pode ser vista como um ajuste fino, cujo objetivo é satisfazer restrições flexíveis.

Outra consideração refere-se ao fato de que a taxa de *slots* ociosos foi reduzida e o número de cirurgias maximizado. Empiricamente, essa análise permite inferir que os custos operacionais serão reduzidos, à medida que a ocupação das salas é mais bem ajustada na direção da sua disponibilidade máxima. Pode-se avaliar que, ao reduzir a ociosidade e aumentar o número de cirurgias em cada sala cirúrgica, é possível obter maior produtividade e rentabilidade, embora não se tenham mensurados os custos dos procedimentos cirúrgicos.

Outro indicador em análise refere-se ao tempo de espera do paciente na fila por procedimentos cirúrgicos. Deve-se notar que, quanto maior for a “vazão” do sistema de cirurgias, um número maior de pacientes pode ser atendido no sistema. Assim, quanto melhor for a alocação dos recursos, maiores serão as chances de redução das filas de espera por procedimentos cirúrgicos. Nesta proposta, o aumento do número de atendimentos pode acontecer quando este se maximiza, em particular, às restrições flexíveis. Dedicar atenção às restrições flexíveis proporciona à abordagem proposta maior qualidade da solução, que, na prática, consiste na readequação das cirurgias conforme as preferências dos cirurgiões.

Principais limitações: não atendimento na plenitude das restrições flexíveis com base, unicamente, na estratégia de reparação implementada. As soluções obtidas são consideradas viáveis e não ótimas. Obviamente, pode-se implementar outras estratégias de buscas heurísticas para elevar a taxa de otimização do agendamento de procedimentos cirúrgicos.

5.5 Resumo do capítulo

Neste capítulo apresentaram-se as fases metodológicas 3, 4 e 5 do projeto PM4SOS, respectivamente, desenvolvimento, demonstração e avaliação. Estabeleceram-se, no contexto da RSL (Revisão Sistemática da Literatura), as fases 1 e 2 do PM4SOS, que compreendem a identificação do problema e a definição dos objetivos. A fase 6 corresponde à comunicação e contempla as publicações de artigos científicos cf. Seção 3.8. Descreveram-se, na Figura 35, a representação do fluxo de entradas e saídas, controles e regras geradas com base em cada fase do PM4SOS.

Revela-se, na Figura 37, o fluxo de atividades do PM4SOS para solucionar o problema do agendamento. O ponto de partida é fundamental para o sucesso do processo é a obtenção de um *log* de eventos bem-formatado. Deve-se também notar que atividade de avaliação da qualidade do *log* é fundamental e a sua automatização reduz os esforços iniciais em relação à adequação ou não da continuação da execução das próximas atividades do PM4SOS. Nessa mesma direção, os elementos de informação, extraídos com auxílio da mineração de processos, permitem construir, de forma automatizada, ou seja, sem a intervenção, o modelo de simulação e possibilita também gerar, de forma automatizada, o modelo de execução do processo em operação.

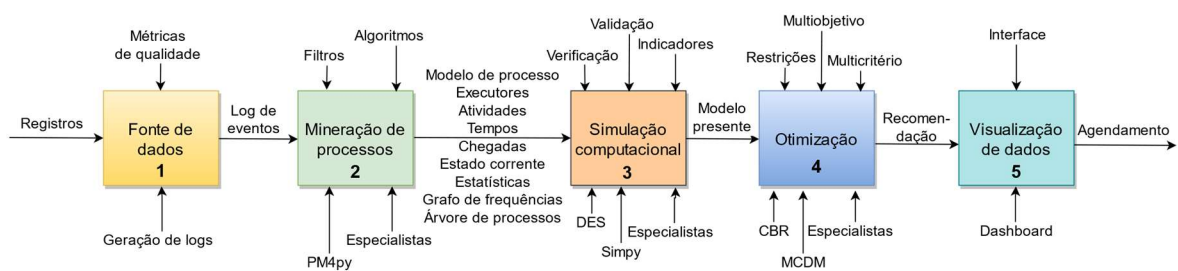


Figura 37 – fluxo de atividades em IDEF0 do PM4SOS – o autor 2022

Deve-se destacar a utilização de indicadores, com suas medidas limítrofes, para indicar que a operação está dentro do esperado e para servir de gatilho para decidir pelo lançamento ou não do procedimento de otimização com vistas a sugerir ajustes no processo que o traga novamente à estabilidade, ou seja, que seus indicadores retornem para dentro dos limites previamente estabelecidos.

Deve-se, também, destacar a utilização de experiências passadas. Estas permitem reaproveitar conhecimentos para resolver novos problemas similares. E a aprendizagem de novas experiências acontece de forma natural, à medida que novas interações acontecem. Entretanto, não se colocou em prática nenhuma estratégia de esquecimento de uma experiência passada ou caso. No caso particular do PM4SOS, apresentou-se a tarefa de adaptação de caso — ou de uma solução passada —, para resolver um problema atual, com ajuda de um algoritmo genérico particular denominado AMGA2. A escolha da abordagem evolucionária para a tarefa de adaptação não foi apenas para implementar as regras de adaptação de casos, mas ela deveria também buscar otimizar a solução adaptada. E, nessa mesma linha, a tarefa reparação de uma solução, quando necessária, foi implementada com ajuda do algoritmo *Simulated Annealing*, que busca melhorar a solução final em termos globais, ou seja, evitando ficar presa em máximo local. A etapa final contempla a recomendação da solução obtida para o problema de agendamento.

A seguir apresentam-se as conclusões deste trabalho.

6 CONCLUSÕES

*Conhecimento não é aquilo que você sabe,
mas o que você faz com aquilo que você sabe.*

(Aldous Huxley)

Esta tese de doutorado tratou da forma como os sistemas de informação podem dar suporte para o processo decisório de gestão operacional com baixo esforço cognitivo e tempo reduzido. O objetivo principal da pesquisa consistiu em projetar, desenvolver, demonstrar e avaliar um *framework* para colocar em prática um conjunto de mecanismos de apoio ao suporte operacional.

A execução do *framework* PM4SOS foi realizada com base em um *log* de eventos e aplicação de técnicas de mineração de processos, simulação computacional, otimização multiobjetivo e com base em resolução de restrições, uso do método multicritério para apoio à tomada de decisão e uso de experiências passadas na forma canônica de raciocínio baseado em casos.

Os objetivos específicos e seus respectivos resultados são os seguintes:

- a) conceber um *framework* para suporte operacional. A consecução deste encerra o desenho do PM4SOS. Esse último, proposto na Seção 5.1, integra eventos e recursos extraídos de logs de eventos de processo de negócio e permite apoiar a tomada de decisão operacional de forma realista. A demonstração da sua adequação foi montada sobre uma aplicação para o problema de planejamento e agendamento de salas de cirurgias.
- b) Construir um ambiente computacional integrando: a mineração de processos, simulação computacional, otimização multiobjetivo e método multicritério de tomada de decisão para permitir gerar suporte operacional com baixo esforço cognitivo e tempo reduzido. Nesse sentido, destaca-se que o conjunto de mecanismos necessários para a automatização do ciclo do processo, que envolve a criação, execução e validação do modelo de simulação, foi colocado em prática na forma de *pipeline*. E a saída desse *pipeline* alimenta um segundo *pipeline* voltado à otimização do uso de recursos em um processo de negócio, no caso específico, agendamento de cirurgias.

- c) Construir uma base de casos: a otimização de processo de negócio baseou-se no ciclo canônico da abordagem CBR. Devem-se destacar as tarefas de adaptação e reparação de casos face às maiores contribuições na estruturação do *pipeline* de otimização do processo, usado para teste, agendamento de cirurgias. A tarefa de adaptação centrou-se na satisfação de restrições rígidas ou *hards*, apoiando-se na abordagem evolucionária do algoritmo genético. E a tarefa de reparação centrou-se na satisfação de restrições flexíveis ou *softs*, usando uma implementação padrão do algoritmo *Simulated Annealing*. A representação e indexação dos casos baseou-se em um conjunto limitado de pares de <atributo, valor>.
- d) Criar uma interface amigável: um *dashboard* foi desenvolvido para permitir que as informações/resultados do processo em análise sejam visualizadas de forma visual e rápida, o que pode contribuir para se obter *insights* da parte do usuário, em particular, na direção de suporte operacional.
- e) Recomendar ações para a tomada de decisões: o resultado principal da operacionalização do *framework* PM4SOS concerne ao agendamento de horários para alocação de recursos em centros cirúrgicos. A demonstração do protótipo operacional, ilustrada na Seção 5.3, mostrou-se adequada, na medida em que a formatação dos resultados obtidos segue o mesmo padrão implementado no dia a dia de um centro cirúrgico.

Destaca-se que os resultados obtidos permitem concluir que o uso de técnicas de processamento de *logs* de eventos, incluindo a avaliação da qualidade dos *logs* de eventos e a mineração de processos, são fundamentais para as etapas subsequentes da criação de modelo de simulação.

Em termos de aplicação, o esforço humano necessário para obter a solução para o agendamento das cirurgias é mínimo. Obviamente, é mínimo quando se considera que os parâmetros necessários para tal já foram fornecidos antes da execução propriamente do PM4SOS, ou seja, a necessidade de interação do usuário é baixa. E, posteriormente à avaliação do usuário, em relação ao modelo do processo, inserem-se os parâmetros de limites mínimos e máximos para os indicadores e informam-se as preferências sobre os critérios da aplicação avaliados no processo de recuperação de casos, e o *pipeline* de execução do PM4SOS não demanda mais nenhuma interação com ele.

O fator tempo também deve ser considerado, visto que, com a automatização de todo o processo de agendamento a partir de *logs* de eventos, reduz-se o esforço do usuário quanto à organização manual dos agendamentos. Se os dados de entrada, como o log de eventos e cadastros de cirurgias, forem fornecidos previamente, o tempo para operacionalização do PM4SOS é adequado à demanda de agendamento diário.

O problema de pesquisa situa-se na classe do planejamento e agendamento. Essa classe de problemas pode ser generalizada para outras tarefas semelhantes, como o agendamento de salas de aulas em uma escola. A classe de problemas contempla os executores (ou agentes) de tarefas, as atividades a serem alocadas (ou procedimentos), os locais onde as atividades serão executadas (locais) e a união de agentes que executam procedimentos em locais específicos com tempo de duração definida e outras preferências a serem determinadas previamente.

A contribuição da mineração de processos face à obtenção automática do modelo de simulação, para a consecução de um processo obtido a partir de um log de eventos bem-formado, representou, em nossa pesquisa, aproximadamente, 90% de fidelidade ao processo original.

A execução do modelo de simulação, com os parâmetros obtidos com a mineração de processos, permitiu replicar o processo original de forma virtual e obter indicadores de gestão de filas. Em modo automático, o PM4SOS verifica se os indicadores de filas estão dentro dos seus limites de aceitação, e, caso não o estiverem, o gatilho de otimização é colocado em ação.

Nessa linha de automatização de tarefas e gatilhos de otimização, o PM4SOS encerrou um ciclo contínuo de operação de agendamento automatizado, cuja ilustração foi feita realizada tese, sobre um caso simples de um centro cirúrgico. E, finalmente, o nível proposto de automatização sugere, claramente, uma redução significativa do esforço cognitivo e de tempo por parte do usuário do PM4SOS.

7 CONSIDERAÇÕES FINAIS

O conhecimento tem um começo, mas não tem um fim.

(Geeta Iyengar)

Neste Capítulo, encerra-se o trabalho de pesquisa com as considerações sobre a relevância do estudo, as contribuições, limitações e trabalhos futuros.

7.1 Relevância do estudo

O planejamento e agendamento cirúrgico são problemas complexos por natureza, na medida em que incluem cirurgias e procedimentos diversos, variando de níveis simples a mais elaborados. E em termos de gestão, a administração hospitalar necessita organizar as salas especializadas para alocar de forma eficiente as cirurgias e diluir os custos e gerar receitas.

A gestão eficiente passa pela organização das agendas cirúrgicas para evitar cancelamentos de procedimentos. Os cancelamentos consomem recursos e tempo do ambiente hospitalar. Outra problemática refere-se ao tempo de espera dos pacientes para o atendimento cirúrgico. Em muitas situações, a espera pode provocar complicações na saúde dos pacientes e até mesmo levá-lo a óbito.

A otimização do processo de planejamento e agendamento cirúrgico é fundamental para obtenção de taxas elevadas de ocupação das salas cirúrgicas e, conseqüentemente, aumentar o número de atendimentos de pacientes que, em geral, aguardam sua vez na fila de espera. Adiciona-se, também, à otimização dos agendamentos cirúrgicos a priorização dos pacientes com cirurgias críticas e que precisam ser executadas de imediato. As preferências dos cirurgias por dias específicos também podem ser contempladas com a automatização e otimização das restrições flexíveis. Deve-se notar que os profissionais estão na categoria de recursos limitados, na medida em que eles, também, exercem outras atividades, como atendimentos em consultórios e/ou docência no Ensino Superior dentre outras.

A relevância desta pesquisa justifica-se pela aplicação prática, com o uso de dados prontamente disponíveis em sistemas de informação, para solucionar o problema do agendamento cirúrgico com base em indicadores do desempenho de filas. A demonstração do PM4SOS revela que se pode maximizar a ocupação dos

centros cirúrgicos, considerando a alocação de cirurgias segundo tempos pré-definidos e atendidas observando otimização multiobjetivo.

A simulação foi aplicada como uma ferramenta para a tomada de decisão operacional. A mineração de processos auxilia a obtenção dos parâmetros, com base em dados históricos, para a composição do modelo de simulação. E os casos em execução podem ser analisados continuamente, permitindo: detectar desvios, promover otimização dos recursos e recomendar ações de acordo com as preferências do usuário.

O suporte à tomada de decisão operacional foi colocado em prática de forma a reduzir o esforço das pessoas que executam a tarefa de agendamento semanal de um centro cirúrgico. O método de tomada de decisão multicritério auxilia a reutilização de casos similares face a otimização de novos agendamentos.

7.2 Contribuições da pesquisa

A pesquisa desenvolvida contribui, tecnicamente, ao propor um ambiente integrado para apoiar a tomada de decisão. A abordagem proposta reduz esforços cognitivos e de tempo na operacionalização do problema de planejamento e agendamento de recursos.

As principais contribuições do PM4SOS são: a) criação do modelo de simulação de forma automatizada com base na árvore de processos — obtida por meio da tarefa de descoberta de modelo de instâncias de processo; b) extração do estado corrente e utilização de cada caso incompleto para apoiar a detecção de desvios, predição de tempos restantes e recomendação de ações; c) integração de método MCDM (e.g., TOPSIS) para recuperar casos semelhantes de uma base de casos; d) padronização da fonte de dados — formato XES, que permite, dado um novo *log* de eventos, gerar um modelo de simulação realista e com baixos esforços cognitivos; e) definição de um fluxo para a simulação e otimização integrando fonte de dados, mineração de processos, simulação e otimização, bem como operacionalizando de tal fluxo em termos de ferramentas computacionais, i.e., colocando em sinergia, em um único ambiente operacional, as necessidades postas em (ROZINAT et. al., 2008b).

7.3 Limitações da pesquisa

Identificaram-se, nesta tese, as seguintes limitações :

- a abrangência do experimento utilizou, apenas, duas salas cirúrgicas. O número de cirurgias utilizadas para o agendamento foi limitado à capacidade máxima de alocação de procedimentos cirúrgicos para cada sala. É notável a necessidade em ampliar o número de salas do experimento para viabilizar a alocação de recursos em centro cirúrgico contemplando todas as salas disponíveis;
- a parametrização da priorização de pacientes pode ser melhorada para melhor eleger pacientes na fila de espera;
- a aplicação do PM4SOS para generalizar a sua aplicação em domínios diversos, como por exemplo, a criação de calendários de alocação de disciplinas em escolas, gerenciamento de palestras em eventos e outras aplicações que necessitem de planejamento e agendamentos de recursos. Obviamente, deve-se observar que o foco é realizar o agendamento partindo de dados de eventos do processo alvo.

7.4 Trabalhos futuros

Esta pesquisa contribui, tecnicamente, para a utilização de dados a fim de gerar soluções assertivas de planejamento e agendamento cirúrgico. No entanto, existem desafios e oportunidades que podem ser investigados em trabalhos futuros.

A automatização e otimização de processos são áreas em constante evolução na direção da eficiência operacional. Maximizar a utilização de recursos disponíveis com redução de custos, baixo esforço e tempo são características necessárias para sobreviver em um ambiente competitivo.

Pretende-se, em futuras pesquisas, ampliar o PM4SOS para operar com operações complexas da árvore de processos, como as operações síncronas. Deve-se, também, explorar melhor o sistema de suporte operacional e tratamento dos casos em modo gráfico e em tempo real.

Levar o trabalho na direção de um “*Digital Twin*”, considerado um gêmeo digital que representa virtualmente um ambiente complexo. O ambiente hospitalar engloba essa complexidade e a possibilidade de simular de forma a granular a realidade. Isso deve permitir tomar decisões antecipadas para evitar desperdícios e aumentar a qualidade do atendimento.

REFERÊNCIAS

- AAMODT, A.; PLAZA, E. Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. **AI Communications**, v. 7, n. 1, p. 39–59, 2001.
- AARTS, E.; KORST, J.; MICHIELS, W. Simulated Annealing. Em: **Search Methodologies: Introductory Tutorials in Optimization and Decision Support Techniques**. Boston, MA: Springer US, 2005. p. 187–210.
- ABDALKAREEM, Z. A. et al. Healthcare scheduling in optimization context: a review. **Health and Technology**, v. 11, n. 3, p. 445–469, 2021.
- ABUKHOUSA, E. et al. Simulation and Modeling Efforts to Support Decision Making in Healthcare Supply Chain Management. **The Scientific World Journal**, v. 2014, p. 1–16, 2014.
- ÁLVAREZ, P. A. et al. A Multi-objective Genetic Algorithm for Inferring Inter-criteria Parameters for Water Supply Consensus. Em: **Lecture Notes in Computer Science**. Switzerland: Springer International Publishing, 2015. v. 9019p. 218–233.
- AMPONSAH, S. K.; AGYEMAN, E.; OKRAH, K. G. Graph colouring, an approach to nurses scheduling, case study: Ejura District Hospital, Ashanti Region, Ghana. **Pioneer Journal of Mathematics and Mathematical Sciences**, v. 1, abr. 2011.
- AUGUSTO, V. et al. **Evaluation of discovered clinical pathways using process mining and joint agent-based discrete-event simulation**. 2016 Winter Simulation Conference (WSC). **Anais...IEEE**, dez. 2016. Disponível em: <<http://ieeexplore.ieee.org/document/7822256/>>
- ÁVILA, S. L. **Otimização paramétrica com computação evolutiva**. 1. ed. Florianópolis: IFSC, 2020.
- BADALLÓ, P. et al. A comparative study of genetic algorithms for the multi-objective optimization of composite stringers under compression loads. **Composites Part B: Engineering**, v. 47, p. 130–136, 1 abr. 2013.
- BALAKHONTCEVA, M. A. et al. **Holistic modeling of chronic diseases for recommendation elaboration and decision making**. *Procedia Computer Science*. **Anais...Elsevier B.V.**, jan. 2018.
- BANKS, J. et al. **Discrete-Event System Simulation**. 5. ed. Harlow: Pearson, 2013.
- BARBALETTA, R.; MARK, W. **Explanation-Based Indexing of cases**. **AAAI**. **Anais...The MIT Press**, 1988.
- BATTITI, R.; PASSERINI, A. Brain-Computer Evolutionary Multi-objective Optimization: A Genetic Algorithm Adapting to the Decision Maker. **IEEE Transactions on Evolutionary Computation**, v. 14, n. 5, p. 671–687, 2010.
- BAZERMAN, M. H.; MOORE, D. A. **Judgment in Managerial Decision Making**. Nova York: John & Wiley & Sons, 2012. v. 8
- BECK, A. Simulation: the practice of model development and use. **Journal of Simulation**, v. 2, n. 1, p. 67, 2008.
- BORGES, A. P. **Uma contribuição para geração de políticas de ações para condução de trens de carga usando raciocínio baseado em casos**. Curitiba: Pontifícia Universidade Católica do Paraná, 2015.
- BORGES, M. E. N. Crença e tomada de decisão: perspectiva de análise do comportamento gerencial para o estudo de uso de informação. **Ciência da Informação**, 2019.
- BORSHCHEV, A.; FILIPPOV, A. **From System Dynamics and Discrete Event to Practical Agent Based Modeling: Reasons, Techniques, Tools**. out. 2004.
- BOTTANI, E.; RIZZI, A. A fuzzy TOPSIS methodology to support outsourcing of logistics services. **Supply Chain Management: An International Journal**, v. 11, n. 4, p. 294–308, jul. 2006.
- BRANS, J.-P.; MARESCHAL, B. Multiple criteria decision analysis: state of the art surveys. Em: New York, NY: Springer, 2005.

- BRASIL. Manual Técnico do Sistema de Informação Hospitalar. **Normas e Manuais Técnicos**, p. 1–198, 2007.
- CAMARGO, M.; DUMAS, M.; GONZÁLEZ-ROJAS, O. Automated discovery of business process simulation models from event logs. **Decision Support Systems**, v. 134, n. March, p. 113284, 2020.
- CARDOEN, B.; DEMEULEMEESTER, E.; BELIËN, J. Operating room planning and scheduling: A literature review. **European Journal of Operational Research**, v. 201, n. 3, p. 921–932, 2010.
- CHO, M. et al. An Evidence-Based Decision Support Framework for Clinician Medical Scheduling. **IEEE Access**, v. 7, p. 15239–15249, 2019.
- CHOUERI, A. C. et al. An extended model for remaining time prediction in manufacturing systems using process mining. **Journal of Manufacturing Systems**, v. 56, p. 188–201, 1 jul. 2020.
- CHWIF, L.; MEDINA, A. C. **Modelagem e simulação de eventos discretos. Teoria e aplicações**. 4. ed. Rio de Janeiro: Elsevier, 2015.
- COELLO, C. A. C.; PULIDO, G. T. A Micro-Genetic Algorithm for Multi-objective Optimization. **Evolutionary Multi-Criterion Optimization**, p. 126–140, 2001.
- DEB, K. Multi-objective optimization using evolutionary algorithms. 2001.
- DEB, K. et al. A fast and elitist multi-objective genetic algorithm: NSGA-II. **IEEE Transactions on Evolutionary Computation**, v. 6, n. 2, p. 182–197, 2002.
- DER AALST, W. VAN; ARCIERI, F.; ADRIANSYAH, A. **Process Mining Manifesto**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. v. 99
- DRESCH, A.; LACERDA, D.; JÚNIOR, J. A. V. **Design science research: método de pesquisa para avanço da ciência e tecnologia**. Porto Alegre: Bookman, 2020.
- DUMAS, M. et al. **Fundamentals of Business Process Management**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013.
- EL MOKHTAR, E.-N.; ZOUHAIR, A. Intelligent dynamic case-based reasoning using multi-agents system in adaptive e-service, e-commerce and e-learning systems. v. 11, p. 42–57, 2016.
- ELHADDAD, Y.; SALLABI, O. A New Hybrid Genetic and Simulated Annealing Algorithm to Solve the Traveling Salesman Problem. **Lecture Notes in Engineering and Computer Science**, 1 jun. 2010.
- FERBER, J. **Multi-Agent Systems: An Introduction to Distributed Artificial Intelligence**. 1st. ed. USA: Addison-Wesley Longman Publishing Co., Inc., 1999.
- FERRONATO, J. J.; SCALABRIN, E. E. **PM2Sim: The Automated Creation of a Simulation Model from Process Mining**. Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics. **Anais...IEEE**, 17 out. 2021.
- FERRONATO, J. J.; SCALABRIN, E. E.; CARVALHO, D. R. **PM4SOS: low-effort resource allocation optimization in a dynamic environment**. 2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC). **Anais...Praga: IEEE**, 2022.
- FISCHER, D. A. et al. Enhancing Event Log Quality: Detecting and Quantifying Timestamp Imperfections. **Business Process Management**, p. 309–326, 2020.
- FONSECA, C. M.; FLEMING, P. J. **Multi-objective optimization**. 1. ed. Berlin: Springer Heidelberg, 2004.
- FU, M. Optimization for simulation: Theory vs. **Practice**. **INFORMS Journal on Computing**, 2002.
- GAREY, M. R.; JOHNSON, D. S. **Computers and intractability: a guide to the theory of NP-completeness**. 1. ed. New York: W. H. Freeman & Co., 1979. v. 37
- GATTO, M. A. F. **Análise da utilização das salas de operações**. Thesis—São Paulo: USP, 1995.
- GEN, M.; CHENG, R. **Genetic Algorithms and Engineering Design**. Ashikaga, Japan: Wiley, 1997.
- GHAZALBASH, S. et al. Operating Room Scheduling in Teaching Hospitals. **Advances in Operations Research**, v. 2012, p. 548493, 2012.

- GOLDBERG, D. E. **Genetic Algorithms in Search, Optimization and Machine Learning**. 1st. ed. USA: Addison-Wesley Longman Publishing Co., Inc., 1989.
- GOLDBERG, D. E.; DEB, K. A Comparative Analysis of Selection Schemes Used in Genetic Algorithms. v. 1, p. 69–93, 1 jan. 1991.
- GUO, Y.; HU, J.; PENG, Y. Research on CBR system based on data mining. **Applied Soft Computing**, v. 11, n. 8, p. 5006–5014, 2011.
- HALAŠKA, M.; ŠPERKA, R. Is there a Need for Agent-based Modelling and Simulation in Business Process Management? **Organizacija**, v. 51, n. 4, p. 255–269, 2018.
- HARREL, C. et al. **Simulation Using ProModel**. McGraw-Hill, 2000.
- HEVNER, A.; CHATTERJEE, S. **Design Research in Information Systems**. 1. ed. New York, NY: Springer, 2010.
- HILLIER, F. S. Introduction to operations research. 1967.
- HSU, P.-Y. et al. Scheduling of Anaesthesia Operations in Operating Rooms. **Healthcare**, v. 9, n. 6, p. 1–640, 2021.
- HWANG, C.-L.; YOON, K. Methods for Multiple Attribute Decision Making. Em: **Multiple Attribute Decision Making**. Berlin, Heidelberg: Springer, 1981. v. 186p. 58–191.
- JACKSON, I.; TOLUJEVS, J.; REGGELIN, T. The Combination of Discrete-Event Simulation and Genetic Algorithm for Solving the Stochastic Multi-Product Inventory Optimization Problem. **Transport and Telecommunication Journal**, v. 19, 2018.
- JACOBSON, R. **Information Design**. Massachusetts: MIT Press, 1999.
- JARDZIOCH, A.; SKOBIEJ, B. Simulated Hardening Algorithm Application in Advanced Planning and Scheduling. **Entrepreneurship and Management**, p. 105–119, 2016.
- JESTON, J.; NELIS, J. **Business Process Management Practical Guidelines to Successful Implementations**. 2. ed. Burlington: Elsevier, 2008. v. 1542
- JUNIOR, F. R. L.; CARPINETTI, L. C. R. Uma comparação entre os métodos TOPSIS e Fuzzy-TOPSIS no apoio à tomada de decisão multicritério para seleção de fornecedores. **Gestão & Produção**, p. 17–34, 2015.
- KATSALIAKI, K.; MUSTAFEE, N. Applications of simulation within the healthcare context. **Journal of the Operational Research Society**, v. 62, n. 8, p. 1431–1451, 2011.
- KHERBOUCHE, M. O.; LAGA, N.; MASSE, P.-A. **Towards a better assessment of event logs quality**. 2016 IEEE Symposium Series on Computational Intelligence (SSCI). **Anais...**2016.
- KHODYREV, I.; POPOVA, S. Discrete modeling and simulation of business processes using event logs. **Procedia Computer Science**, v. 29, p. 322–331, 2014.
- KIRKPATRICK, S.; GELATT, C.; VECCHI, M. Optimization by Simulated Annealing. **Science (New York, N.Y.)**, v. 220, p. 671–680, maio 1983.
- KOBBACY, K. A. H.; VADERA, S.; RASMY, M. AI and OR in management of operations: history and trends. **Journal of the Operational Research Society**, v. 58, maio 2007.
- KOLODNER, J. L. **Case-based Reasoning**. 1. ed. San Francisco, CA: Morgan Kaufmann Publishers, 1993. v. 1
- ŁAWRYNOWICZ, A. Genetic Algorithms for Solving Scheduling Problems in Manufacturing Systems. **Foundations of Management**, v. 3, maio 2011.
- LEEMANS, S. J. J.; FAHLAND, D.; VAN DER AALST, W. M. P. **Discovering block-structured process models from event logs - A constructive approach**. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). **Anais...**2013. Disponível em: <http://link.springer.com/10.1007/978-3-642-38697-8_17>
- LEEMANS, S. J. J.; FAHLAND, D.; VAN DER AALST, W. M. P. Scalable process discovery and conformance checking. **Software and Systems Modeling**, v. 17, n. 2, p. 599–631, 2018.

- LEEMANS, S. J. J.; POPPE, E.; WYNN, M. T. **Directly Follows-Based Process Mining: Exploration amp; a Case Study**. 2019 International Conference on Process Mining (ICPM). **Anais...**2019.
- LI, K.; PAN, Y.; CHENG, B. The Setting and Optimization of Quick Queue. **J Optim Theory Appl**, v. 178, p. 1014–1026, 2018.
- LINDEN, R. **Algoritmos genéticos**. Rio de Janeiro: Brasport, 2008.
- LIU, S. **Integrating Process Mining with Discrete-Event Simulation Modeling**. Theses. **Anais...Provo**: Brigham Young University, 1 nov. 2015. Disponível em: <<https://scholarsarchive.byu.edu/etd/5735>>. Acesso em: 23 mar. 2021
- MÁNTARAS, R. et al. Retrieval, reuse, revision and retention in case-based reasoning. **Knowledge Eng. Review**, v. 20, p. 215–240, 1 set. 2005.
- MARCH, S. T.; SMITH, G. F. Design and natural science research on information technology. **Decision Support Systems**, v. 15, n. 4, p. 251–266, dez. 1995.
- MARIN-CASTRO, H. M.; TELLO-LEAL, E. Event Log Preprocessing for Process Mining: A Review. **Applied Sciences**, v. 11, n. 22, 2021.
- MARQUEZ-CALVO, O. O. **Advancing Robust Multi-Objective Optimisation Applied to Complex Model-Based**. 1st. ed. London: CRC Press, 2020.
- MARR, B. **25 Need-To-Know Key Performance Indicators**. United Kingdom: FT Press, 2014.
- MARTINS, J. R. R. A.; NING, A. **Engineering Design Optimization**. Cambridge: Cambridge University Press, 2021.
- MELVILLE, P.; SINDHWANI, V. Recommender systems. **Encyclopedia of machine learning**, v. 1, p. 829–838, 2010.
- MESABBAH, M.; MCKEEVER, S. **Presenting a hybrid Process Mining framework for automated simulation model generation**. 2018 Winter Simulation Conference (WSC). **Anais...IEEE**, dez. 2018. Disponível em: <<https://ieeexplore.ieee.org/document/8632467/>>
- MIELCZAREK, B. Review of Modelling Approaches for Healthcare Simulation. **Operations Research and Decisions**, v. 26, out. 2016.
- MUN, J. **Advanced Analytical Models: Over 800 Models and 300 Applications from the Basel II Accord to Wall Street and Beyond (Wiley Finance)**. Har/DVD ed. New Jersey: Wiley Publishing, 2008.
- NAKATUMBA, J.; VAN DER AALST, W. M. P. Analyzing Resource Behavior Using Process Mining. Em: **Lecture Notes in Business Information Processing**. Germany: Springer Verlag, 2010. v. 43 LNBIpp. 69–80.
- NEPOTE, M. H. A.; ILZA URBANO MONTEIRO; ELLEN HARDY. Association between operational indexes and the utilization rate of a general surgery center. **Revista Latino-Americana de Enfermagem**, p. 529–534, 2009.
- PATIL, M. B. **Using External Archive for Improved Performance in Multi-Objective Optimization**. Disponível em: <<https://arxiv.org/abs/1811.09196>>. Acesso em: 11 jun. 2022.
- PEFFERS, K. et al. A Design Science Research Methodology for Information Systems Research. **Journal of Management Information Systems**, v. 24, n. 3, p. 45–77, 2007.
- PETROVIC, S.; YANG, Y.; DROR, M. Use of Case Based Reasoning In Solving Examination Timetabling Problems. **European Journal of Operational Research - EJOR**, 1 jan. 2004.
- POURBAFRANI, M.; VAN ZELST, S. J.; VAN DER AALST, W. M. P. **Scenario-based prediction of business processes using system dynamics**. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). **Anais...Springer**, out. 2019. Disponível em: <https://link.springer.com/chapter/10.1007/978-3-030-33246-4_27>
- PRADO, D. **Teoria das Filas e da Simulação**. 4. ed. Belo Horizonte: Editora de Desenvolvimento Gerencial, 2009.

- PRAVILOVIC, S.; APPICE, A.; MALERBA, D. **Process Mining to Forecast the Future of Running Cases**. New Frontiers in Mining Complex Patterns. *Anais...Cham*: Springer International Publishing, 2014.
- PRESLEY, A.; LILES, D. The Use of IDEF0 for the Design and Specification of Methodologies. 1998.
- RAMOS-GUTIÉRREZ, B. et al. **A NLP-Oriented Methodology to Enhance Event Log Quality**. Enterprise, Business-Process and Information Systems Modeling. *Anais...Cham*: Springer International Publishing, 2021.
- REIJERS, H. A.; AALST, W. M. P. VAN DER. Short-Term Simulation: Bridging the Gap between Operational Control and Strategic Decision Making. **Proceedings of the IASTED International Conference on Modelling and Simulation**, p. 417–421, 1999.
- RIESBECK, C. K.; SCHANK, R. C. **Inside Case-Based Reasoning**. USA: L. Erlbaum Associates Inc., 1989.
- ROLDÁN, R. Eduardo; NEGY, Stéphane; ROBLES, Guillermo; LANN, Jean. **Improvement of online adaptation knowledge acquisition and reuse in case-based reasoning: Application to process engineering design**. Engineering Applications of Artificial Intelligence. 41. 10.1016/j.engappai.2015.01.015. 2015.
- ROZINAT, A. et al. **Workflow Simulation for Operational Decision Support Using Design, Historic and State Information**. Business Process Management. *Anais...Berlin, Heidelberg*: Springer Berlin Heidelberg, 2008a.
- ROZINAT, A. et al. Workflow Simulation for Operational Decision Support using YAWL and ProM Table of Contents. **Information Systems**, 2008b.
- ROZINAT, A. et al. Discovering colored Petri nets from event logs. **International Journal on Software Tools for Technology Transfer**, v. 10, n. 1, p. 57–74, 2008c.
- ROZINAT, A. et al. Discovering simulation models. **Information Systems**, v. 34, n. 3, p. 305–327, 2009.
- RUSSELL, S. J.; NORVIG, P. **Artificial Intelligence: A Modern Approach**. 4th. ed. Boston: Pearson, 2020.
- SAGIROGLU, S.; SINANC, D. **Big data: A review**. 2013 International Conference on Collaboration Technologies and Systems (CTS). *Anais...2013*.
- SÀNCHEZ-MARRÈ, M. Principles of Case-Based Reasoning. **Cycle**, p. 1–13, 1994.
- SATALOFF, R. T.; JOHNS, M. M.; KOST, K. M. **BPM CBOOK**. 3. ed. Chicago: ABPMP, 2009.
- SCHOLL, W. et al. **A multi-stage discrete event simulation approach for scheduling of maintenance activities in a semiconductor manufacturing line**. Proceedings - Winter Simulation Conference. *Anais...2012*.
- SIMON, H. A. **The Sciences of the Artificial**. 3. ed. Cambridge, MA, USA: MIT Press, 1996.
- SLACK, N.; LEWIS, M. **The Blackwell Encyclopedic Dictionary of Operations Management**. 2. ed. Australia: Wiley Blackwell, 2005. v. 10
- SOLTI, A. et al. **Repairing event logs using stochastic process models**. Germany: Potsdam University, 2013.
- SOMAYEH, G. et al. Operating Room Scheduling in Teaching Hospitals. abr. 2012.
- SONG, M.; GÜNTHER, C. W.; VAN DER AALST, W. M. P. **Trace clustering in process mining**. Lecture Notes in Business Information Processing. *Anais...Springer Verlag*, 2009.
- SOUZA, V. D. M. **Otimização Multiobjetivo Aplicada em Raciocínio Baseado em Casos**. Curitiba: Pontifícia Universidade Católica do Paraná, 2018.
- STROPARO, J. R.; BICHINHO, G. L.; PROTIL, R. M. **Estudo da taxa de ocupação do centro cirúrgico através da modelagem e simulação de sistemas**. Curitiba: Pontifícia Universidade Católica do Paraná, 2005.
- THABET, D.; GHANNOUCHI, S. A.; BEN GHÉZALA, H. H. **Towards business process model extension with cost perspective based on process mining: Petri net model case**. ICEIS

- 2014 - Proceedings of the 16th International Conference on Enterprise Information Systems. **Anais...SciTePress**, 2014.
- THOMPSON, J. M.; DOWSLAND, K. A. Variants of simulated annealing for the examination timetabling problem. **Annals of Operations Research**, v. 63, n. 1, p. 105–128, 1996.
- TIWARI, S. et al. **AMGA: an archive-based micro genetic algorithm for multi-objective optimization**. Proceedings of the 10th annual conference on Genetic and evolutionary computation - GECCO '08. **Anais...New York, New York, USA: ACM Press**, 2008. Disponível em: <<http://portal.acm.org/citation.cfm?doid=1389095.1389235>>. Acesso em: 9 jun. 2022
- TIWARI, S. et al. **Performance assessment of the hybrid Archive-based Micro Genetic Algorithm (AMGA) on the CEC09 test problems**. 2009 IEEE Congress on Evolutionary Computation. **Anais...2009**.
- TIWARI, S.; FADEL, G.; DEB, K. AMGA2: Improving the performance of the archive-based micro-genetic algorithm for multi-objective optimization. **Engineering Optimization - ENG OPTIMIZ**, v. 43, p. 377–401, maio 2011.
- VAN DER AALST, W. **Process mining: Data science in action**. 2. ed. Berlin Heidelberg: Springer-Verlag, 2016.
- VAN DER AALST, W. et al. A Tour in Process Mining: From Practice to Algorithmic Challenges. Em: KOUTNY, M.; POMELLO, L.; KRISTENSEN, L. M. (Eds.). **Transactions on Petri Nets and Other Models of Concurrency XIV**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2019. p. 1–35.
- VAN DER AALST, W. M. P. Business alignment: Using process mining as a tool for Delta analysis and conformance testing. **Requirements Engineering**, v. 10, n. 3, p. 198–211, 2005.
- VAN DER AALST, W. M. P. et al. **Business Process Simulation: How to get it right?** Handbook on Business Process Management, International Handbooks on Information Systems. **Anais...Berlin: Springer**, out. 2010.
- VAN DER AALST, W. M. P. et al. Soundness of workflow nets: classification, decidability, and analysis. **Formal Aspects of Computing**, v. 23, n. 3, p. 333–363, 2011.
- VAN DER AALST, W. M. P. Business process simulation survival guide. **Handbook on Business Process Management 1: Introduction, Methods, and Information Systems**, p. 337–370, 2015.
- VAN DER AALST, W. M. P. Process mining and simulation: A match made in heaven! **Simulation Series**, v. 50, n. 10, p. 39–50, 2018.
- VAN DER AALST, W.; WEIJTERS, T.; MARUSTER, L. Workflow mining: Discovering process models from event logs. **IEEE Transactions on Knowledge and Data Engineering**, v. 16, n. 9, p. 1128–1142, set. 2004.
- VAN ZELST, S. J. et al. Repairing Alignments of Process Models. **Business & Information Systems Engineering**, v. 62, n. 4, p. 289–304, 2020.
- VERHULST, R. **Evaluating quality of event data within event logs: an extensible framework**. Master's thesis—Eindhoven: Eindhoven University of Technology, 2016.
- WEIJTERS, A. J. M. M.; VAN DER AALST, W. M. P.; ALVES DE MEDEIROS, A. K. Process Mining with the Heuristics Miner Algorithm. **Information Systems IE&IS**, 2006.
- WYNN, M. et al. Process Mining and Simulation. Em: **Modern Business Process Automation**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. p. 437–457.
- WYNN, M. T. et al. **Business process simulation for operational decision support**. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). **Anais...Springer, Berlin, Heidelberg**, 2008. Disponível em: <https://link.springer.com/chapter/10.1007/978-3-540-78238-4_8>
- YANG, X.-S. **Nature-Inspired Optimization Algorithms**. Oxford: Elsevier, 2014.

Anexo I

LAUDO PARA SOLICITAÇÃO DE AUTORIZAÇÃO DE INTERNAÇÃO HOSPITALAR - AIH			
Nº DO PEDIDO		Nº DO LAUDO	
1 - IDENTIFICAÇÃO DO ESTABELECIMENTO DE SAÚDE			
UNIDADE SOLICITANTE		CNES	
UNIDADE EXECUTANTE		CNES	
2 - IDENTIFICAÇÃO DO PACIENTE			
NOME DO PACIENTE			Nº PRONTUÁRIO
CARTÃO NACIONAL DE SAÚDE - CNS	DATA NASCIMENTO	SEXO <input type="checkbox"/> MASC. <input type="checkbox"/> FEM.	RAÇA / COR 1 - BRANCA 2 - NEGRA 3 - PARDA 4 - AMARELA 5 - INDÍGENA ETNIA
NOME DA MÃE			TELEFONE PARA CONTATO
NOME DO RESPONSÁVEL		NATURALIDADE	Nº IDENTIDADE - ORGÃO EMISSOR
ENDEREÇO - RUA, Nº, BAIRRO			
MUNICÍPIO	CÓD. MUNICÍPIO	UF	CEP
3 - DADOS DA INTERNAÇÃO			
DATA DA INTERNAÇÃO		CPF DO DIRETOR CLÍNICO	
4 - JUSTIFICATIVA DA INTERNAÇÃO			
PRINCIPAIS SINAIS E SINTOMAS CLÍNICOS			
CONDIÇÕES QUE JUSTIFICAM A INTERNAÇÃO			
PRINCIPAIS RESULTADOS DE PROVAS DIAGNÓSTICAS (RESULTADOS DE EXAMES REALIZADOS)			
DIAGNÓSTICO INICIAL	CID PRINCIPAL	CID SECUNDÁRIO	CID CAUSAS ASSOCIADAS
5 - PROCEDIMENTO SOLICITADO			
DESCRIÇÃO DO PROCEDIMENTO SOLICITADO			CÓD. DO PROCEDIMENTO
CLÍNICA	CARÁTER DE INTERNAÇÃO	DOCUMENTO <input type="checkbox"/> CNS <input type="checkbox"/> CPF	Nº DOCUMENTO DO PROFISSIONAL SOLICITANTE
NOME PROFISSIONAL SOLICITANTE	CONSELHO DE CLASSE	DATA SOLICITAÇÃO	ASSINATURA - CARIMBO - REGISTRO CONSELHO
6 - PREENCHER EM CASO DE CAUSAS EXTERNAS (ACIDENTES OU VIOLÊNCIA)			
<input type="checkbox"/> ACIDENTE DE TRÂNSITO	CNPJ SEGURADORA	Nº BILHETE	SÉRIE
<input type="checkbox"/> ACIDENTE DE TRABALHO TÍPICO	CNPJ EMPRESA	CNAER DA EMPRESA	CBOR
<input type="checkbox"/> ACIDENTE DE TRABALHO TRAJETO			
VÍNCULO COM A PREVIDÊNCIA <input type="checkbox"/> EMPREGADO <input type="checkbox"/> EMPREGADOR <input type="checkbox"/> AUTÔNOMO <input type="checkbox"/> DESEMPREGADO <input type="checkbox"/> APOSENTADO <input type="checkbox"/> NÃO SEGURADO			
7 - AUTORIZAÇÃO - SUPERVISÃO HOSPITALAR			
SITUAÇÃO DO LAUDO <input type="checkbox"/> APROVADO <input type="checkbox"/> REPROVADO <input type="checkbox"/> MUDANÇA DE PROCEDIMENTO			JUSTIFICATIVA
DOCUMENTO <input type="checkbox"/> CNS <input type="checkbox"/> CPF	Nº DOCUMENTO DO PROFISSIONAL SOLICITANTE	CONSELHO DE CLASSE	
DATA AUTORIZAÇÃO	ASSINATURA E CARIMBO - Nº REGISTRO DO CONSELHO		

