

ANDRÉ PINZ BORGES

DESCOBERTA DE REGRAS DE CONDUÇÃO DE
TRENS DE CARGA

Dissertação apresentada ao Programa de Pós-Graduação em Informática da Pontifícia Universidade Católica do Paraná como requisito para obtenção do título de mestre em Informática Aplicada.

Curitiba
Julho/2009

ANDRÉ PINZ BORGES

DESCOBERTA DE REGRAS DE CONDUÇÃO DE
TRENS DE CARGA

Dissertação apresentada ao Programa de Pós-Graduação em Informática da Pontifícia Universidade Católica do Paraná como requisito para obtenção do título de mestre em Informática Aplicada.

Área de Concentração: Agentes de Software

Orientador: Prof. Dr. Edson Emílio Scalabrin
Co-orientador: Prof. Dr. Fabrício Enembreck

Curitiba

Julho/2009

Borges, André Pinz

Descoberta de Regras de Condução de Trens de Carga. Curitiba. 10-07-2009

Dissertação – Pontifícia Universidade Católica do Paraná. Programa de Pós-Graduação em Informática.

1. Classificação 2. Condução de Trens 3. Mineração de Dados. I. Pontifícia Universidade Católica do Paraná. Centro de Ciências Exatas e Tecnologia. Programa de Pós-Graduação em Informática II - t

FOLHA DE APROVAÇÃO

André Pinz Borges

DESCOBERTA DE REGRAS DE CONDUÇÃO DE TRENS DE CARGA

Dissertação apresentada ao Programa de Pós-Graduação em Informática da Pontifícia Universidade Católica do Paraná como requisito para obtenção do título de mestre em Informática Aplicada.

Banca examinadora

Prof. Dr. _____

Instituição: _____

Julgamento: _____

Assinatura: _____

Prof. Dr. _____

Instituição: _____

Julgamento: _____

Assinatura: _____

Prof. Dr. _____

Instituição: _____

Julgamento: _____

Assinatura: _____

Dedicatória

Aos meus pais, incentivadores
e motivos pela busca ao meu
crescimento.

Agradecimentos

Á Deus, por ter me fornecido capacidade para desenvolver este trabalho e fazer escolhas corretas até o presente momento.

Aos meus pais, por terem sempre me apoiado nas minhas decisões.

Aos professores Dr. Edson Emílio Scalabrin e Dr. Fabrício Enembreck pelas horas dedicadas à realização deste projeto. Em especial ao professor Edson, que me repasou seus conhecimentos e me incentivou ao longo desta pesquisa, sendo um orientador exemplar.

Aos demais colegas integrantes do projeto PAI-L, pelas idéias e colaborações para que este trabalho produzisse frutos e atingisse os objetivos propostos.

E a todos que me ajudaram direta ou indiretamente para que esse trabalho fosse realizado.

Sumário

AGRADECIMENTOS.....	II
SUMÁRIO	III
LISTA DE FIGURAS	V
LISTA DE TABELAS	VII
LISTA DE SÍMBOLOS.....	VIII
LISTA DE TERMOS.....	IX
RESUMO.....	X
ABSTRACT	XI
CAPÍTULO 1 - INTRODUÇÃO.....	1
1.1 MOTIVAÇÃO	3
1.2 OBJETIVOS.....	4
1.3 HIPÓTESE	4
1.4 CONTRIBUIÇÕES.....	5
1.5 ESTRUTURA DO DOCUMENTO.....	5
CAPÍTULO 2 - CONDUÇÃO DE TRENS	6
2.1 MOVIMENTAÇÃO DO TREM.....	8
2.2 FREIOS DE UMA LOCOMOTIVA	10
2.3 TRECHOS ONDULADOS.....	11
2.4 TRECHOS EM CRISTAS	12
2.5 PARTIDA E PARADA DO TREM.....	13
2.6 EQUAÇÕES DE DAVIS	18
2.6.1 Resistências Normais e Acidentais	18
2.6.2 Esforço Trator e Força de Aceleração.....	20
2.6.3 Deslocamento.....	21
2.6.4 Consumo.....	23
2.6.5 Equações de Davis Modificadas	24
2.6.6 Exemplo de aplicação	26
2.7 CONSIDERAÇÕES FINAIS.....	28
CAPÍTULO 3 - DESCOBERTA AUTOMATIZADA DO CONHECIMENTO	30
3.1 O PROCESSO DE DESCOBERTA DO CONHECIMENTO.....	30
3.2 PRÉ-PROCESSAMENTO	31
3.2.1 Transformação dos dados.....	31
3.2.1.1 Remoção de ruídos	32
3.2.1.2 Seleção de atributos	32
3.3 MINERAÇÃO DE DADOS.....	33
3.3.1 Software WEKA	34
3.4 APRENDIZAGEM DE MÁQUINA.....	34
3.4.1 Aprendizagem Simbólica de Máquina.....	35
3.4.2 Aprendizagem Indutiva	35
3.5 MÉTODOS DE CLASSIFICAÇÃO.....	36
3.5.1 Algoritmo C4.5.....	37
3.5.1.1 Algoritmo JRIP.....	45
3.5.2 Método BAGGING com o algoritmo C4.5	46
3.5.3 Método BOOSTING com o algoritmo C4.5.....	50
3.6 VALIDAÇÃO CRUZADA	55
3.7 TESTE ESTATÍSTICO.....	57
3.8 TRABALHOS RELACIONADOS	58
3.9 CONSIDERAÇÕES FINAIS	59

CAPÍTULO 4 - METODOLOGIA	61
4.1 MODELAGEM DOS DADOS	61
4.2 ORIGEM E FORMATO DOS DADOS	62
4.2.1 <i>Remoção de Ruídos</i>	64
4.2.2 <i>Inclusão de Novos Atributos</i>	65
4.2.3 <i>Seleção de Atributos</i>	72
4.2.4 <i>Transformação de Atributos</i>	72
4.3 APLICAÇÃO DOS ALGORITMOS DE MINERAÇÃO	73
4.4 CONSIDERAÇÕES FINAIS	75
CAPÍTULO 5 - SIMULADOR	76
5.1 DESCRIÇÃO GERAL DO SIMULADOR	76
5.2 EXEMPLOS DE CÁLCULOS	78
5.3 USO DOS CLASSIFICADORES	81
5.4 PROCEDIMENTOS DE REFINAMENTOS	83
5.4.1 <i>Controle de Patinação</i>	83
5.4.2 <i>Falta de Potência</i>	85
5.4.3 <i>Marcha Excessiva</i>	86
5.5 EXECUÇÃO DE UMA VIAGEM	87
5.6 CONSIDERAÇÕES FINAIS	93
CAPÍTULO 6 - RESULTADOS	95
6.1 FASE DE APRENDIZAGEM	97
6.2 FASE DE APLICAÇÃO DE CLASSIFICADORES	101
6.2.1 <i>Aplicação do Conhecimento Descoberto</i>	103
6.2.1.1 <i>Viagem A</i>	105
6.2.1.2 <i>Viagem B</i>	107
6.2.2 <i>Análise da Similaridade da Condução</i>	109
6.3 CONSIDERAÇÕES FINAIS	114
CAPÍTULO 7 - CONCLUSÕES	116
7.1 LIMITAÇÕES	117
7.2 TRABALHOS FUTUROS	118
REFERÊNCIAS BIBLIOGRÁFICAS	119

Lista de Figuras

FIGURA 1. BOLETIM DE VIA. (ALL, 2008)	8
FIGURA 2. BOLETIM DE SERVIÇO. (ALL, 2008)	9
FIGURA 3. EXEMPLO DE COMUNICAÇÃO PADRÃO. (ALL, 2008).....	14
FIGURA 4. ETAPAS DO PROCESSO DE KDD.....	31
FIGURA 5 - EXEMPLO DE ÁRVORE DE DECISÃO.	41
FIGURA 6 - ÁRVORE DE DECISÃO NO FORMATO DE REGRAS.....	42
FIGURA 7 - CLASSIFICADOR 1. C4.5 COM ATRIBUTOS CONTÍNUOS.	43
FIGURA 8 - CLASSIFICADOR 1. C4.5 COM ATRIBUTOS CONTÍNUOS (VERSÃO GERADA PELO SOFTWARE WEKA).	43
FIGURA 9 - CÁLCULO DE LIMIARES DO ATRIBUTO <i>PRESSÃO DOS FREIOS</i>	44
FIGURA 10 - MATRIZ DE CONFUSÃO.	44
FIGURA 11 - CLASSIFICADOR 2. BAGGING – L1 E T1.....	49
FIGURA 12 - CLASSIFICADOR 3. BAGGING – L2 E T2.....	49
FIGURA 13 - CLASSIFICADOR 4. BAGGING – L3 E T3.....	49
FIGURA 14 - CLASSIFICADOR 5. BOOSTING – L-1 E T-1.....	53
FIGURA 15 - CLASSIFICADOR 6. BOOSTING – L-2 E T-2.....	54
FIGURA 16 - MODELO CONCEITUAL PARCIAL PARA REPRESENTAR UMA VIAGEM DE TREM EM UMA VIA FÉRREA.	62
FIGURA 17 - EXEMPLO DO ARQUIVO DE VIAGEM.	63
FIGURA 18. DADOS PARCIAIS DO PERFIL DE UMA VIA FÉRREA.....	68
FIGURA 19 – SITUAÇÕES DE CRISTA.....	69
FIGURA 20 – SITUAÇÃO DE DEPRESSÃO.....	69
FIGURA 21 – PONTOS CANDIDATOS PARA CARACTERIZAR CRISTAS E DEPRESSÕES AO LONGO DO TREM.....	69
FIGURA 22. SEGMENTAÇÃO DE UM TREM EM TRÊS PONTOS PARA A IDENTIFICAÇÃO DE CRISTAS E DEPRESSÕES.	70
FIGURA 23. EXEMPLO DE REGRAS NO FORMATO WEKA.	77
FIGURA 24. MODELO CONCEITUAL DO INTERPRETADOR DE REGRAS.	77
FIGURA 25. FORMA DE VOTAÇÃO DOS CLASSIFICADORES.	82
FIGURA 26- EXEMPLO DE PONTO DE MEDIDA NO ARQUIVO XML	88
FIGURA 27- MODELO REPRESENTATIVO DO ALGORITMO DE VIAGEM.....	90
FIGURA 28 - FASE DE APRENDIZAGEM.	96
FIGURA 29 - FASE APLICAÇÃO DO CONHECIMENTO APRENDIDO.....	96
FIGURA 30 – NÚMERO DE REGISTROS VS CLASSES: CN.	97
FIGURA 31 – NÚMERO DE REGISTROS VS CLASSES: C4.....	97

FIGURA 32 – TRECHO REAL DA VIA FÉRREA: DE LONDRINA (PR) ATÉ PAIÇANDU (PR).....	101
FIGURA 33 – USO DOS PONTOS DE ACELERAÇÃO PELA MAQUINISTA SER HUMANO.	102
FIGURA 34 - RESULTADO DA SIMULAÇÃO PARA A CONFIGURAÇÃO A USANDO AS REGRAS DESCOBERTAS POR MEIO DO MÉTODO JRIP+BOOSTING.....	104
FIGURA 35 - RESULTADO DA SIMULAÇÃO PARA A CONFIGURAÇÃO B USANDO AS REGRAS DESCOBERTAS POR MEIO DO MÉTODO JRIP+BOOSTING.....	104
FIGURA 36 - RESULTADO DA SIMULAÇÃO PARA A CONFIGURAÇÃO C USANDO AS REGRAS DESCOBERTAS POR MEIO DO MÉTODO JRIP+BOOSTING.....	104
FIGURA 37 - USO DOS PONTOS DE ACELERAÇÃO NAS VIAGENS 1, 3 E 6 PARA AS CONFIGURAÇÕES A, B E C...	105
FIGURA 38 – DADOS DA VIAGEM A USANDO UM CLASSIFICADOR JRIP.	106
FIGURA 39 - REGRAS FREQUENTEMENTE UTILIZADAS PELO CLASSIFICADOR	106
FIGURA 40 – EXEMPLO DE VIAGEM ANTES (A) E APÓS (B) A REMOÇÃO DAS PARADAS.	107
FIGURA 41 – DADOS DA VIAGEM B USANDO UM CLASSIFICADOR JRIP.	108
FIGURA 42 - REGRA UTILIZADA PELO CLASSIFICADOR JRIP SEM DADOS DE PARADAS.	108
FIGURA 43 – DIFERENÇAS ENTRE PONTOS DE ACELERAÇÃO SUGERIDOS PELO CLASSIFICADOR E APLICADOS PELO MAQUINISTA – CONJUNTO DE TREINAMENTO CN E MÉTODO JRIP.	110
FIGURA 44 - DIFERENÇAS ENTRE PONTOS DE ACELERAÇÃO SUGERIDOS PELO CLASSIFICADOR E APLICADOS PELO MAQUINISTA – CONJUNTO DE TREINAMENTO CN E MÉTODO JRIP+BAGGING.....	110
FIGURA 45 - DIFERENÇAS ENTRE PONTOS DE ACELERAÇÃO SUGERIDOS PELO CLASSIFICADOR E APLICADOS PELO MAQUINISTA – CONJUNTO DE TREINAMENTO CN E MÉTODO JRIP+BOOSTING.....	111
FIGURA 46 - DIFERENÇAS ENTRE PONTOS DE ACELERAÇÃO SUGERIDOS PELO CLASSIFICADOR E APLICADOS PELO MAQUINISTA – CONJUNTO DE TREINAMENTO CN E MÉTODO TODOS.	112

Lista de Tabelas

TABELA 1. TEMPO NECESSÁRIO PARA REDUÇÃO DA PRESSÃO DOS FREIOS.....	13
TABELA 2. PROCEDIMENTOS PARA A PARADA DO TREM.	15
TABELA 3. PROCEDIMENTOS TOMADOS DURANTE A PARTIDA DO TREM.	16
TABELA 4. CONJUNTO DE TREINAMENTO. (EXEMPLO ADAPTADO DE (QUINLAN, 1993))	39
TABELA 5. CONJUNTO DE TREINAMENTO COM VALORES CONTÍNUOS. (EXEMPLO ADAPTADO DE (QUINLAN, 1993)).....	43
TABELA 6. CONJUNTO DE TREINAMENTO. (EXEMPLO ADAPTADO DE (QUINLAN, 1993)).....	47
TABELA 7. CONJUNTO DE APRENDIZAGEM L_1	48
TABELA 8. CONJUNTO DE APRENDIZAGEM L_2	48
TABELA 9. CONJUNTO DE APRENDIZAGEM L_3	48
TABELA 10. CONJUNTO DE TREINAMENTO. (EXEMPLO ADAPTADO DE (QUINLAN, 1993)).....	53
TABELA 11. PESOS ATUALIZADOS.	54
TABELA 12. CONJUNTO DE TREINAMENTO, (EXEMPLO ADAPTADO DE (QUINLAN, 1993)).....	54
TABELA 13. RESUMO DE DADOS DE DIFERENTES VIAGENS.	64
TABELA 14. ATRIBUTOS DERIVADOS POR MEIO DE CÁLCULOS.....	66
TABELA 15 - VALORES CALCULADOS A PARTIR DA SEGMENTAÇÃO DE UM TREM EM ONZE PONTOS PARA A IDENTIFICAÇÃO DE CRISTAS E DEPRESSÕES.....	70
TABELA 16 - FUNÇÕES DE CÁLCULO, PERCEPÇÃO, ATUAÇÃO E CLASSIFICAÇÃO.....	91
TABELA 17 – TAXAS DE ACERTO DOS CLASSIFICADORES OBTIDOS POR MEIO DO MÉTODO JRIP.	98
TABELA 18 - TAXAS DE ACERTO DOS CLASSIFICADORES OBTIDOS POR MEIO DO MÉTODO JRIP+BAGGING.....	98
TABELA 19 - TAXAS DE ACERTO DOS CLASSIFICADORES OBTIDOS POR MEIO DO MÉTODO JRIP+BOOSTING....	99
TABELA 20 - TESTE DE FRIEDMAN PARA OS CLASSIFICADORES OBTIDOS PARA CONJUNTOS DE TREINAMENTOS DE TAMANHOS DIFERENTES.	100
TABELA 21 - CONFIGURAÇÕES DOS EXPERIMENTOS ENTRE TERMOS DE COMPOSIÇÃO E COMPRIMENTO DO TRECHO DE VIA.....	103
TABELA 22 - RESULTADOS USANDO CLASSIFICADORES OBTIDOS A PARTIR DO CONJUNTO DE TREINAMENTO CN.	112
TABELA 23 - RESULTADOS USANDO CLASSIFICADORES OBTIDOS A PARTIR DO CONJUNTO DE TREINAMENTO C4.	113
TABELA 24 - RESULTADOS USANDO CLASSIFICADORES OBTIDOS A PARTIR DO CONJUNTO DE TREINAMENTO SN.	113
TABELA 25 - RESULTADOS USANDO CLASSIFICADORES OBTIDOS A PARTIR DO CONJUNTO DE TREINAMENTO S4.	114

Lista de Símbolos

BAGGING	Bootstrap Aggregating
---------	-----------------------

BOOSTING	Meta-algoritmo de mineração de dados
----------	--------------------------------------

CBL	Computador de bordo das locomotivas
-----	-------------------------------------

HP	Horse Power
----	-------------

JRIP	Algoritmo RIPPER implementado na linguagem Java
------	---

KDD	Knowledge discovery in databases
-----	----------------------------------

Kgf	Quilograma por tonelada
-----	-------------------------

LTKB	Litros por tonelada bruta transportada (medida de consumo das locomotivas)
------	--

PAI-L	Piloto Automático Inteligente para Locomotivas
-------	--

PSI	libras por polegada quadrada (<i>pound force per square inch</i>)
-----	---

Lista de Termos

Ponto de aceleração	Componente análogo à marcha de um carro/caminhão
Ponto de aceleração aberto	Corresponde a aplicação de qualquer ponto de aceleração entre 1 e 8
Ponto de aceleração fechado	Corresponde a não aplicação de um ponto de aceleração
Licença	Conjunto de normas e procedimentos que garantem a circulação segura de um trem
Cluster	Agrupamento formado por dados com características em comum
Outlier	Observação que não corresponde ao modelo dos dados gerado pelo classificador
Ponto de aceleração	Componente análogo à marcha de um carro/caminhão
Ponto de aceleração aberto	Corresponde a aplicação de qualquer ponto de aceleração entre 1 e 8
Ponto de aceleração fechado	Corresponde a não aplicação de um ponto de aceleração
Licença	Conjunto de normas e procedimentos que garantem a circulação segura de um trem

Resumo

Este trabalho visa descobrir padrões para ajudar o maquinista na condução de um trem. A consecução deste objetivo passa pela aplicação de técnicas de aprendizagem de máquina e descoberta de conhecimento a partir de conjuntos de dados de viagens de trens de cargas. O principal resultado é a definição de uma metodologia para a obtenção de padrões de condução segura (e.g. velocidade máxima permitida) e econômica (tempo máximo para uma viagem, consumo mínimo de combustível). O problema de escalabilidade em mineração de dados foi tratado por meio do uso de métodos robustos, tais como, BAGGING e BOOSTING. Os padrões descobertos foram testados de forma objetiva. O processo consistiu, de um lado, aplicar tais padrões para selecionar ações, por meio de um simulador, a serem aplicadas à medida que o trem movia-se, e de outro lado, quantificar a similaridade entre a condução executada pelo simulador e a condução executada pelo maquinista ser humano. Foi observada uma similaridade em torno de 85% entre a condução simulada e a real.

Palavras-chave: Classificação, Condução de Trens, Mineração de Dados.

Abstract

This work aims to find patterns to help the machinist when driving a train. Achieving this objective through by implementation machine learning techniques and knowledge discovery from data sets of travel train loads. The main result is the definition of a methodology to obtain patterns of safe driving (e.g. maximum speed permitted) and economical (maximum time for a trip, less fuel consumption). The problem of scalability in data mining has been addressed through the use of robust methods such as BAGGING and BOOSTING. The patterns discovered were tested in an objective way. The process consisted, of a part, in apply these patterns to select actions, using a simulator, to be applied as the train moved up, and also to quantify the similarity between the conduct performed by driving simulator and implemented the human driver. There was a similarity around 85% between the simulated and real driving.

Keywords: Classification, Drive trains, Data Mining.

Capítulo 1

Introdução

Esta dissertação enquadra-se no contexto do Projeto PAI-L (Piloto Automático Inteligente para Locomotivas). O objetivo geral do PAI-L é a pesquisa & desenvolvimento de um software de condução assistida para locomotivas de trens de carga. Para tal, as locomotivas devem ser dotadas de um computador apropriado, orientando a condução de uma viagem e buscando reduzir o consumo de combustível, o número de acidentes, o congestionamento da malha ferroviária, e o tempo de viagem. O PAI-L é um projeto em desenvolvimento no LAS (Laboratório de Agentes de Software) da PUCPR (Pontifícia Universidade Católica do Paraná), em parceria com empresas que atuam no setor e financiado pela FINEP¹ (Financiadora de Estudos e Projetos). O projeto objeto desta dissertação é um subprojeto do PAI-L. A sua abordagem principal concerne à utilização de técnicas de aprendizagem de máquina para a descoberta de padrões relevantes de condução segura e econômica de trens a partir de bases de dados que incluem os perfis de vias, as características dos trens, os históricos das viagens (dados lidos dos equipamentos de medidas: pressão de freio, velocidade, posição, etc.).

Em termos gerais, utilizar técnicas computacionais para auxiliar o processo de tomada de decisão é uma prática cada vez mais aplicada por empresas de diversos setores. Aprimoramentos ocorridos nos últimos anos, na área informática, propiciam o avanço em áreas que utilizam sistemas informatizados, bem como sistemas embarcados, possibilitando uma integração mais ampla de diferentes áreas do conhecimento (ex: mecatrônica, computação, inteligência artificial). Tal integração propicia, por exemplo, a obtenção de dados confiáveis por meio de sistemas embarcados, sendo o computador de bordo de um trem um exemplo de sistema embarcado. Os dados lidos são confiáveis à medida que se tem a disposição sensores que permitem gerar grandes quantidades de dados com baixo nível de ruído, disponibilizando-os na forma de uma base de dados informatizada.

¹ Convênio 3560/06.

Grandes volumes de dados podem viabilizar inúmeras descobertas. Todavia, a tarefa de análise destes grandes volumes de dados não é trivial. A forma como uma máquina é operada revela a experiência do operador (maquinista), bem como as suas atitudes corretas ou não, que em algumas vezes são imperceptíveis ao olho do ser humano. Tais sutilezas podem ser percebidas/descobertas por meio de técnicas de *mineração de dados*, como por exemplo, a descoberta de padrões de comportamentos de pilotos de aeronaves (ISAAC, et al., 2003) (STOLZER, et al., 2007). Analogamente, no contexto desta dissertação, têm-se o interesse em analisar dados de viagens de trens coletados por diferentes sensores. A análise dos dados visa identificar, por exemplo, diferentes comportamentos executados ou atitudes tomadas pelos maquinistas em diferentes situações rotineiras de condução de um trem de transporte de carga. É importante salientar que a tarefa de conduzir um trem exige habilidades empíricas do operador à medida que as diferentes composições de trens e as diferentes características da via férrea fazem com que uma viagem *A* seja diferente de uma viagem *B*, mesmo quando tais viagens são realizadas em um mesmo trecho.

Canonicamente, por tratar-se de um trabalho de pesquisa de descoberta de padrões, cujas tarefas de mineração foram realizadas sobre dados reais nunca antes submetidos a tal processo, a dificuldade primeira foi à preparação da base de dados. Desta forma, a primeira etapa do trabalho foi um estudo dos modelos dos dados sobre viagens de transporte ferroviário de cargas. Tais dados foram coletados a partir de um conjunto de sensores instalados, em particular, na locomotiva mestra de um trem. A segunda etapa foi o enriquecimento da base de dados. Tal processo gerou uma nova base de dados, contendo dados filtrados por meio de técnicas de seleção e inserção de novos atributos relevantes ao problema, com o objetivo de subsidiar diferentes experimentos de descobertas de padrões e indicar os atributos relevantes ao problema em estudo.

A partir da aplicação de algoritmos de aprendizagem supervisionada, como por exemplo, C4.5 e de métodos de aprendizagem baseados na combinação de classificadores, como BAGGING e BOOSTING, um conjunto de regras foi gerado com base em amostras que representaram proporcionalmente os dados. A utilização de técnicas de amostragem foi necessária devido à complexidade dos dados em termos de número de exemplos, atributos e valores de cada atributo. O interesse primeiro no uso do C4.5, BAGGING e BOOSTING portou essencialmente na geração de regras de fácil compreensão. A escolha das boas regras de condução de trens foi, em um primeiro momento, realizada pelo próprio algoritmo de aprendizagem de máquina, na medida em que ele

gerou classificadores com pequenas taxas de erro, da ordem de 19.5%. Esta última serviu como parâmetro de filtragem de regras aplicáveis à condução dos trens.

A validação dos padrões descobertos foi realizada por meio da aplicação da validação cruzada, do cálculo do cosseno e do Teste de Friedman. Outra forma de avaliação consiste em testar as regras obtidas em um simulador de viagens de trens; para tal foi definido um simulador de condução, que tenta reproduzir o comportamento do maquinista, por meio da aplicação das regras de condução descobertas. No interior do simulador, as ações sugeridas por tais regras são sempre testadas antes de serem aplicadas; caso o teste indique que a ação selecionada produzirá uma situação potencialmente indesejada (e.g. patinação das rodas), um processo iterativo é colocado em prática para corrigir-la. A eficiência do simulador foi medida por meio do cálculo do cosseno, comparando os pontos de aceleração resultantes da classificação com os pontos de aceleração aplicados de pelo maquinista. O cálculo do cosseno foi usado para medir o quão próxima foi a condução realizada pelo simulador e pelo maquinista ser humano. Por fim, a aplicação do Teste de Friedman visou verificar as diferenças entre os classificadores, co-validando a medida de similaridade.

1.1 Motivação

A competitividade no mundo globalizado é altamente acirrada. A eficiência e a eficácia dos processos são mandatárias. O custo do transporte de cargas no Brasil é, em geral, um fator que reduz significativamente os valores repassados, por exemplo, aos produtos alimentícios. Em outras palavras, os elevados consumos de combustíveis geram elevados custos de fretes, o que impede maiores investimentos diretos das empresas neste tipo de transporte, além de dificultar indiretamente as importações e exportações de produtos.

O estilo de condução de um trem influi diretamente no consumo de combustível. Tecnicamente, um ponto de aceleração demora em média de 2 a 3 segundos para gerar a potência no motor (ALL, 2008). Não há um ponto específico que possa ser sempre engrenado num determinado ponto da via, pelo fato de que cada viagem é única; as frequentes mudanças de velocidade máxima permitida em um mesmo trecho de uma via férrea, a distribuição da carga nos veículos ao longo do trem, as condições meteorológicas e as características das locomotivas tornam a repetição de uma configuração quase impossível.

Entretanto, pôde-se constatar que a análise cuidadosa das regras obtidas em um processo de descoberta de conhecimentos permitiu a elaboração de diretrizes de condução potencialmente aplicáveis; testadas em um ambiente computacional simulado.

1.2 Objetivos

O objetivo geral deste trabalho é descobrir padrões a partir de dados coletados por meio de diferentes sensores instalados em um trem. Tais padrões devem servir como forma de reduzir os esforços empregados na escolha do melhor ponto de aceleração.

Desta forma este trabalho compreende em estudar os dados históricos das viagens de trens de carga e empregar técnicas computacionais propostas por Quinlan (1993), Freund e Shapire (1996) e Breiman (1996).

Os objetivos específicos são:

- Realizar um estudo detalhado sobre dados históricos de viagens de trens de cargas, visando compreendê-los e enriquecê-los;
- Extrair, a partir dos dados estudados e enriquecidos, conhecimentos/regras úteis à elaboração de políticas de ações à condução de trens de cargas;
- Validar as regras obtidas;
- Definir um módulo de software de condução simulada que mostre a aplicabilidade potencial das regras descobertas vis-à-vis a elaboração de uma política realista de ações.

1.3 Hipótese

O presente trabalho propõe a extração de regras confiáveis e compreensíveis de condução férrea a partir dos dados históricos de viagens e perfis de via, que evidenciem uma condução econômica e segura de um trem. Acredita-se que a aplicação de técnicas de descoberta de conhecimento, sobretudo algoritmos de mineração de dados gerem soluções eficazes para o problema. A aplicação de algoritmos de mineração de dados mostra-se eficaz na descoberta de regras em bases de dados de diferentes áreas de aplicação. As regras descobertas devem permitir a construção de um simulador que *imite* a condução de trens realizada por um maquinista ser humano. Tal imitação deve ser quantificada por meio de uma medida de similaridade de condução.

1.4 Contribuições

As contribuições científicas do presente trabalho são: (i) a obtenção e a validação de regras de condução de locomotivas úteis à elaboração de políticas de ações potencialmente realistas; e (ii) a definição de um simulador de condução, mostrando a aplicabilidade potencial das regras descobertas vis-à-vis à elaboração de políticas de ações potencialmente realistas.

1.5 Estrutura do Documento

As seções subsequentes estão organizadas da seguinte forma: o Capítulo 2 apresenta um estudo sobre algumas regras de condução que são estudadas pelos maquinistas durante um treinamento. O Capítulo 3 apresenta a fundamentação teórica vis-à-vis o objeto principal do estudo, a mineração de dados. São descritos em detalhes o algoritmo C4.5 e métodos de combinação de classificadores BAGGING e BOOSTING, bem como o método de validação cruzada. O Capítulo 4 apresenta em detalhes a metodologia. O Capítulo 5 apresenta o ambiente de simulação de viagens desenvolvido com o objetivo de usar o conhecimento descoberto, cuja concretização é um software de simulação. O Capítulo 6 apresenta os resultados da pesquisa e uma discussão sobre os mesmos, e por fim, as nossas considerações finais sobre o andamento do projeto de pesquisa são apresentadas no Capítulo 7.

Capítulo 2

CONDUÇÃO DE TRENS

Conduzir um trem é uma tarefa complexa. O conjunto de conhecimentos e habilidades para uma condução eficiente e segura, no particular, é diferente daquele observado na condução de veículos de passeio e caminhões; a aplicação de freios e tempo de aceleração são exemplos destas diferenças.

O estilo de condução de um trem influi diretamente no consumo de combustível. Cada ponto de aceleração (componente análogo à marcha de um caminhão) gera uma determinada força e um determinado consumo. A título de ilustração, uma locomotiva C30 oferece ao seu condutor dez pontos diferentes de aceleração, respectivamente com as seguintes potências, medidas em HP (*horse power*) e consumos, medidos em LTKB (Litros por Tonelada Bruta Transportada): <-1, 0, 0.317> <0, 0, 0.3168> <1, 100, 0.567> <2, 275, 1.0668> <3, 575, 1.95> <4, 960, 3.033> <5, 1440, 4.533> <6, 1930, 6.183> <7, 2500, 7.6998> <8, 2940, 9.4002>. A troca de pontos de aceleração não é obrigatoriamente seqüencial. Esta liberdade de manuseio dos pontos de aceleração não simplifica a tarefa. Há, em função de um conjunto de características particulares e dinâmicas, uma grande dificuldade para definir uma política de condução padrão. Em outras palavras, não há, por exemplo, um ponto específico de aceleração que possa sempre ser empregado, de forma eficiente, em um determinado ponto de uma via. Tal particularidade decorre do fato que cada viagem é única; mudanças de velocidades máximas permitidas, a distribuição da carga nos veículos ao longo do trem, a curvatura da via, as condições meteorológicas e as características das locomotivas dificilmente se repetem. Um bom estilo de condução deve observar os seguintes princípios: (i) evitar danos a via, ao meio ambiente, aos equipamentos e a carga transportada; (ii) realizar uma viagem no menor tempo possível; e (iii) minimizar o consumo de combustível.

Para realizar uma boa viagem é necessário que alguns procedimentos sejam tomados, de modo a aproveitar da melhor forma possível o combustível convertido em energia mecânica, gerando o movimento. Estes procedimentos são:

- Planejar as operações a serem tomadas buscando sempre aproveitar o perfil do terreno e observar atentamente as restrições de velocidades impostas na via. O conhecimento e a antecipação de uma ação são essenciais, por exemplo, não se deve acelerar demasiadamente o trem se logo em seguida há uma restrição de redução de velocidade ou uma parada obrigatória;
- Dar partida no trem ou acelerar apenas quando os freios de todo trem estejam completamente soltos. Aumentar gradativamente os pontos de aceleração em intervalos de dois a três segundos, para o trem absorver melhor a potência do ponto de aceleração engrenado antes de avançar ao próximo;
- Procurar manter velocidade constante ao passar por uma crista, i.e., uma rampa ascendente seguida de uma longa rampa descendente. Este procedimento visa fazer com que quando a maior parte do trem já esteja na rampa descendente o acelerador já esteja fechado (sem aplicar um ponto de aceleração) e o freio dinâmico escolhido para encolher suavemente a folga (intervalo que permanece entre dois vagões e/ou locomotivas) e controlar a velocidade;
- Utilizar o mínimo possível o acelerador aberto (aplicando qualquer ponto de aceleração entre 1 e 8) com os freios aplicados. Este procedimento é realizado apenas quando as folgas precisam estar esticadas. As folgas em um trem não podem ser eliminadas. Cabe ao maquinista controlá-las durante a viagem por meio da utilização adequada do acelerador e dos freios.

Os procedimentos e as regras de condução mencionadas ao longo deste capítulo foram descritos com base em (ALL, 2008). Algumas regras de condução aqui mencionadas não foram implementadas devido à falta de recursos.

As folgas ocorrem quando uma parte do trem se move mais rápida ou mais lentamente do que a outra parte do mesmo. As causas são, por exemplo, o tempo decorrido entre a aplicação do freio ao longo do trem, a variação de atrito, as diferenças na frenagem de vagões vazios e carregados e as variações no perfil do trecho. Um trem com aproximadamente cinquenta veículos tende a ter sete metros de folga, número este que pode chegar a quinze metros e quarenta centímetros para trens com cem veículos e dezenove metros e oitenta centímetros para trens com cento e trinta veículos. Para evitar os impactos resultantes das folgas deve-se manter a aplicação do freio dinâmico até que

os freios não estejam mais sendo aplicados. A forma de aplicação de freios e seus tipos serão discutidos ao longo deste capítulo.

2.1 Movimentação do trem

A movimentação do trem faz-se pelo maquinista e com base em dois boletins. O *boletim A* é destinado para anunciar qualquer situação que interfira na condução segura de um trem, como por exemplo, restrições temporárias da via, velocidade máxima permitida decorrente de algum acidente no trecho ou desnivelamento (ver Figura 1). Tais restrições aparecem também nas licenças repassadas ao maquinista durante a execução de uma viagem. A licença consiste de um conjunto de normas e procedimentos que garantem a circulação segura de um trem. Neste boletim também estão presentes informações referentes às entradas e saídas nos pátios. O *boletim B* é utilizado para anunciar a presença de pessoas e equipamentos trabalhando na via. Este boletim de serviço substitui qualquer instrução especial ou regulamentação que discorde dele sendo numerado, emitido ou cancelado pelo Diretor de Operações da via (ver Figura 2).

FERROVIA SUL- ATLÂNTICO S.A.	
BOLETIM DE VIA FORMULÁRIO A	
N.º _____	Subdivisão _____ 199 _____
A: _____	Em: _____
Correto: _____ horas. Controlador de Tráfego : _____	
Copiado por: _____	

Figura 1. Boletim de via. (ALL, 2008)

As restrições de velocidade são dinâmicas: uma vez encontrado um problema na via que necessite de uma redução de velocidade, uma mensagem é automaticamente enviada a todos os maquinistas que passarão por aquele trecho. Se o maquinista estiver passando pelo ponto que teve velocidade alterada, uma mensagem é mostrada no CBL (Computador de Bordo das Locomotivas), e o mesmo tem 1 minuto para reduzir a velocidade do trem abaixo da velocidade máxima permitida.

FERROVIA SUL- ATLÂNTICO S.A.

BOLETIM DE VIA FORMULÁRIO B

N.º _____ Subdivisão _____ 199 _____

A: _____ Em: _____

_____ regida pela norma 121 dentro dos seguintes limites:

VIA ANULADA	LINHA N.º	Km	LIMITES A	Km	DESDE	ATÉ	VIA(S)	PARE	CHEFE/TURMA
	1.								
	2.								
	3.								
	4.								
	5.								
	6.								
	7.								
	8.								
	9.								
	10.								

Correto: _____ horas. Controlador de Tráfego : _____

Copiado por: _____

Figura 2. Boletim de serviço. (ALL, 2008)

É recomendado que o maquinista mantenha o trem sempre de dois a três km/h abaixo da velocidade linear/cruzeiro para evitar perder velocidade e, conseqüentemente, diminuir consumo. É desejado também que o transporte de uma carga seja feito no melhor tempo possível; vagões e locomotivas são recursos escassos e de alta monta. Deve-se então buscar um compromisso, sem desprezar as restrições constantes nos boletins, entre menor tempo de viagem e menor consumo de combustível.

A velocidade máxima de um trecho é intangível dependendo das características do trem. Isto pode ocorrer, por exemplo, para satisfazer restrições de velocidade de um modelo específico de locomotiva. Há trechos onde um determinado tipo de locomotiva não pode mover-se mais rápido que trinta e cinco km/h, mas a velocidade máxima permitida é de cinquenta km/h.

2.2 Freios de uma Locomotiva

Uma locomotiva possui três tipos diferentes de freios, cuja pressão – medida em *psi* (*pound force per square inch*, em português, libra por polegada quadrada) – é indicada ao maquinista por meio de um dispositivo telemétrico. São eles:

- **Independente:** a ação do freio ocorre somente nas locomotivas, não sendo replicado aos vagões. É recomendável sua utilização de forma isolada para a redução de velocidade. Deve-se observar para tanto que a velocidade esteja abaixo de vinte km/h. O seu uso inadequado pode ocasionar acidentes, bem como rodas deformadas e trincadas. A resposta deste tipo de freio é praticamente instantânea, diferente dos freios presentes nos vagões que possuem um tempo de resposta maior (cerca de um segundo para cada vagão do trem). A restrição de velocidade é imposta para evitar que se busque melhorar o consumo de combustível em detrimento do equipamento. Normalmente, a primeira locomotiva é responsável por sessenta e cinco por cento do emprego do ar para o freio, a segunda locomotiva com vinte por cento e a terceira com quinze por cento; valores definidos pelo fabricante e não configuráveis.
- **Dinâmico:** pode ser comparado com o freio motor de um veículo de passeio. Tem sua utilização recomendada quando a velocidade do trem está entre vinte e trinta km/h. Sua aplicação requer uma espera de dez segundos para a aplicação do freio dinâmico. Se a velocidade do trem for acima de dez km/h não deve ser utilizado juntamente com o freio independente. Deve-se aplicá-lo, em situação de parada em descida e com velocidade abaixo de dez km/h, juntamente com o freio independente para compensar a perda de eficácia do freio dinâmico. A aceleração durante a frenagem dinâmica tem por objetivo fazer com que o motor trabalhe mais, gerando ar para refrigerar os motores de tração. Da mesma forma que o amperímetro mostra a potência gerada para mover uma locomotiva, duran-

te o uso do freio dinâmico, o amperímetro também serve como parâmetro para medir a potência do freio dinâmico.

- **Automático:** corresponde ao freio aplicado aos vagões, resultante da soltura do ar no encanamento dos freios. A primeira aplicação de freios deve variar de cinco a oito *psi*, visto que reduções com valor inferior a cinco *psi* podem resultar em situações como freios agarrados ou alívios indesejados. Tal aplicação de freios deve ser efetuada por vinte segundos, e completada com a aplicação requerida pela situação, sempre tendo a intensidade medida por meio do manômetro. Uma vez aplicado, a soltura do freio automático não deve ser realizada logo após uma aplicação do mesmo. Cabe ao maquinista a ação de aumentar a redução dos freios para dez *psi* antes de fazer a soltura, evitando que o freio fique agarrado.

Um maquinista não pode fazer mais do que certo número de aplicações de freio ao longo da viagem. Por exemplo, não pode fazer mais que dezessete aplicações de freio no trecho entre a cidade Curitiba (PR) e a cidade Morretes (PR). Caso o número de aplicações supere este limite, uma averiguação é realizada para descobrir o motivo de tais aplicações. Caso ele faça uma aplicação de freios errada, antes da terceira aplicação o trem deve ser parado por motivo de segurança. O que determina a frenagem do trem é o perfil no qual ele está movendo-se. É recomendado que utilize cinquenta por cento da capacidade de *psi* para manter a velocidade do trem e a outra metade para uma eventual parada, por exemplo: se há disponível vinte seis *psi* (maioria das locomotivas utiliza este valor), então treze *psi* podem ser usados para manter a velocidade e outros treze *psi* para uma eventual parada. Se observar que o trem precisa de mais de treze *psi* ou mais que quinze *psi* para manter a velocidade, o procedimento padrão é parar e verificar o que pode estar acontecendo.

2.3 Trechos Ondulados

Um trecho de via férrea ondulado caracteriza-se por mudanças freqüentes nas suas rampas, geralmente ocorrendo três ou mais mudanças ascendentes ou descendentes. Neste tipo de terreno, o maquinista deve observar atentamente a relação ao ajuste das folgas, uma vez que parte do trem tende a estar esticado enquanto a outra parte está com as folgas encolhidas. Ao percorrer trechos fortemente ondulados, ou seja, trechos

com rampas maiores que um por cento, as forças aplicadas nos engates tendem a serem maiores; a experiência do maquinista faz-se aqui fundamental para movimentar o trem.

A velocidade do trem em trecho ondulado deve ser controlada acelerando ou desacelerando o trem, fazendo uma redução mínima quando a locomotiva começar a descer o primeiro declive. Quando a cauda do trem passar por uma bacia e começar a subir, a folga do trem tenderá a distender-se, ocasionando aumento nas forças internas do trem. Em geral, para evitar danos, a velocidade do trem deve ser mantida constante em trechos ondulados, observando a aplicação dos freios e em pontos de aceleração mais baixos possíveis. Caso necessite desacelerar o trem, então se deve utilizar o freio dinâmico quando disponível, ou pelo acelerador juntamente com o freio automático.

Quando o trem estiver na parte ascendente da rampa devem-se utilizar pontos de aceleração mais baixos, caracterizando o efeito atrasados dos aclives. Caso o freio automático tenha sido aplicado, então se deve reduzir a velocidade quando o trem começar a descer o primeiro declive, reduzindo ainda de dois a três *psi* se necessário. Não é recomendável utilizar o freio dinâmico ao desacelerar um trem em rampas onduladas. É recomendável iniciar o trecho ondulado em uma velocidade mais baixa do que a velocidade requerida do trecho.

2.4 Trechos em Cristas

As cristas são trechos em que há um longo aclave seguido de um longo declive, não sendo possível especificar uma medida de um trecho longo. Segundo especialistas em condução, um trecho de três quilômetros pode ser considerado longo quando um trem de um quilômetro de extensão o percorre, sendo que o mesmo trecho pode ser considerado curto para um trem de três quilômetros de extensão.

Durante a aproximação de uma crista deve-se reduzir o ponto de aceleração quando as locomotivas ultrapassarem o topo da crista. O acelerador e o freio automático podem ser usados simultaneamente para manter um nível seguro em relação aos engates dos veículos do trem. Outra abordagem possível é a utilização apenas do freio dinâmico ao entrar numa crista. Para isto deve-se reduzir gradualmente o ponto de aceleração, observando atentamente à velocidade das locomotivas ao ultrapassar o topo da crista. Após colocar o trem ponto zero é recomendável esperar dez segundos até acionar o freio dinâmico. O trem deve se inscrever na crista a uma velocidade tal que, quando o mesmo estiver na descendente, a velocidade máxima do trecho não seja ultrapassada.

É possível também utilizar o freio dinâmico juntamente com o freio automático. Neste caso deve-se aliviar o acelerador gradativamente, completando com reduções nos pontos de aceleração quando as locomotivas estiverem ultrapassando o topo da crista. Após ajustadas às folgas do trem deve-se incrementar a frenagem de cinquenta por cento a setenta e cinco por cento da capacidade para então reduzir o freio automático. Se necessário, deve-se fazer reduções de dois a três *psi* para controlar a velocidade do trem.

Pode acontecer que trem esteja numa velocidade baixa e com ponto de aceleração alto. Isto significa que o mesmo está em um trecho de rampa. Tal situação se explica pelo fato do trem não conseguir alcançar a velocidade máxima permitida, apesar de utilizar um alto ponto de aceleração. Quando metade do trem estiver na crista, o acelerador já deve estar no ponto de aceleração neutro para começar a frear o trem.

2.5 Partida e Parada do Trem

A partida do trem é o procedimento que mais consome combustível. Ela depende de fatores como: tipo de locomotiva utilizado, comprimento do trem, distribuição da carga, estados das folgas (esticadas ou encolhidas), condições meteorológicas (alteração do coeficiente de atrito) e perfil da via. Ela segue um protocolo bem-definido. O maquinista (L) deve, ao iniciar a movimentação do trem, informar ao controlador de tráfego (CT) sua posição, o código da locomotiva e receber como retorno informações do trecho por onde passará o trem (ver exemplo, Figura 3).

É fundamental durante a partida relaxar todos os freios do trem. O esforço trator neste início dos movimentos é elevado. A liberação completa dos freios depende do comprimento do trem e pode variar de sete segundos a dez segundos, quando há, por exemplo, oitenta veículos. Dependendo da posição do retentor do freio (componente responsável por restringir a saída de ar) o tempo de redução do cilindro de freio de cinquenta *psi* para cinco *psi* (equivalente a vinte *psi* no encanamento geral) pode chegar a cento e quarenta segundos (cf. Tabela 1).

Tabela 1. Tempo necessário para redução da pressão dos freios.

Posição do Retentor	Tempo
Direita	20 segundos
Restrita	55 segundos
Super-restrita	140 segundos

Um trem de oitenta vagões parado com uma redução de vinte *psi* no encanamento geral e com retentores na posição *super-restrita* (onde há uma forte pressão no encanamento de freios), após posicionar o manipulador do freio automático para a posição de alívio deverá aguardar no mínimo cento e cinquenta segundos para arrancar o trem.

CIRCULAR DE INSTRUÇÕES ESPECIAIS

Procedimento de Comunicações (Norma 121)

O seguinte é o procedimento que deve ser seguido quando se usa o rádio para ingressar em limites protegidos com formulário B:

“FSA, Locomotiva ...(N°)... a Chefe de Turma(nome)...., câmbio.”
 “Aqui Chefe de Turma(nome)...., na escuta locomotiva...(N°)... , câmbio.”

L: “Locomotiva ...(N°)... aproximando-se aos limites de seu formulário B N°.... em km , solicito instruções para prosseguir. Câmbio.”

CT: “Chefe de Turma(nome)...., responsável pelo Boletim de Via N°... entre km e km na Subdivisão,locomotiva ...(N°)... pode passar bandeira vermelha localizada no km..... sem a necessidade de parar (ou “pode ingressar nos limites de trabalho entre km e km”).
 Locomotiva ...(N°)... pode ultrapassar os limites do Boletim de Via N°..... a km/h (ou “à velocidade máxima autorizada”). Câmbio.”

L: “Locomotiva...(N°)... , compreendido (repete as instruções), câmbio.”

CT: “Correto. Chefe de Turma(nome)...., câmbio final.”

L: “Obrigado. Locomotiva ...(N°)...., câmbio final.”

Observação:

Se o Boletim de Via Formulário B não indicar detenção (PARE) ou não existir bandeira vermelha e depois de ter sido tentado estabelecer contato pelo rádio, o trem pode prosseguir à Velocidade Restrita.

Enquanto o trem transita pelo setor, as turmas da via devem fazer a proteção de acordo com a Norma 326 do Regulamento de Operações. Posteriormente, o Chefe de Turma deverá despedir-se do trem em seu território da seguinte forma :

CT: “Chefe de Turma(nome).... a locomotiva ...(N°)...., câmbio.”

L: “Locomotiva ...(N°)...., câmbio.”

CT: “Locomotiva ...(N°)... : seu trem está bem e vai completo. Boa viagem, Chefe de Turma(nome)...., câmbio final”.

L: “Recebido conforme, obrigado. Locomotiva ...(N°)...., câmbio final.”

Figura 3. Exemplo de comunicação padrão. (ALL, 2008)

Em locais com curvas e contracurvas é recomendável que se utilize à mesma potência necessária para a partida do trem. Uma potência maior pode trazer danos à carga, diminuindo a possibilidade de estiramento devido a excessivo esforço de tração em curvas que podem descarrilar ou tombar um trem. A medição de quanta potência está sendo

empregada pode ser inferido por meio de um componente da locomotiva, chamado amperímetro.

Os procedimentos tomados para a parada dos trens são resumidos na Tabela 2. Em um trecho em nível, um trem pode ser parado empregando o método das folgas esticadas ou encolhidas. Para tal, devem ser consideradas as seguintes informações: formação do trem, velocidade do trem, condições do equipamento de freios e folgas.

Tabela 2. Procedimentos para a parada do trem.

Trecho em nível (folgas esticadas)	<ol style="list-style-type: none"> 1. Utilize o acelerador para manter as folgas esticadas; 2. Reduza a pressão de freios do encanamento geral de 5 a 8 <i>psi</i>, aplicando o freio automático; 3. Mantenha os freios das locomotivas aliviados; 4. Reduza os pontos de aceleração, fechando o acelerador e controlando assim a velocidade e forças internas do trem; 5. Aplique de 2 a 3 <i>psi</i> para complementar a parada, se necessário; 6. Aplique o freio independente; 7. Coloque o acelerador em vazio quando o trem parar.
Trecho em nível (folgas encolhidas)	<ol style="list-style-type: none"> 1. Se o trem estiver em tração então: <ol style="list-style-type: none"> a. Feche gradualmente o acelerador; b. Aguarde 10 segundos; c. Selecione a frenagem dinâmica; d. Encolha as folgas usando o freio dinâmico. 2. Se há distância suficiente para parar, então reduza o encanamento de freios geral de 5 a 8 <i>psi</i>, mantendo os freios das locomotivas aliviados; 3. Caso necessário, reduza de 2 a 3 <i>psi</i> para aumentar a desaceleração e reduzir as forças sobre as locomotivas. 4. Ao aplicar o passo 3, deve-se manter aliviados os freios das locomotivas e a frenagem dinâmica incrementada. 5. Quando a velocidade cair ao ponto do freio dinâmico perder a eficiência, use o freio independente para manter as locomotivas “encostadas” no trem; 6. À aproximadamente 60 metros do local da parada deve-se reduzir o encanamento geral, mantendo as folgas encolhidas usando o freio independente.
Rampas ascendentes	<ol style="list-style-type: none"> 1. Reduza de 5 a 8 <i>psi</i> os freios, mantendo-os aliviados; 2. Decremente o acelerador à medida que a velocidade diminui, para prevenir o incremento da amperagem; 3. Reduza o encanamento geral de freios de 2 a 3 <i>psi</i> conforme necessário; 4. À aproximadamente 15 metros do local de parada aplique areia para as locomotivas pararem sobre a areia; 5. Quando paradas, aplique o freio independente nas locomotivas; 6. Posicione o acelerador em vazio; 7. Não aplique menos que 10 <i>psi</i> para evitar que os freios fiquem agarrados durante o alívio dos mesmos.
Rampas descendentes leves	<p>Neste perfil de via, é possível tanto a aplicação do freio automático juntamente com o freio dinâmico ou de forma isolada.</p> <p>Freio Dinâmico e Automático:</p> <ol style="list-style-type: none"> 1. Encolha as folgas do trem usando o freio dinâmico; 2. Reduza de 5 a 8 <i>psi</i> o encanamento geral dos freios; 3. Reduza o encanamento geral de freios de 2 a 3 <i>psi</i> conforme necessário; 4. Quando a velocidade alcançar 10 km/h aplique o freio independente para manter o trem encolhido; 5. Quando alcançar uma distância suficiente para parar reduza o encanamento geral, mantendo as folgas encolhidas por meio do uso do freio independente. <p>Freio automático:</p> <ol style="list-style-type: none"> 1. Reduza o ponto de aceleração até alcançar uma distância suficiente para

	<p>parada;</p> <ol style="list-style-type: none"> 2. Reduza no mínimo de 5 a 8 <i>psi</i> no encanamento geral independente das locomotivas; 3. Reduza 2 a 3 <i>psi</i> se necessário; 4. Antes da parada faça uma redução final necessária para a parada impedindo que as folgas do trem se estiquem; 5. Aplique o freio independente.
Fortes rampas descendentes	<p>Os procedimentos abaixo devem ser tomados sempre cuidando para que não ocorram deslizamentos.</p> <ol style="list-style-type: none"> 1. Aplique o freio dinâmico para controlar a velocidade; 2. Quando alcançar uma distância suficiente para parar reduza de 5 a 8 <i>psi</i>, mantendo o freio independente da locomotiva totalmente aliviado; 3. Reduza 2 a 3 <i>psi</i> se necessário; 4. Mantenha o freio dinâmico totalmente aplicado; 5. Mantenha o freio independente aliviado; 6. Quando o freio dinâmico perder a eficiência <ol style="list-style-type: none"> a. Se a velocidade for baixa, aplique gradualmente o freio mantendo o trem encolhido.
Depressão	<p>Tal condição depende da folga e da localização do trem em relação à depressão.</p> <p>Ao aproximar-se de uma depressão desacelere o trem, mantendo as folgas esticadas por meio da utilização do freio automático e do acelerador.</p> <p>Se a aproximação ocorrer já em rampa descendente com freio dinâmico ativado, uma desaceleração com folga encolhida pode ser feita por meio do freio automático.</p>
Trechos ondulados	<p>Se a folga do trem estiver esticada:</p> <ol style="list-style-type: none"> 1. Mantenha as folgas esticadas controlando a tração do trem; 2. Aplique o freio automático; 3. Reduza o ponto de aceleração gradualmente para controlar a velocidade e as forças internas do trem; <p>Se a folga do trem estiver encolhida:</p> <ol style="list-style-type: none"> 1. Aplique o freio dinâmico; 2. Aplique freio automático até que a velocidade do trem possa atingir um valor onde o freio dinâmico não seja mais eficiente; 3. Aplique então o freio independente para manter a cabeceira do trem encolhida.

A Tabela 3 resume os diferentes procedimentos que devem ser tomados durante a partida do trem com base no perfil da via.

Tabela 3. Procedimentos tomados durante a partida do trem.

Trechos em nível	<ol style="list-style-type: none"> 1. Solte o freio automático; 2. Alivie o freio independente 3. Mova o manípulo do acelerador para o primeiro ponto, acompanhando sempre o incremento da corrente no amperímetro; 4. Aguarde alguns segundos; 5. Avance para o segundo ponto de aceleração, caso precise de mais potência para movimentar o trem; 6. Se necessário avance até o terceiro ou quarto ponto de aceleração; 7. Se o trem não se movimentar, retorne para a posição “vazio” e inspecione o trem; 8. Não ultrapasse os 2 km/h até que todo o trem esteja se movimentando.
------------------	---

Rampas ascendentes (folgas esticadas)	<ol style="list-style-type: none"> 1. Posicione o freio automático na posição de “alívio”; 2. Incremente o ponto de aceleração até que exista potência suficiente para não deixar o trem recuar; 3. Alivie então o freio independente das locomotivas; 4. Se necessário avance até o terceiro ou quarto ponto de aceleração; 5. Reduza a aceleração caso ocorra uma indicação de patinação; 6. Despeje areia nos trilhos caso o passo 5 ocorra; 7. Se o trem não partir com o quinto ponto de aceleração então <ol style="list-style-type: none"> a. Reduza a aceleração até o trem permanecer parado por si só; b. Aplique o freio independente; c. Se necessário aplique o freio automático; d. Verifique porque o trem não parte.
Rampas descendentes (folgas encolhidas)	<ol style="list-style-type: none"> 1. Aplique totalmente o freio independente; 2. Mova o manípulo do freio automático até a posição de alívio; 3. Aguarde o carregamento do ar do trem e gradualmente; 4. Alivie o freio independente somente o suficiente para que o trem comece a se movimentar gradualmente; 5. Quando todo o trem estiver em movimento, utilize o freio independente para controlar a velocidade e folgas durante a partida;
Fortes rampas descendentes	<ol style="list-style-type: none"> 1. Certifique-se que o sistema de freio esteja totalmente recarregado à pressão de trabalho; 2. Mantenha as folgas encolhidas e freio independente totalmente aplicado; 3. Aplique o máximo possível do freio dinâmico; 4. Observe o incremento da aceleração; 5. Solte o freio automático, acompanhando sempre a recuperação do ar; 6. Controle a velocidade com o freio independente ao iniciar o movimento da seguinte forma: <ol style="list-style-type: none"> a. Mantenha as folgas encolhidas até a atuação do freio dinâmico; b. Complete a operação reduzindo o serviço no freio automático. 7. Finalize a aplicação do freio automático para controlar a velocidade do trem antes de atingir a velocidade estipulada para o trecho.
Depressão	<p>Tal condição depende da folga e da localização do trem em relação à depressão.</p> <ol style="list-style-type: none"> 1. Mantenha a velocidade da locomotiva baixa até que todo o trem esteja em movimento; <ol style="list-style-type: none"> a. Para isto, controle a aceleração e, se necessário, utilize o freio independente. 2. Acelere o trem aliviando gradualmente o freio independente e incrementando o ponto de aceleração.
Aceleração após desaceleração em trechos de depressões	<ol style="list-style-type: none"> 1. Reduza a aceleração do trem ao aproximar-se da bacia da depressão, reduzindo os pontos de aceleração ou usando o freio automático; 2. Por meio da modulação do acelerador, evite que o trem acelere após a cabeceira entrar na bacia; 3. Aumente a aceleração antes que a cabeceira do trem entre na rampa ascendente; 4. Aumente a aceleração até a cauda do trem alcance o fundo da bacia, impedindo assim o encolhimento da folga; 5. Reduza a aceleração quando a cauda começar a subir a rampa ascendente de modo a ajustar a folga.
Trechos ondulados	<p>Com a cabeceira em declive:</p> <ol style="list-style-type: none"> 1. Alivie o freio independente lentamente, controlando para que todo o trem esteja em movimento; 2. Se o trem não movimentar, deve-se aplicar o mínimo de tração até que todo o trem esteja em movimento; <ol style="list-style-type: none"> a. A aceleração deve ser suave de modo a evitar a geração de forças internas nas folgas dos trens. <p>Com a cabeceira em acentive:</p> <ol style="list-style-type: none"> 1. Aplique a tração gradualmente, impedindo que a cabeceira do trem recue quando o freio for aliviado; 2. Alivie totalmente o freio independente; 3. Avance o acelerador até o trem todo esteja em movimento.

	<p>Aceleração do trem:</p> <ol style="list-style-type: none"> 1. Mantenha a velocidade constante enquanto o trem estiver percorrendo trechos ondulados.
Cristas	<ol style="list-style-type: none"> 1. Deve-se primeiro verificar o sentido da força resultante após o alívio do freio automático; 2. Alivie gradualmente o freio independente das locomotivas, permitindo o movimento lento para frente; 3. Aumente gradativamente o acelerador; 4. Quando a cauda começar a movimentar-se, acelere gradativamente o trem.

Basicamente, todos os procedimentos de aceleração envolvem a aplicação de esforços tratores vis-à-vis a superar forças contrárias ou resistências. A formalização primeira para cálculos de tais resistências foi publicada por Davis em 1926.

2.6 Equações de Davis

As equações de Davis, publicadas em 1926, descrevem os cálculos das resistências que um trem deve superar para deslocar-se de um ponto a outro na via férrea.

Para facilitar a compreensão das equações é necessário observar as seguintes equivalências nas unidades de medidas aqui enumeradas:

- 1t (tonelada numérica) = 1,1 *short-ton* (tonelada inglesa);
- 1lb/ton (libra/*short-ton*) = 0,5 Kg/t ou Kgf (quilograma por tonelada ou quilograma força);
- 1 km = 0.622 mi (milhas).

2.6.1 Resistências Normais e Acidentais

Segundo Brina (1982), a resistência de um trem em Kg/t consiste em um conjunto de forças contrárias ao movimento. Ela é considerada no ponto de contato entre as rodas e os trilhos. Ela é oposta ao esforço trator nas rodas da locomotiva. A Equação 1 define tal resistência.

$$R_{Total} = n_l \cdot P_l (R'_{nl} + R'_{cl} + R'_{\gamma} + R'_i) + n_v \cdot P_v (R'_{nv} + R'_{cv} + R'_{\gamma} + R'_i) \quad (1)$$

A resistência pode ser decomposta em:

- R_{Total} é a resistência total (em Kgf);
- n_l é o número de locomotivas;

- n_v é o numero de vagões;
- P_v é o peso do vagão (em toneladas);
- P_l é o peso locomotiva (em toneladas);
- R'_{nl} é a resistência normal da locomotiva (em Kgf);
- R'_{nv} é a resistência normal do vagão (em Kgf);
- R'_{cl} é a resistência de curva da locomotiva (em Kgf);
- R'_{cv} é a resistência de curva do vagão (em Kgf);
- R'_γ é a resistência de inércia (em Kgf);
- R'_i é a resistência de rampa (em Kgf);

As resistências se dividem em dois grupos: normais e acidentais. As normais são aquelas presentes quando o veículo encontra-se em nível e em reta. Já as acidentais decorrem de certas circunstâncias específicas, tais como: início de movimento, curvas e rampas.

A resistência normal da locomotiva é determinada pela Equação 2.

$$R_{nl} = 1,3 + \frac{29}{w} + 0,03 \times V + \frac{0,0024 \times A \times V^2}{w \times n} \quad (2)$$

Onde:

- w é o peso por eixo da locomotiva (em *short-ton*),
- V a velocidade (em milhas/hora),
- A é a área frontal da locomotiva (em *square-feet*) e
- n o número de eixos da locomotiva.

A resistência normal de cada vagão é determinada pela Equação 3.

$$R_{nv} = 1,3 + \frac{29}{w} + 0,045 \times V + \frac{0,0024 \times A \times V^2}{w \times n} \quad (3)$$

Onde:

- w é o peso por eixo do vagão em toneladas (*short-ton*),
- V é a velocidade (em milhas/hora),
- A é a área frontal do vagão (*square-feet*) e

- n é o número de eixos do *vagão*.

As resistências de curvas para locomotivas e vagões são definidas respectivamente pela Equação 4 e Equação 5.

$$R_{cl} = 0,2 + \frac{100}{R} \times (p + b + 3,8) \quad (4)$$

$$R_{cv} = \frac{500 \times b}{R} \quad (5)$$

Onde:

- R corresponde ao valor do raio da curva (em metros),
- p ao valor da base rígida e
- b o tamanho da bitola (em metros).

A resistência inercial é calculada com base na Equação 6. Ela fornece a força que deve ser aplicada no veículo para que o mesmo modifique seu estado de movimento atual (e.g. partir de repouso para movimento).

$$R_\gamma = 4 \frac{V_F^2 - V_I^2}{\ell} \quad (6)$$

Onde:

- V_F e V_I correspondem às velocidades finais e iniciais (em milhas/hora); e
- ℓ corresponde ao deslocamento desejado (em metros).

Por último, a Equação 7 nos fornece o valor da resistência de rampa.

$$R_i = 10 \times i \quad (7)$$

Onde:

- i corresponde ao percentual de inclinação de certo ponto da via.

2.6.2 Esforço Trator e Força de Aceleração

O esforço trator é a força total que um veículo tem disponível para aplicar na sua movimentação, conforme Equação 8. O resultado é uma força em *kgf*.

$$F_t = \frac{273.24 \times 0.82 \times HP}{V} \quad (8)$$

Onde:

- V é a velocidade final (em Km/h);
- HP é a potência de uma locomotiva (em *horse power*);
- 273.24 é uma constante definida empiricamente; e
- 0.82 é o valor da perda de rendimento entre os motores diesel e elétrico.

Para que ocorra o deslocamento do trem sem patinagem uma força tratora máxima aplicada em um determinado instante deve ser estabelecida. Esta força tratora máxima é determinada pela Equação 9.

$$F_{tm} = \frac{P \times f}{1 + (0,01 \times V)} \quad (9)$$

Onde:

- P é o peso (em toneladas);
- V é velocidade média (em Km/h); e
- f é coeficiente de aderência;

A força disponível para aceleração é determinada pela diferença entre a força de tração máxima e o conjunto das resistências. Esta relação é dada pela Equação 10. O valor resultado é expresso em Kgf .

$$F_{ac} = F_{tm} - R_{Total} \quad (10)$$

2.6.3 Deslocamento

O deslocamento de um trem existe quando a potência superior às somas das resistências contrárias sem ultrapassar um limite de aderência entre as rodas e os trilhos, evitando assim patinagens. Para que esse movimento ocorra, o esforço trator efetivo não deve superar o esforço trator aderente. Analogamente, o esforço trator efetivo não poderá ser menor que a soma de todas as resistências. O cálculo de deslocamento de um trem está expresso na Equação 11. O valor resultado é expresso em metros.

$$\Delta \ell = 4 \times \frac{P \times (V_F^2 - V_I^2)}{F_{ac}} \quad (11)$$

Onde:

- P é o peso do trem (em toneladas);
- V_F é a velocidade final (em Km/h);
- V_I velocidade inicial (em Km/h); e
- F_{ac} é a força aceleradora (em Kgf).

A variação do tempo é determinada pela Equação 12. O valor resultado é expresso em horas.

$$\Delta t = 7,2 \times \frac{\Delta \ell}{V_F + V_I} \quad (12)$$

Onde:

- $\Delta \ell$ é a variação do deslocamento (em quilômetros);
- V_F é a velocidade final (em Km/h); e
- V_I velocidade inicial (em Km/h).

A velocidade média do trem pela razão entre duas grandezas: distância percorrida expressa em quilômetros e tempo gasto expresso em horas (Equação 13). Multiplicando o valor obtido por 3.6 tem-se a velocidade em quilômetros por hora.

$$V_m = \frac{\Delta \ell}{\Delta t} \quad (13)$$

Onde:

- $\Delta \ell$ é a variação do deslocamento (em metros);
- Δt é a variação do tempo (em segundos);

O cálculo do deslocamento final de um trem, em um trecho de n quilômetros, pode ser efetuado em k frações deste trecho. Analogamente, pode-se aplicar o mesmo procedimento para o cálculo do tempo gasto final. É também possível obter a força traçadora máxima, bem como a resistência total e a força aceleradora para sub-trecho k . Esta última é necessária para alcançar uma determinada velocidade para sub-trecho k . Desta forma, o deslocamento final do trem é determinado pela soma dos k deslocamentos (Equação 14), do mesmo modo, o tempo gasto final é determinado pela soma dos tempos

parciais gastos (Equação 15). A aceleração γ pode ser calculada com base na Equação 16.

$$\Delta \ell_{final} = \sum_{k=1}^n \Delta \ell_k \quad (14)$$

$$\Delta t_{final} = \sum_{k=1}^n \Delta t_k \quad (15)$$

$$\gamma = F_{ac} \times \frac{g}{P} \quad (16)$$

Onde:

- $\Delta \ell_k$ é a variação de deslocamento de sub-trecho k (em metros);
- Δt_k é a variação do tempo de deslocamento de sub-trecho k (em minutos);
- F_{ac} é a força de aceleração (em Kgf);
- P é o peso do trem (em toneladas);
- g é a aceleração da gravidade (em metros/segundo²).

Apesar da aceleração γ poder ser calculada em função da força de aceleração, da aceleração da gravidade e do peso do trem, a mesma não é aplicada, pois o valor da aceleração buscada é sempre de dois km/h.

As diferentes equações vistas anteriormente são importantes para os cálculos de resistência de um trem sobre um via férrea e também para os cálculos de esforços para mover este trem em certa velocidade. Uma variável importante no transporte de cargas comerciais é o consumo de combustível da(s) locomotiva(s) usada(s) para mover um trem em situações diversas: partida, em aceleração, em desaceleração, etc.

2.6.4 Consumo

O consumo acumulado de uma viagem de trem é calculado com base na Equação 17. O valor resultado é expresso em LTKB².

$$LTKB = \frac{CA}{P \times \left(\frac{DA}{1000} \right)} \times 1000 \quad (17)$$

² A sigla LTKB significa: litros por tonelada bruta transportada.

Onde:

- CA é o consumo acumulado da viagem (em litros);
- P o peso do trem (em toneladas); e
- DA a distância percorrida da viagem (em quilômetros).

Como já foi dito, o estilo de condução de um trem influi diretamente no consumo de combustível. Cada ponto de aceleração de uma locomotiva gera uma determinada força e um determinado consumo. A título de ilustração, uma locomotiva do tipo C30 disponibiliza um conjunto $PA = \{ \langle p_1, hp_1, cp_1 \rangle, \langle p_2, hp_2, cp_2 \rangle, \dots, \langle p_{10}, hp_{10}, cp_{10} \rangle \}$ com dez pontos diferentes de aceleração, onde p_1 é o identificador do primeiro ponto de aceleração, hp_1 é a potência nominal do primeiro ponto de aceleração, cp_1 é o consumo nominal do primeiro ponto de aceleração, e assim por diante. A Equação 18 fornece o consumo nominal para um intervalo Δt por hora.

$$C = \frac{\Delta t}{60} \times cp \quad (18)$$

Onde:

- cp é o consumo nominal de um determinado ponto de aceleração (em litros);
- Δt é a variação do tempo (em minutos);

2.6.5 Equações de Davis Modificadas

Diferentes modificações as equações de Davis foram efetuadas para atender diferentes propósitos e situações. Em Tuthill (1948, p. 376) apud (AVALLONE, et al., 1996), tem-se uma alteração para o cálculo de resistência de um trem para aceleração até quarenta milhas/hora (Equação 19). Deve-se notar que, para velocidades entre cinquenta e setenta milhas/hora, os valores resultantes das resistências são calculadas com a equação 19.

$$R = 1.3 \times W + 29 \times n + 0.045 \times W \times V + 0.0005 \times A \times V^2 \quad (19)$$

Onde:

- W é o peso por eixo de uma locomotiva (em *short-ton*);
- V é a velocidade (em milhas/hora);

- A é a área frontal da locomotiva (em *square-feet*); e
- n é o número de eixos de uma locomotiva.

A Equação 20 é derivada da equação precedente. A modificação situa-se no seu último termo, onde a área frontal é desconsiderada. Entretanto, a constante deste termo é multiplicada por noventa.

$$R=1.3\times W + 29\times n + 0.045\times W \times V + 0.045\times V^2 \quad (20)$$

A equação de Davis também foi modificada para se adaptar a ferrovia canadense, onde também não é considerada a área frontal (Equação 21). A Equação 22 é uma adaptação para *trailers* e *containeres*. Ambas as equações apresentam as mesmas variáveis da Equação 20, mas os valores das constantes sofrem ajustes.

$$R=0.6\times W + 20\times n + 0.01\times W \times V + 0.07\times V^2 \quad (21)$$

$$R=0.6\times W + 20\times n + 0.01\times W \times V + 0.2\times V^2 \quad (22)$$

Em Toten (1937) apud (AVALLONE, et al., 1996) o autor propôs modificações da equação de Davis para carros de passageiros, onde a particularidade porta sobre o tamanho de cada veículo L em pés. A Equação 23 é para trens de passageiros considerados modernos na época e a Equação 24 é para trens de passageiros considerados antigos.

$$R=1.3\times W + 29\times n + 0.045\times W \times V + [0.00005 + 0.060725(L/100)^{0.88}] \times V^2 \quad (23)$$

$$R=1.3\times W + 29\times n + 0.045\times W \times V + [0.00005 + 0.1085(L/100)^{0.7}] \times V^2 \quad (24)$$

Onde:

- L é o comprimento de um veículo de passageiros (em pés).

As equações de Davis apresentadas anteriormente são a base de muitas outras equações para o mesmo propósito, mas com pequenas modificações para atender algumas particularidades.

Na seqüência ilustrar-se-á uma aplicação das principais equações de Davis, sem modificações, visando simular a movimentação de um trem.

2.6.6 Exemplo de aplicação

O exemplo de aplicação das equações de Davis segue a mesma abordagem mostra em (BRINA, 1982) e (PIRES, et al., 2005). O cenário consiste em deslocar em linha reta e em nível um trem com a seguinte configuração:

- 1 locomotiva
- 100 toneladas cada locomotiva;
- 10 vagões
- 50 toneladas cada vagão;
- [0, 4] é intervalo de velocidades (em Km/h);
- 8º ponto de aceleração da locomotiva desempenha a potência de 1957.5 *kgf*;
- 600 toneladas é peso total do trem (1 x 100 + 10 x 50).

O primeiro passo da simulação consiste em movimentar o trem, variando a velocidade de zero a dois quilômetros/hora.

Velocidade média em *Km/h*: 1

Velocidade média em *milhas/hora*: $1 \times 0,622 = 0,622$

Peso por eixo da locomotiva: $\frac{100}{4} \times 1.1 = 27,5$ toneladas/eixo

Peso por eixo do vagão: $\frac{50}{4} \times 1.1 = 13.75$ toneladas/eixo

$$F_{trator} = \frac{273.24 \times 0.82 \times 1957,5}{2} = 219295.6 \text{ kgf}$$

$$R_{nl} = 1,3 + \frac{29}{27.5} + 0,03 \times 0.622 + \frac{0,0024 \times 110 \times 0.622^2}{27.5 \times 4} = 2.374134 \text{ libras/ton}$$

$$R_{nl} = 2.374134 \times 0,5 = 1.187067 \text{ kg/ton}$$

$$R_{nv} = 1.3 + \frac{29}{13.75} + 0,045 \times 0.622 + \frac{0,0024 \times 85 \times 0.622^2}{13.75 \times 4} = 3.438516 \text{ libras/ton}$$

$$R_{nl} = 3.438516 \times 0,5 = 1.719258 \text{ kg/ton}$$

$$R_{Total} = 1 \times 100 \times (1.187067 + 0 + 0 + 0) + 10 \times 50 \times (1.719258 + 0 + 0 + 0)$$

$$R_{Total} = 978.3357 \text{ kgf}$$

Na seqüência são calculados os valores para a força de aceleração, deslocamento, tempo gasto e consumo:

$$F_{ac} = 219295.6 - 978.3357 = 218317.3 \text{ kgf}$$

$$\Delta \ell = 4 \times \frac{600 \times (2^2 - 0^2)}{218317.3} = 0.043973 \text{ metros}$$

$$\Delta t = 7,2 \times \frac{0.043973}{2 + 0} = 0.158302 \text{ segundos}$$

$$\gamma = \frac{218317.3}{\frac{600000}{9.80665}} = 3,568269$$

$$Consumo = \frac{0.158302}{60} \times 7.78 = 0.020526 \text{ litros/minutos}$$

O segundo passo da simulação consiste em movimentar o trem, variando a velocidade de dois a quatro quilômetros/hora.

Velocidade média em *Km/h*: 3

Velocidade média em *milhas/hora*: $3 * 0,622 = 1.244$

Peso por eixo da locomotiva: $\frac{100}{4} \times 1.1 = 27,5$ toneladas/eixo

Peso por eixo do vagão: $\frac{50}{4} \times 1.1 = 13.75$ toneladas/eixo

$$F_{trator} = \frac{273.24 \times 0.82 \times 1957,5}{4} = 109647.8$$

$$R_{nl} = 1,3 + \frac{29}{27.5} + 0,03 \times 1.244 + \frac{0,0024 \times 110 \times 1.244^2}{27.5 \times 4} = 2.39558 \text{ libras/ton}$$

$$R_{nl} = 2.39558 \times 0,5 = 1.19779 \text{ kg/ton}$$

$$R_{nv} = 1,3 + \frac{29}{13.75} + 0,045 \times 1.244 + \frac{0,0024 \times 85 \times 1.244^2}{13.75 \times 4} = 3.470811 \text{ libras/ton}$$

$$R_{nv} = 3.470811 \times 0,5 = 1.735405 \text{ kg/ton}$$

$$R_{Total} = 1 \times 100 \times (1.19779 + 0 + 0 + 0) + 10 \times 50 \times (1.735405 + 0 + 0 + 0)$$

$$R_{Total} = 987.4817 \text{ kgf}$$

Na seqüência são calculados os valores para a força de aceleração, deslocamento, tempo gasto e consumo e consumo acumulado (LTKB):

$$F_{ac} = 109647.8 - 987.4817 = 108660.3 \text{ kgf}$$

$$\Delta l = 4 \times \frac{600 \times (4^2 - 2^2)}{108660.3} = 0.265046 \text{ metros}$$

$$\Delta t = 7,2 \times \frac{0.265046}{4 - 2} = 0.318055 \text{ segundos}$$

$$\gamma = \frac{108660.3}{\frac{600000}{9.80665}} = 17.75989$$

$$Consumo = \frac{0.318055}{60} \times 7.78 = 0.041241 \text{ litros/minutos}$$

$$LTKB = \frac{(0.020526 + 0.041241)}{600 \times \left(\frac{0.309019}{1000} \right)} \times 1000 = 0.333138$$

Para os cálculos dos dois passos mostrados, têm-se os seguintes resultados:

- Tempo gasto total é de 0.476357 segundos, sendo 0.158302 segundos para passar de 0 até 2 km/h e 0.318 segundos para passar de 2 até 4 km/h;
- Deslocamento total é de 0.309019 metros, sendo 0.043973 metros para passar de 0 até 2 km/h e 0.265 metros) para passar de 2 até 4 km/h; e
- O consumo para passar 0 até 4 km/h é 0,333 LTKB.

2.7 Considerações finais

A condução de um trem de carga é uma tarefa complexa. Esta complexidade manifesta-se nos seguintes termos: um grande conjunto de variáveis precisa ser levado em conta para dar partida, bem como para manter os veículos e os seus engates em equilíbrio, geralmente, sobre vias férreas com perfis verticais e horizontais sinuosos. Este equilíbrio é assegurado basicamente por meio do correto uso dos diferentes tipos freios (independente, dinâmico e automático) e acelerador. Dada esta complexidade, a forma-

ção de um bom maquinista é dispendiosa, porém necessária. É importante salientar que é uma formação que visa o desenvolvimento de habilidades específicas, o que requer grande número de conhecimentos práticos; e os resultados melhoram à medida que as experiências são enriquecidas ao longo dos anos.

A disponibilidade de experiências na forma de um sistema baseado em conhecimentos se justifica plenamente, em particular, para funcionar como auxílio à tomada de decisão para um iniciante ou memória de situações raras para um maquinista experiente.

Os conhecimentos advindos das experiências de condução de um trem podem ser adquiridos, basicamente, por meio de entrevistas com especialistas em condução ou por meio da exploração dos registros de dados sobre as ações tomadas por um maquinista. Esta exploração pode ser feita por meio da aplicação de técnicas de descobertas automáticas de conhecimentos a partir de bases de dados. É sobre este segundo viés que os próximos capítulos são delineados.

Capítulo 3

DESCOBERTA AUTOMATIZADA DO CONHECIMENTO

Neste capítulo são examinados em detalhes alguns métodos de aprendizagem de máquina simbólica, técnicas de validação e análise do desempenho de classificadores. Cada classificador pode assumir a forma de uma árvore de decisão ou a forma de um conjunto ordenado de regras. De forma pragmática, os estudos portam essencialmente sobre os seguintes temas:

- Aprendizagem de máquina simbólica realizada por meio do algoritmo C4.5 e os métodos de combinação de classificadores BAGGING e BOOSTING;
- Seleção de atributos como uma forma de redução da complexidade dos dados por meio da filtragem de atributos irrelevantes;
- Avaliação de classificadores realizada por meio da técnica de validação cruzada;
- Avaliação dos conhecimentos aplicados na forma de um simulador de condução, que deve imitar o comportamento de um maquinista ser humano.

Os melhores classificadores são incorporados ao simulador de condução de trens para auxiliar na definição das políticas de ações.

3.1 O processo de descoberta do conhecimento

O processo de descoberta de conhecimento KDD (Knowledge Discovery in Databases) visa extrair conhecimentos de maneira automatizada e útil a partir de dados sobre certo domínio de problema (FAYYAD, 1996) (MAIMON, *et al.*, 2005). Ele requer, em geral, etapas tais como: pré-processar, minerar os dados a partir de conjunto de fatos e pós-processar o conhecimento obtido. Obtém-se como resultado um padrão, onde a partir dele pode-se criar um modelo que represente os dados de modo geral; i.e., a criação de uma descrição do conjunto de dados por meio de um processamento dos da-

dos. A Figura 4 ilustra as etapas do processo de KDD, as quais são discutidas ao longo do Capítulo 4.

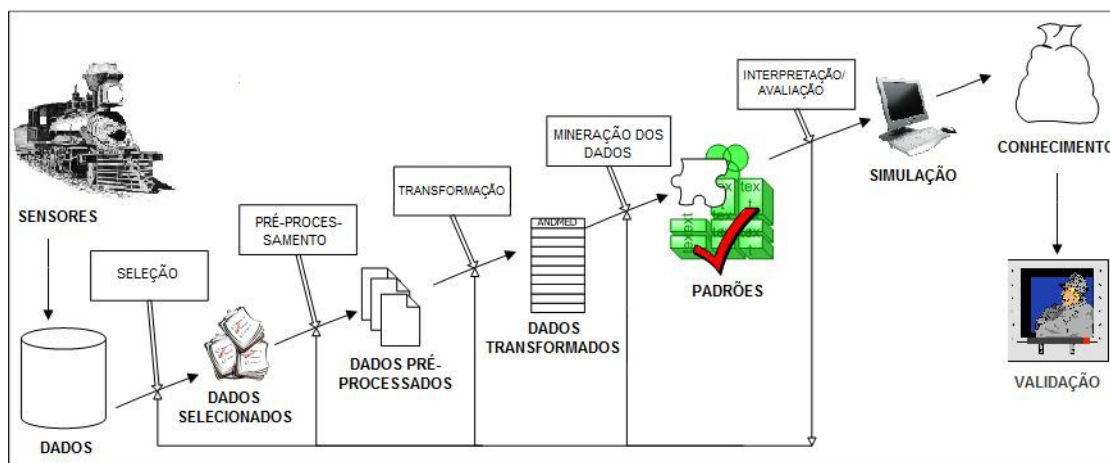


Figura 4. Etapas do processo de KDD.

O processo de descoberta de conhecimentos, em nossos experimentos, abordou todas as suas etapas. Os fatos registrados na base de dados são inerentes a medidas obtidos por meio de sensores instalados em locomotivas.

3.2 Pré-processamento

Segundo (FAYYAD, et al., 1996) o pré-processamento é dividido em: remover ruídos do conjunto de dados – obtendo somente as informações úteis no processo de construção do modelo, decidir a forma de tratamento para dados faltantes, selecionar os atributos de modo a reduzir a dimensionalidade do problema e incluir novas características úteis no processo de descoberta do conhecimento.

3.2.1 Transformação dos dados

Nesta etapa os dados são transformados ou consolidados para a etapa de mineração. Uma das ações possíveis é normalizar os dados, onde são criadas escalas para os dados com pequenos intervalos, reduzindo assim a discrepância entre os valores (HAN, et al., 2006).

3.2.1.1 Remoção de ruídos

Ainda segundo (FAYYAD, et al., 1996), uma grande quantidade de dados ruidosos torna difícil a tarefa de identificar padrões que representem os dados. Um ruído é um erro ou variância aleatória de uma variável cujos valores são conhecidos (HAN, et al., 2006), podendo ser gerado por erro na leitura dos dados ou simplesmente erro na entrada manual dos dados.

Uma das técnicas utilizadas na literatura para identificação de ruídos é o algoritmo de agrupamento (*clustering*). Neste, valores similares de um atributo são identificados e organizados em grupos, também chamados de *clusters*. Um algoritmo de agrupamento difundido na literatura é o *k-means* (MACQUEEN, 1967). Nele, são formados grupos de dados (conjunto de *clusters*) com valores semelhantes, onde os que possuírem valores fora do conjunto de *clusters* são considerados ruídos ou *outliers*.

3.2.1.2 Seleção de atributos

A seleção de atributos pode ser vista como uma etapa de filtragem de dados ou como um método para aprimorar o modelo gerado pelo algoritmo utilizado. É importante notar que o objetivo é encontrar o melhor subconjunto, mas isto requer escolher quais atributos serão utilizados durante a etapa de classificação, e dentre estes, quais são os melhores segundo um determinado critério. Uma vez feita esta distinção, algumas características (atributos) podem ser removidas da base de dados ou desconsideradas durante o processo de aprendizagem, o que deve aumentar o desempenho do classificador e possibilitar também uma menor taxa de erro (HUAN, et al., 1998) (LIU, et al., 2005).

A busca pelos subconjuntos gerados pode ser completa, ou exaustiva (*Exhaustive Search*), garantindo que o melhor resultado seja encontrado, desde que um resultado completo seja encontrado. Um resultado é considerado completo quando a busca abrange todas as combinações de atributos possíveis. Isto pode ser realizado de forma sequencial inserindo ou removendo atributos, bem como de forma aleatória gerando um conjunto inicial aleatório e, a partir dele, edificando conjuntos de modo sequencial ou aleatório (LIU, 1998).

O primeiro passo no processo de seleção de atributos é gerar, a partir de um conjunto de dados, um subconjunto por meio da adição (*forward*) ou remoção (*backward*) atributos e também medir a sua qualidade. Caso o subconjunto gerado seja satisfatório, atendendo o critério de parada, ele é considerado como o resultado da geração, caso

contrário um novo subconjunto é gerado e o processo é repetido. A qualidade de um subconjunto pode ser medida por três métodos distintos:

- Filtro: a avaliação é feita de acordo com um critério de parada que é uma medida independente do algoritmo de aprendizagem e aplicada em um único atributo por vez. Em problemas que utilizam bases de dados reais e com grande quantidade de atributos, o método de filtragem mostra-se computacionalmente menos custoso em relação, por exemplo, ao modelo *Wrapper*;
- *Wrapper*: a avaliação do subconjunto é realizada por meio de um algoritmo de aprendizagem selecionado como critério de parada e aplicado sobre um conjunto de atributos (ZHU, et al., 2007); e
- Híbrido: neste modelo, o subconjunto é avaliado primeiramente por meio de uma medida independente e na seqüência por um algoritmo de aprendizagem (SEBBAN, et al., 2001).

O critério de parada depende do método de seleção de atributos utilizado, conforme mencionado anteriormente. O critério pode ser satisfeito: (i) quando a busca pelo melhor subconjunto foi alcançada, (ii) quando um limite – número de atributos ou número de iterações – for atingido, (iii) quando a adição ou remoção de um certo atributo ocasiona uma melhora no subconjunto ou (iv) quando um subconjunto suficientemente bom foi gerado – produzindo uma taxa de erro aceitável para determinada tarefa.

A seleção de atributos é uma etapa importante porque, a partir dela, é possível obter o conjunto reduzido de atributos a serem usados na geração de classificadores. Na próxima seção são apresentados os algoritmos de classificação utilizados neste trabalho.

3.3 Mineração de Dados

A mineração de dados busca extrair informações válidas e interessantes que se encontram ocultas em conjunto de dados de algum domínio (CIOS, et al., 2007) (HAN, et al., 2006). Diferente de uma consulta a um banco de dados, onde o critério para a realização da consulta é estabelecido antes de efetivá-la. Por exemplo: o gerente de uma loja deseja saber se o produto B é vendido juntamente com o produto A, onde uma consulta no banco de dados provê a resposta. Porém, para conhecer a relação entre vendas de dois produtos quaisquer, a aplicação de algoritmos de mineração de dados é mais indicada.

As informações são válidas e interessantes se forem (i) de fácil compreensão, (ii) válidas no caso de aplicação em novos dados, (iii) potencialmente utilizáveis e (iv) novas. Uma medida interessante na análise do grau de interesse de uma regra é o suporte, que representa a porcentagem de transações da base de dados que satisfaz a regra. Outra medida interessante é o grau de confiança, que expressa a probabilidade de que uma regra contendo o atributo X contenha outro atributo Y (HAN, et al., 2006). A aceitação ou não de uma regra baseada em tais valores pode ser controlada pelo usuário, que pode estabelecer, por exemplo, somente a aceitação de regras com suporte maior que 60%.

Neste trabalho foram aplicados algoritmos de mineração de dados visando à descoberta de regras de condução de trens de carga. O uso da mineração de dados tem se mostrado eficiente em diversas áreas do conhecimento humano, como na química (KALOS, et al., 2005), classificação de músicas (CHEN, et al., 2009), exploração de conhecimento médico (RODDICK, et al., 2003), entre outras.

3.3.1 Software WEKA

Os experimentos foram realizados com o software WEKA. Ele compreende uma coleção implementações de algoritmos de aprendizagem para tarefas de mineração de dados (WITTEN, et al., 2005).

3.4 Aprendizagem de Máquina

A aprendizagem de máquina pode ser vista como a inferência automática de alguns conceitos partindo de exemplos rotulados sobre o conceito a ser aprendido (MITCHELL, 1997).

Existem algumas diferenças entre a aprendizagem de máquina e a mineração de dados. Segundo Prati, em (PRATI, 2006), a mineração de dados é uma fonte de aplicação de algoritmos de aprendizagem de máquina, por prover dados reais e de grande volume, sendo parte do processo de aprendizagem de máquina. Na origem, os algoritmos de aprendizagem de máquina operavam apenas sobre pequenas bases de dados (com atributos previamente selecionados para facilitar o aprendizado). Como a mineração de dados é parte de um processo maior, que vai desde a preparação dos dados até a utilização dos resultados obtidos, ela permitiu que a aprendizagem de máquina explorasse maiores conjuntos de dados.

3.4.1 Aprendizagem Simbólica de Máquina

A aprendizagem simbólica de máquina é fundamental à construção de um sistema de previsão. Um sistema de previsão é um programa de computador capaz de tomar decisões aplicando conhecimentos obtidos a partir de experiências ou de soluções de problemas, geralmente armazenadas em grandes bases de dados ao longo do tempo. Ou seja, um sistema de aprendizagem de máquina simbólica realiza a tarefa de aprender e construir representações simbólicas de um conceito a partir da análise de exemplos e contra-exemplos disponíveis na forma de um conjunto de treinamento. Tais representações assumem, em geral, a forma de árvores de decisão ou regras de produção e podem ser consideradas muito eficientes quanto comparadas a outras abordagens de aprendizagem, como por exemplo, as redes neurais.

A aprendizagem de máquina simbólica é utilizada nas situações em que o modelo obtido assume uma forma compreensível. Tal compreensibilidade é, em alguns casos, fundamental. O sistema ID3 desenvolvido por Quinlan (1986) é um exemplo de sistema de aprendizagem simbólica, onde o conhecimento é obtido por meio da indução de uma árvore de decisão. Em Quinlan (1987) tem-se também a realização de um sistema voltado à geração de regras de produção a partir de uma árvore de decisão.

Além destas importantes contribuições para a área de aprendizagem simbólica de máquina, o leitor encontra em Quinlan (1993), versões mais recentes e mais eficientes desses algoritmos, denominadas de C4.5 e C5.0. O algoritmo C4.5 será detalhado na seqüência deste documento, bem como os métodos de combinação de classificadores BAGGING e BOOSTING.

Em resumo, um sistema de aprendizagem de máquina simbólica corresponde à automatização de um processo de aprendizagem, enquanto que a aprendizagem equivale à obtenção de regras baseada em observações de estados ambientais e transições.

3.4.2 Aprendizagem Indutiva

Na aprendizagem indutiva, o sistema de aprendizagem deduz o conhecimento pela observação do seu ambiente. Existem duas estratégias principais para realizar tal tarefa de dedução. A estratégia de *aprendizagem supervisionada* que corresponde à aprendizagem por meio de exemplos previamente classificados por um supervisor; auxilia o sistema de aprendizagem a montar um modelo de previsão para cada classe. I.e., o sistema reúne as propriedades comuns dos exemplos de cada classe, definindo uma des-

crição simbólica para cada classe. Para cada descrição formulada, uma regra de classificação pode ser usada para prever a classe de um determinado evento, cuja classe é desconhecida ou não informada. A estratégia de *aprendizagem não supervisionada* é regida pela aprendizagem por meio de observações não classificadas. A busca da descrição de cada classe é cega à medida que se tem que reconhecer padrões por si só, examinando os exemplos. O resultado da execução de tal processo de busca é um conjunto de descrições de classe, uma para cada classe descoberta.

A aprendizagem indutiva consiste na extração de padrões a partir de um universo de exemplos. Um método de aprendizagem indutiva pode produzir um modelo cuja qualidade é tal que o mesmo poderia ser usado para prever o resultado de situações futuras. A indução é uma forma de inferência lógica que permite a utilização de premissas para obter conclusões genéricas a partir de exemplos particulares. A indução pode ser caracterizada como uma forma de raciocínio que parte de um conceito específico e o generaliza (MALOOF, et al., 2000).

Um sistema de aprendizagem simbólica pode construir vários modelos a partir dos dados de um ambiente, onde alguns destes modelos são mais simples que outros. Neste caso, opta-se normalmente pelos mais simples. Tal decisão adere à teoria de Ockham (OCKHAM, 1999), segundo a qual a pluralidade não deve ser posta sem necessidade ou se existem inúmeras explicações igualmente válidas para um fato, então se deve escolher a explicação mais simples. Trata-se de uma diretriz lógica reducionista em ciência. Porém, ela nos ajuda a escolher entre várias hipóteses a serem verificadas, aquela que contém o menor número de afirmações não demonstradas.

3.5 Métodos de Classificação

Primeiramente, deve-se observar que a classificação é uma das tarefas mais populares da aprendizagem de máquina e ela visa encontrar uma função que mapeie um conjunto de registros em um conjunto de rótulos pré-definidos. Estes rótulos são denominados de classes. Uma vez obtida esta função, ela pode ser aplicada a novos registros para prever a classe em que estes se enquadrariam ou que se enquadram (QUINLAN, 1996). Em outras palavras, um método de classificação equivale à obtenção de regras baseado em observações de estados ambientais e transições. Por outro lado, um sistema de aprendizagem simbólica de máquina corresponde à automatização de um processo de

aprendizagem. O algoritmo C4.5 é um exemplo de extrator de padrões de bases de dados.

3.5.1 Algoritmo C4.5

O algoritmo C4.5 gera, a partir de um conjunto de treinamento, classificadores no formato de árvores de decisão e também no formato regras do tipo *se <condição> então <conclusão>*, as quais têm como objetivo representar os conhecimentos sobre determinado assunto (MITCHELL, 1997) (HAN & KAMBER, 2006). Tal método é supervisionado, ou seja, as regras obtidas a partir de amostras presentes no conjunto de treinamento são previamente rotuladas (MITCHELL, 1997).

O método de aprendizagem por árvore de decisão caracteriza-se por ser robusto quanto há ruídos nos dados (QUINLAN, 1996). Ele pode ser utilizado para classificar tanto valores discretos (ex: alta, média e baixa) como valores contínuos; esses últimos precisam ser discretizados.

A árvore de decisão é formada por um nó raiz e vários nós-folha, onde o nó raiz é o atributo que melhor separa por si só os exemplos a serem classificados. Cada nó da árvore representa um atributo. Os ramos são formados pelos valores dos atributos e as folhas são as classificações dos exemplos de acordo com os nós e os ramos. A árvore de decisão pode ser também representada por um conjunto de regras no formato *se-então*. (MITCHELL, 1997) (QUINLAN, 1993) (KOHAVI, et al., 2002). A classificação de uma determinada instância é feita à medida que a árvore é percorrida, sempre de modo descendente e *guloso*, o que significa que uma árvore nunca retorna ao nível superior para testar novamente um determinado atributo (MITCHELL, 1997).

O processo de criação da árvore de decisão, a partir de um conjunto contendo n exemplos de treinamento $T = \{t_1, \dots, t_n\}$, onde t corresponde a um exemplo de treinamento, contendo um conjunto $A = \{a_1, \dots, a_i\}$ formado por i atributos, cujos exemplos são rotulados com as m classes $C = \{c_1, \dots, c_m\}$. Todo este processo ocorre da seguinte forma:

1. Caso o conjunto de treinamento T seja vazio:
 - a. Forma-se apenas um nó folha cuja classe associada a ele é determinada pela informação mais freqüente existente em C ;

2. Se T contém um ou mais exemplos de treinamento:
 - a. Se todos os elementos de T pertencerem à mesma classe c_j , então uma folha é gerada, cujo rótulo pertence à classe c_j ;
3. Se T contém mais de um exemplo que pertençam a diferentes classes do conjunto C :
 - a. deve-se escolher um atributo a_i , que possua um ou mais resultados possíveis O , e então particionar T em p subconjuntos $\{T_1, \dots, T_p\}$, onde T_p contém todos os exemplos que possuem os k valores possíveis para o atributo a_i , executando recursivamente o processo para cada subconjunto T_k pertencente ao conjunto T .

O processo repete-se até que todos os exemplos de treinamento possam ser classificados pela árvore.

Durante a criação da árvore de decisão nós que refletem anomalias devido a ruídos ou *outliers* podem ser gerados. Um ruído, ou *outlier*, consiste em uma observação que não corresponde ao modelo dos dados gerado pelo classificador, gerando a suspeita de que tal observação tenha sido gerada por outro mecanismo (HAN & KAMBER, 2006) (HAWKINS, 1980). O processo de poda tende a resolver este problema, usando medidas estatísticas para remover nós e folhas que prejudiquem a performance da árvore de decisão. No processo de podagem, a árvore é percorrida e para cada nó é calculado o erro do nó e a soma dos erros dos nós descendentes; se o erro do nó for menor ou igual à soma dos erros dos nós descendentes então o nó é transformado em folha (BREIMAN, et al., 1984).

A busca pela informação desejada de um atributo, a qual pertence uma determinada classe, é calculada pela função *Entropia(S)*, que busca a frequência de exemplos que resultam em casos positivos p_i presentes no conjunto de treinamento multiplicando-a pelo logaritmo na base dois dos mesmos exemplos positivos, somando com os exemplos negativos, que são obtidos da mesma forma, conforme mostra a Equação 25, que resulta na quantidade de informação necessária para codificar a classificação de um determinado caso pertencente conjunto de treinamento S (MITCHELL, 1997).

$$\text{Entropia}(S) = - \sum_{j=1}^m p_i \log_2(p_i) \text{ bits} \quad (25)$$

Escolher aleatoriamente um atributo para dividir a informação do conjunto de treinamento T em subconjuntos não trás vantagens ao algoritmo. Para melhor separar os subconjuntos de T , o critério de ganho de informação é utilizado. Tal critério consiste em medir a frequência $freq(C_j, S)$ na qual um conjunto qualquer de exemplos S pertence à mesma classe C_j , sendo que $|S|$ é o número de exemplos em S . Tomando como exemplo a base de dados da Tabela 4, a frequência de exemplos positivos (cujo valor da classe é “Aumenta”) é $9/14$ ou 0.642 . Já a frequência de exemplos negativos (valor da classe “Não aumenta”) é $5/14$ ou 0.357 .

Tabela 4. Conjunto de Treinamento. (Exemplo adaptado de (QUINLAN, 1993))

Exemplo	Tempo	Velocidade	Pressão dos freios	Depressão	Classe
1	Ensolarado	Média	Normal	Sim	Aumenta
2	Ensolarado	Alta	Alta	Sim	Não aumenta
3	Ensolarado	Alta	Alta	Não	Não aumenta
4	Ensolarado	Média	Alta	Não	Não aumenta
5	Ensolarado	Baixa	Normal	Não	Aumenta
6	Nublado	Alta	Alta	Não	Aumenta
7	Nublado	Alta	Normal	Não	Aumenta
8	Nublado	Baixa	Normal	Sim	Aumenta
9	Nublado	Média	Alta	Sim	Aumenta
10	Chovendo	Baixa	Normal	Sim	Não aumenta
11	Chovendo	Média	Alta	Sim	Não aumenta
12	Chovendo	Baixa	Normal	Não	Aumenta
13	Chovendo	Média	Alta	Não	Aumenta
14	Chovendo	Média	Normal	Não	Aumenta

O critério de ganho de informação, mostrado na Equação 27, utilizado na seleção de um teste que maximiza o ganho de informação, é medido pela entropia (número de bits necessários para categorizar a classe como positiva ou negativa) menos a informação de T do atributo a . A informação do atributo a é medida de acordo com a equação 26, onde o conjunto de dados de treinamento T é particionado em subconjuntos, de acordo com o conjunto de valores v possíveis $dom(a)$ ao domínio do atributo a , sendo T_{a_v} o subconjunto formado por exemplos que possuem valor v para o atributo a .

$$info(T, a) = \sum_{v \in dom(a)} \frac{|T_{a_v}|}{T} \times Entropia(T_{a_v}) \quad (26)$$

$$ganho(T, a) = entropia(T) - info(T, a) \quad (27)$$

A entropia dos dados é calculada por meio de $Entropia(T)$:

$$Entropia(T) = -\frac{9}{14} \times \log_2\left(\frac{9}{14}\right) - \frac{5}{14} \times \log_2\left(\frac{5}{14}\right) = 0.940 \text{ bits}$$

Pode-se, desta forma, calcular o ganho para cada atributo da base de dados, escolhendo assim o atributo que melhor separa por si só as informações. O exemplo a seguir mostra o ganho calculado em detalhes para o atributo *Tempo* e também os valores do ganho de informação dos atributos *Velocidade*, *Pressão dos Freios* e *Depressão*.

$$\begin{aligned} info(T, tempo) &= \frac{5}{14} \times \left(-\frac{3}{5} \times \log_2\left(\frac{3}{5}\right) - \frac{2}{5} \times \log_2\left(\frac{2}{5}\right) \right) \\ &+ \frac{4}{14} \times \left(\frac{4}{4} \times \log_2\left(\frac{2}{4}\right) \right) \\ &+ \frac{5}{14} \times \left(-\frac{3}{5} \times \log_2\left(\frac{3}{5}\right) - \frac{2}{5} \times \log_2\left(\frac{2}{5}\right) \right) \\ &= 0.694 \end{aligned}$$

$$info(T, velocidade) = 0.890$$

$$info(T, pressaodos freios) = 0.788$$

$$info(T, depressão) = 0.892$$

Após calcular a informação para todos os atributos, o próximo passo é calcular o ganho para os mesmos.

$$ganho(Tempo) = 0,940 - 0,694 = 0,246$$

$$ganho(Velocidade) = 0,940 - 0,890 = 0,05$$

$$ganho(Pressão Freios) = 0,940 - 0,788 = 0,152$$

$$ganho(Depressão) = 0,940 - 0,892 = 0,048$$

O atributo X que possuir o valor mais alto de $ganho(X)$ é então escolhido como sendo a raiz da árvore, neste caso o atributo *Tempo*. O critério de ganho possui um *bias* que favorece atributos com muitos valores possíveis, sendo um grande defeito do algoritmo. Quando há diferentes funções que possuem valores em comum para um determinado exemplo, deve haver a seleção de uma função que se encaixe para ambos os pontos *a priori*, tal seleção do valor *a priori* é chamada *bias*. Sem o valor do *bias*, um aprendizado útil se torna impossível (NILSSON, 1996). Tal fato ocorre quando há um

atributo como código, que possui valor único para cada instância do conjunto de dados, ocorrendo uma divisão que gera inúmeros subconjuntos com apenas um valor possível para cada um deles, onde o valor de $info(S)$ é igual a zero, portanto, mínimo. A correção para este problema consiste em ajustar os resultados possíveis, sendo que a potencial divisão da informação computa a subtração do conjunto de treinamento T e de cada subconjunto n , conforme Equação 28. A divisão da informação resulta em informações úteis na classificação, eliminando assim atributos que podem por si só classificar amostras (i.e. o código de um paciente, que é único e possui ganho de informação máximo. Apesar disto, não é recomendável sua utilização por não tornar genérica a árvore de decisão).

$$info_div(a_i) = - \sum_{i=1}^n \frac{|T_i|}{|T|} \times \log_2 \left(\frac{|T_i|}{|T|} \right) \quad (28)$$

A proporção de ganho de um atributo a_i é medida com base na informação relevante, Equação 29:

$$ganho_medio(a_i) = \frac{ganho(a_i)}{Info_div(a_i)} \quad (29)$$

A estrutura gerada a partir da execução do algoritmo C4.5 (implementação JRIP) sobre os dados da Tabela 4 pode ser encontrada na Figura 5. Cada valor entre parênteses corresponde ao número de exemplos bem classificados / número de exemplos mal classificados.

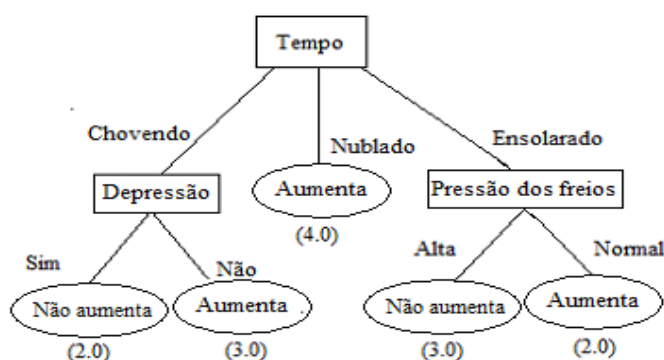


Figura 5 - Exemplo de árvore de decisão.

Uma versão simplificada da árvore também é gerada. Observe que o atributo *Velocidade* não é testado neste caso, sendo apenas necessário o teste dos atributos *Tempo*, *Pressão de freios* e *Depressão* para classificar um exemplo.

A mesma árvore pode ser vista no formato de regras (cf. Figura 6), sendo que ao lado da classificação da regra é mostrada a precisão da regra em porcentagem. Por exemplo, a Regra 4, pode ser lida da seguinte forma: SE *o tempo estiver chovendo* e *estiver em depressão* ENTÃO tem-se a indicação que não aumentará o ponto de aceleração com uma taxa de acerto de 50%. Caso nenhuma regra consiga classificar um determinado exemplo, o mesmo é associado à classe padrão *Aumenta*.

Se atribuídos valores contínuos a um atributo, conforme Tabela 5, então o ponto que melhor separa os exemplos de treinamento será dado por meio do limiar t . Se, para um determinado conjunto de exemplos existem N valores distintos para certo atributo A , então há $N - 1$ limiares possíveis que podem ser utilizados para testar o valor do atributo, sendo que cada limiar possível possui um determinado subconjunto de exemplos, e o valor que separa as amostras é calculado com base no limiar (QUINLAN, 1996).

```

Regra 1:
  SE      Tempo = Ensolarado E Pressão dos freios = Alta
  ENTÃO classe Não Aumenta [63.0%]

Regra 2:
  SE      Tempo = chovendo E Depressão = Sim
  ENTÃO classe Não Aumenta [50.0%]

Regra 3:
  SE      Tempo = Nublado
  ENTÃO classe Aumenta [70.7%]

Regra 4:
  SE      Tempo = Chovendo E Depressão = Sim
  ENTÃO classe Não Aumenta [50.0%]

Regra 5:
  SE      Pressão dos freios = normal
  ENTÃO classe Aumenta [66.0%]

Default classe:
  Aumenta

```

Figura 6 - Árvore de decisão no formato de regras.

A árvore de decisão da Figura 7 foi gerada pelo C4.5 a partir dos dados da Tabela 5, tomando como base o limiar cujo valor é igual a 25 para o atributo pressão dos freios.

Regra 1:	SE Pressão dos freios > 15 E Pressão dos freios <=70
	ENTÃO classe Aumenta [73.1%]
Regra 2:	SE Pressão dos freios > 70
	ENTÃO classe Não Aumenta [50%]
Classe Default:	Não Aumenta

Figura 7 - Classificador 1. C4.5 com atributos contínuos.

O mesmo conjunto de regras pode ser gerado conforme a Figura 8. A interpretação é a mesma, diferenciando apenas o software usado na criação da árvore de decisão.

PRESSÃO DOS FREIOS > 15 PRESSÃO DOS FREIOS <= 70 : Aumenta [73.1%] PRESSÃO DOS FREIOS > 70 : Não Aumenta [50%] => Não Aumenta

Figura 8 - Classificador 1. C4.5 com atributos contínuos (versão gerada pelo software WEKA).

Tabela 5. Conjunto de treinamento com valores contínuos. (Exemplo adaptado de (QUINLAN, 1993))

Exemplo	Tempo	Velocidade	Pressão dos freios	Depressão	Classe
1	Ensolarado	12	40	Sim	Aumenta
2	Ensolarado	15	80	Sim	Não Aumenta
3	Ensolarado	20	15	Não	Não Aumenta
4	Ensolarado	22	15	Não	Não Aumenta
5	Ensolarado	25	45	Não	Aumenta
6	Nublado	25	50	Não	Aumenta
7	Nublado	30	70	Não	Aumenta
8	Nublado	34	20	Sim	Aumenta
9	Nublado	20	50	Sim	Aumenta
10	Chovendo	22	80	Sim	Não Aumenta
11	Chovendo	26	70	Sim	Não Aumenta
12	Chovendo	30	30	Não	Aumenta
13	Chovendo	35	25	Não	Aumenta
14	Chovendo	36	10	Não	Aumenta

O maior ganho de informação foi para o atributo *Pressão dos Freios*, sendo este o escolhido como raiz da árvore. Caso o valor do atributo *Pressão dos Freios* seja maior que 15 e menor ou igual a 70 a classe associada é Aumenta, O valor do atributo *Pressão dos Freios* foi calculado em função do limiar t , visto que se trata de um atributo contínuo. Os valores de t_1 e t_2 são calculados da seguinte forma: ordenados os valores do atributo em ordem crescente, o próximo passo consiste em identificar pontos em que a classe associada ao atributo tem seu valor alterado. São identificados a quantidade de limiares e somados os valores antes e após a mudança da classe (10 + 15, 15 + 20 e 70 +

70) para serem divididos por 2. Após a divisão obtêm o valor do limiar. São calculados então o ganho de informação para o atributo com cada valor do limiar, sendo o melhor escolhido, conforme mostra a Figura 9.

Pressão dos freios	10	15	15	20	25	30	40	45	50	70	70	80
<i>Classe</i>	A	NA	NA	A	A	A	A	A	A	A	NA	NA
<i>Limiares</i>	$t_1 = (10 + 15) / 2 = 12.5$ $t_2 = (15 + 20) / 2 = 17.5$ $t_3 = (70 + 70) / 2 = 70.0$											

Figura 9 - Cálculo de limiares do atributo *pressão dos freios*.

Deste modo, o algoritmo atribuiu o menor valor mais próximo de 12.5, calculado com base no ganho de informação do atributo pressão dos freios.

Em certas situações o valor de um determinado atributo pode estar faltando, ou ser um valor desconhecido. Nestes casos, a medida de ganho é alterada. O ganho pode ser medido segundo a Equação 30, onde pA representa a probabilidade do atributo A possuir o valor conhecido, $n(T)$ o ganho de informação calculado pela função $info(X)$ e pnA a probabilidade do atributo A não possuir valor conhecido.

$$gain(X) = pA \times (n(T) - n_x(T)) \quad (30)$$

A divisão do conjunto de treinamento é realizada de modo probabilístico. Se o exemplo possuir uma saída cujos valores forem conhecidos, um exemplo do conjunto de treinamento deve possuir tal valor com probabilidade 1. Caso o valor da saída seja desconhecido, um peso representando a probabilidade de a amostra pertencer a cada subconjunto é atribuído. De forma análoga, caso o algoritmo deva classificar um exemplo cujo valor é desconhecido, todos os possíveis resultados de classificação serão explorados, onde a classe com maior probabilidade de classificação é a escolhida.

O resultado do C4.5 é mostrado por uma matriz de confusão. Nela são mostrados como o exemplo foi classificado e sua classificação correta (Figura 10).

=== Confusion Matrix ===		
a	b	classified as
8	1	a = aumenta
1	4	b = não aumenta

Figura 10 - Matriz de confusão.

Na Figura 10, 8 exemplos do tipo ‘a’ foram classificadas corretamente como ‘a’, e 4 exemplos do tipo ‘b’ foram classificadas corretamente. Um exemplo do tipo ‘a’ foi classificado como sendo do tipo ‘b’, havendo erro do classificador na classificação de uma amostra. E outro exemplo do tipo ‘b’ foi classificado como sendo do tipo ‘a’, também havendo erro do classificador na classificação de uma amostra.

A cada regra gerada pelo classificador (no formato *se X então Y*), o fator de suporte e o fator de confiança são atribuídos (HAN, et al., 2006). O fator de suporte é igual ao número de registros contendo a característica X e Y dividido pelo número total de registros. A confiança é igual ao número de registros com X e Y dividido pelo número de registros com X. Por exemplo, durante a condução de uma locomotiva do ponto de medida A para o ponto de medida B, o maquinista aumentou o ponto de aceleração 400 vezes. Em 40 destes 400 registros a locomotiva estava em subida e destes 40 aumentos, 20 aumentaram o ponto. Assim a regra “SE subida ENTÃO aumentar o ponto de aceleração” teria um suporte de $40/400 = 10\%$ e confiança de $20/40 = 50\%$.

3.5.1.1 Algoritmo JRIP

Conforme mencionado anteriormente, neste trabalho utilizaremos o algoritmo RIPPER (COHEN, 1995), chamado de JRIP na implementação do software WEKA por ser implementado em linguagem Java. O algoritmo RIPPER (*Repeated Incremental Pruning to Produce Error Reduction*), constrói um conjunto de regras que modelam um conjunto de dados.

O algoritmo C4.5 pode criar um conjunto de regras a partir de uma árvore de decisão não-podada. O processo de aperfeiçoamento remove ou adiciona regras de modo a reduzir a dimensionalidade do conjunto de regras, repetindo o processo para subconjuntos de diferentes tamanhos e selecionando o melhor subconjunto ao término do processo. Por sua vez o algoritmo RIPPER gera o primeiro modelo adicionando uma regra por vez. O passo seguinte consiste em remover todos os exemplos classificados pela regra do conjunto base de geração das regras, processo chamado de *pruning*. O processo é repetido até que não existam mais exemplos a serem classificados positivamente, ou até que uma regra gerada possua uma taxa de erro alta.

3.5.2 Método BAGGING com o algoritmo C4.5

Outra forma de obter uma árvore de decisão sugerida em (BREIMAN, 1996) é por meio do método BAGGING (*Bootstrap aggregating*). Este método baseia-se na geração e combinação de múltiplos classificadores obtidos a partir de diferentes amostras de um conjunto de treinamento (SILLA, et al., 2005). Pode-se imaginar que utilizar a mesma base de treinamento para a geração de diferentes classificadores seja algo incoerente, porém, uma pequena mudança no conjunto de dados de treinamento pode fazer com que um atributo, que a princípio se mostre insignificante, torne-se um atributo importante. Tal fato faz com que diferentes classificadores gerem classificações diferentes para uma mesma instância.

Dado um conjunto de dados consistentes $L = \{(y_n, x_n), n = 1, \dots, N\}$, onde os valores de y correspondem às classificações, pode-se prever y a partir de uma entrada x , aplicando o preditor $\varphi(x, L)$. A partir do momento em que são gerados vários conjuntos i (amostras) de dados de aprendizagem $\{L_i\}$ cada um consistindo de N diferentes observações da mesma distribuição, tem-se também diferentes preditores $\varphi(x, L_i)$ para um conjunto de classes $j \in \{1, \dots, J\}$ (BREIMAN, 1996).

Caso o rótulo da classe seja um valor numérico, então o valor atribuído ao preditor é a média dos valores de $\varphi(x, L_i)$, caso contrário é feita uma votação. A votação pode ser feita com base no peso do atributo ou com base na média dos valores do atributo, caso o mesmo seja numérico (WITTEN & FRANK, 2005). Para o método *Bagging* o peso é sempre o mesmo para diferentes modelos.

É importante notar que o método de BAGGING deixa dúvida quanto à melhora da estabilidade de um procedimento construído por φ (BREIMAN, 1996). Por exemplo, se ocorrer alguma mudança no conjunto de dados, como uma replicação do conjunto, poucas mudanças irão ocorrer nos preditores (BREIMAN, 1996). Algumas melhoras ocorrem em algoritmos instáveis, onde uma pequena mudança no conjunto de treinamento ocasiona grandes mudanças nos preditores.

O método de BAGGING consiste em executar o seguinte procedimento cada vez que o critério de parada não é satisfeito (i.e. o número de classificadores não é alcançado):

1. O primeiro passo consiste em dividir o conjunto de dados em 2 subconjuntos: um para testes (T) um para aprendizagem (L).

2. É gerado então um classificador φ a partir do conjunto de aprendizagem, obtendo uma taxa de erro estimada $e_S(L, T)$ do classificador. A taxa de erro é calculada utilizando o conjunto de testes para fazer inferências sobre o classificador φ .
3. Uma amostra L_B é gerada a partir do conjunto L gerando um preditor φ_B por i vezes, tendo $\varphi_1, \dots, \varphi_i$.
4. Se a classe j associada ao elemento x pertence ao conjunto de testes T , então a classe de x é aquela que mais está presente entre $\varphi_1(x), \dots, \varphi_k(x)$. A proporção de vezes em que a classe é diferente da classe correta é dada pela taxa de erro de BAGGING $e_B(L, T)$.
5. A divisão aleatória dos dados em conjuntos de testes e conjunto de aprendizagem é repetida k vezes e gerados então e_S, e_B sobre a média das k iterações.

Uma ilustração do funcionamento do algoritmo de BAGGING é mostrada a seguir. Os dados da Tabela 6 são utilizados para tal ilustração.

A geração de cada amostra é feita de forma aleatória. As tabelas 7, 8 e 9 definem os conjuntos de aprendizagem: L_1, L_2 e L_3 respectivamente. Estes conjuntos foram obtidos a partir da Tabela 6.

Tabela 6. Conjunto de treinamento. (Exemplo adaptado de (QUINLAN, 1993))

Exemplo	Tempo	Velocidade	Pressão dos freios	Depressão	Classe
1	Ensolarado	Média	Normal	Sim	Aumenta
2	Ensolarado	Alta	Alta	Sim	Não Aumenta
3	Ensolarado	Alta	Alta	Não	Não Aumenta
4	Ensolarado	Média	Alta	Não	Não Aumenta
5	Ensolarado	Baixa	Normal	Não	Aumenta
6	Nublado	Alta	Alta	Não	Aumenta
7	Nublado	Alta	Normal	Não	Aumenta
8	Nublado	Baixa	Normal	Sim	Aumenta
9	Nublado	Média	Alta	Sim	Aumenta
10	Chovendo	Baixa	Normal	Sim	Não Aumenta
11	Chovendo	Média	Alta	Sim	Não Aumenta
12	Chovendo	Baixa	Normal	Não	Aumenta
13	Chovendo	Média	Alta	Não	Aumenta
14	Chovendo	Média	Normal	Não	Aumenta

Tabela 7. Conjunto de aprendizagem L₁.

Exemplo	Tempo	Velocidade	Pressão dos freios	Depressão	Classe
1	Ensolarado	Média	Normal	Sim	Aumenta
2	Ensolarado	Alta	Alta	Sim	Não Aumenta
3	Ensolarado	Alta	Alta	Não	Não Aumenta
4	Ensolarado	Média	Alta	Não	Não Aumenta
5	Ensolarado	Baixa	Normal	Não	Aumenta
6	Nublado	Alta	Alta	Não	Aumenta
7	Nublado	Alta	Normal	Não	Aumenta
8	Nublado	Baixa	Normal	Sim	Aumenta
9	Nublado	Média	Alta	Sim	Aumenta
10	Chovendo	Baixa	Normal	Sim	Não Aumenta
11	Chovendo	Média	Alta	Sim	Não Aumenta
12	Chovendo	Baixa	Normal	Não	Aumenta
13	Chovendo	Média	Alta	Não	Aumenta
14	Chovendo	Média	Normal	Não	Aumenta

Tabela 8. Conjunto de aprendizagem L₂

Exemplo	Tempo	Velocidade	Pressão dos freios	Depressão	Classe
1	Ensolarado	Média	Normal	Sim	Aumenta
2	Ensolarado	Alta	Alta	Sim	Não Aumenta
3	Ensolarado	Alta	Alta	Não	Não Aumenta
4	Ensolarado	Média	Alta	Não	Não Aumenta
5	Ensolarado	Baixa	Normal	Não	Aumenta
5	Ensolarado	Baixa	Normal	Não	Aumenta
7	Nublado	Alta	Normal	Não	Aumenta
8	Nublado	Baixa	Normal	Sim	Aumenta
8	Nublado	Baixa	Normal	Sim	Aumenta
9	Nublado	Média	Alta	Sim	Aumenta
11	Chovendo	Média	Alta	Sim	Não Aumenta
12	Chovendo	Baixa	Normal	Não	Aumenta
13	Chovendo	Média	Alta	Não	Aumenta
14	Chovendo	Média	Normal	Não	Aumenta

Tabela 9. Conjunto de aprendizagem L₃

Exemplo	Tempo	Velocidade	Pressão dos freios	Depressão	Classe
1	Ensolarado	Média	Normal	Sim	Aumenta
2	Ensolarado	Alta	Alta	Sim	Não Aumenta
3	Ensolarado	Alta	Alta	Não	Não Aumenta
4	Ensolarado	Média	Normal	Sim	Aumenta
7	Nublado	Baixa	Normal	Sim	Aumenta
7	Nublado	Alta	Normal	Não	Aumenta
8	Nublado	Baixa	Normal	Sim	Aumenta
8	Nublado	Baixa	Normal	Sim	Aumenta
9	Nublado	Média	Alta	Sim	Aumenta
9	Nublado	Média	Alta	Sim	Aumenta
11	Chovendo	Média	Alta	Sim	Não Aumenta
12	Chovendo	Baixa	Normal	Não	Aumenta
13	Chovendo	Baixa	Normal	Não	Aumenta
14	Chovendo	Média	Normal	Não	Aumenta

A partir dos três diferentes conjuntos de aprendizagem e testes, mostrados nas tabelas 7, 8 e 9 foram geradas as árvores de decisão representadas nas figuras 11, 12 e 13. Cabe ressaltar que os conjuntos de treinamento e testes são iguais para cada classifi-

gador, visto que durante o processo são geradas amostras aleatórias com reposição. Para o conjunto de testes e aprendizagem L_1 a seguinte árvore foi gerada:

```

TEMPO = Ensolarado
| PRESSÃO DOS FREIOS = Normal: Aumenta (2.0)
| PRESSÃO DOS FREIOS = Alta: Não_ Aumenta (3.0)
TEMPO = Nublado: Aumenta (4.0)
TEMPO = Chovendo
| DEPRESSÃO = Sim: Não_ Aumenta (2.0)
| DEPRESSÃO = Não: Aumenta (3.0)

```

Figura 11 - Classificador 2. BAGGING – L1 e T1.

A taxa de erro para o Classificador 2 foi de zero por cento, sendo todos os exemplos classificados corretamente pela árvore em questão.

Para o conjunto de testes T_2 e conjunto de aprendizagem L_2 foi gerada a seguinte árvore:

```

PRESSÃO DOS FREIOS = Normal: Aumenta(8.0)
PRESSÃO DOS FREIOS = Alta: Não_Aumenta(6.0/2.0)

```

Figura 12 - Classificador 3. BAGGING – L2 e T2.

A Figura 12 mostra o Classificador 3 que por sua vez possui uma taxa de classificação incorreta de 28,57% e uma taxa de classificação correta de 71,43%. Três exemplos rotulados com a classe *Aumenta* foram classificados como sendo da classe *Não Aumenta* e um exemplo da classe *Não Aumenta* foi também mal classificado, conforme mostra a matriz de confusão a seguir. Novamente, neste caso, o conjunto de treinamento e testes foi o mesmo, porém as amostras foram sorteadas com reposição.

```

a b <-- classificado como
6 3 | a = Aumenta
1 4 | b = Não_Aumenta

```

Finalmente, para o terceiro conjunto de aprendizagem a árvore de decisão foi gerada (Figura 13).

```

PRESSÃO DOS FREIOS = Normal: Aumenta (9.0)
PRESSÃO DOS FREIOS = Alta
| TEMPO = Ensolarado: Não_ Aumenta (2.0)
| TEMPO = Nublado: Aumenta (2.0)
| TEMPO = Chovendo: Não_ Aumenta (1.0)

```

Figura 13 - Classificador 4. BAGGING – L3 e T3.

O Classificador 4 apresentou uma taxa de classificação incorreta de 14,29% e uma taxa de classificação correta de 85,71%. A matriz de confusão é mostrada a seguir.

a	b	<--	classificado como
8	1		a = Aumenta
1	4		b = Não_ Aumenta

A classificação de um exemplo $x = \langle 15, \text{Ensolarado}, \text{Alta}, \text{Normal}, \text{Não}, \text{Aumenta} \rangle$ é realizada submetendo tal exemplo aos três classificadores gerados anteriormente. Para o caso em questão, os três classificadores geraram a mesma classificação, a classe *Aumenta*. Os três classificadores acertaram a classe.

Para a classificação de outra instância $y = \langle 16, \text{Nublado}, \text{Alta}, \text{Alta}, \text{Não}, \text{Aumenta} \rangle$, o Classificador 2 e o Classificador 4 classificaram o evento y corretamente como *Aumenta*. O Classificador 3 classificou erroneamente o evento y . Neste caso, utiliza-se um mecanismo de votação que classificará o evento y com *Aumenta*. Vence a classe que obtiver o maior número de votos. No caso em questão, a classe *Aumenta* vence por 2 votos contra 1.

Em trabalhos já realizados por Lopes (2007), cujas bases de dados possuíam um grande número de atributos, o método BAGGING mostrou-se eficiente quando utilizado com amostras de dados pequenas, em torno de 10% a 20% da base de dados original. Quando aplicado em bases com um elevado número de atributos, BAGGING mostrou-se superior ao algoritmo C4.5 (LOPES, 2007).

3.5.3 Método BOOSTING com o algoritmo C4.5

A proposta do método BOOSTING é melhorar a precisão dos algoritmos de aprendizagem, reduzindo a taxa de erro por meio de uma técnica que busca combinar classificadores. Apesar dos potenciais benefícios que o método promete pelos resultados teóricos, a avaliação do mesmo só pode ser realizada na prática, em problema de aprendizagem reais, visto que os resultados teoricamente prometidos devem ser comprovados por meio de experimentos em diversas bases de dados (FREUND, et al., 1996) (LOPES, 2007).

O método BAGGING é eficiente quando é feita a combinação de classificações a partir de modelos diferentes, explorando desta forma a instabilidade do método de aprendizagem, à medida que um modelo complementa outro modelo (WITTEN, et al., 2005). Por sua vez, o método BOOSTING explora modelos que são complementares reciprocamente. Ele faz isto de forma análoga ao método BAGGING por meio de pro-

cesso de votação para classificar eventos cujas classes são discretas ou a realização do cálculo da média dos previsores numéricos.

O objetivo do método BOOSTING é gerar com eficiência hipóteses precisas usando um algoritmo de aprendizagem que gera hipóteses de baixa precisão, também chamado algoritmo de aprendizagem fraco (FREUND, et al., 1996). Esta abordagem é útil em problemas nos quais existem diferentes níveis de dificuldade, particularmente em problemas do mundo real, onde o método de aprendizagem tende a gerar distribuições que se concentram nos exemplos mal-classificados. Este comportamento faz com que um algoritmo melhore seu desempenho nestes conjuntos de exemplos. O método BOOSTING também faz com que o algoritmo fraco se torne sensível a mudanças no conjunto de treinamento, gerando diferentes hipóteses para diferentes conjuntos. Estas últimas tendem a serem complementares.

A mecânica do método BOOSTING é a seguinte: ele recebe m exemplos com o conjunto de treinamento $L = \{(x_1, y_1), \dots, (x_m, y_m)\}$ onde x_i é uma instância do espaço X e y_i é a classe pertencente ao conjunto de classes $Y = \{1, \dots, k\}$ associada à instância x_i e k é o número de possíveis valores da classe. O algoritmo é executado por T vezes, e em cada iteração t , é gerada uma distribuição de probabilidade D_t sobre o conjunto de treinamento L , resultando uma hipótese h_t sobre o espaço X para as possíveis classificações. O objetivo é encontrar a hipótese que minimize o erro do próximo classificador gerado. No método BOOSTING esta geração de distribuição de probabilidade é feita com base em uma distribuição uniforme (BAUER & KOHAVI, 1999). O processo continua até que sejam combinadas todas as hipóteses h_1, \dots, h_T em uma única hipótese final h_{fin} .

A primeira distribuição de probabilidade D_1 é gerada com base no conjunto de treinamento L , sendo que o peso do primeiro elemento é igual à $\frac{1}{m}$, onde m é o número elementos de conjunto de treinamento. A próxima distribuição de probabilidade é gerada por meio da multiplicação do peso do exemplo i por algum valor $B_t \in [0,1)$, se h_t classifica corretamente x_i , caso contrário o peso muda. Em seguida é feita a normalização dos pesos dividindo-os pela constante Z_t . A hipótese final h_{fin} tem um peso mais importante que as demais hipóteses, i.e., para certa instância x , h_{fin} gera um rótulo y que maximiza a soma dos pesos das hipóteses fracas pré-estabelecidas pelo rótulo. O peso da hipótese h_t influencia o classificador de tal modo que um peso alto gera um erro baixo, e é calculado por meio da seguinte fórmula: $\log(\frac{1}{B_t})$. (FREUND & SCHAPIRE, 1996).

Os passos abaixo ilustram a mecânica do método BOOSTING (WITTEN, et al., 2005).

1. Para cada exemplo de treinamento m atribua um peso igual.
2. Gere o classificador γ utilizando uma distribuição D_t .
3. Para cada uma das t, \dots, T vezes em que o algoritmo é executado faça
 4. Aplique o algoritmo de aprendizagem ao conjunto L com respectivo peso e armazene a hipótese $h_t : X \times Y \rightarrow [0,1]$ resultante
5. Calcule o erro do modelo resultante do passo 4 com base na equação 31

$$erro_t = \sum_{i:h_t(x_i) \neq y_i} D_t(i) \quad (31)$$

6. Se o erro for igual à zero, ou maior igual a 0.5 atribua $T = t - 1$ e cancele o laço
7. Calcule o valor de $B_t = erro_t / (1 - erro_t)$
8. Gere uma nova distribuição

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \times B_t \text{ se a hipótese for classificada corretamente ou}$$

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \times 1 \text{ caso contrário}$$

Sabendo que o valor de Z_t é uma constante utilizada para normalizar o peso das amostras.

9. Gere a hipótese de classificação final com base na equação 32

$$h_{fin}(x) = \arg \max_{y \in Y} \sum_{t:h_t(x)=y} \log \frac{1}{B_t} \quad (32)$$

O funcionamento do método BOOSTING é ilustrado a seguir. Primeiramente, uma amostra é gerada a partir da base de dados da Tabela 10 onde o número de iterações é Z .

Tabela 10. Conjunto de treinamento. (Exemplo adaptado de (QUINLAN, 1993))

Exemplo	Tempo	Velocidade	Pressão dos freios	Depressão	Classe
1	Ensolarado	Média	Normal	Sim	Aumenta
3	Ensolarado	Alta	Alta	Não	Não Aumenta
3	Ensolarado	Alta	Alta	Não	Não Aumenta
4	Ensolarado	Média	Alta	Não	Não Aumenta
4	Ensolarado	Média	Alta	Não	Não Aumenta
6	Nublado	Alta	Alta	Não	Aumenta
7	Nublado	Alta	Normal	Não	Aumenta
8	Nublado	Baixa	Normal	Sim	Aumenta
8	Nublado	Baixa	Normal	Sim	Aumenta
9	Nublado	Média	Alta	Sim	Aumenta
10	Chovendo	Baixa	Normal	Sim	Não Aumenta
10	Chovendo	Baixa	Normal	Sim	Não Aumenta
12	Chovendo	Baixa	Normal	Não	Aumenta
12	Chovendo	Baixa	Normal	Não	Aumenta

O classificador 5 apresentou uma taxa de classificação incorreta de 14,29% e uma taxa de classificação correta de 85,71%. Três exemplos rotulados com a classe *Aumenta* foram classificados como sendo da classe *Aumenta* e um exemplo da classe *Não Aumenta* foi também mal classificado, conforme mostra o conjunto de regras da Figura 14.

```

TEMPO = Ensolarado: Não_Aumenta (5.0/1.0)
TEMPO = Nublado: Aumenta (5.0)
TEMPO = Chovendo
|  DEPRESSÃO = Sim: Não_ Aumenta (2.0)
|  DEPRESSÃO = Não: Aumenta (2.0)

```

Figura 14 - Classificador 5. BOOSTING – L-1 e T-1.

Como o valor do erro foi menor que 0.5, então é calculado a importância do classificador, dada pela razão: $B_i = \frac{0.142}{1-0.142} = 0.165$ e a distribuição atualizada. I.e., como a taxa de erro não foi superior a 0.5, são recalculados os pesos das instâncias j da nova amostra. A normalização do peso faz com que a soma de todos os pesos seja igual a 1, onde é escolhido aleatoriamente um valor dentro da variação entre 0 a 1. Supondo que uma determinada instância tenha como peso 0.5. Esta instância tem uma probabilidade maior de ser escolhida na geração da nova amostragem do que uma instância com o peso 0.1.

Tabela 11. Pesos atualizados.

j	w _j inicial	w _j passo 1
1	0,071	0,100
2	0,071	0,017
3	0,071	0,017
4	0,071	0,017
5	0,071	0,100
6	0,071	0,017
7	0,071	0,017
8	0,071	0,017
9	0,071	0,017
10	0,071	0,017
11	0,071	0,017
12	0,071	0,017
13	0,071	0,017
14	0,071	0,017

Após recalculer o peso das instâncias, uma nova amostra é gerada (Tabela 12) e para ela um classificador é gerado (Figura 15).

Tabela 12. Conjunto de treinamento, (Exemplo adaptado de (QUINLAN, 1993))

Exemplo	Tempo	Velocidade	Pressão dos Freios	Depressão	Classe
2	Ensolarado	Alta	Alta	Sim	Não Aumenta
2	Ensolarado	Alta	Alta	Sim	Não Aumenta
3	Ensolarado	Alta	Alta	Não	Não Aumenta
4	Ensolarado	Média	Alta	Não	Não Aumenta
5	Ensolarado	Baixa	Normal	Não	Aumenta
5	Ensolarado	Baixa	Normal	Não	Aumenta
6	Nublado	Alta	Alta	Não	Aumenta
8	Nublado	Baixa	Normal	Sim	Aumenta
9	Nublado	Média	Alta	Sim	Aumenta
10	Chovendo	Baixa	Normal	Sim	Não Aumenta
10	Chovendo	Baixa	Normal	Sim	Não Aumenta
12	Chovendo	Baixa	Normal	Não	Aumenta
13	Chovendo	Média	Alta	Não	Aumenta
14	Chovendo	Média	Normal	Não	Aumenta

A Figura 15 mostra a árvore de decisão gerada para segunda iteração do algoritmo, cujos exemplos tiveram seus pesos atualizados conforme a Tabela 11.

```

TEMPO = Ensolarado
| PRESSÃO DOS FREIOS = Normal: Aumenta(2.0)
| PRESSÃO DOS FREIOS = Alta: Não_Aumenta(4.0)
TEMPO = Nublado: Aumenta(3.0)
TEMPO = Chovendo
| DEPRESSÃO = Sim: Não_Aumenta(2.0)
| DEPRESSÃO = Não: Aumenta(3.0)

```

Figura 15 - Classificador 6. BOOSTING – L-2 e T-2.

Para este segundo classificador, se aplicarmos como conjunto de testes os exemplos da Tabela 4, podemos observar que todos os exemplos foram corretamente classificados, encerrando assim o laço, pois o número de iterações foi definido como 2.

A taxa de acerto do classificador pode ser utilizada na avaliação de desempenho. Porém, tal avaliação precisa ser simples e eficiente. Entretanto, é inaceitável utilizar, por exemplo, toda a base de dados sobre um assunto pesquisado para ambos os processos: treinamento e teste. A principal falha é que se pode gerar um resultado tendencioso por utilizar o mesmo conjunto de dados em ambos os processos. Neste contexto surge o processo de validação cruzada que será discutido na seqüência.

3.6 Validação Cruzada

Existem inúmeros métodos para a avaliação de um classificador. Um dos métodos mais simples proposto na literatura por (TAN, et al., 2006), chamado de *holdout com re-amostragem aleatória*, divide o conjunto de dados em dois subconjuntos: um de treinamento e um de conjunto de testes. As classes presentes nestes conjuntos devem ser representativas, ou seja, devem estar presentes na mesma proporção da base de dados original para que o classificador gerado possa ser eficiente. Caso uma classe não esteja presente no conjunto de treinamento, o classificador não será capaz de prever o valor de um exemplo que pertença a tal classe. O mesmo ocorre caso a base de treinamento possua muitos exemplos de uma mesma classe, onde o classificador também não será capaz de classificar amostras de uma classe diferente. Para eliminar tal problema, o método de estratificação é utilizado.

O método de estratificação consiste em treinar e testar um classificador inúmeras vezes, com a mesma distribuição dos dados originais, sempre com diferentes amostras aleatórias. Assim são gerados inúmeros classificadores e a taxa de erro total consistirá na média das taxas de erro dos classificadores gerados, método chamado de *repeated holdout* (WITTEN, et al., 2005).

No processo de *holdout* simples, um conjunto de treinamento e um conjunto de testes são gerados. O classificador gerado pelo conjunto de treinamento é testado utilizando as amostras presentes no conjunto de testes. Já o classificador gerado pelo conjunto de testes é avaliado utilizando as amostras do conjunto de treinamento. Este método é eficiente quando as mesmas proporções de tamanho são utilizadas para ambos os conjuntos: teste e treinamento.

A validação cruzada surge como uma alternativa para o processo de avaliação de desempenho. A base de dados com n instâncias é dividida em f amostras (chamados de *fold*s) aleatórias e com tamanho de amostra igual ao número de instâncias dividido pelo número de amostras (TAN, et al., 2006), (DIAMANTIDIS, et al., 2000). Em cada f amostras é utilizado $f - 1$ amostras para treinamento. O número de vezes em que o processo de treinamento e teste será executado é igual ao número de amostras, retornando uma taxa de acerto para cada amostra.

Por exemplo, em uma base de dados contendo 100 instâncias, cujo número de amostras é 5, teremos para cada amostra 20 instâncias. Em cada um dos 5 passos, serão utilizadas 4 amostras como conjunto de treinamento, que irá conter 80 instâncias. As outras 20 instâncias não utilizadas no treinamento serão utilizadas no processo de teste. A taxa de acerto a média final dos acertos das 5 amostras utilizadas como teste.

A precisão de um classificador pode ser superestimada ou subestimada, expressa por meio do valor denominado de *bias*. Este último é calculado da seguinte forma: a expectativa do avaliador é subtraída da exatidão da classificação ($bias = E[\hat{o}] - o$, onde E denota a expectativa de acerto do classificador, \hat{o} o avaliador e o a exatidão do classificador). Já a variância estima qual a variação do avaliador (DIAMANTIDIS, et al., 2000).

A validação cruzada com amostragem estratificada busca distribuir em cada amostra a mesma proporção de instâncias do conjunto de dados inicial, dando estimativas melhores para os valores de *bias* e *variância*, sendo aplicada quando as classes não são distribuídas uniformemente no conjunto de treinamento (GU, et al., 2000).

Já na validação cruzada *leave-one-out*, o número de amostras (n) é igual ao número de exemplos, sendo o classificador construído n vezes. Neste tipo de validação, todos os dados são utilizados para treinamento, não havendo sub-amostras aleatórias. Este processo possui a desvantagem de ser computacionalmente oneroso.

A escolha do número de amostras na validação cruzada é uma tarefa não trivial, uma vez que as seguintes situações são possíveis:

- um número pequeno de amostras pode resultar em uma variância pequena e um valor de *bias* elevado; e
- um número grande de amostras pode resultar em uma variância alta e um valor de *bias* baixo.

O método de validação cruzada é pessimista com relação à exatidão, na medida em que cada iteração apenas um subconjunto de instâncias é utilizado para teste. É pos-

sível que instâncias pertencentes a uma determinada classe estejam presentes com muita frequência em uma amostra gerada pelo processo de validação, o que resultaria em uma tendência do algoritmo em ser específico a tal classe.

Com a aplicação de algoritmo de aprendizagem é obtida uma taxa de erro cada vez que o mesmo é executado, independente do algoritmo utilizado. Neste trabalho, para sugerir determinada ação ao maquinista, como por exemplo, aumentar a velocidade atual na mudança de ponto de aceleração, o sistema deve ter a “certeza” do que está sugerindo. Erros como um aumento na velocidade em lugar proibido pode acarretar sérios problemas, ocasionando perdas financeiras e até de vidas.

3.7 Teste Estatístico

Testes estatísticos tendem a verificar se houve ou não melhoria no processo (DEMSAR, 2006). Devido às características dos experimentos optamos por aplicar testes não-paramétricos porque são mais prováveis de rejeitar a hipótese nula.

Conforme Callegari (2003), nos testes paramétricos os valores da variável estudada devem ter distribuição normal ou aproximação normal. Já os testes não-paramétricos (i.e. Teste de Friedman), também chamados por testes de distribuição livre, não têm exigências quanto ao conhecimento da distribuição da variável na população.

O Teste de Friedman é um teste não-paramétrico, ou seja, não têm exigências quanto ao conhecimento da distribuição da variável na população, testando associações, dependência/independência e modelos ao invés de parâmetros (CALLEGARI, et al., 2003). Os algoritmos são ranqueados para cada conjunto de dados separadamente, onde o algoritmo com melhor desempenho ocupa a primeira posição do ranking e assim sucessivamente. Em caso de repetições destes valores é feita a média dos rankings.

Deste modo, o Teste de Friedman compara a média dos algoritmos segundo a Equação 33, onde r_j^i é o rank do j -ésimo de k algoritmos no i -ésimo de N conjuntos de dados e é realizado de acordo com a Equação 34.

$$R_j = \frac{1}{N} \sum_i r_i^j \quad (33)$$

$$\chi_F^2 = \frac{12}{k(k+1)} \left[\sum_j R_j^2 - \frac{k(k+1)}{4} \right] \quad (34)$$

O objetivo deste teste é verificar se os classificadores gerados apresentam diferenças significativas, não caracterizando a hipótese nula. Caso a hipótese nula seja caracterizada, todos os algoritmos são equivalentes uma vez que possuem rankings iguais. Ao considerarmos como hipótese nula a inexistência de diferenças entre as condições dos k algoritmos, teríamos amostras bem distribuídas, não havendo co-relação entre elas. Porém, para verificar se há ou não correlação entre as condições, deve-se fazer o somatório das variâncias (Q) dos ranques. Obtido o valor de Q , calcula-se o p -valor como a probabilidade do valor ser superior ou igual à variância obtida utilizando a distribuição qui-quadrada com $k-1$ graus de liberdade. O resultado numérico do teste de Friedman fornece um nível de significância (p -valor), caso este seja menor que 0.05 (valor adotado nesta dissertação, que representa 5% de tolerância de aceitação) então é recomendável rejeitar a hipótese nula.

3.8 Trabalhos relacionados

Diversas técnicas de aprendizagem de máquina têm sido adotadas como opção no tratamento de problemas que exigem controle, regularidade e segurança.

Benenson (et al., 2008) tratou o problema de navegação autônoma de um carro-robô em ambientes urbanos, os quais consistiam de geometrias heterogêneas e com obstáculos móveis, tornando complexa a navegação autônoma do robô. A solução encontrada pelo autor foi desenvolver um agente capaz de perceber as mudanças no ambiente e planejar ações as quais considera explícitas à natureza dinâmica do veículo e do ambiente enquanto garante que as restrições de segurança sejam obedecidas.

Kolski, em (KOLSKI, et al., 2006), apresentou um sistema de navegação híbrido que combina os benefícios das abordagens existentes para condução em ambientes estruturados (i.e. rodovias) e não estruturados (i.e. estacionamentos). O sistema utiliza detecção visual em tais passagens e faixas de laser para gerar um mapa local do ambiente, o qual é processado por um planejador local para guiar o veículo enquanto desvia os obstáculos. Quando andando em ambientes não estruturados, o sistema emprega um mapa global e planejador para gerar uma trajetória eficiente até o objetivo.

Bertolazzi (et al., 2008) desenvolveram um veículo autônomo de tamanho reduzido e focado na estratégia de controle a qual é baseada no algoritmo “*Nonlinear Receding Horizon Control*” (NRHC). O planejador é invocado por um gerenciador de alto nível para resolver uma seqüência de problemas de controles. O movimento

planejado provê uma seqüência de conjunto de pontos referenciados até que um novo plano esteja disponível.

O uso de outras técnicas de inteligência artificial como a lógica *fuzzy* foi objeto de estudo para o desenvolvimento de sistemas de controle automático de trens, conforme proposto por (RACHEL, 2006). O autor conclui que é possível utilizar lógica *fuzzy* no controle metroviário, mesmo se tratando de uma aplicação crítica de controle. O sistema metroviário possui uma característica de ser veloz e, por outro lado, as distâncias entre estações não são tão grandes como ocorre no sistema ferroviário, acelerando e freando com maior taxa possível e ocasionando maior manutenção do sistema de propulsão e frenagem. Para a implantação do sistema *fuzzy* de controle, foram feitas modificações na sinalização fixa da via, uma vez que o controle de espaçamento dos trens é feito pelo próprio controlador. As ações executadas pelo controlador proposto consistem apenas em acelerar ou desacelerar o trem, tomando como base a posição e a velocidade relativa de dois trens na via.

Outro método de aprendizagem proposto na literatura por (PASQUIER, et al., 2008) consiste no desenvolvimento de um sistema de condução inteligente capaz de aprender com base em exemplos passados. Neste, o autor modela as habilidades de condução humana utilizando regras para mapear as entradas de informações. A aprendizagem é feita com o auxílio de subsistemas baseados em regras *fuzzy*, sendo validadas pelo autor em simulação utilizando um carro equipado com um microprocessador capaz de controlar as ações do sistema. Algumas vantagens foram mostradas pelo autor, tais como: capacidade de fazer curvas, frear o veículo, conduzir o veículo em linha reta com sucesso em velocidades de 50 a 80 km/h. Apesar de tais vantagens, houve a necessidade da produção de um hardware específico para validar os experimentos além da escolha dos parâmetros adequados ser feita com base no processo de tentativa e erro, visto que as características do problema não eram conhecidas *a priori*.

3.9 Considerações Finais

Ao longo deste capítulo, foram examinados os métodos que foram utilizados em nossos experimentos. O C4.5, além de possuir a capacidade de gerar modelos de classificação a partir de exemplos com valores faltantes, apresenta bons resultados em bases de dados ruidosas. Durante a análise dos dados obtidos e que foram utilizados, amostras ruidosas foram encontradas, i.e. velocidade atual fora dos padrões ou acima da máxima

permitida e códigos que indicavam a perda de sinal do GPS. Frente estas características o C4.5 tende a apresentar resultados bastante satisfatórios, principalmente quando combinado com métodos BAGGING e BOOSTING.

Foi possível observar que durante experimentos realizados por Breiman (1996) o método de BAGGING reduziu a taxa de erro do classificador C4.5 quando aplicado em bases de dados com características diferentes. Em Lopes (2007) BAGGING obteve resultados satisfatórios quando selecionada uma quantidade pequena de dados.

Tanto o método de BAGGING quanto o método de BOOSTING foram utilizados por Lopes (2007) para a obtenção de conhecimento em bases de dados da área da saúde, onde a quantidade de atributos era significativamente grande. BOOSTING também obteve bons resultados quando filtrados os dados da base original, não havendo perda na taxa de acerto após o processo de filtragem ou quando aumentado o número de classificadores gerados.

A validação cruzada complementa o nosso ferramental. Ela é particularmente interessante por proporcionar aos classificadores conjuntos de dados com a mesma proporção do conjunto de dados original, eliminando o problema da proporção de exemplos. Tal método proporciona uma confiança quanto aos modelos gerados por ser um método confiável na avaliação de classificadores.

Concluindo, trabalhos já desenvolvidos e citados neste capítulo contribuem para a elaboração deste trabalho, servindo como base para a compreensão da complexidade da condução de veículos autônomos. Observamos que os resultados apresentados por estes trabalhos são relacionados a veículos com características diferentes do proposto neste trabalho, considerando, por exemplo, o tamanho, capacidade de transporte, funcionalidade e peso.

No próximo capítulo serão descritos os procedimentos que serão realizados durante o trabalho para a obtenção das regras de condução.

Capítulo 4

Metodologia

O trabalho proposto visa descobrir padrões, a partir de dados coletados por meio de diferentes sensores instalados em um trem de carga, para ajudar no planejamento e execução de uma boa política de condução. A consecução deste objetivo inclui diferentes tarefas não triviais, a saber: (i) um estudo detalhado sobre os dados coletados a partir de dados por meio de diferentes sensores instalados em um trem; (ii) a execução de inúmeros experimentos visando extrair conhecimentos úteis para o processo de condução; e (iii) uma análise criteriosa para validar as regras, a qual será feita por meio da comparação das taxas de acerto dos classificadores, similaridade da condução e Teste de Friedman.

Os conhecimentos descobertos devem ajudar realizar uma boa condução de um trem de carga. Relembrando, segundo a ALL (2008), uma boa condução caracteriza-se por: (i) economia de combustível, (ii) minimização de esforços internos entre veículos e destes sobre a via, (iii) redução dos danos ao equipamento e (iv) eliminação de danos à carga. É importante salientar que o foco principal neste trabalho é gerar regras de condução que auxiliem na rápida decisão de qual ponto de aceleração utilizar, sem perder de vista os demais princípios de boa condução.

4.1 Modelagem dos Dados

Para a compreensão da solução proposta foi desenvolvido o modelo de dados de domínio parcial (cf. Figura 16), cujos detalhes serão vistos nas próximas seções. Este modelo representa uma viagem de trem que reúne basicamente um maquinista (operador), um trem e um trecho de via férrea a percorrer. O trem é formado por uma ou mais locomotivas, dependendo do trecho a percorrer, e por zero ou mais vagões. Cada locomotiva possui identificador único. Cada vagão possui um determinado peso. Os equipamentos do trem fornecem as seguintes informações: velocidade, pressão dos freios, potência empregada por ponto de aceleração utilizado, deslocamento realizado durante a viagem, entre outros.

A viagem é realizada por um maquinista ser humano. Esta viagem transcorre em um trecho de via férrea, representada por conjunto ordenado de pontos de medida, com a indicação do quilômetro de origem e quilômetro de destino. Cada ponto de medida é uma *tupla* formada pelo identificador do quilômetro, raio da curva e percentual de rampa. Para cada ponto de medida são realizadas leituras periódicas nas saídas dos sensores e os valores recuperados são persistidos. Alguns relês, instalados na locomotiva, fornecem informações referentes ao ponto de aceleração empregado (*PAAtual*), bem como informações de freio (*pressaoDoFreio*), entre outras. Cada ponto de aceleração tem associado um consumo e uma potência dependendo da locomotiva.

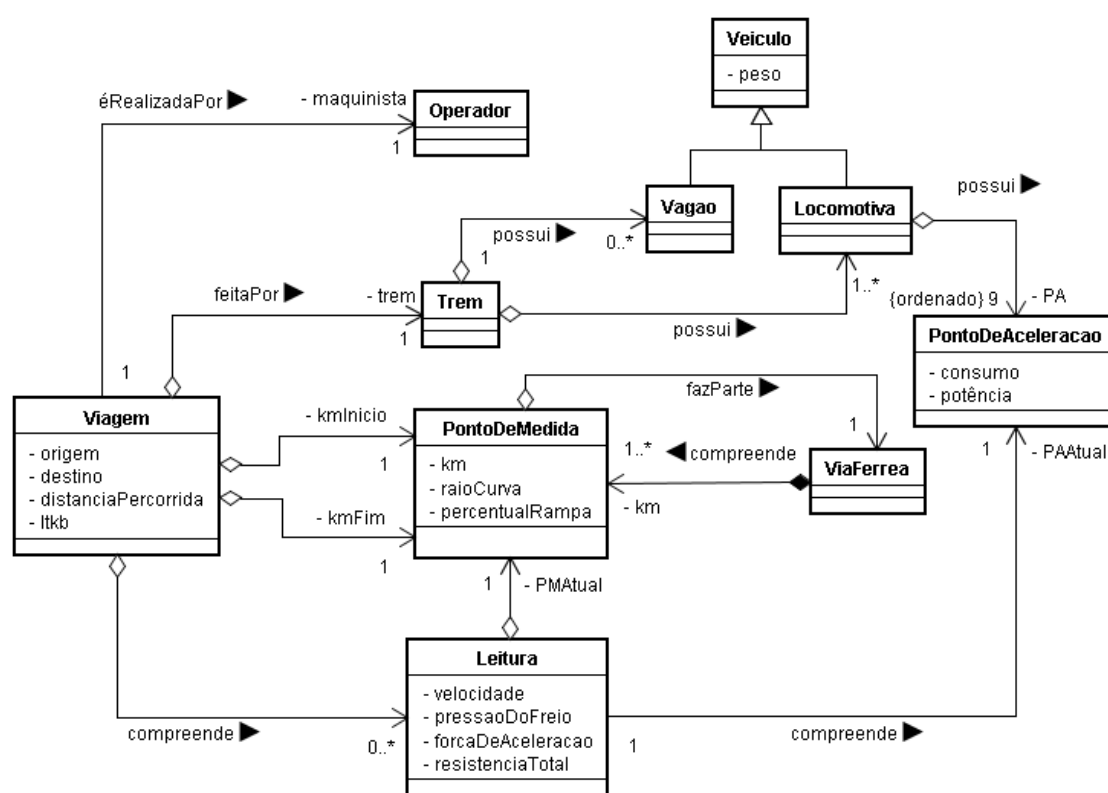


Figura 16 - Modelo conceitual parcial para representar uma viagem de trem em uma via férrea.

4.2 Origem e formato dos dados

Os dados utilizados para a geração da base de dados inicial foram originados de quatro fontes distintas. A Figura 17 ilustra estas fontes de dados.

A FONTE I contém os dados lidos dos sensores com o trem em movimento, como por exemplo: quilômetro (*kilometer*), velocidade (*speed*), velocidade máxima

(*maximum_speed*), pressão de freios (*pressure*), hora (*hour*), minuto (*minute*), segundo (*second*), e dados de relês que informam o ponto de aceleração das locomotivas (*av/bv/cv/dv*), posição da manopla reversora de movimento do trem para frente ou para trás (*ffor/rer*), indicação de intervenção do maquinista (*ths1_8*), freio dinâmico ligado ou desligado (*br*), seqüencial de identificação de evento (*event_sequence*), indicação de patinagem (*pat*) e relê terra (*gr*). A FONTE II contém os dados referentes as data de início e fim da viagem, bem como as identificações das localidades de início de término da viagem. Por exemplo, “LMG-” e “LAP-” identificam as sub-estações onde a viagem iniciou e terminou. Cada sub-estação tem um tipo associado (*type*): principal, não principal (ou desviada). A FONTE III contém dados gerais da viagem, tais como: peso do trem em toneladas (*tons*), distância percorrida em metros (*distance*), consumo final em toneladas transportadas por quilômetro (*tkb*), identificação do trem (*train_code*), identificação do operador (*operator_code*). A FONTE IV contém apenas as identificações das locomotivas da viagem. Desta fonte são obtidos os seguintes dados: número de locomotivas usadas e o tipo de cada uma delas.

FONTE I: LOG da viagem: SEQ,KILOMETER,SPEED,MAXIMUM_SPEED,PRESSURE,HOUR,MINUTE,SECOND,THS1_8,AV,BV,CV,DV,BR,EVENT_SEQUENCE,FFOR,RER,GR,PAT,LATITUDE,LONGITUDE,DI 1,339495,9,40,85,9,38,8,0,0,0,0,0,1,1,0,0,0,843906,1871598,0 2,339491,9,40,85,9,38,13,0,0,0,0,0,0,2,1,0,0,0,843901,1871598,0 3,339481,9,40,85,9,38,19,0,0,0,0,0,0,3,1,0,0,0,843897,1871597,0 4,339477,9,40,85,9,38,23,0,0,0,0,0,0,4,1,0,0,0,843893,1871597,0 5,339470,9,40,85,9,38,28,1,0,0,0,0,0,5,1,0,0,0,843888,1871597,0	
FONTE II: data e local da viagem: SEQ,DATA,HORARIO,LOCAL,TYPE 1,18/10/2006,09:38:05,LMG- ,1 2,18/10/2006,14:26:00,LAP- ,2	FONTE IV: identificação e número de locomotivas SEQ, LOCOMOTIVE_CODE 1,7678 2,7683 3,7686 4,7687
FONTE III: consumo da viagem: SEQ,TONS,DISTANCE,TKB,TRAIN_CODE,OPERATOR_CODE 1,6278,64.206000,6.190000,M58,910009970	

Figura 17 - Exemplo do arquivo de viagem.

O conjunto de dados utilizados durante foram de oito viagens realizadas no trecho entre Londrina (PR) e Paiçandu (PR). Cada viagem foi realizada por um trem com configurações diferentes, em particular, em termos de peso, número de vagões e locomotivas. O número total de registros da base de dados, disponibilizada para os experimentos, foi de aproximadamente 17164, onde aproximadamente 14258 instâncias foram utilizadas nos experimentos. O número de registros obtidos, em cada viagem, ficou entre 1600 a 2800 exemplares para um trecho de 69500m, em média, um registro para cada 33m. É importante salientar que o número de registros é diretamente proporcional ao tempo levado para percorrer o percurso da viagem (cf. Tabela 13).

Tabela 13. Resumo de dados de diferentes viagens.

ID da Viagem	Quilômetro		Número de Registros (NR)		Metros Percorridos (NP)	NP ÷ NR	Tempo de Viagem (min)
	Inicial	Final	Original	Filtrado			
1	339495	268014	2154	1999	71481	33	184
2	335693	268837	2350	1695	66856	28	134
3	336858	269919	1763	1694	66939	38	144
5	340627	268443	1817	1666	72184	39	150
6	339779	269160	2061	1962	70619	34	172
7	335874	268370	1699	1566	67504	39	176
8	341434	268108	2514	1541	73326	29	208
9	339980	268289	2806	2135	71691	25	197

A próxima fase dos estudos concerne à remoção de ruídos e também o enriquecimento dos dados. A atenção portará em particular ao processo de enriquecimento que agregou muitas novas características ao conjunto de dados original.

4.2.1 Remoção de Ruídos

Como dito anteriormente, os dados das viagens são obtidos por meio de sensores instalados nas locomotivas. Estes sensores, às vezes, fornecem valores inconsistentes, por exemplo, devido à perda de comunicação entre satélite e GPS, responsável por gerar os dados de quilômetro e velocidade. Esta perda de comunicação ocorre principalmente em trechos de túneis, onde o sinal não pode ser enviado devido ao fato do trem não estar sendo “visto” pelo satélite.

A remoção de dados ruidosos visa obter um conjunto de dados consistente, de modo a facilitar a extração de padrões válidos no processo de aprendizagem. É possível citar como exemplo de valor claramente inconsistente o deslocamento calculado entre a posição A e a posição B, quando ele gera um valor muito elevado (e.g., centenas de quilômetros). Outra informação ruidosa presente no conjunto de dados refere-se à velocidade, onde certos trechos da viagem apresentavam velocidade de 255 km/h. A remoção desta informação foi realizada para tornar os cálculos de tempo do deslocamento, consumo e velocidade média consistentes para todo o conjunto de dados. A remoção de tais ruídos tornou o conjunto de dados consistente e resultou na melhoria da qualidade das regras de condução geradas. Antes da remoção, as regras geradas utilizavam os dados ruidosos, principalmente os da quilometragem, durante a classificação de uma nova instância. Apesar de presente nas regras, esta informação não resultava numa boa classificação da instância, motivo pelo qual foi removida.

A remoção de dados ruidosos teve uma contribuição significativa para o enriquecimento da base de dados.

4.2.2 Inclusão de Novos Atributos

Classicamente, a etapa de pré-processamento consiste em enriquecer a base de dados original. Em nossos experimentos, parte de tal enriquecimento foi feito por meio da adição de novos atributos; resistência total do trem calculada em cada ponto de leitura. Para realizar os cálculos certas informações foram adicionadas àquelas geradas durante a viagem (passadas como parâmetros e identificadas pelo prefixo “P_”). Segue alguns exemplos:

- P_CIDADE_ORIGEM, P_CIDADE_DESTINO: cidades de origem e destino da viagem. Estes valores são derivados a partir dos valores dos seguintes atributos: LOCAL_INICIO e LOCAL_FIM;
- P_PESO_VAGAO: peso médio dos vagões (em toneladas);
- P_PESO_LOCO: peso médio das locomotivas (em toneladas);
- P_NRO_VAGUES: número de vagões do trem;
- P_AREA_FRONTAL_VAGAO: área frontal do vagão (em *square-feet*);
- P_AREA_FRONTAL_LOCO: área frontal da locomotiva (em *square-feet*);
- P_COMPRIMENTO_VAGAO: comprimento do vagão (em metros);
- P_COMPRIMENTO_LOCO: comprimento da locomotiva (em metros);
- P_NRO_EIXOS_VAGAO: número de eixos do vagão;
- P_NRO_EIXOS_LOCO: número de eixos da locomotiva;
- P_BITOLA: tamanho da bitola (em metros);
- P_COEF_ADER: coeficiente de aderência;

Os enriquecimentos realizados por meio de atributos calculados são prefixados por ‘C_’. Todos os cálculos tiveram como base as fórmulas descritas em (BRINA, 1982) e (AVALLONE, et al., 1996) e também os dados de FONTE I-IV. A Tabela 14 apresenta os campos adicionados à base de experimentos por meio destes calculados.

Tabela 14. Atributos derivados por meio de cálculos

Campo/Atributo (Unidade)	Formula
C_VELOCIDADE_INICIAL (km/h)	VELOCIDADE do registro atual
C_VELOCIDADE_FINAL (km/h)	VELOCIDADE do próximo registro
	TKB
C_LTKB_PARCIAL (LTKB)	número de registros contidos no arquivo da viagem
C_NRO_LOCO	$\sum_{i=1}^5 \text{Locomotiva}_i$
C_PPE_LOCO	$\frac{P_PESO_LOCO}{P_NRO_EIXOS_LOCO} \times 1.1$
C_PPE_VAGAO	$\frac{P_PESO_VAGAO}{P_NRO_EIXOS_VAGAO} \times 1.1$
C_PONTO_ACEL	Se $DI=1$ então -1 Se $AV=0 \wedge BV=0 \wedge CV=0 \wedge DV=0$ então 1 Se $AV=1 \wedge BV=0 \wedge CV=0 \wedge DV=0$ então 2 Se $AV=0 \wedge BV=0 \wedge CV=1 \wedge DV=0$ então 3 Se $AV=1 \wedge BV=0 \wedge CV=1 \wedge DV=0$ então 4 Se $AV=0 \wedge BV=1 \wedge CV=1 \wedge DV=1$ então 5 Se $AV=1 \wedge BV=1 \wedge CV=1 \wedge DV=1$ então 6 Se $AV=0 \wedge BV=1 \wedge CV=1 \wedge DV=0$ então 7 Se $AV=1 \wedge BV=1 \wedge CV=1 \wedge DV=0$ então 8
C_POTENCIA (HP)	A potência do ponto de aceleração da locomotiva, i- dentificado pelo valor de C_PONTO_ACEL.
C_CONSUMONOPONTO (LTKB)	Consumo da locomotiva para o ponto de aceleração, identificado pelo valor de C_PONTO_ACEL.
C_LIMITE_ADERENCIA	(idem equação 9)
C_ESF_TRATOR (Kgf)	$\frac{273,24 \times 0.82 \times C_POTENCIA}{VELOCIDADE}$
C_VEL_MILHAS (Milhas/hora)	VELOCIDADE x 0.622
C_RES_NOR_LOCO (Kgf)	(idem equação 2)
C_RES_NOR_VAGAO (Kgf)	(idem Equação 3)
C_R' LOCO (Kgf)	$C_RES_NOR_LOCO \times 0,5$
C_R' VAGAO (Kgf)	$C_RES_NOR_VAGAO \times 0,5$
C_CURVA (metros)	valor do raio da curva, identificado pelo valor de C_QUILÔMETRO.
C_RES_CURVA_LOCO (Kgf)	(idem equação 4)
C_RES_CURVA_VAGAO (Kgf)	(idem equação 5)
C_RES_RAMPA_LOCO (Kgf)	(idem equação 7)
C_RES_RAMPA_VAGAO (Kgf)	(idem equação 7)
C_RES_INERCIAL_LOCO e C_RES_INERCIAL_VAGAO (Kgf)	(idem equação 6)
C_RES_TOTAL (Kgf)	(idem equação 1)
C_DESLOCAMENTO (metros)	$Quilometro_{atual} - Quilometro_{atual-1}$
C_TEMPO_GASTO (segundos)	$\frac{7.2 \times l}{V_F - V_I}$
C_CONSUMO	(idem equação 17)

C_DESLO_ACUM	$\sum_{n=1}^n \text{deslocamento}$
	sendo n o número de registros
C_CONS_ACUM	$\sum_{n=1}^n \text{Consumo}$
	sendo n o número de registros
C_LTKB	$\left(\frac{\text{ConsumoAcumulado}}{\text{PesoTotal} \times \frac{\text{DeslocamentoAcumulado}}{100}} \right) \times 1000$
C_TAMANHO_COMPOSICAO	$C_NRO_LOCO * P_COMPRIMENTO_LOCO + P_NRO_VAGOES * P_COMPRIMENTO_VAGAO$
C_QUILÔMETRO	Campo que resulta da associação do quilômetro atual com um ponto de medida da via que possua valor de quilômetro semelhante, com variação de aproximadamente 20 metros.
C_DESCONSIDERADO	Campo que indica se determinado registro foi ou não desconsiderado. Para que um registro seja desconsiderado o deslocamento deve ser maior que 9999, fazendo com que “ruídos” no banco de dados sejam desconsiderados. O valor 9999 foi empiricamente considerado, uma vez que é praticamente impossível que uma locomotiva se desloque este valor em metros em 3 segundos (tempo de intervalo da leitura). Caso o quilômetro atual seja igual a 0 ou 16777215 ou a velocidade 255 o registro também é desconsiderado, visto que tais valores indicam perda de sinal do GPS.
C_DIRECAO_INSCRICAO	Valor que indica a direção a qual a locomotiva líder está indo, podendo assumir os valores “direita” ou “esquerda”
C_ACAO_MAQUINISTA	Valor que indica a ação tomada pelo maquinista em relação a mudanças no ponto de aceleração em determinado ponto de medida. Valores possíveis são: “aumentou”, “diminuiu”, “manteve”
C_ACAO_LOCOMOTIVA	Valor que indica se a locomotiva está “arrancando”, “parando” ou simplesmente “andando”

Outra fonte de dados importante para a montagem da base de experimentos foi os dados sobre as características físicas da via férrea. A Figura 18 mostra um exemplo destes dados. Estes dados de perfil referem-se às plantas reais do projeto de uma via férrea, que inclui: raio de curva, corda de grau de 20m, percentual de rampa, entre outros.

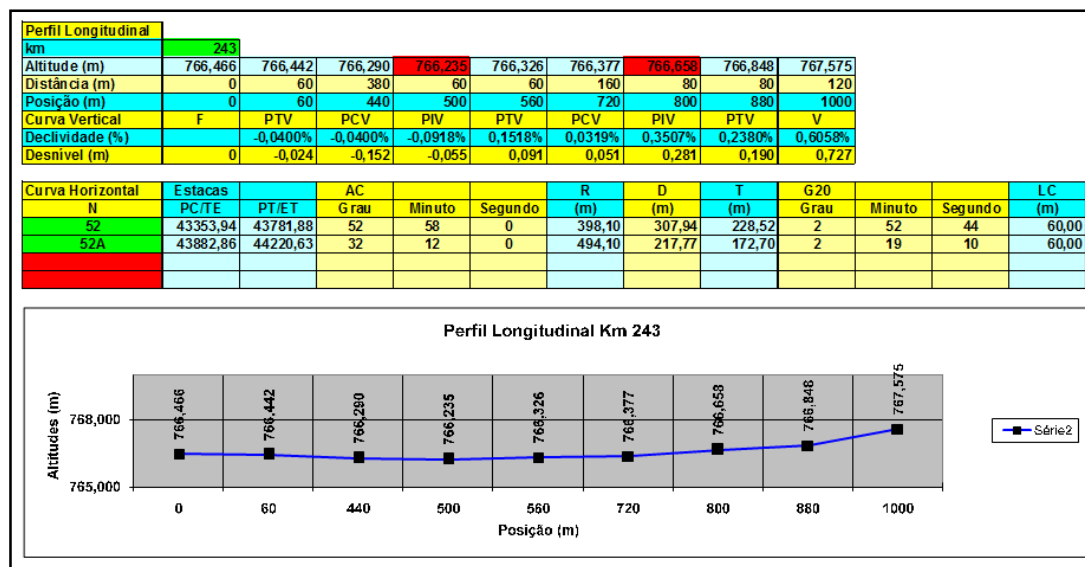


Figura 18. Dados parciais do perfil de uma via férrea.

O elevado número de dados de descrição física de cada trecho de via férrea (ex: Figura 18) e a alta dimensionalidade da base de dados – fez surgir à necessidade do desenvolvimento de um software que realizasse todos os cálculos anteriormente mencionados e inserisse automaticamente os seus resultados na base de dados. Este módulo de software foi um importante componente para expandir a aplicação da metodologia a outros trechos de vias férreas de forma facilitada. Outros dados importantes na tomada de decisão do maquinista durante a condução de trem dizem respeito ao reconhecimento e antecipação de cristas e depressões. Tais situações estão diretamente relacionadas ao perfil da via, comprimento e peso do trem, devido ao esforço trator existente nos engates ou choques no trem.

A Figura 19 (a) mostra uma situação típica de crista no ponto *C*, caracterizado quando a locomotiva líder ultrapasse o ponto *C* e em função do comprimento e do peso do trem obrigue manter o ponto de aceleração alto. Nesta situação, o esforço no engate do vagão que está no ponto de inflexão é a soma do esforço trator das locomotivas mais o esforço do peso do trem que já passou pela rampa. O valor de *h*, neste caso, não influencia na identificação de cristas segundo a opinião do especialista em condução ferroviária. A situação da Figura 19 (b) pode ser ou não caracterizada como uma crista. Caso o comprimento do trem seja maior que o trecho *i* então existe crista, caso contrário, o intervalo entre os pontos *C* e *D* deve ser considerado como um trecho em nível. O trecho em nível é aquele em que as rampas tenham valor entre -1% e 1%, percentuais atribuídos empiricamente.

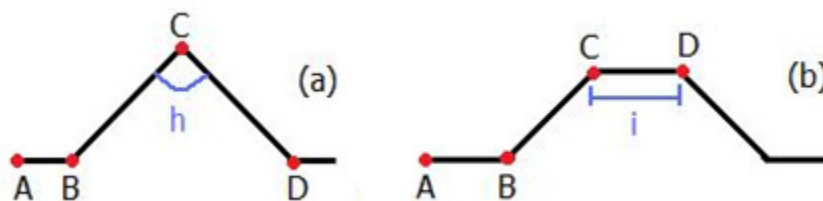


Figura 19 – Situações de crista

A Figura 20 ilustra a situação de depressão, que exige do maquinista atenção redobrada para evitar que a cauda do trem exerça uma força sobre a locomotiva líder causando acidente. De forma análoga à crista, o ângulo β do ponto C não influencia na identificação de uma depressão.

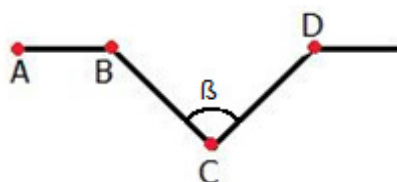


Figura 20 – Situação de depressão

Identificar as situações de crista e depressão não é uma tarefa trivial. Durante a análise, várias idéias surgiram, tais como identificar apenas o início, meio e final do trem. Esta hipótese foi descartada pelo fato de não abordar as diversas situações que podem ocorrer ao longo do trem devido ao seu comprimento. A solução encontrada foi identificar 11 pontos ao longo do trem (cf. Figura 21), considerando que o tamanho médio de um trem varia entre 2 e 3 km, ou seja, tem-se um ponto a cada 200 ou 300 m, sendo esta considerada uma distância aceitável.

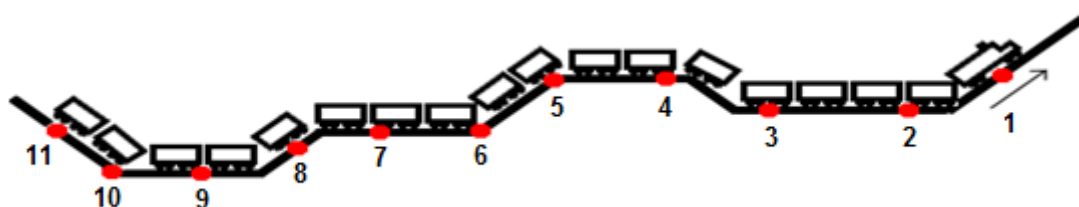


Figura 21 – Pontos candidatos para caracterizar cristas e depressões ao longo do trem

Após identificar os 11 pontos, sendo um primeiro ponto correspondendo sempre ao início do trem, um segundo ponto a cauda do trem e um terceiro ponto o meio do

trem. Cada ponto identificado é então subdividido em outros três pontos. A distância de cada ponto pode ser de 70m dependendo do tamanho do trem; valor atribuído devido ao fato de um vagão ter em média 20m, sendo assim a ponta do próximo/anterior vagão é alcançado com 70m a partir do meio do vagão atual (cf. Figura 22). A escolha de uma faixa maior que 70m mostrou-se mais eficaz ao longo dos testes.

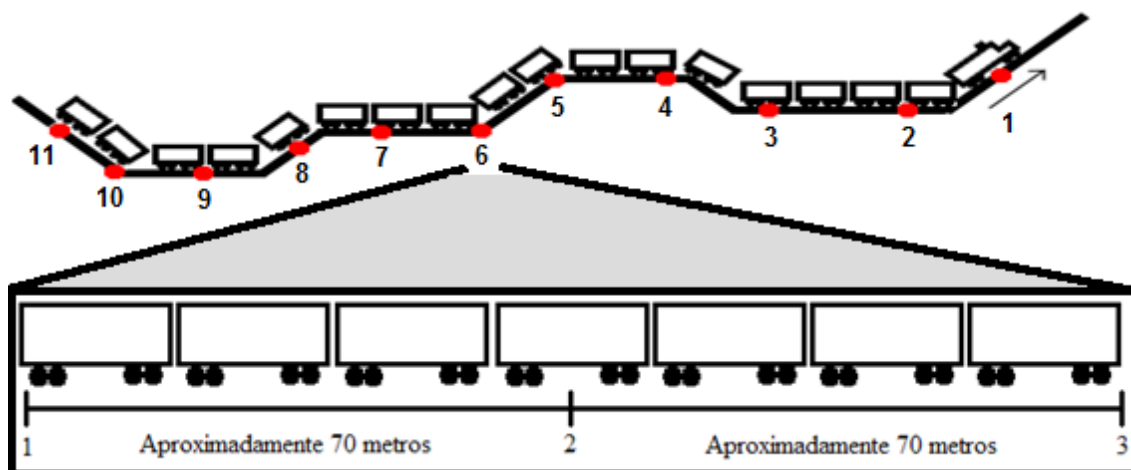


Figura 22. Segmentação de um trem em três pontos para a identificação de cristas e depressões.

Os valores calculados a partir dos 11 pontos, mostrados na Tabela 15, são também incluídos na base de experimentos de forma complementar a Tabela 15.

Tabela 15 - Valores calculados a partir da segmentação de um trem em onze pontos para a identificação de cristas e depressões.

C_KM_PONTO	Fila que armazena o quilômetro de 11 pontos ao longo do trem
C_CABECEIRA	Fila que armazena o status da cabeceira de cada um dos 11 pontos
C_MEIO	Fila que armazena o status do meio de cada um dos 11 pontos
C_CAUDA	Fila que armazena o status da cauda de cada um dos 11 pontos
C_PERCENT_ENT_INCLINACAO_INSCRICAO	Fila que armazena o percentual de rampa da cabeceira de cada um dos 11 pontos
C_PERCENT_SAI_INCLINACAO_INSCRICAO	Fila que armazena o percentual de rampa da cauda de cada um dos 11 pontos
C_PERCENT_CURVA_VENCIDA	Fila que armazena o percentual de curva de cada um dos 11 pontos já percorrida
C_ONDULACAO	Indica se houve ou não uma ondulação. Caso ocorram 3 ou mais situações de crista e/ou depressão ao longo dos 11 pontos o valor é "sim", caso contrário é "não".
C_RAIO_CURVA	Fila que armazena cada raio de curva de cada um dos 11 pontos.

O percentual de rampa em cada um dos três pontos mostrados na Figura 22 é levado em consideração para caracterizar as situações de nível, rampa ascendente (subida), rampa descendente (descida), crista e depressão. Para caracterizar uma destas situações é preciso associar o quilômetro de leitura com o quilômetro mais próximo na via convertida distâncias de 20 metros; a exatidão muitas vezes não é possível.

Feita tal associação, o próximo passo consiste em analisar o percentual de rampa do quilômetro. Se o percentual de rampa for maior-igual que -1 e menor-igual a 1 então ao ponto é atribuída à situação “em nível”. Valores menores que -1 são considerados “descidas” e valores maiores que 1 são consideradas “subidas”. Uma situação de crista ocorre quando há uma situação de subida longa, seguida por uma situação de descida também longa. Já a depressão ocorre quando há uma descida longa, seguida por uma longa subida (ALL, 2008), não havendo um consenso sobre o que é um trecho longo por parte dos especialistas.

Outra situação particular de condução de trens é em trechos ondulados. Um trecho ondulado é um perfil de via onde há mudanças freqüentes nas rampas, podendo haver três ou mais subidas ou descidas (ALL, 2008). Numa situação de crista ou depressão há alternância entre subida e descida. Sendo assim, caso ao longo dos onze pontos de um trem ocorram três ou mais situações de crista/depressão tem-se então a caracterização de uma ondulação.

A partir da associação do quilômetro atual com o quilômetro mais próximo no perfil de via foi possível identificar as situações na qual o trem está em reta ou em curva. Uma situação em reta é aquela em que o ângulo central da curva está entre -1.0 e 1.0. Situação em curva é aquela verificada quando, ângulo central do próximo quilômetro é diferente do ângulo atual; tanto para frente quando para trás do ponto atual. Realizada a identificação, então é efetuado o cálculo ângulo (percentual) da curva que já foi percorrido pelo trem (equação 35). Tal ângulo é o valor absoluto (*abs*) do quilômetro atual menos o quilômetro anterior, multiplicado por 100 e dividido pelo valor absoluto do próximo quilômetro menos o anterior.

$$percentual = \frac{abs(kmAtual - kmAnterior) * 100}{abs(kmPosterior - kmAnterior)} \quad (35)$$

Onde:

- *kmAtual* é o quilômetro atual;
- *kmAnterior* é o quilômetro anterior; e

- *kmPosterior* é o quilômetro após ao atual.

4.2.3 Seleção de Atributos

Segundo Lee (1999) em (LEE, et al., 1999), muitos algoritmos de mineração de dados não funcionam bem com grandes quantidades de dados, deste modo, a seleção de atributos tende a melhorar o desempenho de tais algoritmos. Outro fator que justifica a seleção de atributos é a melhor compreensão do problema com uma menor quantidade de atributos, resultando num conhecimento induzido na forma de classificadores com compreensão mais fácil.

Em nossos experimentos as motivações para a seleção de atributos foram para reduzir o número de atributos e, em consequência, reduzir o tempo de aprendizagem e também melhorar a qualidade dos classificadores gerados. A redução foi feita por meio da aplicação dos métodos *CfsSubSetEval* utilizando o algoritmo *BestFirst* como mecanismo de busca, *GainRatio* utilizando o algoritmo *Ranker* como mecanismo de busca. O primeiro método reduziu em 95% o número de atributos e o segundo método reduziu em 12% o número de atributos.

Segundo Hall em (HALL, 2000) e Koller (KOLLER, et al., 1996) deve-se considerar a redundância dos atributos durante o processo de seleção, visto que atributos redundantes tendem a afetar a qualidade da classificação. O problema da redundância ocorreu neste trabalho após o enriquecimento dos dados com os seguintes atributos: quatro sensores (AV, BV, CV e DV) são responsáveis por fornecer o ponto de aceleração atual de uma locomotiva. Para facilitar a compreensão das regras e também gerar regras com menor quantidade de atributos, os valores dos quatro sensores foram combinados resultando num único atributo que mostra o ponto de aceleração utilizado.

4.2.4 Transformação de Atributos

Relembrando, o nosso principal objetivo é descobrir regras de condução que minimizem o consumo de combustível, sendo o atributo meta o ponto de aceleração. Este atributo é um valor contínuo. Ele é calculado para cada leitura de dados dos sensores. A discretização dos valores é alvo de experimentos neste trabalho, à medida que os algoritmos usados não geram classificadores quando o atributo meta é um valor contínuo. Foi feita então a discretização das variáveis contínuas, avaliando quão bom foi o desempenho dos classificadores na base de dados discretizada dinamicamente pelo próprio

algoritmo de classificação comparado ao desempenho da base de dados discretizada a priori. A discretização a priori permite, por exemplo, que sejam escolhidos automaticamente a quantidade de intervalos (*bins*), o tamanho dos intervalos e a frequência de itens em cada intervalo.

4.3 Aplicação dos Algoritmos de Mineração

A atividade de descoberta de conhecimento teve início por meio da seleção dos dados de viagens. Os dados selecionados foram transformados e sobre o resultado foi feita a filtragem dos dados mais importantes. Uma vez filtrados os dados a mineração foi iniciada por meio da aplicação do C4.5, e na seqüência também foram aplicados métodos de combinação de classificadores BAGGING e BOOSTING.

O passo seguinte, após o enriquecimento e filtragem, foi a aplicação e a análise de um processo de amostragem. Este processo foi realizado por meio do algoritmo *Resample*³. Este último realiza uma amostragem levando em consideração a classe a qual a amostra está associada, ou seja, mantendo-se a proporção de distribuição da base original. A implementação do algoritmo permite informar o tamanho da amostra que se deseja obter em relação ao tamanho do conjunto de dados original; esta informação é passada por meio do parâmetro *sampleSizePercent*. Os tamanhos das amostras geradas variaram de 10 a 30%; tais tamanhos de amostras foram analisados de acordo com a taxa de acerto dos classificadores testados. A utilização de amostras com tamanhos maiores não foi possível devido a falta de recursos computacionais para gerar os classificadores.

Em termos práticos, a amostragem foi usada para gerar dois conjuntos de dados: um para treinamento e um para testes. O primeiro tem por objetivo ser utilizado pelos algoritmos de classificação durante a etapa de geração das árvores de decisão, ou seja, durante a construção do classificador. O segundo é utilizado para testar quão eficiente é o classificador gerado, produzindo um percentual de acerto (TAN, et al., 2006). O tamanho de ambos os conjuntos é de 70% do número total de amostras para o conjunto de treinamento e 30% destinado ao conjunto de testes.

Após a definição dos conjuntos de dados e das configurações dos experimentos, os algoritmos de mineração foram aplicados. Para tal, foram utilizados dois atributos alvo diferentes:

³ Este algoritmo é uma implementação existente no pacote de software WEKA.

- **Ponto de Aceleração:** para que a locomotiva se desloque de um ponto a outro, uma força de aceleração deve ser realizada para vencer um conjunto de resistências. As variáveis presentes nestes deslocamentos possuem valores que mudam de acordo com a situação que a locomotiva se encontra. Cada situação exige que certo ponto de aceleração seja utilizado. Assim, os classificadores que utilizam tal atributo como classe-alvo são capazes de indicar ao condutor qual ponto empregar para deslocar o trem.
- **Consumo no Quilômetro:** para todos os deslocamentos realizados durante as viagens reais foram calculados os valores de resistência total e potência, na etapa de enriquecimento dos dados, com o objetivo de classificar o consumo de determinada ação tomada pelo maquinista. Deste modo, é possível verificar se o consumo resultante da aplicação de certo ponto de aceleração foi similar ao obtido numa situação semelhante de uma viagem real.

A utilização dos classificadores somente é possível com um software capaz de calcular todas as variáveis envolvidas na viagem, simulando assim uma viagem real. O software desenvolvido ao longo deste trabalho, com este propósito juntamente com a forma que os classificadores são utilizados, são apresentados no próximo capítulo.

Foram utilizadas diferentes formas de avaliar a qualidade dos classificadores, visto que as taxas de acerto geradas pelo C4.5, BAGGING e BOOSTING não condizem com a real taxa de aplicabilidade das classes sugeridas. Assim, optou-se por utilizar as técnicas de validação cruzada, distância do cosseno e Teste de Friedman para avaliar a qualidade dos classificadores.

A validação cruzada também foi usada na geração e avaliação dos classificadores. Ela tem por objetivo gerar amostras utilizadas durante as etapas de treinamento e teste do classificador, obtendo assim uma taxa de acerto para cada uma das amostras geradas e ao final fazendo a média aritmética das taxas obtidas, resultando assim na taxa de acerto do classificador. A geração dos classificadores foi realizada com cinco e dez amostras (*folds*). Segundo (WITTEN, et al., 2005), dez é o número de *folds* que obtém melhores taxas de acerto.

Para medir a similaridade entre as ações tomadas pelo maquinista e as ações tomadas pelo simulador de condução, foi usada a fórmula do cosseno. Ela traduz de certa forma a real capacidade do classificador indicar ações aplicáveis que foram abstraídas a

partir do processo de mineração de dados. Em outras palavras, o cosseno traduz o grau de “imitação” na forma de condução entre o simulador de condução e o maquinista.

4.4 Considerações Finais

A metodologia apresentada segue, em linhas gerais, as etapas clássicas do processo de descobrimento de conhecimento. A principal contribuição ao domínio de conhecimento estudado concerne à etapa de enriquecimento dos dados. Deve-se também destacar a inclusão de uma métrica para quantificar a similaridade entre as tomadas de ações reais realizadas por um maquinista e as tomadas de ações realizadas pelo simulador de condução. Esta métrica mede o grau de “imitação” entre as tomadas de ações de um ser humano maquinista e de um sistema computacional; deve-se salientar que tal sistema computacional coloca em prática as regras abstraídas a partir de uma base de dados histórica de viagens de trens conduzidas por seres humanos maquinistas.

Capítulo 5

Simulador

A metodologia delineou um estudo clássico de descoberta de conhecimento, onde o principal esforço foi o enriquecimento dos dados; isto é, na fase de pré-processamento. Entretanto, o grande interesse da pesquisa concerne a descoberta e aplicação de padrões de tomadas de ações no tocante a condução de um trem. Nestes termos foram realizados vários estudos e experimentos computacionais na direção de construção de um sistema computacional de condução automática de trens; o corpo central deste sistema é descrito neste capítulo.

O fluxo básico do simulador concerne: (i) à percepção de evento, (ii) a classificação deste evento para a obtenção de um ponto de aceleração, (iii) o cálculo do consumo com base em tal ponto de aceleração, (iv) a avaliação da ação a ser empregada, (v) reconsideração de tal ação ou aplicação da mesma, (vi) retorna para (i).

Este capítulo foi organizado em várias seções. A seção 5.1 descreve alguns recursos computacionais utilizados. A seção 5.2 ilustra um exemplo dos cálculos realizados pelo sistema. A seção 5.3 apresenta a forma como os classificadores são utilizados. A seção 5.4 apresenta como as camadas de refinamento das tomadas de ações são empregadas. A seção 5.5 descreve em detalhes o algoritmo responsável pela execução de uma viagem.

5.1 Descrição Geral do Simulador

O simulado é composto por três módulos principais. O primeiro módulo é responsável por interpretar as regras geradas pelos classificadores. O segundo módulo é responsável pelo fluxo principal que gera e avalia as alternativas de políticas de tomadas de ações, doravante chamado de *Sistema de Decisão*. O terceiro é um módulo para cálculos especializados nas equações de Davis, doravante chamado de *Calculador*.

O módulo interpretador de regras de condução recebe como entrada um arquivo texto contendo tais regras (cf. Figura 23). O núcleo deste módulo permite instalar diferentes conjuntos ordenados de regras em uma mesma camada, bem como diferentes

conjuntos ordenados de regras em diferentes camadas. Por exemplo, um evento e é submetido aos classificadores da primeira camada. O evento e é enriquecido tornando-se e' . Este último é submetido aos classificadores da segunda camada. A Figura 24 mostra o modelo conceitual deste interpretador.

```
Options: -S 10

JRIP rules:
=====

(PRESSURE <= 88) and (C_PERCENTUAL_RAIO_CURVA_4 <= 402.850006) =>
C_PONTO_ACEL=1.0 (23.0/3.0)

(C_ACAO_MAQUINISTA_PONTO = diminuiu) and (C_PERCENTUAL_CURVA_VENCIDA_7 <=
4.087943) and (C_MEIO_3 = subindo) => C_PONTO_ACEL=4.0 (13.0/3.0)

(C_LIMITE_ADERENCIA >= 135593.2188) and (C_ACAO_MAQUINISTA_PONTO = aumentou)
and (C_PERCENTUAL_RAIO_CURVA_3 >= 601) and (MAXIMUM_SPEED >= 53) and
(C_CABECEIRA_10 = subindo) => C_PONTO_ACEL=4.0 (16.0/2.0)

(C_LIMITE_ADERENCIA >= 132231.4063) and (C_ACAO_MAQUINISTA_PONTO = aumentou)
and (C_PERCENTUAL_CURVA_VENCIDA_5 <= 6.987162) and (SPEED >= 20) =>
C_PONTO_ACEL=3.0 (9.0/0.0)

=> C_PONTO_ACEL=0 (2417.0/710.0)
```

Figura 23. Exemplo de regras no formato Weka.

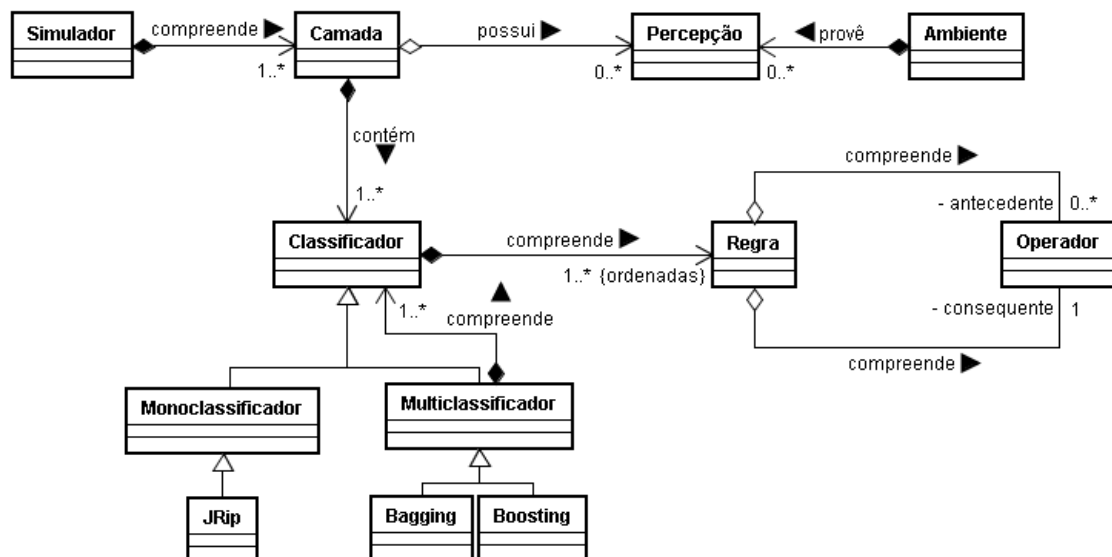


Figura 24. Modelo conceitual do interpretador de regras.

O modelo exibido, na Figura 24, define um conjunto de elementos. De forma prática, podemos lê-lo da seguinte maneira: um Simulador pode ter diferentes camadas de decisão. Cada camada pode ter diferentes classificadores. Cada classificador pode ter diferentes regras. Cada regra tem um conseqüente e um antecedente. O antecedente de uma regra pode ser vazio para o caso da regra *default*. Cada cláusula do antecedente de uma regra é formada por um operador relacional, envolvendo sempre uma variável e uma constante. O conseqüente de uma regra é uma constante. Cada camada pode ter também uma base de exemplos. Cada base de exemplo tem várias colunas (ou atributos). Cada coluna tem uma lista de valores. Cada linha da base de dados é um exemplo. Esta base de exemplos pode ser usada para gerar a matriz de confusão. Um Simulador tem também uma base de percepções. Cada percepção é um evento a ser classificado.

O módulo *Sistema de Decisão* lê um conjunto de informações referentes ao perfil via (rampas e curvas), ao trem (peso, quantidade de locomotivas e vagões, entre outros). Ele usa estas informações nas trocas com os demais módulos; ele controla o fluxo principal de dados e ações do simulador. Este módulo gerencia um LOG das regras aplicadas pelo simulador. Cada registro deste LOG mapeia a regra aplicada e os dados da viagem daquele instante. Estes registros são úteis para analisar quais regras foram aplicadas com mais freqüência e qual classificador foi mais eficiente em certos trechos da viagem. Este módulo é detalhado nas próximas seções.

O módulo *Calculador* é peça fundamental para o enriquecimento de uma percepção e validação de uma tomada de decisão. Este módulo implementa, dentre outras funcionalidades, um conjunto de procedimentos de cálculos baseados nas equações de Davis. Dado a importância das equações de Davis, a próxima seção mostra um exemplo parcial dos cálculos realizados com dados reais de viagem.

5.2 Exemplos de Cálculos

As fórmulas das equações de Davis usadas nesta seção foram apresentadas na seção de fundamentação. O cenário de ilustração possui:

- Um trem com 3 locomotivas de 160 toneladas cada e 58 vagões de 100 toneladas cada, a área frontal de 120 pés, o peso total do trem é de 6280 toneladas: $(160 * 3) + (100 * 58)$.
- Um trecho de via férrea entre Londrina (PR) e Paiçandu (PR), começando do quilômetro 339.495.

- A bitola da via ferra em questão é de 1.6 metros, o coeficiente de aderência de 0.22;

Para o cenário, deseja-se testar se ponto de aceleração quatro é suficiente para movimentar o trem. Os cálculos são para duas iterações, sendo a primeira para fazer o trem passar de 0 km/h para 2 km/h e de 2 km/h para 4 km/h.

- a) Primeira iteração: mudar a velocidade do trem de 0 km/h para 2 km/h, usando o ponto de aceleração quatro, cujo qual desempenha 960 HP e consome 1.95 litros/min.

Percentual de rampa: 0.956

Raio de curva: 3089.0 metros

Velocidade média: 1 km/h

Velocidade média em milhas/hora: $1 * 0,622 = 0,622$ milhas/hora

Peso por eixo de cada locomotiva: $\frac{160}{4} \times 1.1 = 44.0$ toneladas/eixo

Peso por eixo do vagão: $\frac{100}{4} \times 1.1 = 27.5$ toneladas/eixo

$$F_t = \frac{273.24 \times 0.82 \times 960}{2} = 107547.3 \text{ kgf}$$

$$R_{nl} = 1.3 + \frac{29}{44} + 0.03 \times 0.622 + \frac{0.0024 \times 120 \times 0.622^2}{44 \times 4} = 1.978384 \text{ libras/ton}$$

$$R_{nl} = 1.978384 \times 0.5 = 0.989192 \text{ kg/ton}$$

$$R_{nv} = 1.3 + \frac{29}{27.5} + 0.045 \times 0.622 + \frac{0.0024 \times 120 \times 0.622^2}{27.5 \times 4} = 2.383548 \text{ libras/ton}$$

$$R_{lv} = 2.383548 \times 0.5 = 1.191774 \text{ kg/ton}$$

$$R_{cl} = 0.2 + \frac{100}{3089} (2.4 + 1.6 + 3.8) = 0.452509$$

$$R_{cv} = \frac{500 \times 1.6}{3089} = 0.258983$$

$$R_i = 10 * 0.956 = 9.56$$

$$R_{\text{Total}} = 3 \times 0.160 \times (0.989192 + 0.452509 + 0 + 9.56) + 58 \times 0.100 \times (1.191774 + 0.258983 + 0 + 9.56)$$

$$R_{\text{Total}} = 65013.29 \text{ kgf}$$

$$F_{ac} = 107547.36 - 65013.29 = 42533.97 \text{ kgf}$$

$$\Delta l = 4 \times \frac{6280 \times (2^2 - 0^2)}{42533.97} = 2.362347 \text{ metros}$$

$$\Delta t = 7,2 \times \frac{2.362347}{2 + 0} = 8.50 \text{ segundos}$$

$$\gamma = \frac{42533.97}{\frac{628000}{9.80665}} = 0.664197$$

$$\text{Consumo} = \frac{8.50}{60} \times 1.95 = 0.276395 \text{ litros/min}$$

- b) Segunda iteração: mudar a velocidade do trem de 2 km/h para 4 km/h, usando o ponto de aceleração quatro cujo qual desempenha 960 HP e consome 1.95 litros/min.

Percentual de rampa: 0.956

Raio de curva: 3089.0 metros

Velocidade média: 1 km/h

Velocidade média em milhas/hora: $2 * 0,622 = 1.866$ milhas/hora

Peso por eixo de cada locomotiva: $\frac{160}{4} \times 1.1 = 44.0$ toneladas/eixo

Peso por eixo do vagão: $\frac{100}{4} \times 1.1 = 27.5$ toneladas/eixo

$$F_t = \frac{273.24 \times 0.82 \times 960}{4} = 53773.63$$

$$R_{nl} = 1,3 + \frac{29}{44} + 0,03 \times 1.866 + \frac{0,0024 \times 120 \times 1.866^2}{44 \times 4} = 2.020769 \text{ libras/ton}$$

$$R_{ll} = 2.020769 \times 0,5 = 1.010384 \text{ kg/ton}$$

$$R_{nv} = 1,3 + \frac{29}{27.5} + 0,045 \times 1.866 + \frac{0,0024 \times 120 \times 1.866^2}{27.5 \times 4} = 2.447632 \text{ libras/ton}$$

$$R_{lv} = 2.447632 \times 0,5 = 1.223816 \text{ kg/ton}$$

$$R_{cl} = 0,2 + \frac{100}{3089} (2.4 + 1.6 + 3,8) = 0.452509$$

$$R_{cv} = \frac{500 \times 1.6}{3089} = 0.258983$$

$$R_i = 10 * 0.956 = 9.56$$

$$R_{\text{Total}} = 3 \times 160 \times (1.010384 + 0.452509 + 0 + 9.56) + \\ 58 \times 100 \times (1.223816 + 0.258983 + 0 + 9.56)$$

$$R_{\text{Total}} = 65209.31 \text{ kg f}$$

$$F_{ac} = 53773.63 - 65209.31 = -11435.7 \text{ kgf}$$

$$\Delta l = 4 \times \frac{6280 \times (2^2 - 0^2)}{-11435.7} = -26.3596 \text{ metros}$$

$$\Delta t = 7,2 \times \frac{-26.3596}{4 + 2} = -31.6315 \text{ segundos}$$

$$\gamma = \frac{-11435.7}{\frac{628000}{9.80665}} = -0.17858$$

$$\text{Consumo} = \frac{-31.6315}{60} \times 1.95 = -1.02803 \text{ litros/min}$$

A aplicação do ponto de aceleração quatro foi suficiente para progredir o trem de 0 km/h para 2 km/h. Na segunda iteração, a utilização do mesmo ponto de aceleração não foi capaz de deslocar o trem, visto que a força de aceleração é negativa; isto resulta em um deslocamento negativo. Uma solução seria, por exemplo, aumentar o ponto de aceleração para cinco e verificar com tal aumento de potência se a força de aceleração seria positiva. Deve-se ressaltar que não foram consideradas as resistências inerciais nos exemplos mostrados e foi calculada apenas a resistência do primeiro vagão.

Em termos práticos, os cálculos são feitos pelo módulo *Calculador*. Entretanto, para o *Calculador* realizar os cálculos de forma completa, ele precisa de dois conjuntos de dados: um estático e um dinâmico. O conjunto estático corresponde ao perfil do trecho da via, sobre o qual o trem se encontra, e características do trem. O conjunto dinâmico corresponde aos valores lidos dos sensores, doravante chamado de *percepção*. Por exemplo, os valores lidos podem ser: quilômetro e velocidade atual do trem.

No simulador, uma percepção enriquecida é submetida a um classificador no intuito de obter o ponto de aceleração candidato.

5.3 Uso dos classificadores

A concepção inicial foi utilizar apenas um classificador e no final avaliar seu desempenho, medindo a quantidade de acertos e erros na escolha do ponto de aceleração. Porém, ao longo de experimentos notou-se que uma regra era aplicada com mais fre-

qüência que as demais, apesar de não possuir valor de suporte alto. Outro critério de escolha do ponto de aceleração foi à escolha do menor ou maior valor sugerido pelos classificadores por meio de votação. Criar este procedimento de votação busca, além de escolher a regra com maior suporte, observar qual classificador possuiu regras que melhor se encaixaram nas situações da viagem.

Uma das abordagens utilizadas na votação foi classificar a percepção em diferentes classificadores e escolher o ponto de aceleração com base no valor de suporte. A Figura 25 ilustra o processo. O *Sistema de Decisão* recebe uma percepção do simulador, e a submete a três diferentes classificadores: um contendo as regras geradas pelo JRIP e outros dois contendo as regras geradas pelos metaclassificadores BAGGING e BOOSTING. A regra utilizada para previsão de uma classe que possui maior suporte é escolhida e enviada ao simulador para que o mesmo possa aplicá-la.

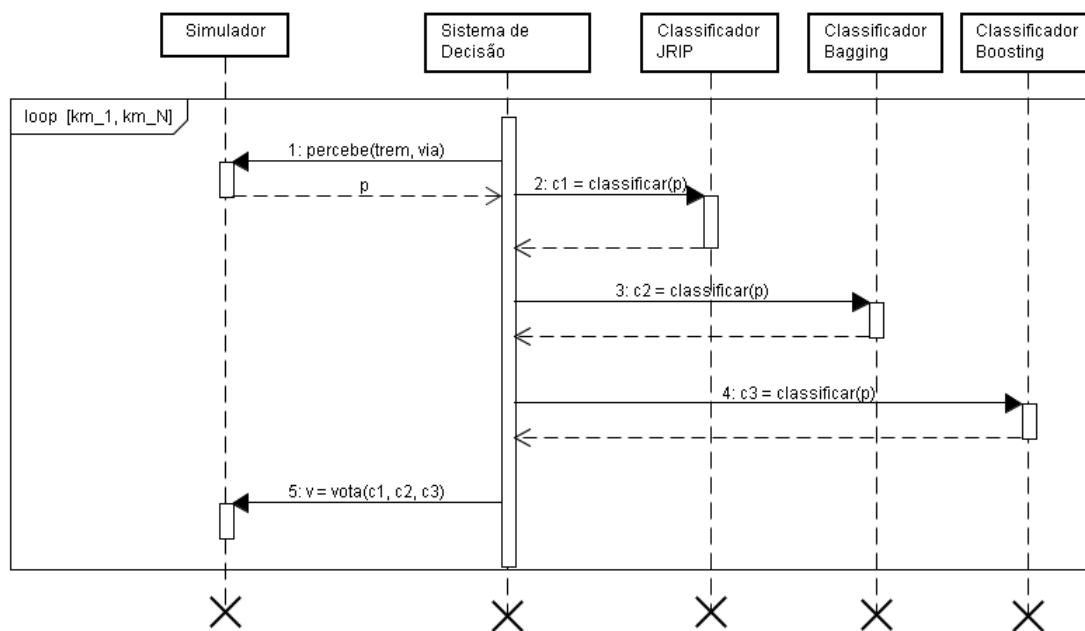


Figura 25. Forma de votação dos classificadores.

Para submeter um exemplo a qualquer um dos classificadores, o exemplo deve estar no mesmo formato dos exemplos da base de treinamento. A utilização de um classificador ou a escolha de um ponto de aceleração entre vários disponíveis pode não garantir uma ação passível de execução, pois pode ocorrer patinação ou falta de potência para percorrer o percurso, necessitando de intervenções na ação por meio de refinamentos sucessivos.

5.4 Procedimentos de Refinamentos

Os procedimentos de refinamentos são necessários para interferir nas ações resultantes do processo de classificação. Relembrando, um classificador pode sugerir um ponto de aceleração (ação) que não se adere ao perfil da via onde está localizado o trem. Além do mais, não é possível garantir que o classificador sugira sempre o ponto de aceleração que resulte em menor consumo, pois não há certeza que os dados obtidos foram de uma viagem ótima. Por exemplo, sugerir o uso de um ponto abaixo do necessário, resultando em falta de potência para percorrer o trajeto e ocasionando a parada do trem. Desta forma, para garantir que o trem progrida foram desenvolvidas três procedimentos, cujas ações básicas são respectivamente:

- Controlar patinação;
- Aumentar o ponto de aceleração em caso de pouca potência; e
- Reduzir o ponto de aceleração em caso de excesso de consumo.

A intervenção destes procedimentos é mostrada mais a frente, em ciclo de execução que consiste em perceber um evento, classificar tal percepção, validar tal classificação e aplicar a ação validada.

5.4.1 Controle de Patinação

A patinação de um trem é uma das situações que deve ser evitada ao longo da viagem. Quando ela ocorre, ela ocasiona sérios danos às rodas das locomotivas e a via (ALL, 2008). A escolha do ponto de aceleração implica na determinação de um esforço trator efetivo ee para cada locomotiva do trem. Cada locomotiva possui também um esforço trator aderente ea , que depende do peso e da velocidade do trem. A soma de todos os esforços tratores efetivos não pode ser maior que a soma dos esforços tratores aderentes multiplicados pelo número de locomotivas nl , caso contrário ocorre patinação (PIRES, et al., 2005).

Quando ocorre patinação, o primeiro procedimento tomado é reduzir o ponto de aceleração pa sugerido, o que gera um esforço trator efetivo menor que a soma dos esforços aderentes. Em algumas situações ocorre do ponto de aceleração ser totalmente reduzido, chegando a zero, resultando na aplicação de um segundo procedimento: reduzir a velocidade desejada. Esta redução pode fazer com que a força de aceleração seja positiva, porém a velocidade desejada seja menor que a velocidade atual, ou seja, dese-

ja-se reduzir a velocidade sem que haja uma aceleração negativa. Neste caso a ação é desconsiderada, fixando o deslocamento previsto dp em 50m e buscando a velocidade final vf de acordo com a Equação 36, onde vi é a velocidade atual, fa a força de aceleração e pe o peso do trem.

$$vf = \sqrt{\left| vi^2 + \frac{fa \times dp}{4 \times pe} \right|} \quad (36)$$

A velocidade desejada passa a ser a velocidade retornada pela Equação 34 mais 2 km/h. Se a força de aceleração não for positiva e a velocidade estimada for maior que a velocidade do trem, o procedimento adotado é reduzir a velocidade desejada em 0.5 km/h.

Para cada mudança de valores ocorre nova invocação do *Calculador*, visando atualizar o vetor de dados a ser submetido ao classificador, o qual deve retornar um novo ponto de aceleração. Esta atualização ocorre por meio da chamada da função *atualiza*, do algoritmo descrito abaixo. ω é uma via férrea e τ é um trem, .

Para cada mudança de valores ocorre nova invocação do *Calculador*, visando atualizar o vetor de dados a ser submetido ao classificador, o qual deve retornar um novo ponto de aceleração. Esta atualização ocorre por meio da chamada da função *atualiza*, do algoritmo descrito abaixo. ω é uma via férrea, τ é um trem, p é a percepção, pa é o ponto de aceleração, Ei o quilômetro atual.

Algoritmo cPatinagem()Entrada: E_I , ω , τ ; var p, pa, vi, vf, ea, ee, dp, fa, nl;

```

1.  enquanto ( ee > ( ea * nl ) )
2.      se ( pa > 0 )
3.          Pa--;
4.          p := atualiza(  $E_I$ ,  $\omega$ ,  $\tau$ , p, pa );
5.      senão
6.          se ( fa > 0 e vf < Vi )
7.              
$$vf = \sqrt{vi^2 + \frac{fa \times dp}{4 \times pe}}$$

8.              se ( vf > 0 )
9.                  vi := vf;
10.                 vf := vf + 2;
11.                 p := atualiza(  $E_I$ ,  $\omega$ ,  $\tau$ , p, pa );
12.             senão
13.                 vi := 0;  vf := 2;
14.                 p := atualiza(  $E_I$ ,  $\omega$ ,  $\tau$ , p, pa );
15.             fim-se
16.         senão
17.             vf := vf - 0.5;
18.             vi := vi - 0.5;
19.             p := atualiza(  $E_I$ ,  $\omega$ ,  $\tau$ , p, pa );
20.         fim-se
21.     fim-se
22. fim-enquanto

```

A falta de potência também requer um procedimento exclusivo. Neste caso para evitar a aplicação de um ponto de aceleração que gere esforço negativo.

5.4.2 Falta de Potência

O procedimento para gerir a falta de potência tem por objetivo aumentar o ponto de aceleração sugerido pelo classificador a fim de evitar com que o trem pare de se movimentar ao longo da viagem. Caso o ponto de aceleração sugerido seja menor que oito (maior ponto de aceleração das locomotivas), então o ponto é incrementado. Esta mudança no ponto de aceleração faz com que ocorra a reavaliação de falta de potência; caso o ponto de aceleração seja máximo, então a velocidade final é reduzida em 0.5 km/h. De forma análoga, o exemplo modificado passa por uma nova verificação de falta de potência. As iterações param quando uma ação válida é encontrada.

Este procedimento também evita que deslocamentos previstos maiores que 500 metros sejam considerados durante a viagem. Durante a fase de experimentos foi possível notar que deslocamentos maiores que 500 metros resultavam num tempo de deslocamento muito alto, fora dos padrões das viagens reais, implicando em alto valor de consumo à medida que o mesmo é calculado com base no deslocamento e no tempo necessário para deslocar o trem. Caso o deslocamento da ação supere o limite, assume-

se o valor de 500 metros. Os valores do exemplo a classificar permanecem os mesmos, mas a posição do trem na próxima ação será 500 metros além da posição da ação atual.

O algoritmo *vFaltaPotencia()* é descrito a seguir. As variáveis de entrada são: ponto de aceleração (*pa*), velocidade inicial (*vi*), velocidade final (*vf*), força de aceleração (*fa*), deslocamento previsto (*dp*), percepção (*p*), quilômetro atual (*Ei*).

Algoritmo *vFaltaPotencia()*

```
Entrada:  $E_I$ ,  $\omega$ ,  $\tau$ ; var p, pa, vi, vf, dp, fa;
1.  enquanto( fa <= 0 ou dp > 500 )
2.      se( pa < 8 )
3.          pa := pa + 1;
4.          p := atualiza( $E_I$ ,  $\omega$ ,  $\tau$ , p, pa );
5.      senão
6.          vf := vi - 0.5;
7.          vi := vf - 2.0;
8.          se( vi < 1.0 ) então pára fim-se
8.          p := atualiza( $E_I$ ,  $\omega$ ,  $\tau$ , p, pa );
10.     fim-se
11. fim-enquanto
```

Caso não haja potência suficiente para percorrer um trecho e a velocidade do trem seja zero, a viagem é interrompida. Isto ocorrerá se o peso do trem estiver acima do peso máximo permitido para percorrer o trecho com a maior rampa da viagem. Em outras palavras, isto resulta numa resistência impossível de ser vencida pelo esforço tração efetivo naquele ponto.

5.4.3 Marcha Excessiva

Ao longo de experimentos foi observado que o ponto de aceleração sugerido pelo classificador em algumas situações era alto. Estas situações ocorreram principalmente quando a regra *default* era aplicada. Normalmente, a regra *default* indica o valor 7 ou 8. Estes pontos são os empregados durante uma viagem. A aplicação de tais pontos de aceleração gera um alto valor para o consumo.

O procedimento proposto busca reduzir o valor de consumo da viagem, examinando se o ponto de aceleração sugerido pelo classificador pode ser reduzido. Para tal, o ponto deve ser maior que um. Para cada verificação o exemplo a ser classificado é modificado por meio da invocação do *Calculador*. Durante a execução destes procedimentos a força de aceleração é analisada, devendo esta ser maior que zero e o deslocamento previsto menor que 500 metros, similar ao que ocorre no procedimento de verificação de falta de potência.

O algoritmo *vMarchaExcessiva* é mostrado a seguir. Ele recebe os seguintes dados de entrada: o ponto de aceleração das locomotivas (pa), a força de aceleração (fa), o deslocamento previsto (dp), uma percepção (p), uma via férrea (\mathfrak{W}), um trem (τ), um quilômetro atual (E_i).

```

Algoritmo vMarchaExcessiva()
Entrada:  $E_i$ ,  $\mathfrak{W}$ ,  $\tau$ , var  $p$ ,  $pa$ ,  $dp$ ,  $fa$ ;
1.  enquanto (  $pa > 1$  E  $pa \leq 8$  )
2.       $pa := pa - 1$ ;
3.       $p := atualiza(E_i, \mathfrak{W}, \tau, p, pa)$ ;
4.      enquanto (  $fa \leq 0$  E  $dp > 500$  )
5.           $pa := pa + 1$ ;
6.           $p := atualiza(, \mathfrak{W}, \tau, p, pa)$ ;
7.      fim-enquanto
8.  fim-enquanto

```

A próxima seção descreve o algoritmo de geração de políticas de ações para uma viagem. Este algoritmo depende dos procedimentos descritos anteriormente.

5.5 Execução de uma Viagem

A execução de uma viagem começa pela interpretação dos boletins de bordo, que informam as restrições da via e eventuais obras no trecho. Por questões práticas, em nossas simulações foram levadas em conta apenas as informações do Boletim A. Este último é passado ao simulador por meio de um arquivo XML que reúne um grande conjunto de características estáticas de um via férrea real (Figura 26).

Cada via férrea é expressa por meio de um conjunto de pontos de medida. Cada ponto de medida representa um trecho de via férrea de 20 metros. Ele é definido por um conjunto de atributos:

- um identificador (id);
- uma velocidade máxima permitida ($velocidadeMax$);
- um quilômetro (km);
- um percentual de rampa ($rampa$);
- um raio de curva ($raioCurva$);
- um ângulo central de curva (ac);
- uma corda de 20 metros ($g20$);
- uma altitude;

- uma latitude; e
- uma longitude.

Estes pontos foram obtidos a partir plantas das vias férreas. É importante salientar que nem todas as informações constantes em cada ponto de medida foram usadas, como por exemplo, latitude e longitude.

```
<pontoDeMedida>
  <!-- Pasta de origem rampa = km204 -->
  <!-- Pasta de origem curva = km204 -->
  <id>1</id>
  <velocidadeMax>60</velocidadeMax>
  <km>204</km>
  <rampa ini="0.0" fim="20.0">-0.98642856</rampa>
  <raioCurva ini="0.0" fim="20.0">0.0</raioCurva>
  <ac ini="0.0" fim="20.0">0.0</ac>
  <g20 ini="0.0" fim="20.0">0.0</g20>
  <altitude ini="0.0" fim="20.0">527.573</altitude>
  <localizacao>
    <latitude>0</latitude>
    <longitude>0</longitude>
  </localizacao>
</pontoDeMedida>
```

Figura 26- Exemplo de ponto de medida no arquivo XML

As características dos trens usadas nas simulações são realistas. Entretanto, a virtualização da condução nos levou a incluir por de meio da fórmula abaixo, o peso aproximado de cada vagão, calculado de acordo com a equação 37.

$$pv = \frac{pt - (nl \times pl)}{nv} \quad (37)$$

Onde:

- nv é o número de vagões;
- nl é o número de locomotivas;
- pt é o peso total do trem (em toneladas); e
- pl é o peso de cada locomotiva (em toneladas).

Os quilômetros de início de fim da viagem também são fornecidos, bem como as configurações específicas da locomotiva (área frontal, número de eixos e consumos em cada ponto de aceleração).

Como já dito, a arquitetura do simulador inclui o seguintes módulos (Figura 27): *Sensor*, *Atuador*, *Calculador* e *Decisão*. O *Sensor* realiza a captura de diferentes medidas, tais como: quilômetro e velocidade atual, quilômetro atual, aderência e resistência. O *Atuador* age por meio da execução de uma ação. Esta última, em geral, altera a velocidade do trem. A atuação limita-se em aumentar, manter ou reduzir a velocidade. O *Calculador* encapsula as fórmulas dos cálculos para as diferentes resistências e esforços de tração.

O fluxo principal de controle é executado pelo módulo de *Decisão*. Este último recebe, na forma de uma percepção p , os valores lidos do *Sensor*. Os valores de p são convertidos para o formato de um vetor de características v . Esse vetor v é submetido ao classificador A . Este último retorna um ponto de aceleração pa . A partir de pa pode-se calcular os valores de: *resistência*, *esforço trator*, *deslocamento previsto* e *força de aceleração*. Estes valores são inseridos no vetor u . Se o valor da *resistência* é menor que o valor do *esforço trator*, então o estado é registrado em arquivo de LOG e o ponto de aceleração pa é aplicado por meio do *Atuador*. Caso contrário, a política de ação é modificada. Tal modificação consiste em aumentar em uma unidade o ponto de aceleração pa . Este incremento ocorre até que o *esforço trator* resultante seja maior que as resistências a serem vencidas. Caso não seja mais possível incrementar o ponto de aceleração, então a estratégia é decrementar em 0.5 a *velocidade estimada* e a *velocidade corrente*. O vetor v é atualizado com os novos valores de ambas as velocidades. O vetor v modificado é re-submetido ao classificador A . Desta forma, o processo se repete até que seja encontrado um valor de pa que movimente o trem de forma satisfatória.

Antes de detalharmos, na forma de um algoritmo, o esquema da Figura 27, é importante definirmos um pouco mais o módulo *Calculador*. Este último toma como entrada uma percepção, realiza inúmeros cálculos e devolve um conjunto de valores. Além da percepção, ele se utiliza de grande número de outros valores relativos ao perfil da via férrea e do trem para realizar tais cálculos, a saber: posição inicial, posição final, posição atual, velocidade inicial, velocidade atual, velocidade desejada, velocidade final, coeficiente de atrito, direção, janela de velocidade máxima projetada, janela de distância máxima de frenagem, deslocamento fixo de frenagem máxima, quantidade de locomotivas, número de pontos de aceleração. Tem-se também um conjunto de valores calculados/derivados, para uma dada percepção, envolvendo trecho de via férrea e trem. Eles são: deslocamento previsto, tempo gasto da ação, força aceleradora, resistência das locomotivas, resistência dos vagões, velocidade máxima permitida, velocidade máxima

projetada, velocidade desejada, pressão de freios utilizada, esforço aderente, esforço efetivo e consumo da ação. Este módulo não foi desenvolvido porque ele não faz parte do contexto deste trabalho, mas ele faz parte do Projeto PAI-L.

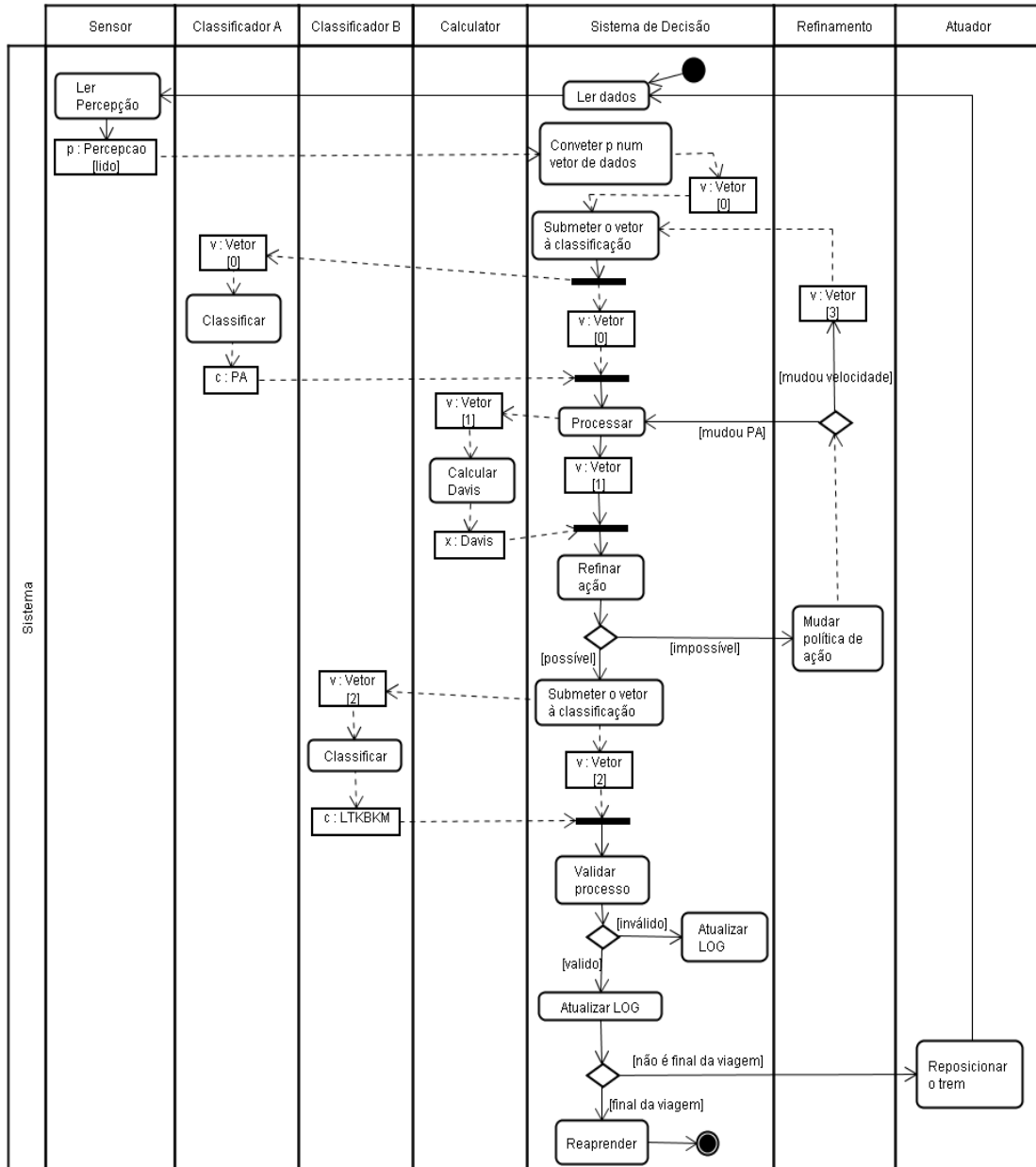


Figura 27- Modelo representativo do algoritmo de viagem

O fluxo das atividades desenhadas no diagrama da Figura 27 está descrito no algoritmo *ExecutaViagem()*, apresentado a seguir. A Tabela 16 descreve, de forma resumida, as funções usadas em tal algoritmo.

Tabela 16 - Funções de cálculo, percepção, atuação e classificação.

Operação	Descrição
$X:=cCamada(c)$	Cria uma camada e instanciá-a com um classificador c , onde c é o nome do arquivo que contém um determinado conjunto de regras.
$pt:=pTrem(\tau)$	Retorna o peso total de um trem τ
$nl:=nLocomotivas(\tau)$	Retorna o número de locomotivas de um trem τ
$p:=percebe()$	Realiza a função de percepção.
$v:=converte(p)$	Converte uma percepção p em um vetor v de informações.
$pa:=classifica(A, v)$	Submete v ao classificador A . Este último retorna um ponto de aceleração.
$u:=calcula(v, pa)$	Realiza os cálculos de esforço trator, deslocamento previsto, etc.
$va:=vAtual(p)$	Retorna a velocidade atual contida na percepção p .
$vd:=vDesejada(p)$	Retorna a próxima velocidade desejada. O valor padrão é 2 km/h maior que a velocidade atual, exceto para os casos que são necessários valores menores para evitar patinagem.
$ea:=eAderente(p)$	Retorna o esforço trator aderente do trem.
$ee:=eEfetivo(p)$	Retorna o esforço trator efetivo total do trem.
$dp:=dPrevisto(p)$	Retorna o deslocamento previsto da ação, em metros; ele não pode ser superior a 500m.
$fa:=fAceleração(p)$	Retorna a força de aceleração total do trem
$atua(E_I, direção, dp)$	O <i>Atuador</i> executa tal função, reposicionando o trem na via. A nova posição é dada em função do quilômetro atual E_I , da direção da movimentação do trem na via e o deslocamento previsto dp ,
$ca:=cAcumulado(p)$	Retorna o consumo acumulado de uma viagem a cada nova percepção p .
$da:=dAcumulado(p)$	Retorna o deslocamento acumulado de uma viagem a cada nova percepção p .
$ltk:=cFinal(ca, da, pt)$	Retorna o consumo final de uma viagem, em LTKB.
$aLOG(v)$	Atualiza um arquivo de LOG com o vetor v , incluindo os dados da ação aplicada. Este arquivo de LOG poderá ser usado em processo de reaprendizagem, o qual não foi examinado no contexto deste trabalho.

Os dados de entrada do algoritmo são, respectivamente, dois classificadores A e B , um trem τ , uma via férrea ω , um quilômetro inicial E_0 , um quilômetro atual/intermediário E_I e um quilômetro final E_N indicando o local de término da viagem.

Na seqüência os classificadores A e B são carregados em suas respectivas camadas de contexto. Nesta mesma seção do algoritmo, determina-se também o peso total do trem, o número total de locomotivas, ajusta o ponto de aceleração inicial em zero e a direção da movimentação. Ajusta-se também o quilômetro atual com o valor do quilômetro inicial.

```

Algoritmo ExecutaViagem()
Entrada: A := {regra1, regra2, ..., regrax}
        B := {regra1, regra2, ..., regray}
        τ := {locomotiva1, ..., locomotivam} U {vagão1, ..., vagãon}
        ω := {pontoMedida1, pontoMedida2, ..., pontoMedidak}
        E0 := Quilômetro Inicial
        EI := Quilômetro Atual
        EN := Quilômetro Final

Inicio
R := cCamada( A )
S := cCamada( B )
pt := pTrem( τ )
nl := nLocomotivas( τ )
pa := 0
se(EN < E0) então direção := -1 senão direção := 1 fim-se
EI := E0
enquanto EI ≤ EN faça

    {PERCEPÇÃO}
    p := percebe( EI, ω, τ )

    {CLASSIFICAÇÃO}
    v := converte( p )
    pa := classifica( A, v )
    davis := calcula( p, pa )
    p := atualiza( EI, ω, τ, p, pa )
    ltkbC := classifica( B, p )
    p := atualiza( EI, ω, τ, p, ltkbC )

    {CÁLULOS}
    vi := vAtual( p )
    vf := vDesejada( p )
    ea := eAderente( p )
    ee := eEfetivo( p )
    dp := dPrevisto( p )
    fa := fAceleração( p )

    {O símbolo ↑ indica passagem de parâmetro por referência}
    {CONTROLE DE PATINAGEM E FALTA DE PONTÊNCIA}
    cPatinagem( EI, ω, τ, ↑p, pa, vi, vf, ea, ee, dp, fa, pt, nl )
    vFaltaPotencia( EI, ω, τ, ↑p, pa, vi, vf, dp, fa )
    vMarchaExcessiva( EI, ω, τ, ↑p, pa, dp, fa )

    {GERAÇÃO DE LOG}
    aLOG( v )

    {ATUAÇÃO}
    atua( EI, direção, dp ) {movimenta o trem}

end-enquanto

    {RELATÓRIO}
    ca := cAcumulado( p )
    da := dAcumulado( p )
    cf := cConsumoFinal( ca, da, pt )

fim-algoritmo

```

O laço principal do algoritmo se executará enquanto a movimentação do trem não atingir a indicação do quilômetro final. Dentro deste laço, as principais ações são as seguintes: percepção, classificação, cálculos, controle de patinagem, marcha excessiva e falta de potência, geração de *log*, atuação e relatório.

A percepção consiste em ler para p diferentes valores fornecidos por sensores presentes nas locomotivas, a saber: posição atual, velocidade, pressão de freios, entre outros mostrados na Figura 17. A classificação consiste em converter o conjunto de valores percebidos em um vetor de características v e submetê-lo a um classificador A que tem por objetivo fornecer um ponto de aceleração pa . O conjunto de valores da percepção p é enriquecido com pa . A partir de p , diferentes cálculos são efetuados, a saber: velocidade atual, velocidade desejada, esforço trator, deslocamento previsto e força de aceleração. O próximo passo refere-se à execução de três controles básicos, respectivamente: patinagem, falta de potência e marcha excessiva. Estes procedimentos foram descritos anteriormente na forma de algoritmos. Eles atuam essencialmente sobre o ponto de aceleração, aumentando-o ou diminuindo-o, e em caso de insucesso, atua-se também sobre a velocidade. Antes de executar a política de ação, que consiste em aumentar, manter ou reduzir um ponto de aceleração, o vetor v de dados em memória é armazenado em um arquivo de *log*. A iteração encerra-se com a aplicação da política de ação determinada. A seção de relatórios é executada após o encerramento da viagem.

5.6 Considerações finais

De forma pragmática, os nossos esforços foram concentrados em aplicar técnicas de aprendizagem de máquina para extrair padrões de condução a partir de dados históricos e dotar um módulo de software de condução com tais padrões, e examinar o comportamento deste software. Lembrando, que o objetivo do software é conduzir um trem de ponto inicial S a ponto final E . Ele realiza tal objetivo por meio de uma atividade iterativa que consiste em definir e aplicar políticas de ações. Vimos que a política de ações que um condutor pode empregar é: aumentar, manter ou reduzir um ponto de aceleração. Para definir uma política de ação, o simulador dispõe de um conjunto de recursos, tais como perfil da via, perfil do trem e uma base de conhecimento com as regras de condução. Uma boa política de ação é alcançada quando a execução do comportamento associado a tal política realizou uma condução, para um dado trecho, de forma segura e econômica e o mais similar possível da condução do maquinista. Os padrões de condu-

ção são representados por um conjunto de regras ordenadas. A interpretação e a validação de tais padrões são tarefas complexas, devido à dinâmica do domínio de aplicação e a presença de informações incompletas.

A nossa esperança, em termos de resultados com vistas a validar os conhecimentos descobertos, é colocá-los em prática e observar/medir quão próximo é o comportamento do simulador do comportamento do maquinista ser humano.

Capítulo 6

Resultados

Os resultados mostrados, neste capítulo, referem-se aos experimentos realizados em laboratório. Estes experimentos envolveram quatro fases bem-distintas:

(i) a primeira fase consistiu na montagem de uma base de dados reais de viagens de trens, doravante denominada *BDViagens*. Tal base contém 17164 registros e 35 atributos. Após o pré-processamento, a *BDViagens* foi modificada, onde foram removidos 2906 registros e acrescentados 84 atributos, resultando em uma nova base de dados, doravante denominada *BDViagensF*.

(ii) a segunda fase consistiu na obtenção de diferentes conjuntos de treinamentos a partir de *BDViagensF*, a saber: (a) CN: este conjunto de treinamento é a base de dados *BDViagensF* propriamente dita. Ele contém os dados das paradas intermediárias dos trens durante as viagens e o seu conjunto de classes é igual ao número de *pontos de aceleração* de uma locomotiva C30; (b) C4: este conjunto de treinamento é o mesmo que CN, exceto que número de classes foi reduzido para quatro; (c) SN: este conjunto de treinamento é o mesmo que CN, porém sem os dados referentes às paradas intermediárias dos trens durante as viagens; e (d) S4: este conjunto de treinamento é o conjunto C4, porém sem os dados referentes às paradas intermediárias dos trens durante as viagens. Para cada um destes conjuntos de treinamentos (CN, C4, SN e S4) foram geradas três amostras, cujos tamanhos foram de 10%, 20% e 30%, respectivamente.

(iii) a terceira fase consistiu em gerar para cada amostra três classificadores, usando respectivamente os seguintes algoritmos/+métodos de aprendizagem: JRIP, JRIP+BAGGING e JRIP+BOOSTING. Nesta fase, utilizou-se o Teste de Friedman para avaliar as taxas de acerto dos classificadores obtidos pelos diferentes métodos, tomando como grau de significância 5%. A Figura 28 resume esquematicamente esta situação. Deve-se notar que cada classificador obtido a partir de dado conjunto de treinamento, foi avaliado de duas formas: a primeira foi utilizando 30% do conjunto inicial para teste (TT) e a segunda foi utilizando o método de validação cruzada (VC).

(iv) a quarta fase consistiu em empregar os diferentes classificadores, no ciclo <percepção, seleção de uma política de ação, aplicação de ação> (Figura 29), com a finalidade de apoiar a atividade de selecionar a próxima ação a ser aplicada. Nesta fase, utilizou-se a fórmula do cálculo de cosseno para medir a similaridade entre os dados reais de uma condução realizada por um maquinista ser humano e os dados de uma condução gerados pelo simulador.

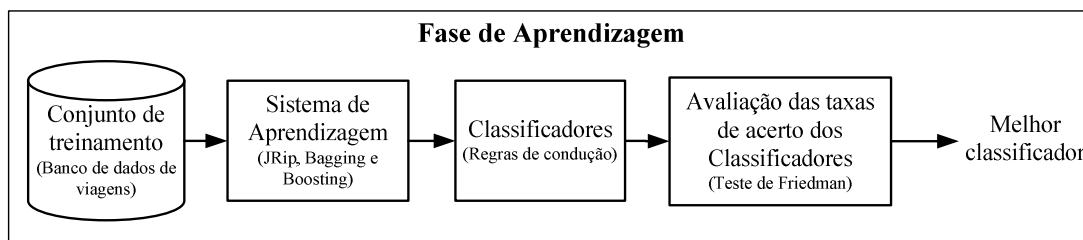


Figura 28 - Fase de aprendizagem.

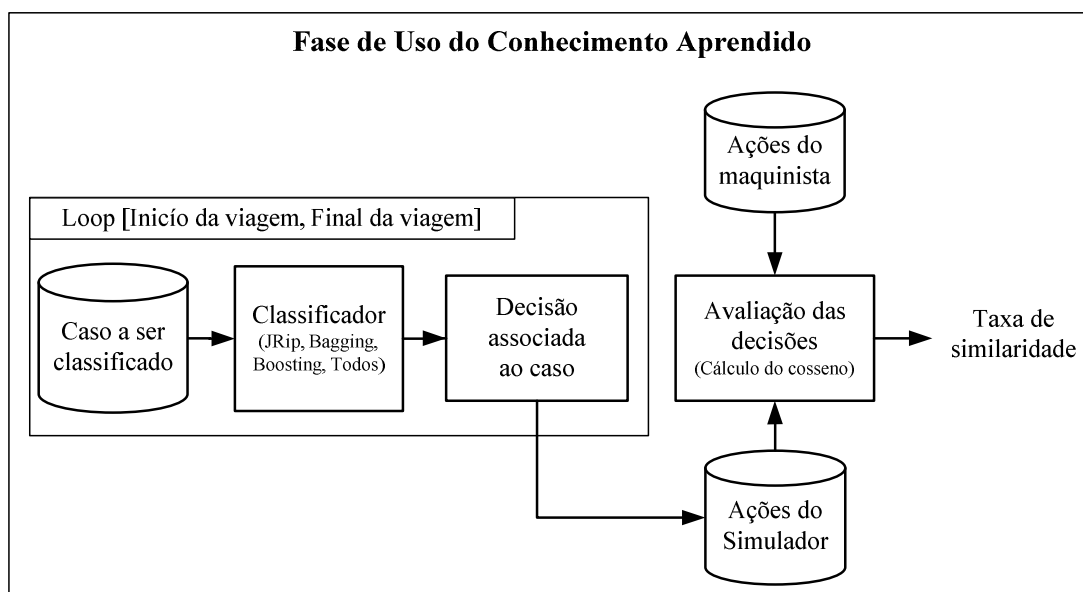


Figura 29 - Fase aplicação do conhecimento aprendido.

A Figura 30 mostra o número de registros associado a cada valor do atributo *meta ponto de aceleração*. Pode-se observar que as classes ou os pontos de aceleração 7 e 8 possuem valores bem superiores às demais classes. Este fato caracteriza um desbalançamento das classes. A Figura 31 mostra a distribuição das classes para os conjuntos de treinamentos: C4 e S4, após a aplicação do procedimento para reduzir tal desbalançamento. Este processo consistiu em rotular todos os registros das classes de 1 a 6 para a classe 3.

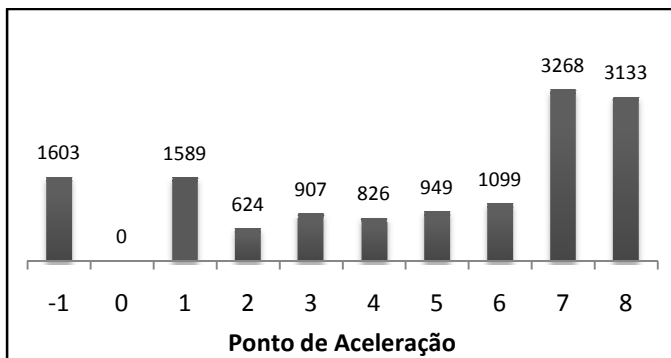


Figura 30 – Número de registros vs classes: CN.

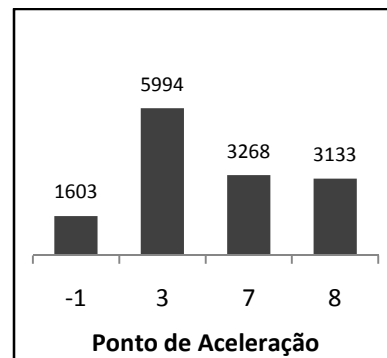


Figura 31 – Número de registros vs classes: C4.

As análises apresentadas nas próximas seções buscam responder as seguintes questões: (a) Quais foram os melhores classificadores obtidos? (b) Qual foi a influência das alterações nos conjuntos de treinamentos no tocante a eficiência dos classificadores obtidos? (c) Qual foi a efetividade, na condução de um trem, dos classificadores empregados pelo simulador em comparação com as ações tomadas pelo maquinista ser humano.

6.1 Fase de Aprendizagem

A fase de aprendizagem contempla basicamente dois momentos: a obtenção dos classificadores e a avaliação das taxas acerto.

A configuração dos experimentos foi estruturada em três tabelas: a Tabela 17 apresenta os dados referentes ao método JRIP; a Tabela 18 apresenta referentes ao método JRIP+BAGGING; e finalmente, a Tabela 19 apresenta os dados referentes ao método JRIP+BOOSTING. Cada valor da coluna *Classificador*, de cada uma destas tabelas, identifica um classificador específico. Por exemplo, o valor *BO_10_S4_VC* é um identificador para um classificador gerado, onde *BO* refere-se ao método de aprendizagem JRIP+BOOSTING, 10 é tamanho da amostra, S4 é conjunto de treinamento, VC é método de validação cruzada. Doravante, os classificadores são referenciados apenas pelo seu identificador, precedido pelo nome do método de aprendizagem para facilitar a leitura.

Como já dito, o foco nesta seção é a análise, de forma comparativa, das taxas de acerto dos diferentes classificadores. Assim, cada tabela subsequente apresenta também a taxa de acerto desses classificadores.

Tabela 17 – Taxas de acerto dos classificadores obtidos por meio do método JRIP.

Método	Tamanho da Amostra	Conjunto de Treinamento	Avaliação	Classificador	Taxa de Acerto
JRIP	10	CN	TT	JR_10_CN_TT	46,9%
			VC	JR_10_CN_VC	76,3%
		C4	TT	JR_10_C4_TT	68,2%
			VC	JR_10_C4_VC	83,0%
		SN	TT	JR_10_SN_TT	48,4%
			VC	JR_10_SN_VC	80,8%
	S4	TT	JR_10_S4_TT	65,9%	
		VC	JR_10_S4_VC	82,3%	
	20%	CN	TT	JR_20_CN_TT	56,4%
			VC	JR_20_CN_VC	78,4%
		C4	TT	JR_20_C4_TT	73,6%
			VC	JR_20_C4_VC	87,6%
		SN	TT	JR_20_SN_TT	56,9%
			VC	JR_20_SN_VC	83,1%
	S4	TT	JR_20_S4_TT	73,1%	
		VC	JR_20_S4_VC	89,1%	
	30%	CN	TT	JR_30_CN_TT	58,5%
			VC	JR_30_CN_VC	77,2%
		C4	TT	JR_30_C4_TT	78,6%
			VC	JR_30_C4_VC	87,6%
		SN	TT	JR_30_SN_TT	60,6%
			VC	JR_30_SN_VC	79,9%
	S4	TT	JR_30_S4_TT	76,9%	
		VC	JR_30_S4_VC	91,2%	

Tabela 18 - Taxas de acerto dos classificadores obtidos por meio do método JRIP+BAGGING.

Método	Tamanho da Amostra	Conjunto de Treinamento	Avaliação	Classificador	Taxa de Acerto
JRIP+BAGGING	10	CN	TT	BA_10_CN_TT	58,9%
			VC	BA_10_CN_VC	93,0%
		C4	TT	BA_10_C4_TT	76,1%
			VC	BA_10_C4_VC	94,7%
		SN	TT	BA_10_SN_TT	58,9%
			VC	BA_10_SN_VC	95,0%
	S4	TT	BA_10_S4_TT	72,5%	
		VC	BA_10_S4_VC	94,8%	
	20%	CN	TT	BA_20_CN_TT	64,4%
			VC	BA_20_CN_VC	93,5%
		C4	TT	BA_20_C4_TT	79,7%
			VC	BA_20_C4_VC	94,3%
		SN	TT	BA_20_SN_TT	64,8%
			VC	BA_20_SN_VC	93,5%
	S4	TT	BA_20_S4_TT	79,9%	
		VC	BA_20_S4_VC	95,0%	
	30%	CN	TT	BA_30_CN_TT	68,6%
			VC	BA_30_CN_VC	92,4%
		C4	TT	BA_30_C4_TT	84,2%
			VC	BA_30_C4_VC	95,7%
		SN	TT	BA_30_SN_TT	69,4%
			VC	BA_30_SN_VC	93,9%
	S4	TT	BA_30_S4_TT	84,7%	
		VC	BA_30_S4_VC	96,2%	

Tabela 19 - Taxas de acerto dos classificadores obtidos por meio do método JRIP+BOOSTING.

Método	Tamanho da Amostra	Conjunto de Treinamento	Avaliação	Classificador	Taxa de Acerto	
JRIP+BOOSTING	10	CN	TT	BO_10_CN_TT	52,6%	
			VC	BO_10_CN_VC	100,0%	
		C4	TT	BO_10_C4_TT	76,9%	
			VC	BO_10_C4_VC	100,0%	
		SN	TT	BO_10_SN_TT	56,8%	
			VC	BO_10_SN_VC	100,0%	
		S4	TT	BO_10_S4_TT	73,6%	
			VC	BO_10_S4_VC	100,0%	
		20%	CN	TT	BO_20_CN_TT	58,5%
				VC	BO_20_CN_VC	99,4%
			C4	TT	BO_20_C4_TT	82,5%
				VC	BO_20_C4_VC	99,4%
	SN		TT	BO_20_SN_TT	61,8%	
			VC	BO_20_SN_VC	99,8%	
	S4		TT	BO_20_S4_TT	81,7%	
			VC	BO_20_S4_VC	99,8%	
	30%		CN	TT	BO_30_CN_TT	63,5%
				VC	BO_30_CN_VC	99,2%
			C4	TT	BO_30_C4_TT	87,4%
				VC	BO_30_C4_VC	99,7%
		SN	TT	BO_30_SN_TT	68,0%	
			VC	BO_30_SN_VC	99,9%	
		S4	TT	BO_30_S4_TT	84,2%	
			VC	BO_30_S4_VC	99,5%	

As melhores taxas de acertos foram obtidas nos classificadores gerados a partir do método JRIP+BOOSTING. Obteve-se, por exemplo, para os classificadores BO_10_C4_VC, BO_10_SN_VC e BO_10_S4_VC a taxa de acerto igual a 100%. A pior taxa de acerto foi obtida para um classificador gerado a partir do método JRIP, a saber: JR_10_CN_TT apresentou a taxa de acerto de 49,6%. Em geral, os classificadores que obtiveram as menores taxas de acerto foram aqueles gerados a partir do método JRIP, seguido, respectivamente, pelos classificadores gerados pelo método JRIP+BAGGING. Em termos teóricos, este resultado já era esperado. Em termos práticos, este resultado também era esperado com base em Lopes (2007).

Do universo de classificadores gerados pelo JRIP+BOOSTING, apenas seis deles tiveram sua taxa de acerto inferior aos demais, a saber: BA_10_SN_TT, BA_20_CN_TT, BA_20_SN_TT, BA_30_CN_TT, BA_30_SN_TT e BA_30_S4_TT. Ou seja, em 75% dos casos, o método JRIP+BOOSTING produziu melhores classificadores em termos de taxa de acerto. O JRIP+BAGGING mostrou-se superior ao JRIP+BOOSTING apenas quando foi utilizado 30% dos dados para teste do classificador e os outros 70% para treinamento.

Foi realizado um aprofundamento da análise para verificar se os diferentes conjuntos de treinamentos poderiam dar origem a classificadores significativamente diferentes. Isto foi feito usando o Teste de Friedman. Tal teste é não-paramétrico. Ele não requer o conhecimento da distribuição da variável na população. Ele permite ranquear os métodos (JRIP, JRIP+BAGGING e JRIP+BOOSTING) para cada conjunto de dados de forma separada. O método com o melhor desempenho toma a primeira posição do *ranking* e assim sucessivamente.

Em termos pragmáticos, tal teste ajuda verificar se os classificadores gerados apresentam diferenças significativas. Caso a hipótese nula seja caracterizada, todos os métodos são equivalentes uma vez que possuem *rankings* iguais. Entretanto, para verificar se há ou não correlação, deve-se fazer o somatório das variâncias dos ranques. A partir deste somatório, pode-se calcular o *p-valor* como a probabilidade do valor ser superior ou igual à variância obtida, utilizando a distribuição *qui-quadrada* com *k-1* graus de liberdade. O resultado numérico do Teste de Friedman fornece um nível de significância (*p-valor*), caso este seja maior que 0.05, então é recomendado rejeitar a hipótese nula.

O valor do *qui-quadrado* do conjunto de valores foi 8.2 e *p-valor* de 0.042. Para efeito comparação apenas o *p-valor* foi considerado. Este valor representa a semelhança entre os classificadores. A Tabela 20 apresenta o *p-valor* de cada configuração para o Teste de Friedman.

Tabela 20 - Teste de Friedman para os classificadores obtidos para conjuntos de treinamentos de tamanhos diferentes.

Tamanho da Amostra	Configuração	p-valor
10%	70% treinamento e 30% teste	0.04200
10%	validação cruzada	0.30802
20%	70% treinamento e 30% teste	0.04206
20%	validação cruzada	0.13443
30%	70% treinamento e 30% teste	0.04206
30%	validação cruzada	0.12176

Para um valor de *p-valor* < 0.05 é possível concluir que há diferença significativa entre as diferentes configurações, ou seja, é rejeitada a hipótese nula, apesar de três testes obtiveram resultados inferiores e muito próximos de 0.05 em três testes. Isto se verificou, em particular, para os casos onde foi usado o método de validação cruzada.

Pouco foi dito até o momento sobre o desempenho dos classificadores em operação. Para medir tal desempenho, a abordagem adotada foi calcular a similaridade entre uma viagem realizada por um maquinista e uma viagem, com a mesma configuração, realizada pelo simulador. A similaridade porta sobre as ações reais tomadas pelo maquinista e as ações tomadas pelo simulador.

6.2 Fase de Aplicação de Classificadores

De forma prática, o principal resultado deste trabalho consistiu em obter classificadores a partir de históricos de viagens de trens e aplicá-los de modo a sugerir rapidamente um ponto de aceleração, visando à definição de uma boa política de condução. Tal política deve gerar uma condução eficiente.

Os parâmetros que indicam a eficiência de uma condução são: o tempo de viagem e o consumo final. Deve-se notar que o consumo final varia de acordo com as características do trem (peso, quantidade de locomotivas/vagões e distância percorrida). Nos experimentos, observou-se que os valores dos consumos variaram de 5.68 a 6.49 LTKB, para os tempos de viagens de 134 e 208 minutos de duração, respectivamente. O trecho foi de aproximadamente 68 km, partindo de pontos próximos ao quilômetro 338 e finalizando próximo ao quilômetro 270, conforme mostrado na Figura 32.

Lembramos que a pressão normal do encanamento de freios de uma locomotiva é em torno de 90 *psi*. A aplicação de freio resulta na redução deste valor, podendo chegar em um mínimo de 64 *psi* por medidas de segurança. Na Figura 32 é possível observar fortes aplicações de freio, principalmente quando o trem necessita parar, alcançando a pressão de freios máxima permitida (observadas no quilômetro 295 e 332).

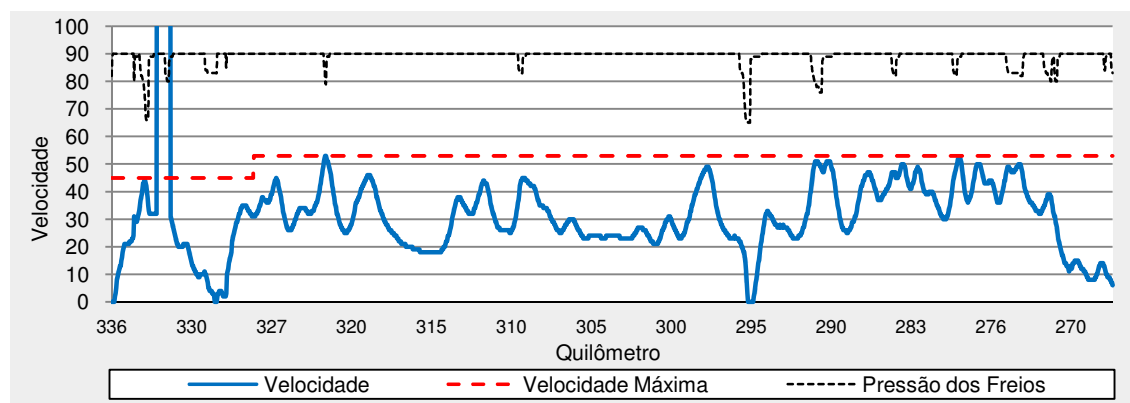


Figura 32 – Trecho real da via férrea: de Londrina (PR) até Paiçandu (PR)

Os motivos das paradas podem ser: (i) a presença de outro trem vindo em sentido contrário, o que obriga o trem a entrar num desvio e parar até que o outro trem passe; (ii) a falta de licença para percorrer o trecho. Como tais paradas são em números pequenos optou-se por gerar classificadores com e sem os dados referentes a tais paradas, no intuito de analisar se as mesmas representavam ruídos significativos na aprendizagem. (iii) a freqüente variação da velocidade do trem. Esta variação é observada quando a velocidade chega muito próxima a velocidade máxima permitida e em seguida retorna a velocidade média do trajeto (conforme trechos entre os quilômetros 323 e 320). O contrário, reduções bruscas e retomadas à velocidade média, é observado somente quando o trem pára. Poucos foram os trajetos onde a velocidade se manteve constante ou muito próxima da velocidade média, a exemplo do que acontece próximo ao quilômetro 315 e entre aos quilômetros 305 e 300. As velocidades de 255 km/h, presentes no quilômetro 332, foram removidas durante o pré-processamento dos dados. Elas indicam ausência de medida exata.

A Figura 33 mostra o número de aplicações de cada ponto de aceleração ao longo de uma determinada viagem. Pode-se observar que há uma grande quantidade de aplicações do ponto 7. Tal ponto de aceleração gera a potência necessária para percorrer trechos de via com alta e baixa resistência. Este ponto é mais utilizado em situações de cruzeiro.

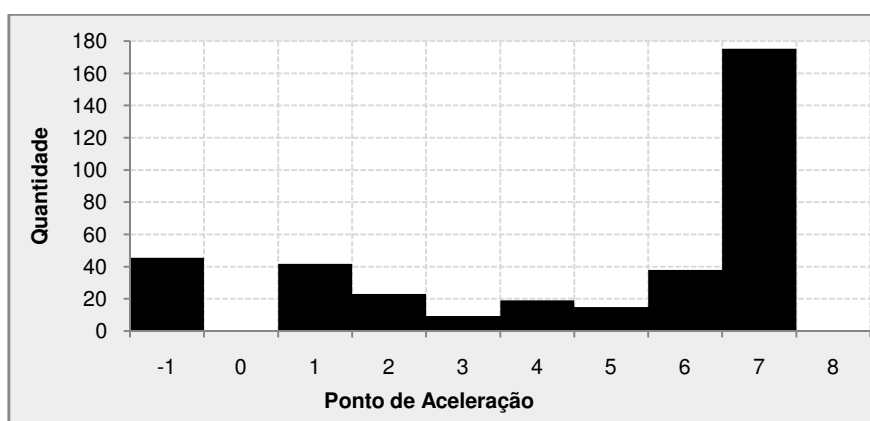


Figura 33 – Uso dos pontos de aceleração pela maquinista ser humano.

Mostrar-se-á, na próxima seção, que foi possível abstrair o conhecimento do maquinista por meio da aplicação métodos de aprendizagem de máquina, onde tal conhecimento foi obtido na forma de classificadores. A abstração conseguida é traduzida na forma de um grau de similaridade entre as tomadas de ações de um maquinista e as

tomadas de ações sugeridas pelo simulador, por meio de seu classificador. Em outras palavras, esta abstração é quantificada. E ela representa de certa forma, o grau de “imitação” que simulador realiza, usando os conhecimentos que lhe foi fornecido na forma de um classificador, com relação ao maquinista.

6.2.1 Aplicação do Conhecimento Descoberto

Como já foi dito anteriormente, a aplicação dos conhecimentos descobertos, no formato de classificadores, foi realizada por meio da inclusão das regras de tais classificadores no ciclo de tomada de decisões do simulador de condução de trens, objeto deste trabalho. Desta forma, o simulador é o responsável por carregar os classificadores gerados, interpretar suas regras e tomar decisões com base nas recomendações sugeridas. Foram simuladas diferentes viagens seguindo as características descritas na Tabela 21.

Tabela 21 - Configurações dos experimentos entre termos de composição e comprimento do trecho de via.

Configuração	Distância Percorrida	Quantidade	
		Locomotivas	Vagões
A	71.48	3	58
B	66.94	4	59
C	70.62	3	47

Da Figura 34 à Figura 36 mostra-se o resultado da aplicação do classificador JRIP+BOOSTING nas viagens A, B e C. A sua aplicação resulta em diferentes comportamentos, visto que as viagens possuem características distintas entre si, como peso e pontos de início e fim. Outros pontos merecem destaque quanto ao comportamento do simulador:

- (i). As restrições de velocidade máxima foram impostas de acordo com os históricos das viagens. As viagens, quando comparadas, seguem as mesmas características de via e trens das viagens reais, assim os limites de velocidades são os mesmos;
- (ii). As velocidades iniciais da viagem até os pontos A1, A2 e A3 são semelhantes. Elas mudam de forma significativa, após estes pontos (entre quilômetros 337 e 325), porque as velocidades máximas são diferentes; e
- (iii). As aplicações de freio são em pontos semelhantes durante as viagens e não apresentam valores abaixo de 80 *psi*. As aplicações são mais frequentes

quando há velocidades máximas muito próximas das velocidades praticadas.

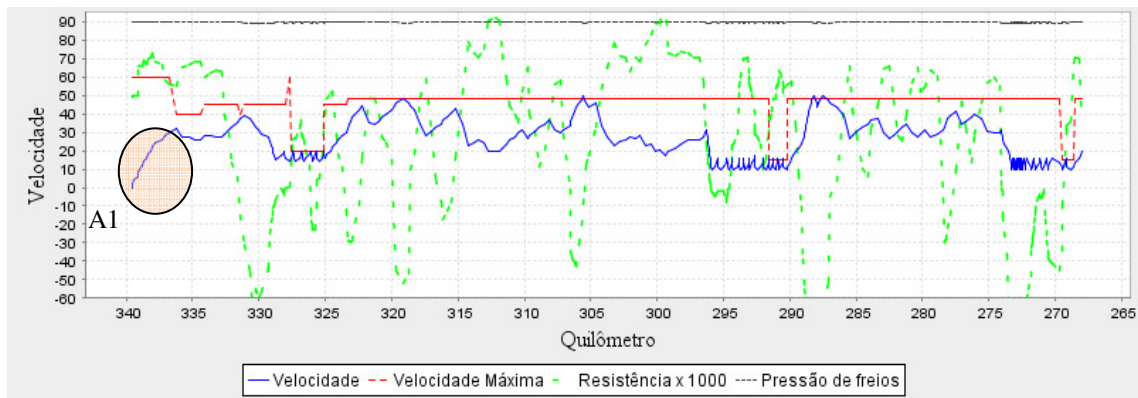


Figura 34 - Resultado da simulação para a configuração A usando as regras descobertas por meio do método JRIP+BOOSTING.

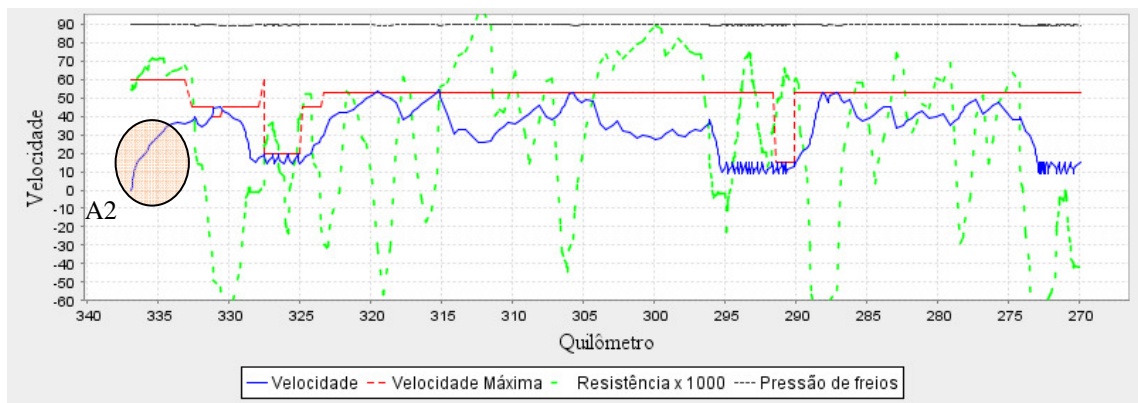


Figura 35 - Resultado da simulação para a configuração B usando as regras descobertas por meio do método JRIP+BOOSTING.

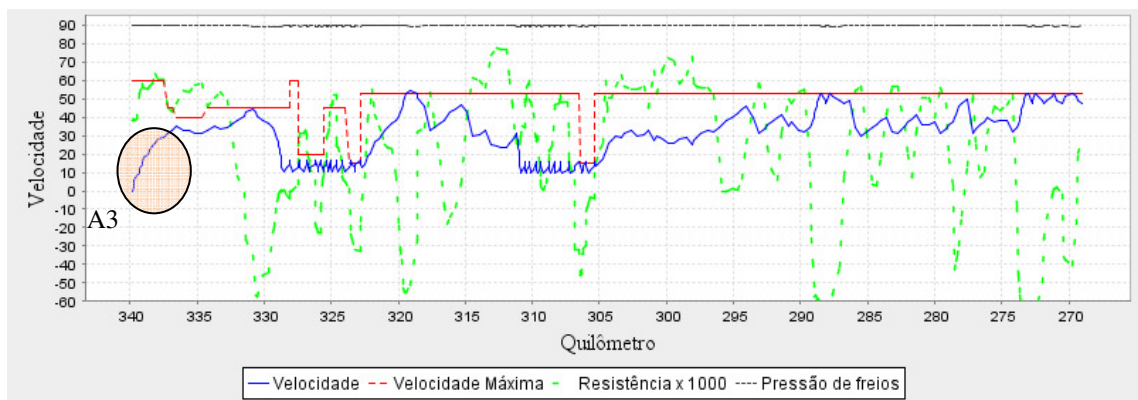


Figura 36 - Resultado da simulação para a configuração C usando as regras descobertas por meio do método JRIP+BOOSTING.

A Figura 37 apresenta a soma das utilizações de cada ponto de aceleração nas viagens A, B e C. Em ambas as viagens, o ponto de aceleração mais utilizado foi o -1, que representa aplicação de freio automático, seguido pelo ponto 1 e o ponto 8. O ponto de aceleração 1 foi mais utilizado que o ponto 8 devido a sobra de potência para vencer o percurso.

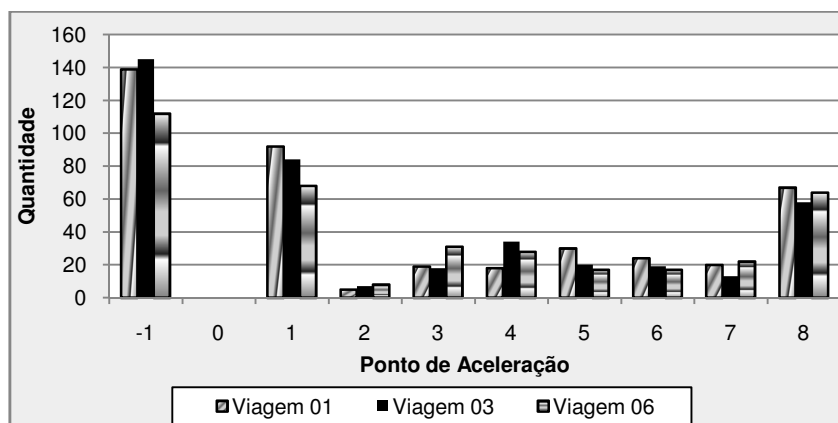


Figura 37 - Uso dos pontos de aceleração nas viagens 1, 3 e 6 para as configurações A, B e C

As próximas seções mostram os resultados obtidos a partir destas duas viagens simuladas, seguindo características definidas para C.

6.2.1.1 Viagem A

Esta viagem foi realizada aplicando todos os classificadores obtidos pelos métodos JRIP, JRIP+BAGGING e JRIP+BOOSTING, a partir do conjunto de treinamento CN. O trecho de ferrovia tinha aproximadamente 68 quilômetros de extensão. O trem era composto por duas locomotivas e trinta e um vagões. O consumo final desta viagem foi de 5,92 LTKB. Este valor é próximo do consumo real no valor de 6.17 LTKB. As durações de 161 minutos e 176 minutos foram observadas, respectivamente, para a condução realizada pelo simulador e pelo maquinista.

A Figura 38 mostra os resultados da execução da viagem usando um classificador gerado pelo JRIP. A velocidade praticada permaneceu na maioria do percurso abaixo da velocidade máxima permitida. Em alguns pontos, tal velocidade foi maior que a máxima, porém não ultrapassou 2 km/h, limite considerado aceitável e seguro. A resistência é inversamente proporcional a velocidade praticada. Quando a resistência aumenta, a velocidade tende a cair devido à falta de força do trem para percorrer o trajeto. Tre-

chos com resistência alta indicam acives e nestas situações a velocidade tende a ser reduzida.

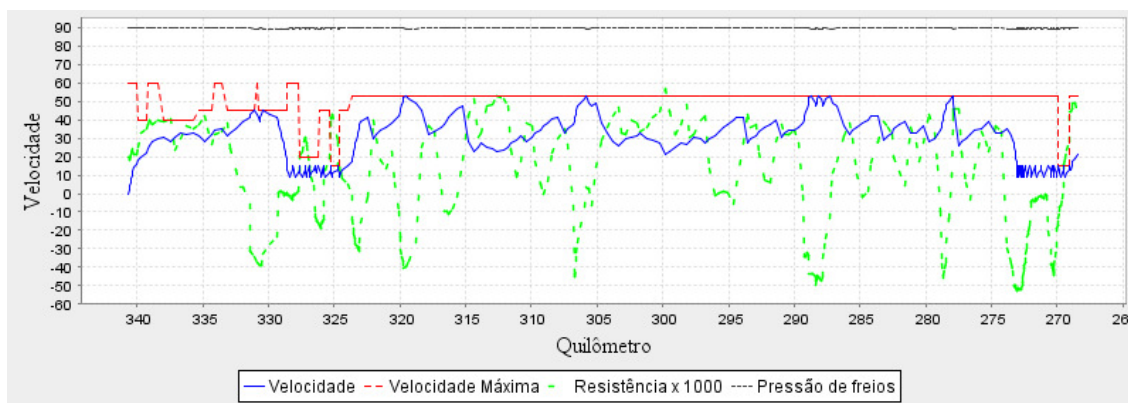


Figura 38 – Dados da viagem A usando um classificador JRIP.

Durante a viagem, (i) várias trocas de ponto de aceleração são realizadas, sendo o ponto -1 o mais utilizado, seguido pelos pontos 1 e 8; (ii) das vinte e cinco regras do classificador onze foram aplicadas. Cada regra indica um ponto de aceleração não exclusivo. A regra mais utilizada foi a DEFAULT, que sugere a aplicação do oitavo ponto de aceleração. Sua aplicação foi de 63% da viagem. A Figura 35 mostra outras duas regras que também foram utilizadas com frequência.

<p>Regra I:</p> <p>$(C_PERCENTUAL_RAIO_CURVA_6 \geq 600.650024)$ and $(C_PERCENT_ENT_INCLINACAO_INSCRICAO_6 \leq 0.501071)$ and $(C_PERCENTUAL_RAIO_CURVA_10 \leq 499)$ $\Rightarrow C_PONTO_ACEL=1 (41.0/4.0)$</p>
<p>Regra II:</p> <p>$(C_PERCENTUAL_RAIO_CURVA_2 \leq 402.850006)$ and $(C_PERCENT_ENT_INCLINACAO_INSCRICAO_5 \leq 0.125223)$ and $(C_PERCENTUAL_CURVA_VENCIDA_1 \leq 90.359055)$ and $(C_CABECEIRA_4 = \text{nível})$ $\Rightarrow C_PONTO_ACEL=7 (35.0/6.0)$</p>

Figura 39 - Regras frequentemente utilizadas pelo classificador

A Regra I sugere a aplicação do ponto de aceleração 1, baseando-se no percentual de raio da curva e no percentual de rampa. A Regra II sugere a aplicação do ponto de aceleração 7, baseado-se no percentual de raio de curva, no percentual de rampa, no percentual de curva vencida e parte da locomotiva encontra-se em nível. Um segundo

classificador foi utilizado para a mesma viagem. Este segundo classificador foi gerado usando o método BOOSTING. A regra padrão foi novamente a mais utilizada, porém em apenas 31% da viagem. Esta taxa foi inferior ao JRIP (63%), JRIP+BAGGING (36.5%).

Finalmente, executou-se uma abordagem combinatória. Em outras palavras, cada percepção foi submetida a três classificadores diferentes, obtidos, respectivamente, pelo JRIP, JRIP+BAGGING e JRIP+BOOSTING, resultando em três diferentes regras sugeridas. A regra com maior suporte foi selecionada para ser aplicada. A execução desta abordagem visou utilizar o conhecimento abstraído pelos três classificadores durante uma condução. Esta abordagem foi nomeada de TODOS.

6.2.1.2 Viagem B

Esta viagem foi realizada aplicando todos os classificadores obtidos pelos métodos JRIP, JRIP+BAGGING e JRIP+BOOSTING, a partir do conjunto de treinamento SN. A particularidade reside na remoção dos registros relacionados às paradas intermediárias do trem. A questão era saber se tais registros interferem na montagem de uma política de condução. Os gráficos (a) e (b) da Figura 40 mostram as velocidades executadas e máximas da viagem antes da remoção e depois da remoção de tais registros. Visualmente, a diferença entre os gráficos reside na ausência de velocidades igual a zero, no gráfico (b).

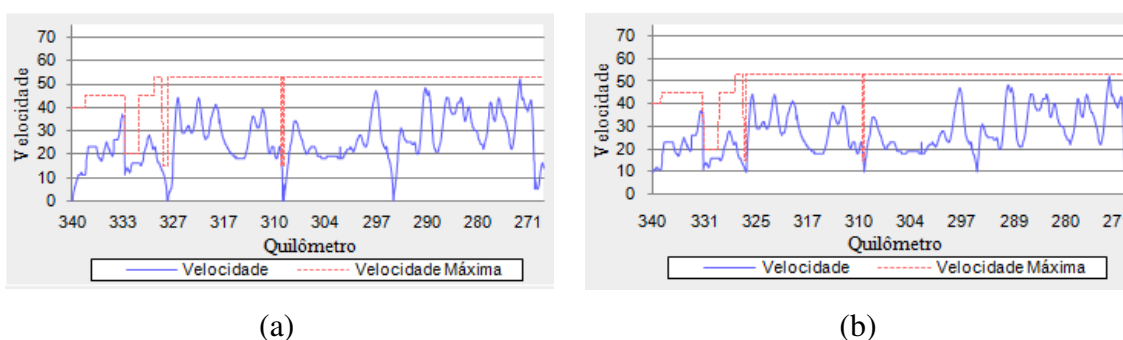


Figura 40 – Exemplo de viagem antes (a) e após (b) a remoção das paradas.

O resultado da execução da viagem utilizando o classificador JRIP é mostrado na Figura 41. Assim como na Viagem Simulada I, as políticas de condução sugeridas pelos classificadores resultaram em comportamentos semelhantes. A velocidade prati-

cada e as frenagens foram similares às praticadas na Viagem Simulada I. Em resumo, a remoção dos registros referentes às paradas intermediárias não foi importante.

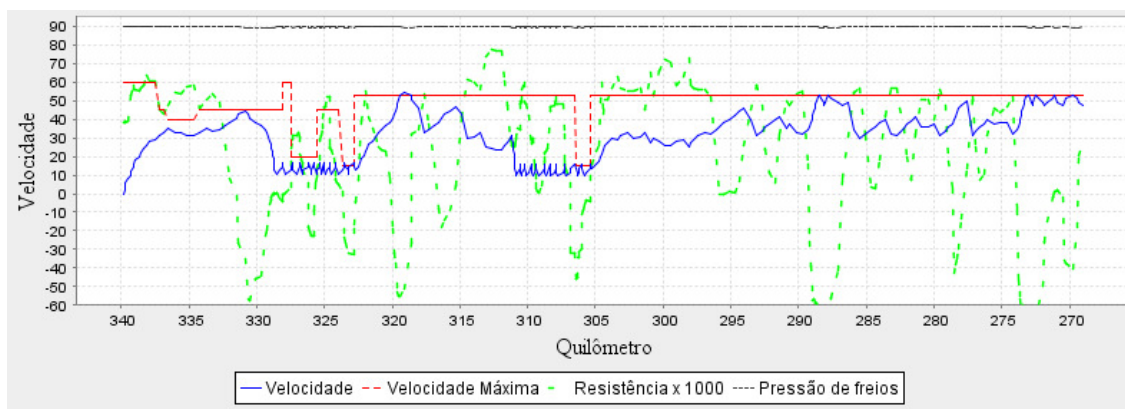


Figura 41 – Dados da viagem B usando um classificador JRIP.

Durante a viagem, alguns pontos de aceleração tiveram a mesma quantidade de aplicações, como, por exemplo, os pontos 5 e 6. Outros, como o ponto 2, foram menos utilizados durante a viagem. O ponto -1 foi o mais aplicado, devido às várias utilizações de freio durante a viagem, necessárias para não deixar o trem ultrapassar a velocidade máxima permitida. Isto se faz necessário porque o Simulador busca sempre aumentar a velocidade do trem, enquanto a velocidade máxima não foi atingida. Como o classificador não foi capaz de abstrair o conhecimento necessário para a utilização de freios, então foi necessário incluir um conjunto de regras específicas para uso dos freios. Essas regras interferem na sugestão do classificador.

A regra padrão foi aplicada de forma muito semelhante em todos os casos estudados (em torno de 30% dos casos), seguida pela regra mostrada na Figura 42. Esta regra baseia-se fundamentalmente no percentual de inclinação de rampa em quatro pontos distintos da composição (0, 1, 4 e 5) e sugere a aplicação do ponto de aceleração 1.

```
(C_PERCENT_SAI_INCLINACAO_INSCRICAO_4 <= -0.0175) and
(C_PERCENT_ENT_INCLINACAO_INSCRICAO_5 <= 0.183347) and
(C_PERCENT_ENT_INCLINACAO_INSCRICAO_0 <= 0.283798) and
(C_PERCENT_ENT_INCLINACAO_INSCRICAO_1 <= 0.811892)
=> C_PONTO_ACEL=1
```

Figura 42 - Regra utilizada pelo classificador JRIP sem dados de paradas.

Na próxima seção será detalhada a análise de similaridade entre as políticas de ações do maquinista ser humano e as políticas de ações do simulador.

6.2.2 Análise da Similaridade da Condução

A forma adotada para avaliar a eficiência de um classificador, na definição de uma política de ação, foi medir a similaridade entre as ações sugeridas pelo classificador e as ações efetivamente realizadas pelo maquinista ser humano. Em termos matemáticos, a métrica foi operacionalizada por meio do cálculo do cosseno. Tal cálculo é dado pela Equação 38, onde o vetor \vec{u} representa os pontos de aceleração usados pelo maquinista ser humano e o vetor \vec{v} representa os pontos de aceleração sugeridos pelo classificador. O número total de ações é igual a n . Os valores resultantes da equação variam entre zero e um, sendo que quanto mais próximo de um, mais similares são os vetores e conseqüentemente melhor o conhecimento obtido.

$$\cos \theta = \cos(\vec{u}, \vec{v}) = \frac{\vec{u} \times \vec{v}}{\sqrt{\sum u_n^2 \times \sum v_n^2}} \quad (38)$$

Na seqüência serão mostrados os resultados obtidos a partir da aplicação de quatro classificadores, obtidos respectivamente, a partir do conjunto de treinamentos CN. A diferença entre os classificadores encontra-se no método utilizado para obtê-los, a saber: JRIP, JRIP+BAGGING, JRIP+BOOSTING e TODOS. Este último é uma combinação dos três primeiros. A configuração do trem corresponde a viagem C da Tabela 21.

A Figura 43 mostra as diferenças entre os pontos de aceleração sugeridos pelo classificador e os pontos aplicados pelo maquinista ser humano. Tais diferenças são computadas da seguinte forma: (i) quando o valor é zero, tem-se que o ponto de aceleração sugerido e aplicado são os mesmos; (ii) quando o valor for maior que zero, tem-se que o ponto de aceleração sugerido é maior que o aplicado; e (iii) quando o valor for menor que zero, tem-se que o ponto de aceleração sugerido é menor que o aplicado. Desta forma, quanto maior for a concentração de pontos em torno do valor zero, melhor foi aplicação do classificador na definição de uma política de ação. A similaridade de condução foi de 83.9%. Este valor foi calculado por meio da fórmula 38.

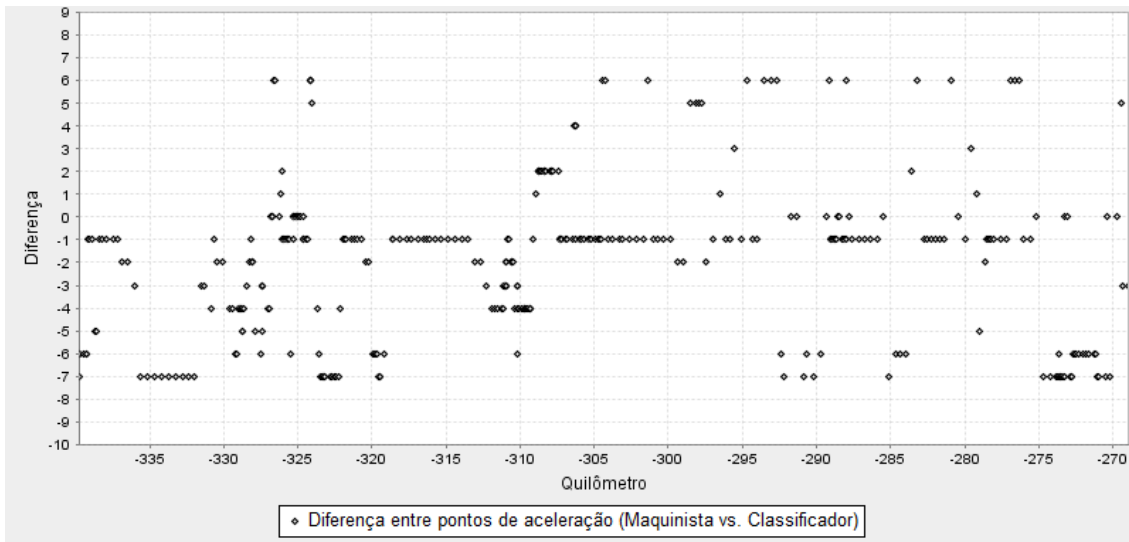


Figura 43 – Diferenças entre pontos de aceleração sugeridos pelo classificador e aplicados pelo maquinista – conjunto de treinamento CN e método JRIP.

De forma análoga ao raciocínio anterior, a Figura 44 mostra as diferenças entre os pontos de aceleração sugeridos pelo classificador obtido por meio do método JRIP+BAGGING. Trechos antes muito similares (entre quilômetros 310 e 300) tiveram uma similaridade reduzida, devido à diferença nas regras de condução, que foram menos eficazes neste trecho. A similaridade de condução aqui foi de 85.6%.

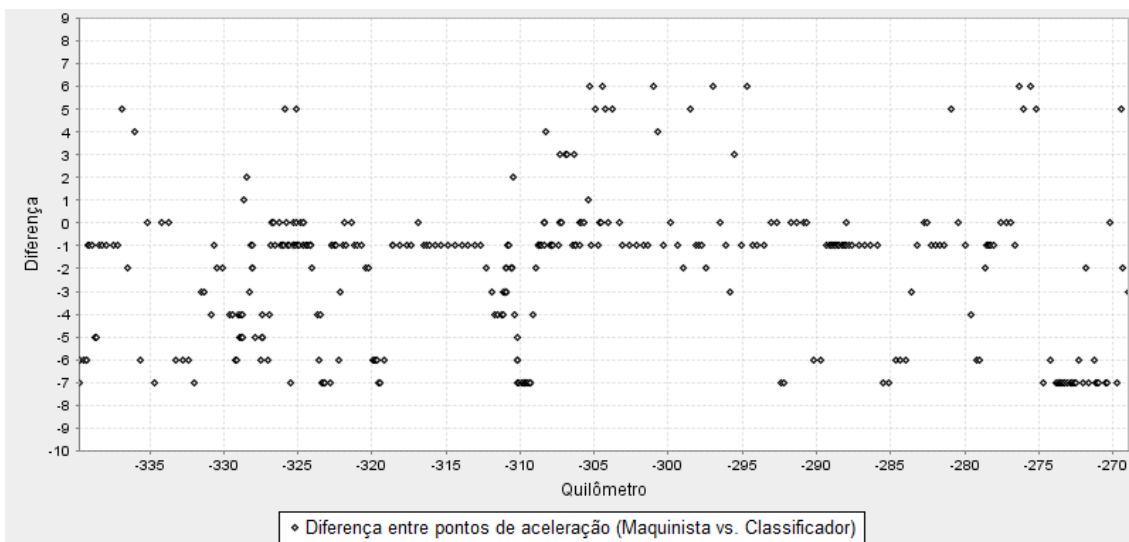


Figura 44 - Diferenças entre pontos de aceleração sugeridos pelo classificador e aplicados pelo maquinista – conjunto de treinamento CN e método JRIP+BAGGING.

A Figura 45 mostra as diferenças entre os pontos de aceleração sugeridos pelo classificador obtido por meio do método JRIP+BOOSTING. Com este conjunto de regras, a similaridade foi de 87,8%, sendo a melhor obtida nos experimentos.

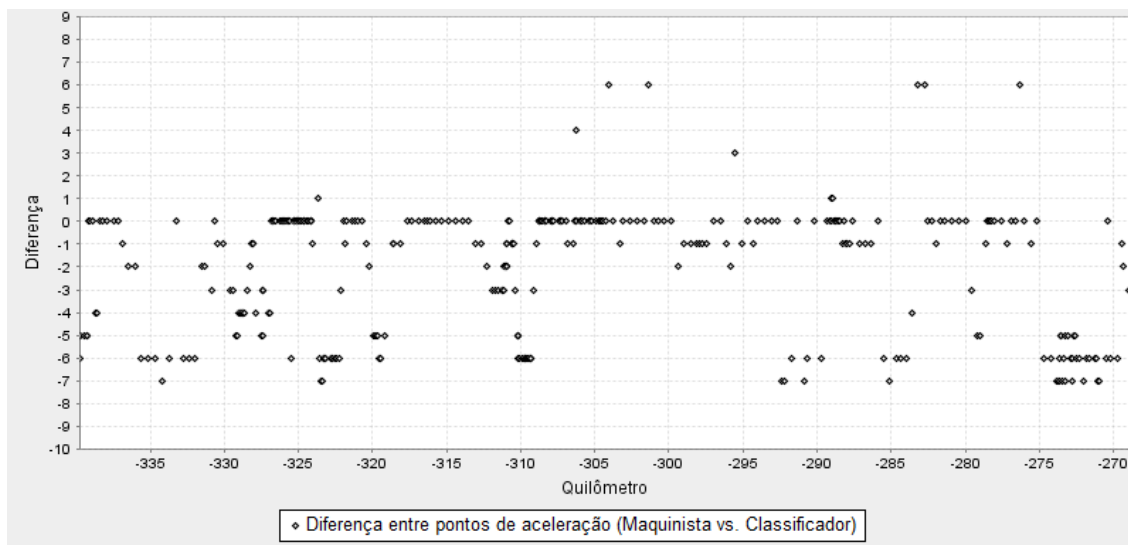


Figura 45 - Diferenças entre pontos de aceleração sugeridos pelo classificador e aplicados pelo maquinista – conjunto de treinamento CN e método JRIP+BOOSTING.

Finalmente, executou-se uma abordagem combinatória. Em outras palavras, cada percepção foi submetida a três classificadores (obtidos por meio da aplicação respectivamente dos métodos JRIP, JRIP+BAGGING e JRIP+BOOSTING ao conjunto de treinamento CN), retornando três diferentes regras. A regra com maior suporte é selecionada para ser aplicada. A condução usando tal abordagem (Figura 46), não mostrou ser melhor que a obtida por apenas um classificador na predição do ponto de aceleração a ser empregado. A similaridade foi de 85.6% contra 87.8% do classificador JRIP+BOOSTING.

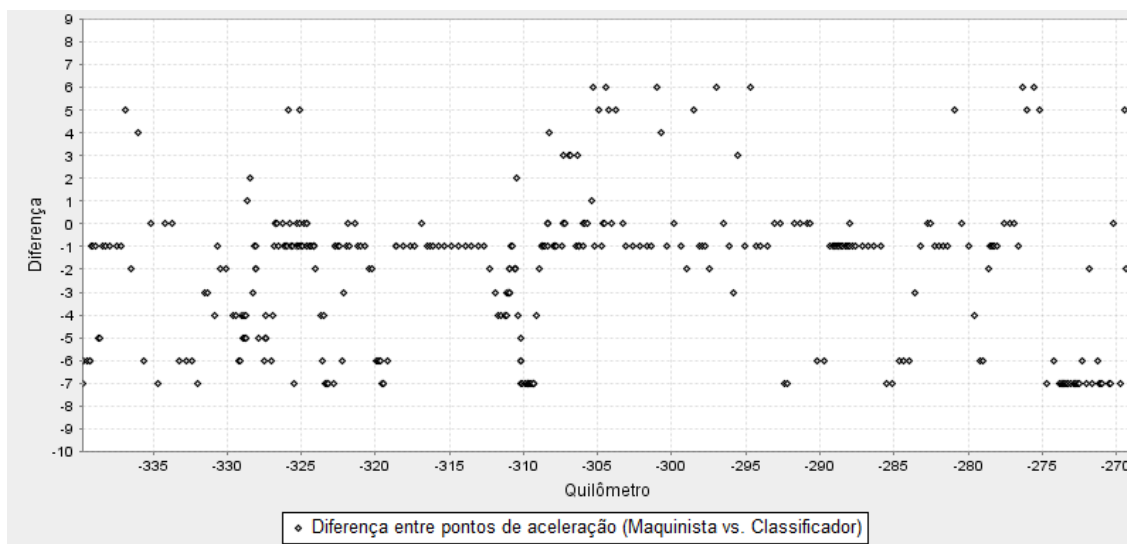


Figura 46 - Diferenças entre pontos de aceleração sugeridos pelo classificador e aplicados pelo maquinista – conjunto de treinamento CN e método TODOS.

Na seqüência são mostrados, na forma de tabelas, os melhores resultados para oito viagens. As diferenças entre as tabelas residem no conjunto de treinamento usado para obter os classificadores.

A Tabela 22 resume os melhores resultados para o conjunto de treinamento CN. Pode-se observar que o classificador que forneceu o melhor resultado foi obtido pelo método JRIP+BOOSTING. Ele foi melhor em cinco das oito diferentes viagens com de taxa de similaridade superior a 85%. As piores taxas de similaridade foram às resultantes da aplicação do JRIP, em seis das oito viagens.

Tabela 22 - Resultados usando classificadores obtidos a partir do conjunto de treinamento CN.

Classificador	Viagem número	Tamanho da Amostra (%)	Resultados Obtidos		
			Similaridade	LTKB	Tempo
JRIP+BAGGING	1	10	0,812	-0,30	37
JRIP+BOOSTING	2	20	0,864	0,10	28
JRIP	3	10	0,847	-0,52	8
JRIP+BOOSTING	5	10	0,855	-0,29	14
JRIP+BOOSTING	6	10	0,878	-0,36	-11
JRIP+BOOSTING	7	10	0,865	-0,07	-44
JRIP+BOOSTING	8	30	0,855	-0,25	-47
JRIP+BAGGING	9	20	0,822	0,12	-12

A Tabela 23 resume os melhores resultados para o conjunto de treinamento C4. A particularidade reside na redução do número de classes de 10 para 4 (classes: -1, 3, 7 e 8). Pode-se observar que o classificador que forneceu o melhor resultado foi obtido

pelo método JRIP. Comparando tais resultados com os valores da Tabela 22, nota-se que houve uma queda na similaridade na condução. A única linha de ambas as tabelas que não sofreram alteração foi à primeira.

Tabela 23 - Resultados usando classificadores obtidos a partir do conjunto de treinamento C4.

Classificador	Viagem Número	Tamanho da Amostra (%)	Similaridade
JRIP+BAGGING	1	30	0,804
JRIP+BAGGING	2	30	0,829
JRIP	3	30	0,813
JRIP	5	10	0,831
TODOS	6	10	0,815
JRIP	7	30	0,809
JRIP	8	10	0,840
JRIP+BOOSTING	9	20	0,817

A Tabela 24 apresenta os melhores resultados para o conjunto de treinamento SN. Assim como nos experimentos realizados com o conjunto de treinamento C4, o classificador obtido pelo método JRIP+BOOSTING, não se sobrepôs em relação aos demais, visto que o classificador com os melhores resultados foi o JRIP. Ele foi melhor em três das oito diferentes viagens com de taxa de similaridade superior a 85%.

Tabela 24 - Resultados usando classificadores obtidos a partir do conjunto de treinamento SN.

Classificador	Viagem Número	Tamanho da Amostra (%)	Similaridade
JRIP	1	20	0,825
JRIP+BOOSTING	2	10	0,849
JRIP	3	20	0,841
JRIP+BAGGING	5	20	0,846
JRIP	6	20	0,876
JRIP	7	20	0,863
JRIP	8	20	0,864
JRIP+BAGGING	9	20	0,818

A Tabela 25 apresenta os melhores resultados para o conjunto de treinamento S4. Diferente do que aconteceu nos experimentos descritos em Tabela 23 e Tabela 24, os resultados mostram que o classificador JRIP+BOOSTING foi melhor em cinco das oito viagens (viagens 3, 5, 7, 8 e 9). As taxas de similaridade permaneceram muito próximas das obtidas nos experimentos anteriores.

Tabela 25 - Resultados usando classificadores obtidos a partir do conjunto de treinamento S4.

Classificador	Viagem Número	Tamanho da Amostra (%)	Similaridade
TODOS	1	30	0,811
JRIP+BAGGING	2	10	0,820
JRIP+BOOSTING	3	30	0,825
JRIP+BOOSTING	5	30	0,842
JRIP	6	30	0,843
JRIP+BOOSTING	7	20	0,835
JRIP+BOOSTING	8	30	0,847
JRIP+BOOSTING	9	30	0,817

6.3 Considerações Finais

Em resumo, todos os classificadores usados nos experimentos foram obtidos a partir de uma mesma base de dados, *BDViagensF*. As diferenças portam sobre algumas variações que foram realizadas nos conjuntos de treinamentos, a saber: redução do número de classes para balancear o conjunto de treinamento e remoção dos registros referentes às paradas intermediárias do conjunto de treinamento. Estas variações resultaram em quatro diferentes conjuntos de treinamentos, que são: a) CN: é base de dados original enriquecida *BDViagensF*; b) C4: é base de dados CN com apenas quatro classes; c) SN: é base de dados CN removidos os registros das paradas intermediárias; e d) S4: é base de dados SN com apenas quatro classes.

Sobre cada uma destas quatro bases de dados foram geradas três amostras, cujos tamanhos foram de 10%, 20% e 30%. Para cada amostra foram gerados três classificadores, usando respectivamente os seguintes métodos, JRIP, JRIP+BAGGING e JRIP+BOOSTING. Cada classificador foi testado usando as seguintes abordagens: validação cruzada ou um conjunto de exemplos específicos de teste, correspondendo 30% do conjunto de treinamento. Em resumo, foram gerados 72 classificadores.

Em conclusão, os classificadores gerados a partir do método JRIP+BOOSTING produziram as melhores taxas de acertos e também as melhores taxas de similaridades, quando comparadas as viagens simuladas e viagens reais realizadas por maquinistas seres humanos. Teoricamente, este resultado já era esperado. Experimentalmente, este resultado também corrobora com o que foi descrito em Lopes (2007).

Para um grau de significância de 5%, o Teste de Friedman mostrou que não há diferenças entre os classificadores em termos de suas taxas de acerto em algumas situa-

ções. Entretanto, as alterações nos conjuntos de treinamentos (C4 e SN) não se traduziram em melhores resultados em termos de taxa de acerto. A efetividade da aplicação dos classificadores, na condução de um trem, por meio do simulador na montagem de uma boa política de condução. Tal efetividade foi traduzida por meio da taxa de similaridade entre o conjunto de ações tomadas pelo maquinista e o conjunto de ações tomadas pelo simulador. As melhores taxas ficaram em torno de 85%.

A principal contribuição de um classificador, na abordagem adotada para a realização do simulador de condução de trens, reside na redução do espaço de busca do ponto de aceleração ideal para uma dada circunstância. A taxa de similaridade de 85% supracitada corrobora com tal afirmação.

Capítulo 7

Conclusões

Conforme mencionadas no Capítulo 2, diferentes técnicas de mineração de dados têm sido utilizadas no processo de descoberta de conhecimento em bases de dados no decorrer dos últimos anos. Algoritmos de aprendizagem de máquina, como por exemplo, o C4.5, são empregados com sucesso em bases de dados com características distintas. Quando combinados em diferentes variantes, tais como BAGGING e BOOSTING, tem-se percebido melhoras significativas no processo de classificação.

Relembramos que o objetivo deste trabalho era descobrir padrões, a partir de dados coletados por meio de diferentes sensores instalados em um trem de carga, para ajudar no planejamento e execução de uma boa política de condução. Sobre os dados coletados foram aplicados processos de descoberta de conhecimento: (i) seleção dos atributos, (ii) inclusão de novos atributos, (iii) transformação de dados, (iv) remoção de dados ruidosos, (v) aplicação de algoritmos tais como: JRIP, JRIP+BAGGING e JRIP+BOOSTING (vi) interpretação das regras e (vii) validação do conhecimento obtido.

Para validar a abordagem, uma vez obtidos os classificadores, era necessário aplicá-los para verificar sua eficiência. Para tal foi desenvolvido um módulo de software que simula a realização das tomadas de ações para uma viagem. O ciclo básico de execução do simulador consiste em perceber um evento, submetê-lo ao classificador que retorna o ponto de aceleração pa , avaliar se o pa é aplicável, se a avaliação for positiva então o pa é aplicada, caso contrário um processo de ajuste é lançado.

Este processo de tomada de decisão resultou em regras de condução que, quando empregadas, obtiveram similaridades em torno de 85% entre as ações do simulador e do maquinista para uma mesma viagem. Acreditamos que estas regras: (i) poderiam ter aumentado tal taxa de similaridade, se o conjunto de dados de treinamentos fosse mais significativo e também se os maquinistas tivessem comportamentos menos específicos no tocante às trocas dos pontos de aceleração; e (ii) podem ser aplicadas como uma

forma de auxiliar o maquinista, sugerindo rapidamente um ponto de aceleração frente uma situação observada.

Os classificadores gerados pelo método JRIP+BOOSTING produziram os melhores resultados, tanto em taxas de acerto como em taxas de similaridade. Tal taxa de similaridade mede o quão próximo foram tomadas ações, na execução de uma viagem, pelo simulador e pelo maquinista.

De modo geral, os classificadores tendem a sugerir pontos de aceleração acima do necessário para vencer as resistências devido à grande quantidade de registros nos conjuntos de treinamentos associados aos pontos sete e oito. Isto pode ser observado uma vez nos gráficos que mostraram a similaridade da condução por das diferenças entre o ponto de aceleração sugerido e aplicado pelo simulador. Acreditamos que, quando tais classificadores, forem usados para conduzir trens com pesos maiores que os simulados neste trabalho podem-se melhorar os resultados, pois a tendência do classificador em sugerir pontos com potência superior à requerida poderá ser mais bem aproveitada.

Para um grau de significância de 5%, o teste de Friedman mostrou diferenças entre os classificadores gerados a partir dos diferentes conjuntos de treinamentos usados. Entretanto, tal diferença teve efeito quase nulo na utilização dos classificadores, no interior do simulador, para sugerir pontos de aceleração e traduzir suas sugestões em viagens mais econômicas. Isto ocorreu devido à influência das regras das camadas de controle. Estas regras ajustam as sugestões dos classificadores, por exemplo, para evitar patinagem. Observou-se que embora cada classificador faça sugestões de pontos de aceleração diferentes, a aplicação das regras das camadas de controle selecionam pontos de aceleração muito próximos ou iguais. Entretanto, deve-se salientar que por mais que sejam aplicados os mesmos pontos de aceleração, como cada classificador sugere pontos de aceleração diferentes, as taxas similaridades são diferentes.

7.1 Limitações

As principais limitações podem ser resumidas nos seguintes itens:

1. Número reduzido de dados históricos de viagens;
2. Dificuldade em gerar classificadores com o conjunto total de exemplos em termos de recursos computacionais;

3. Ausência de modelos matemáticos referentes à frenagem de trem, impossibilitando à realização de procedimentos simulados de frenagem mais aderentes a realidade;
4. Ausência de validação do conjunto de regras por um especialista em condução. Os testes realizados foram objetivos.

7.2 Trabalhos Futuros

Uma das possíveis extensões é utilizar uma quantidade de dados maior na geração dos classificadores. Tais dados poderão ser obtidos a partir de viagens reais ou viagens simuladas em laboratório. A simulação de viagem em laboratório poderia ser realizada pelo próprio simulador.

Outras extensões podem ser executadas no que se referem às alterações quanto à forma de escolher o ponto de aceleração. Até o presente momento, a única abordagem testada foi selecionar as regras com maior suporte. As outras abordagens poderiam escolher, por exemplo, o ponto: *(i)* que resultasse em maior deslocamento, *(ii)* que resultasse em menor consumo e *(iii)* o mais votado dentre três ou mais classificadores.

A utilização de outros algoritmos de mineração de dados também pode ser explorada como possível extensão do trabalho no intuito de suprir as deficiências dos algoritmos usados. Uma alternativa seria a mudança no módulo interpretador de regras para que este operasse com regras geradas pelo software Clementine®, disponível no laboratório de pesquisa. Este software gera um conjunto de regras utilizando o algoritmo See5⁴, uma versão aprimorada do algoritmo C45. Acredita-se poder, desta forma, melhorar a taxa de acerto dos classificadores e conseqüentemente a taxa de similaridade de condução.

⁴ Maiores informações em <http://www.rulequest.com/>

Referências Bibliográficas

- ALL. 2007.** ALL Anuncia Resultados de 2007. [Online] 2007. [Citado em: 24 de 04 de 2008.] Disponível em: http://www.mzweb.com.br/all/web/arquivos/ALL_Resultados_4T07_port.pdf.
- . **2008.** *Procedimentos na condução de trens*. Curitiba : AMÉRICA LATINA LOGÍSTICA DO BRASIL S.A., 2008.
- ANTT, AGÊNCIA NACIONAL DE TRANSPORTES TERRESTRES. 2008.** Relatório Anual de Acompanhamento das Concessões Ferroviárias. [Online] 2008. [Citado em: 24 de 04 de 2008.] <http://www.antt.gov.br/relatorios/ferroviario/concessionarias2006/index.asp>.
- AVALLONE, E. A. e BAUMEISTER, T. 1996.** *Marks' Standard Handbook for Mechanical Engineers*. s.l. : McGraw-Hill Professional, 1996. p. 1792.
- BAUER, ERIC e KOHAVI, RON. 1999.** An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants. *Machine Learning*, 36(1/2). 1999, pp. 105-139.
- BENENSON, R., et al. 2008.** Towards urban driverless vehicles. *International Journal of Vehicle Autonomous Systems*. 2008, Vol. 6, pp. 4-23.
- BERTOLAZZI, E., et al. 2008.** Development of a reduced size unmanned car. *Advanced Motion Control AMC 10th IEEE International Workshop on AMC*. 2008, Vols. 26-28, pp. 763-770.
- BRAGA, A.C. 2000.** Curvas ROC: Aspectos Funcionais e Aplicações. s.l. : Universidade do Minho, 2000. Tese (Doutorado em Engenharia de Produção e Sistemas).
- BREIMAN, L. 1996.** Bagging predictors. *Machine learning*, 24(2). 1996, pp. 123-140.
- BREIMAN, L., et al. 1984.** *Classification and Regression Trees*. 1ª Edição. s.l. : Chapman & Hall/CRC, 1984. p. 368. 0412048418.
- BRINA, HELVÉCIO LAPERTOSA. 1982.** *Estradas de ferro*. Rio de Janeiro : LTC, 1982. Vol. 2.
- CALLEGARI, J. e SIDIA, M. 2003.** *Bioestatística: Princípios e Aplicações*. Porto Alegre : Artmed, 2003.
- CHEN, L., WRIGHT, P. e NEJDL, W. 2009.** Improving music genre classification using collaborative tagging data. *Proceedings of the Second ACM International Conference on Web Search and Data Mining*. 2009, pp. 84-93.
- CIOS, K. J., et al. 2007.** *Data Mining: A Knowledge Discovery Approach*. s.l. : Springer, 2007. p. 606. ISBN: 978-0-387-33333-5.
- COHEN, W. W. 1995.** Fast Effective Rule Induction. *Twelfth International Conference on Machine Learning*. 1995, 12, pp. 115-123.
- DEMSAR, J. 2006.** Statistical Comparisons of Classifiers over Multiple Data Sets. *Journal of Machine Learning Research*. 7, 2006, 7, pp. 1-30.

- DIAMANTIDIS, N.A., KARLIS, D. e GIANKOUMAKIS, E.A. 2000.** Unsupervised stratification of cross-validation for accuracy estimation. *Artificial Intelligence*, 116. 2000, pp. 1-16.
- DOWNING, D. e J., CLARK. 1998.** *Estatística Aplicada*. 1 ed. São Paulo : Saraiva, 1998.
- FAWCETT, T. e PROVOST, F. 2006.** An introduction to ROC analysis. *Pattern Recognition Letters*. 2006, Vol. 227, 8, pp. 861-874.
- FAWCETT, T. 2004.** *ROC Graphs: Notes and Practical Considerations for Researches*. Palo Alto : s.n., 2004. p. 38.
- FAYYAD, U. M. 1996.** Data mining and knowledge discovery: making sense out of data. *IEEE Expert*. 5, 1996, Vol. 11, pp. 20-25.
- FAYYAD, U., PIATETSKI-SHAPIRO, G. e PADHRAIC, P. 1996.** The KDD Process for Extracting Useful Knowledge from Volumes of Data. *Communications of the ACM*. 1996, pp. 27-34.
- FEYYAD, U. M. 1996.** Data mining and knowledge discovery: making sense out of data. *IEEE Expert*. 5, 1996, Vol. 11, pp. 20-25.
- FREUND, Y. e SCHAPIRE, R.E. 1996.** Experiments with a new boosting algorithm. *Proceedings of the 13th International Conference on Machine Learning*. 1996, pp. 148-156.
- GENG, X., LIU, T. e LI, H. 2007.** Feature Selection for Ranking. *Proceedings of 30th Annual International ACM SIGIR Conference*. 2007, 30th.
- GU, BAOHUA, HU, FEIFANG e LIU, HUAN. 2000.** *Sampling and Its Application in Data Mining: A Survey*. Cingapura : s.n., 2000. p. 33. Disponível em: <http://hdl.handle.net/1900.100/1408>. Acesso em: 20 de abril de 2008..
- HALL, M. A. 2000.** Correlation-based feature selection for discrete and numeric class machine learning. *Proc. of the 17th Int. Conf. on Machine Learning*. 2000, pp. 359-366.
- . **1998.** *Correlation-based Feature Selection for Machine Learning*. Hamilton : Department of Computer Science, University of Waikato, 1998. Ph.D. thesis.
- HALL, M. A. e HOLMES, G. 2003.** Benchmarking Attribute Selection Techniques for Discrete Class Data Mining. *IEEE Transactions on Knowledge and Data Engineering*. 2003, Vol. 15, 3.
- HAN, JIAWEI e KAMBER, MICHELINE. 2006.** *Data Mining: Concepts and Techniques*. Second Edition. San Francisco, CA : Morgan Kaufmann, 2006. p. 772.
- HAND, DAVID J. e TILL, ROBERT J. 2001.** A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. *Machine Learning*. 2001, pp. 171-186.
- HAWKINS, D. M. 1980.** *Identification of outliers*. London : Chapman & Hall, 1980. p. 188.
- HUAN, LIU e MOTODA, HIROSHI. 1998.** *Feature Selection for Knowledge Discovery and Data Mining*. s.l. : Kluwer Academic Publishers, 1998.
- HUSSAIN, FARHAD, et al. 1999.** Discretization: An Enabling Technique. *The National University of Singapore. Technical Report*. Junho de 1999, pp. 1-29.

- INMON, W. H. 2002.** *Building the Data Warehouse*. 3 ed. Toronto : Willey Computer Publishing, 2002.
- ISAAC, ANDREW e SAMMUT, CLAUDE. 2003.** Goal-directed Learning to Fly. *Proceedings of the 20th International Conference on Machine Learning*. 2003, pp. 258-265.
- KALOS, A. e REY, T. 2005.** Data mining in the chemical industry. *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*. 2005, pp. 763-769.
- KOHAVI, R. e JOHN, G. 1996.** Wrappers for Feature Subset Selection. *AIK special issue on relevance*. 1996.
- KOLLER, D. e SAHAMI, M. 1996.** Toward optimal feature selection. *Proc. of the 13th Int. Conf. on Machine Learning*. 1996, pp. 284-292.
- KOLSKI, S., et al. 2006.** Autonomous driving in structured and unstructured environments. *IEEE Intelligent Vehicles Symposium*. 2006, pp. 558-563.
- LEE, H. D., MONARD, M. C. e BARANAUSKAS, J. A. 1999.** *Empirical comparison of wrapper and filter approaches for feature subset selection*. São Carlos : ICMC-USP, 1999. Disponível em: ftp://ftp.icmc.usp.br/pub/BIBLIOTECA/rel_tec/RT_094.pdf.
- LIU, HUAN e YU, LEI. 2005.** Toward Integration Feature Selection Algorithms for Classification and Clustering. *IEEE Transactions on Knowledge and Data Engineering*. 2005, pp. 491-502.
- LOPES, LUCELENE. 2007.** *Aprendizagem de máquina baseada na combinação de classificadores em bases de dados da área de saúde*. Curitiba : s.n., 2007. Dissertação – Mestrado – Tecnologia em Saúde – Pontifícia Universidade Católica do Paraná – Centro de Ciências Biológicas e da Saúde.
- MACHADO, FERNANDO e ABREU, MAURICIO. 2004.** *Projeto de Banco de Dados: Uma visão prática*. 11 ed. São Paulo : Érica, 2004.
- MALLOF, A. MARCUS e MICHALSKI, S. RYSZARD. 2000.** Selecting Examples for Partial Memory Learning. *Machine Learning Journal*. 2000, Vol. 41, pp. 27-52.
- MALOOF, MARCUS A. e MICHALSKI, RYSZARD S. 2000.** *Selecting Examples for Partial Memory Learning*. s.l. : Machine Learning Journal, 2000. pp. 27-52. Vol. 41, citeseer.ist.psu.edu/maloof00selecting.html.
- MITCHELL, T. 1997.** *Machine Learning*. New York : McGraw-Hill, 1997.
- MRS, MRS LOGÍSTICA. 2007.** Relatório da Administração. [Online] 2007. [Citado em: 24 de 04 de 2008.] http://www.mrs.com.br/download/Relatorio_Anuual_2007.pdf.
- NIEVOLA, J. C. 2008.** Classificação. [Online] 2008. [Citado em: 09 de 01 de 2008.] <http://www.ppgia.pucpr.br/~nievola/Treinamento/MD/MD-04-Classificacao.pdf>.
- NILSSON, N.J. 1996.** *Introduction to machine learning*. Stanford : Stanford university, 1996.
- OCKHAM, W. 1999.** *Prólogo da Exposição dos Oitos Livros da Física*. São Paulo : Nova Cultural, 1999.

- PASQUIER, M. e OENTARYO, R. 2008.** Learning to frive the human way: a step towards intelligent vehicles. *International Journal Vehicle Autonomous Systems*. 2008, Vol. 6.
- PERLICH, C., PROVOST, F. e SIMONOFF, J. 2003.** Tree Induction vs. Logistic Regression: A Learning-Curve Analysis. *Journal of Machine Learning Research*. 4, 2003, pp. 211-255.
- PERLICH, CLAUDIA, PROVOST, FOSTER e SIMONOFF, JEFFREY S. 2003.** Tree Induction vs. Logistic Regression: A Learning-Curve Analysis. *Journal of Machine Learning Research* 4. 2003, pp. 211-255.
- PINTO, B. G.M., et al.** Controle Automático de Potência em Locomotivas em Tração Múltipla usando Microprocessador.
- PIRES, C. L., NABETA, S. I. e CARDOSO, J. R. 2005.** Simulação de composição ferroviária acionada por motores de indução e inversores de tensão. *Revista Controle & Automação*. 2005, Vol. 16, 1.
- PRATI, R. C. 2006.** *Novas abordagens em aprendizado de máquina para a geração de regras, classes desbalanceadas e ordenação de casos*. São Carlos : s.n., 2006. p. 191. Tese de doutorado.
- PROVOST, FOSTER e FAWCETT, TOM. 1997.** *Analysis and Visualization of Classifier Performance: Comparison under Imprecise Class and Cost Distributions*. Huntington Beach : American Association for Artificial Intelligence, 1997. pp. 43-48.
- QUINLAN, J. R. 1993.** *C4.5: Programs for machine learning*. San Francisco : Morgan Kaufman, 1993.
- . **1987.** *Generation Production Rules from Decision Trees*. s.l. : In Proc. of IJCAI 87, 1987. pp. 304-307.
- . **1996.** Improved Use of Continuous Attributes in C4.5. *Journal of Artificial Intelligence Research*. 1996, Vol. IV, pp. 77-90.
- RACHEL, F. M. 2006.** *Proposta de um controlador automático de trens utilizando lógica nebulosa preditiva*. São Paulo : s.n., 2006. Dissertação apresentada à Escola Politécnica da Universidade de São Paulo para obtenção do título de Mestre em Engenharia..
- RODDICK, J., FULE, P. e GRACO, W. 2003.** Exploratory medical knowledge discovery: experiences and issues. *ACM SIGKDD Explorations Newsletter*. 1, 2003, Vol. 5, pp. 94-99.
- SILLA, C. N., KAESTNER, C. e KOERICH, A. 2005.** Classificação Automática de Gêneros Musicais Utilizando Métodos de Bagging e Boosting. *SBCM - Simpósio Brasileiro de Computação Musical*. Outubro de 2005. Disponível em: <http://gsd.ime.usp.br/sbcm/2005/papers/tech-12438.html>.
- SILVA, F. C. 2008.** Análise ROC. [Online] 2008. [Citado em: 15 de 01 de 2008.] http://www.dpi.inpe.br/~felipe/works/inpe/spr/roc_analyzes.pdf.
- SPIEGEL, M. R. 1974.** *Estatística*. 7 ed. São Paulo : McGraw-Hill do Brasil, 1974.
- STOLZER, ALAN J e HALFORD, CARL. 2007.** Data Mining applied to flight operations quality assurance data: A comparison to standart statistical methods. [Online] 2007. [Citado em: 13 de 03 de 2008.] http://findarticles.com/p/articles/mi_qa5467/is_200701/ai_n21291012.

- SWOKOWSKI, E. 1983.** *Cálculo com Geometria Analítica*. São Paulo : McGraw-Hill, 1983. Vol. II.
- TAN, M., STEINBACH, V. e KUMAR, A.W. 2006.** *Introduction to Data Mining*. Minnesota : Addison Wesley, 2006.
- THOMSEN, E. 2002.** *OLAP: Construindo sistemas de informações multidimensionais*. 2 ed. Rio de Janeiro : Campus, 2002.
- UCI. 2008.** UCI Machine Learning Repository. [Online] 2008. [Citado em: 30 de 03 de 2008.] <http://archive.ics.uci.edu/ml/>.
- WATKINS, C. J.C.H. e DAYAN, P. 1992.** Q-Learning. *Machine Learning*. 1992, Vol. VIII, pp. 279-292.
- WEKA. 2008.** Data Mining with Open Source Machine Learning Software in JAVA. *WEKA*. [Online] 2008. <http://www.cs.waikato.ac.nz/ml/weka/> Acesso em: 15/01/2008.
- WINKLER, S.M., AFFENZELLER, M. e WAGNER, S. 2006.** Sets of receiver operating characteristic curves and their use in the evaluation of multi-class classification. *Proceedings of the 8th annual conference on Genetic and evolutionary computation*. 2006, pp. 1601 – 1602.
- WITTEN, I.H e FRANK, E. 2005.** *Data Mining: Practical machine learning tools and techniques*. 2 ed. San Francisco : Morgan Kaufmann, 2005.
- WREMBEL, R. e C., KONCILIA. 2007.** *Data Warehouses and OLAP: Concepts, Architectures and Solutions*. 1 ed. Singapura : IRW Press, 2007.