

PRISCILA LOUISE LEYSER SANTIN

**ANÁLISE AUTOMÁTICA DE TEXTOS DE
MENSAGENS INSTANTÂNEAS PARA
DETECÇÃO DE ALICIAMENTO SEXUAL DE
CRIANÇAS E ADOLESCENTES**

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

CURITIBA

2013

PRISCILA LOUISE LEYSER SANTIN

**ANÁLISE AUTOMÁTICA DE TEXTOS DE
MENSAGENS INSTANTÂNEAS PARA
DETECÇÃO DE ALICIAMENTO SEXUAL DE
CRIANÇAS E ADOLESCENTES**

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

Área de Concentração: *Ciência da Computação*
Linha de Pesquisa: *Visão Computacional, Reconhecimento de Padrões e Computação Forense*

Orientadora: Prof. Dra. Cinthia O. de A. Freitas

Coorientador: Prof. Dr. Emerson Cabrera Paraiso

CURITIBA

2013

Dados da Catalogação na Publicação
Pontifícia Universidade Católica do Paraná
Sistema Integrado de Bibliotecas – SIBI/PUCPR
Biblioteca Central

| | |
|---------------|---|
| S235a 2013 | <p>Santin, Priscila Louise Leyser</p> <p>Análise automática de textos de mensagens instantâneas para detecção de aliciamento sexual de crianças e adolescentes / Priscila Louise Leyser Santin ; orientadora, Cinthia O. de A. Freitas ; coorientador, Emerson Cabrera Paraiso. – 2013.</p> <p>xiv, 73 [17] f. : il. ; 30 cm</p> <p>Dissertação (mestrado) – Pontifícia Universidade Católica do Paraná, Curitiba, 2013</p> <p>Bibliografia: f. 69-73</p> <p>1. Informática. 2. Crime contra a criança. 3. Crime sexual. 4. Mensagens instantâneas. 5. Aprendizado do computador. I. Freitas, Cinthia Obladen de Almendra. II. Paraiso, Emerson Cabrera. III. Pontifícia Universidade Católica do Paraná. Programa de Pós-Graduação em Informática Aplicada. IV. Título.</p> <p>CDD 20. ed. – 004</p> |
|---------------|---|

*Dedico essa dissertação à família
que construí; meu amigo e marido Altair e
a meus filhos Lucas e Marcelo.*

"A vida não deveria tirar alguém da infância sem antes conseguir para essa pessoa um bom lugar na juventude".
Mafalda (Quino)

Agradecimentos

Se dissesse que este trabalho foi desenvolvido somente por mim estaria mentindo. Este trabalho contou com a colaboração de muitas pessoas, algumas com mais, outras com menos intensidade. Vou citar aqui as pessoas das quais minha memória se lembra no momento, mas sem querer ser injusta com ninguém, desde já agradeço a todos que direta ou indiretamente contribuíram para que este trabalho fosse possível.

Em primeiro lugar meu especial agradecimento à minha orientadora professora Cinthia e ao meu co-orientador professor Emerson pelo apoio, compreensão e orientação do trabalho. Sem dúvidas estes professores contribuíram de maneira definitiva nos momentos mais críticos para que essa dissertação apresentasse o ótimo resultado que obtivemos.

Tenho especial gratidão à Rede Marista de Solidariedade pelo apoio financeiro ao projeto, e em especial a Sra. Jimena, que sempre deu apoio às necessidades do projeto.

Um especial agradecimento aos alunos de PIBIC Caio, Eloara, Fernanda, Lohann, Paola e Tiago pelos estudos realizados e pelas implementações envolvendo as ferramentas relacionadas a este trabalho, pelas várias discussões e pelas sugestões que fizeram ao trabalho.

A Mariza, Andréia, Edenilson, Franciele, Irapuru e demais colegas do PPGIa agradeço pela oportunidade de poder trocar idéias e enriquecer o meu conhecimento e, conseqüentemente, melhorar os resultados alcançados no trabalho. Não poderia deixar de menciona a Cheila, secretária do PPGIa, que sempre me auxiliou nas questões burocráticas e se tornou uma grande amiga.

Ao Altair, meu marido, agradeço por ter me proporcionado a oportunidade de desenvolver este Mestrado. Com certeza se não fosse por ele eu hoje não teria chego até aqui. Obrigada também pelas inúmeras sugestões, revisões do texto, apoio e compreensão, mesmo nas horas em que eu estava mais apreensiva com as atividades do Mestrado.

A meus filhos, Lucas e Marcelo, minhas motivações a fazer o Mestrado. Foi por eles e para eles que larguei meu emprego para me dedicar aos estudos novamente após dez anos de formada.

A meus pais, a quem devo minha vida, e que sempre acreditaram em mim. Sou grata por tudo o que fizeram por mim.

Por fim agradeço a Deus por guiar minha vida, pela saúde e disposição que tenho, e pela oportunidade de poder a cada dia de minha vida ir construindo o meu futuro.

Sumário

| | |
|---|------|
| Agradecimentos | v |
| Sumário..... | vii |
| Lista de Figuras | x |
| Lista de Tabelas | xi |
| Lista de Abreviaturas..... | xii |
| Resumo | xiii |
| Abstract..... | xiv |
| Capítulo 1 | 1 |
| Introdução..... | 1 |
| 1.1. Motivação | 3 |
| 1.2. Hipóteses do Trabalho | 3 |
| 1.3. Objetivo Geral..... | 4 |
| 1.4. Objetivos Específicos | 4 |
| 1.5. Contribuições | 4 |
| 1.6. Estrutura do Documento | 6 |
| Capítulo 2 | 7 |
| Fundamentação Teórica..... | 7 |
| 2.1. Pressupostos Conceituais | 7 |
| 2.2. Aliciamento de Menores (Crianças e Adolescentes) | 8 |
| 2.2.1. O que é Aliciamento? | 8 |
| 2.2.2. O Aliciamento e suas Consequências | 9 |

| | | |
|------------------------|--|----|
| 2.2.3. | Estágios do Aliciamento..... | 9 |
| 2.2.4. | Estágios do aliciamento no contexto computacional..... | 11 |
| 2.3. | Abordagens para Detecção de Aliciamento Sexual..... | 13 |
| 2.3.1. | Baseadas em filtro de palavras..... | 13 |
| 2.3.2. | Baseadas em regras..... | 14 |
| 2.3.3. | Baseadas em classificadores..... | 14 |
| 2.4. | Análise Textual..... | 14 |
| 2.4.1. | Algoritmos de Aprendizagem de Máquina..... | 14 |
| 2.5. | Processos Estocásticos..... | 22 |
| 2.5.1. | Cadeias de Markov..... | 23 |
| 2.5.2. | Modelos Escondidos de Markov..... | 25 |
| 2.6. | Avaliação de Classificadores..... | 30 |
| 2.7. | Considerações Finais..... | 32 |
| Capítulo 3 | | 33 |
| Trabalhos Relacionados | | 33 |
| 3.1. | Detecção dos estágios baseada em filtro de palavras..... | 33 |
| 3.2. | Detecção dos estágios baseada em regras..... | 34 |
| 3.3. | Detecção dos estágios baseada em classificadores..... | 36 |
| 3.4. | Detecção de conversas com aliciamento..... | 39 |
| 3.5. | Detecção do perfil do agressor..... | 44 |
| 3.6. | Distinções entre agressor e vítima..... | 44 |
| 3.7. | Considerações Finais..... | 45 |
| Capítulo 4 | | 47 |
| Método Proposto | | 47 |
| 4.1. | Transcrição de Conversas: Base de Dados..... | 50 |
| 4.1.1. | Análise das Transcrições de Conversas..... | 51 |

| | | |
|------|--|----|
| 4.2. | Preparação da Base de Conversação..... | 52 |
| 4.3. | Extração do Léxico do Agressor..... | 53 |
| 4.4. | Pré-Processamento e Extração de Características..... | 54 |
| 4.5. | Classificação dos Estágios utilizando o <i>Latent Semantic Analysis</i> (LSA) | 56 |
| 4.6. | Classificação dos Estágios utilizando o <i>Support Vector Machine</i> (SVM) | 56 |
| 4.7. | Identificação das Sequências de Observações com o <i>Hidden Markov Model</i> (HMM) 57 | |
| 4.8. | Considerações finais | 57 |
| | Capítulo 5 | 58 |
| | Experimentos Realizados e Análise dos Resultados | 58 |
| 5.1. | Identificando os Estágios de Comunicação com o <i>Latent Semantic Analysis</i> (LSA) | 58 |
| 5.2. | Identificando os Estágios de Comunicação com o <i>Support Vector Machine</i> (SVM) | 60 |
| 5.3. | Criando o Modelo de Aliciamento com o <i>Hidden Markov Model</i> (HMM)..... | 64 |
| 5.4. | Estabelecendo os Níveis de Susceptibilidade ao Abuso Sexual de Crianças e Adolescentes..... | 65 |
| 5.5. | Considerações Finais | 66 |
| | Capítulo 6 | 67 |
| | Conclusão | 67 |
| | Referências | 69 |
| | Anexo I..... | 74 |

Lista de Figuras

| | |
|--|----|
| Figura 2.1 - Adaptação do Modelo de Comunicação [OLS07]..... | 10 |
| Figura 2.2 - Panorama do Protótipo para Detecção de Assédio Moral em e-mails [NUN09] . | 13 |
| Figura 2.3 - Hierarquia de Aprendizado [FAC11] | 16 |
| Figura 2.4 - Adaptação do Hiperplano Ótimo [MAN08] | 17 |
| Figura 2.5 - Conjunto de Dados não Lineares. Adaptado de [LOR03]..... | 17 |
| Figura 2.6 - Fronteira Linear no Espaço de Características. Adaptado de [MUL01]..... | 18 |
| Figura 2.7 - Diagrama de transição da cadeia de Markov | 24 |
| Figura 2.8 - Matriz de transição de uma cadeia de Markov com 4 estados..... | 24 |
| Figura 2.9 - Exemplo de parâmetros probabilísticos de um MEM: s – estados; v – símbolos de observações; a – probabilidade de transição entre estados; b – probabilidade de observação dos símbolos | 26 |
| Figura 2.10 - Modelo Ergótico de MEM: a) representação gráfica; b) representação matricial | 27 |
| Figura 2.11 - Modelo Sequencial de MEM: a) representação gráfica; b) representação matricial..... | 28 |
| Figura 2.12 - Modelo Paralelo de MEM | 28 |
| Figura 4.1 - Sequenciamento para Aplicação do Método Proposto | 48 |
| Figura 4.2 – Método estudado utilizando o LSA..... | 49 |
| Figura 4.3 – Método desenvolvido para a fase de treinamento utilizando o SVM e HMM | 50 |
| Figura 4.4 - Visão geral da página inicial do <i>site www.perverted-justice.com</i> [PER03] | 51 |
| Figura 4.5 – Fragmento de transcrições das conversas de aliciamento [PER03] | 53 |
| Figura 4.6 - Separação dos enunciados do agressor e da vítima | 54 |
| Figura 5.1 - Semáforo representando os níveis de susceptibilidade ao abuso sexual de crianças e adolescentes | 66 |

Lista de Tabelas

| | |
|--|----|
| Tabela 3.1 - Comparativo de trabalhos utilizando SVM..... | 46 |
| Tabela 4.1 - Representação numérica para os estágios de aliciamento | 52 |
| Tabela 4.2 - Distribuição do léxico de palavras do agressor por estágio de comunicação | 54 |
| Tabela 5.1 – Exemplos de enunciados..... | 58 |
| Tabela 5.2 – Matriz de saída do LSA com os estágios de aliciamento encontrados nos enunciados | 59 |
| Tabela 5.3 - Matriz de Confusão em <i>Cross-validation</i> para o 1º Classificador | 61 |
| Tabela 5.4 - Matriz de Confusão para a Base de Teste do 1º Classificador | 62 |
| Tabela 5.5 - Matriz de Confusão em <i>Cross-validation</i> para o 2º Classificador | 62 |
| Tabela 5.6 - Matriz de Confusão para a Base de Teste do 2º Classificador | 63 |
| Tabela 5.7 - Resultados do trabalho de MCGHEE et al. [MCG11] no uso de classificadores para a identificação dos estágios de comunicação | 63 |
| Tabela 5.8 - Taxa de Acerto do HMM para as Sequências de Teste..... | 64 |
| Tabela 5.9 - Taxa de Acerto do HMM para as Sequências de Teste geradas pelo SVM..... | 65 |

Lista de Abreviaturas

| | |
|--------|---|
| APC | Associação Paranaense de Cultura |
| AM | Aprendizado de Máquina |
| EUA | Estados Unidos da América |
| GI | Ganho de Informação |
| HMM | <i>Hidden Markov Model</i> |
| IA | Inteligência Artificial |
| INPI | Instituto Nacional de Propriedade Industrial |
| LIWC | <i>Linguistic Inquiry and Word Count</i> |
| LSA | <i>Latent Semantic Analysis</i> |
| LSI | <i>Latent Semantic Indexing</i> |
| MEM | Modelos Escondidos de Markov |
| MPF | Ministério Público Federal |
| NCMEC | <i>National Center for Missing and Exploited Children</i> |
| NEO | Nenhum Estágio Observado |
| ONG | Organização Não Governamental |
| PIBIC | Programa Institucional de Bolsas de Iniciação Científica |
| PUCPR | Pontifícia Universidade Católica do Paraná |
| RBF | <i>Radial-Basis Function</i> |
| SCI | <i>Suspicious Conversations Identification</i> |
| SVD | <i>Single Value Decomposition</i> |
| SVM | <i>Support Vector Machine</i> |
| TF-IDF | <i>Term Frequency – Inverse Document Frequency</i> |
| VFP | <i>Victim From Predator disclosure</i> |

Resumo

A detecção do aliciamento sexual de crianças e adolescentes em salas de bate-papo *online* pode prevenir que a vítima sofra abuso sexual. As abordagens existentes na literatura ainda apresentam baixa taxa de detecção, além de algumas terem por base técnicas que incorporam na lógica da detecção do aliciamento, características específicas do idioma da conversação. A principal razão identificada para a baixa eficiência destas abordagens é o emprego de um modelo único para agressor e vítima e, também, da dificuldade de encontrar características que definam precisamente o aliciamento. Este trabalho utiliza aprendizado de máquina para identificação dos estágios de aliciamento que compõem a evolução de uma conversa em salas de bate-papo *online* a partir de mensagens instantâneas. Foi utilizado um classificador SVM (*Support Vector Machine*) para realizar a identificação dos estágios de comunicação obtendo-se acurácia similar a trabalhos relacionados. Porém, como se trabalhou apenas com o modelo do agressor e com o SVM se herda a vantagem de melhor desempenho em aplicações de tempo real como salas de bate-papo. Além disso, inovou-se no uso do HMM (*Hidden Markov Model*), com acurácia de 97%, para emissão de alertas de susceptibilidade ao abuso sexual de crianças e adolescentes antes mesmo que estes ocorram.

Palavras-Chave: Aliciamento Sexual *Online* de Crianças e Adolescentes, Salas de Bate-papo *Online*, Criança e Adolescente, Aprendizado de Máquina, Mensagens Instantâneas.

Abstract

The detection of sexual grooming of minors in chat rooms online can prevent that the victim suffers sexual abuse. The existing approaches in the literature still exhibit low detection rate, in addition to being based on some techniques that incorporate the logic of detection of grooming, specific features of the language of conversation. The main reason identified for the low efficiency of these approaches is the use of a single model for offender and victim, and also of the difficulty to find characteristics that define precisely the grooming. This work uses machine learning to identify the stages of grooming that compose the evolution of a conversation in chat rooms online from instant messaging. It was used a classifier SVM (Support Vector Machine) to perform the identification of stages of communication obtaining comparable results to the literature. However, the proposed method only considers the model of the aggressor associated with the SVM it inherits the advantage of better performance in real-time applications such as chat rooms. In addition, the method considers an innovation based on HMM (Hidden Markov Model), achieving an accuracy of 97 %, for issuing alerts for susceptibility to sexual abuse of minors even before they occur.

Keywords: Grooming, Enticement, Chat Rooms, Child and Teenager, Instant Messages, Machine Learning.

Capítulo 1

Introdução

A *Internet* oferece inúmeros recursos para interação entre indivíduos das mais variadas faixas etárias. Crianças (pessoas de até 12 anos de idade incompletos) e adolescentes (pessoas entre 12 e 18 anos) tomam contato com a rede mundial de computadores ainda nos seus primeiros anos de vida. Assim, a partir do uso rotineiro da *Internet*, crianças e adolescentes vão assimilando os diferentes tipos de recursos à medida que passam pelas diferentes fases de seu crescimento. Deste modo, a interatividade fornecida por recursos como salas de bate-papo (*chats*, recurso para troca de mensagens instantâneas *online*) traz para as crianças e adolescentes uma espécie de local onde eles encontram todas as respostas para as suas dúvidas e curiosidades, porém, em sua visão, sem se expor a “censura” tradicional dos familiares.

Neste contexto, indivíduos mal intencionados, conhecendo esta tendência comportamental das crianças e adolescentes, usam as salas de bate-papo para o aliciamento (*grooming/enticement*) com o intuito de praticar o abuso sexual de menores (crianças e adolescentes) [RAS08].

Nos Estados Unidos da América (EUA), o Centro Nacional para Crianças Desaparecidas e Exploradas (*National Center for Missing and Exploited Children – NCMEC*) recebe informações de casos de exploração sexual de crianças e adolescentes, incluindo o aliciamento *online*. Nesta categoria de aliciamento *online*, desde 9 de março de 1998 até o final de 2012, já haviam sido informados 59.544 casos ao NCMEC, sendo que em 2012 a média foi de 97 casos por semana [NCM84].

No *site* sobre a Campanha Nacional de Combate à Pedofilia na *Internet*¹ encontra-se que: “Também constam dados de uma pesquisa realizada nos EUA, dizendo que de cada 5 crianças que navegam na *Internet*, uma recebeu proposta de um pedófilo, e uma a cada 33 já se comunicou, por meio de telefone e recebeu dinheiro ou passagem para se encontrar com um criminoso”. O *site* ainda alerta: “Pais e filhos, inconscientes dos perigos da rede são presas fáceis de pedófilos. Uma criança ingenuamente não identifica um adulto se passando por um amiguinho da mesma idade”.

Em 2006, o Brasil já ocupava o terceiro lugar entre os países do mundo em número de denúncias de crimes *online* que ferem os direitos humanos, tais como: racismo, intolerância religiosa e pornografia infantil. Nesta época, tinham ocorrido em 20 dias um total de 2,25 mil denúncias, como declarado por Thiago Tavares, presidente da SaferNet, uma organização não governamental (ONG) que luta contra crimes virtuais, durante a cerimônia de assinatura do acordo entre a entidade e o Ministério Público Federal (MPF) no dia 29 de março de 2006 em São Paulo². Em 2012, a violência de abuso sexual de menores (crianças e adolescentes), segundo a SaferNet³, representou 24% dos casos reportados.

Diante desta problemática, pesquisadores têm estudado e desenvolvido técnicas para analisar conversas em salas de bate-papo [KON09a] [HOW04] [KON09b] [MCG11], de modo a auxiliar na descoberta dos perfis de agressores e no estabelecimento e identificação dos estágios da conversação que estes agressores utilizam no contato com crianças e adolescentes. A definição de tais perfis tem o intuito de contribuir para detectar a agressividade do aliciamento sexual colaborando na previsão da susceptibilidade ao abuso sexual antes mesmo que este ocorra.

Assim, o presente trabalho apresenta e descreve detalhadamente o método proposto para classificação de textos visando à detecção do aliciamento sexual de crianças e adolescentes em conversações realizadas por meio de salas de bate-papo na *Internet*.

¹ <http://www.censura.com.br/>

² Disponível em: <http://www.safernet.org.br/site/noticias/brasil-ocupa-3%C2%BA-posi%C3%A7%C3%A3o-ciber-crimes-contra-os-direitos-humanos>. Acesso em 03 de junho de 2013. Esse acordo foi reiniciado em 24/11/2010, <http://g1.globo.com/tecnologia/noticia/2010/11/safernet-divulga-nota-sobre-rescisao-de-acordo-com-ministerio-publico.html>

³ <http://www.safernet.org.br/site/indicadores>

1.1. Motivação

Inicialmente, pode-se citar uma motivação de impacto social deste projeto: contribuir para que a troca de mensagens por meio de sala de bate-papo se torne um ambiente virtual mais seguro para crianças e adolescentes.

Como motivação técnico-científica, pode-se apontar resumidamente as seguintes considerações:

- As técnicas de detecção de aliciamento sexual da literatura, baseadas em análise dos elementos gramaticais de uma frase são ineficientes [KON09b] porque os diálogos em salas de bate-papo são muito próximos da linguagem coloquial, a qual não segue o rigor gramatical da escrita;
- As técnicas de detecção baseadas em regras [KON09b] são dependentes da linguagem coloquial analisada e dos vícios de linguagem de seus interlocutores;
- As técnicas baseadas em classificadores sofrem com problemas de seleção das características, palavras ou frases, de modo a formar um léxico válido e representativo do problema em questão. Além disto, uma mesma palavra ou frase pode ser encontrada em diferentes estágios (fase) do processo de aliciamento, dificultando assim a classificação ou identificação do problema [MCG11].

Levando-se em consideração o exposto, tem-se que as abordagens apresentadas na literatura consideram modelos de detecção nos quais as características (léxicos de palavras e frases) encontram-se embutidas nos próprios modelos, fato que dificulta a portabilidade das técnicas desenvolvidas para outros idiomas, tornando a solução dependente do conjunto de dados que está sendo avaliado.

1.2. Hipóteses do Trabalho

As hipóteses deste trabalho são:

- Processar os enunciados (do inglês *utterance*, linhas da transcrição da conversa da sala de bate-papo) do agressor e da vítima separadamente pode remover os estados de indecisão do classificador na detecção do aliciamento;
- Processar os enunciados da vítima e do agressor em separado pode levar a uma melhora na taxa de acerto de técnicas de análise automática de textos;

- Trabalhar com os enunciados usando uma técnica de *aprendizagem de máquina* (AM) é mais eficiente que aplicar uma técnica para tratamento de linguagem natural estatística como o *Latent Semantic Analysis* (LSA) e, finalmente,
- É possível a definição de um método independente do idioma.

1.3. Objetivo Geral

Este trabalho teve como objetivo estabelecer um método de detecção de aliciamento sexual de crianças e adolescentes em sala de bate-papo (*chat*), baseado na identificação dos estágios do diálogo entre um agressor e sua possível vítima, considerando uma abordagem de aprendizado de máquina para a referida identificação. O trabalho utiliza tanto os enunciados do agressor quanto da vítima, visando detectar a ocorrência do sincronismo (*match*) entre os estágios de cada uma das partes com intuito de melhorar a precisão da detecção, usando uma técnica de avaliação estocástica de modo a quantificar de maneira objetiva a exposição (susceptibilidade) das vítimas ao risco do abuso sexual.

1.4. Objetivos Específicos

O trabalho se desdobra nos seguintes objetivos específicos:

- Comparar o método de tratamento de linguagem natural, no caso, *Latent Semantic Analysis* (LSA), com o método de classificação, no caso, *Support Vector Machine* (SVM);
- Aplicar os resultados obtidos com o processamento dos enunciados de modo a estabelecer o nível de susceptibilidade ao abuso sexual de crianças e adolescentes usando o método estocástico de Modelos Escondidos de Markov (*Hidden Markov Models* – HMM);
- Implementar um cenário de teste como prova de conceito para avaliar e validar o método proposto.

1.5. Contribuições

Reconhece-se que a contribuição social deste trabalho é de suma importância, sendo que de um modo geral as contribuições vão além de um método de detecção de aliciamento de crianças e adolescentes, visto que o presente trabalho reveste-se de caráter inovador diante da

não existência no Brasil de nenhum método, mecanismo ou *software* que se proponha enfrentar o problema em questão.

Assim, deve-se mencionar, inicialmente, que este trabalho está diretamente ligado ao projeto intitulado: “Detecção de Assédio Sexual e Pedofilia na *Internet*” aprovado na Chamada de Projetos Sociais No. 01/2010 (vigência 2011-2013), sendo que o mesmo contou com apoio financeiro do Grupo Marista/Centro Social Champagnat mantido pela Associação Paranaense de Cultura (APC).

Além disto, cabe destacar que o *software* desenvolvido encontra-se em procedimento de registro junto ao Instituto Nacional de Propriedade Industrial (INPI) e que o processo de detecção de aliciamento de crianças e adolescentes em mensagens instantâneas encontra-se em procedimento de análise de patenteamento.

O presente projeto foi avaliado pela Dra. Leila Paiva, Coordenadora do Programa Nacional de Enfrentamento da Violência Sexual contra Crianças e Adolescentes, sendo o Parecer apresentado no Anexo I. A Dra. Leila aponta em seu parecer que:

A referida proposta, a princípio, da forma como descrita no material disponibilizado para análise, vem ao encontro das necessárias ações propugnadas nos diferentes instrumentos normativos nacionais e internacionais de enfrentamento da violência sexual contra crianças e adolescentes, já citados neste parecer. Vislumbra-se, portanto, a possibilidade de constituir-se em uma real e necessária estratégia de proteção dos direitos humanos de meninas e meninos brasileiros(as). Para tanto, será condição para seu êxito, a interlocução com o poder público no âmbito do executivo, responsável pela implementação de políticas públicas para a infância e adolescência, e com o Sistema de Garantia de Direitos, por meio dos representantes dos seus três eixos - promoção, defesa e controle.

O trabalho estabelece um método de detecção de aliciamento com base nos diálogos em salas de bate-papo de modo a constituir uma ferramenta de autoproteção no contexto de aliciamento sexual de crianças e adolescentes permitindo a sinalização do nível de susceptibilidade de risco de ocorrência de abuso.

Os resultados do trabalho também são úteis para que a família e a escola possam orientar e até mesmo identificar situações de aliciamento, bem como, tomar as medidas cabíveis de maneira a proteger seus filhos(as) e/ou alunos(as).

1.6. Estrutura do Documento

A estrutura deste documento está organizada de tal forma que o Capítulo 2 traz uma introdução dos conceitos fundamentais, permitindo o entendimento e diferenciação dos termos: pornografia infantil, exploração sexual de menores, abuso sexual de menores e pedofilia. Este Capítulo fornece os pressupostos conceituais do trabalho, ou seja, o embasamento teórico e conceitual necessário ao estudo do tema em questão. O Capítulo 3 apresenta os trabalhos relacionados com o desenvolvimento de métodos e técnicas voltados à detecção de aliciamento de crianças e adolescentes. No Capítulo 4 encontra-se a descrição detalhada do método de detecção de aliciamento de crianças e adolescentes em mensagens instantâneas. Já o Capítulo 5 apresenta os experimentos realizados para comprovação do método seguido de suas respectivas análises. Finalmente, o Capítulo 6 conclui e pondera algumas considerações relevantes, bem como, aponta os caminhos para trabalhos futuros.

Capítulo 2

Fundamentação Teórica

Neste Capítulo são apresentados os conceitos que estão relacionados com o trabalho da análise automática de textos de mensagens instantâneas para detecção de aliciamento sexual de crianças e adolescentes.

Inicialmente serão apresentados alguns pressupostos conceituais. É importante conhecer tais conceitos para compreender o assunto principal desse trabalho de pesquisa. Neste contexto, as primeiras seções deste Capítulo apresentam também a definição de aliciamento de crianças e adolescentes e as abordagens encontradas na literatura científica.

Nas seções seguintes serão apresentados os conceitos de técnicas relacionadas à Aprendizagem de Máquina (AM), importantes para o entendimento dos Capítulos posteriores. Nessas seções serão apresentados o *Support Vector Machine* (SVM) e o *Latent Semantic Analysis* (LSA). Estas seções se encerram com a definição de algumas métricas utilizadas para avaliar o desempenho dos classificadores e também as métricas utilizadas para a criação dos vetores de características necessários com o uso do SVM.

Finalmente, as últimas seções apresentam a definição dos processos estocásticos e os Modelos Escondidos de Markov, que foram utilizados neste trabalho para a obtenção do alerta de susceptibilidade ao abuso sexual de crianças e adolescentes.

2.1. Pressupostos Conceituais

Para que se possa trabalhar no tema principal deste trabalho, seguem pressupostos conceituais que permitem entender e diferenciar os seguintes termos:

- **Pornografia Infantil:** relaciona-se com a representação gráfica, escrita ou sonora, visando à excitação sexual. Segundo as Nações Unidas, é todo tipo de representação, por vários meios de comunicação, de prática sexual real ou simulada, de imagens de órgãos genitais com propósito sexual [TRI10];
- **Exploração Sexual de Menores:** é o crime que envolve o aliciamento de crianças e adolescentes para satisfação da lascívia de outrem mediante obtenção de vantagem econômica [TRI10];
- **Abuso Sexual de Menores:** caracteriza-se pelo aliciamento, assédio, instigação ou constrangimento de criança ou adolescente com o fim de praticar com ela ato libidinoso [TRI10];
- **Pedofilia:** é o nome dado a parafilia ou transtorno sexual, onde o indivíduo envolve-se em atividades sexuais com uma criança pré-púbere (13 anos ou menos), sendo que o indivíduo com pedofilia deve ter 16 anos ou mais e ser pelo menos 5 anos mais velho que a criança. Logo, não se trata de um crime e sim de uma doença. Importante, pois, frisar que nem todo pedófilo comete delito, e nem todo aquele que comete crime é pedófilo [GAU03].

A partir destes conceitos pode-se adentrar de modo apropriado no que é o cerne do presente projeto: o aliciamento sexual de menores (crianças e adolescentes).

2.2. Aliciamento de Menores (Crianças e Adolescentes)

A seguir são abordadas definições teóricas sobre o aliciamento sexual de crianças e adolescentes, visando permitir o entendimento do tema e do problema em questão.

2.2.1. O que é Aliciamento?

O termo “aliciamento” em inglês pode ser definido pelas seguintes palavras: *alluring*, *seduction*, *persuading*, *harassment*. Porém, após uma busca na literatura [MIC10] os termos mais apropriados para aliciamento na área computacional é *grooming* e *enticement*. O aliciamento sexual de crianças e adolescentes se refere ao ato de preparar a vítima, isto é, ter acesso a provável vítima usando salas de bate-papo via *Internet* de modo que a conversação estabelecida possibilite a marcação, por exemplo, de um encontro para cometer abuso sexual [MIC10]. Não há consenso no uso dos termos *grooming* – mais comumente utilizado para se referir à agressão sexual de crianças e adolescentes no ambiente familiar e *entice/enticement* –

mais comumente empregado no contexto de aliciamento usando o ambiente informatizado, assim, ambos são tratados como sinônimos neste trabalho.

Pelas características de interatividade e recursos que os meios computacionais atuais oferecem o *grooming/enticement* ocorre com muita frequência em salas de bate-papo. Neste caso, a detecção do aliciamento precisa ser feita *online*, avaliando as mensagens de texto que trafegam entre o possível agressor e a possível vítima [MIC10]. A detecção do aliciamento por meio da avaliação das mensagens de textos existentes em conversas de salas de bate-papo pode ser feita associando à conversa a indicação de estágios que o agressor utiliza para conquistar sua vítima com o objetivo de se relacionar sexualmente com a mesma.

2.2.2. O Aliciamento e suas Consequências

O aliciamento sexual de crianças e adolescentes pode apresentar diversas consequências às vítimas, sendo a pior delas a ocorrência efetiva do abuso sexual. Dentre as outras principais consequências para a vítima devido ao aliciamento, pode-se citar⁴:

- Vergonha de si mesmo;
- Ansiosidade;
- Insônia;
- Isolamento;
- Depressão;
- Disfunções sexuais.

Em casos extremos, quando ocorre o aliciamento e também o abuso sexual, o menor quando atinge a fase adulta pode também se tornar um agressor, repetindo o comportamento que lhe causou algum dano, seja esse físico ou psicológico [CFM06].

2.2.3. Estágios do Aliciamento

Para entender o aliciamento, deve-se entender como este ocorre, quais seus estágios, em suma, o seu *modus operandi*. Assim, um modelo de comunicação identificando os estágios que um agressor desenvolve no mundo real para atrair suas vítimas foi proposto por OLSON et al. [OLS07] com base nos estudos da Teoria de Comunicação Ludibriante (*Luring Communication Theory*). Este modelo apresenta três estágios mais relevantes: a persuasão (ganho de acesso) da vítima, o envolvimento da vítima numa relação enganosa e a iniciação e manutenção de um relacionamento sexualmente abusivo (Figura 2.1).

⁴ Disponível em www.todoscontraapedofilia.com.br. Acessado em 05 de junho de 2013.

Neste modelo, OLSON et al. (2007) apresenta que o estágio da persuasão (ganho de acesso à vítima) é composto por três divisões: as características individuais do agressor, as características individuais da vítima e o posicionamento estratégico, este definido como um encontro de curto prazo, tais como em *shopping* ou parques, ou um encontro de longo prazo, onde o agressor constrói um relacionamento com a vítima e também com seus familiares [OLS07].

O segundo estágio do modelo de comunicação de OLSON et al. (2007) trata de um ciclo para envolver a vítima numa relação enganosa. A preparação (*grooming/enticement*) consiste nas estratégias de comunicação que o agressor utiliza para que a vítima aceite o contato sexual. O desenvolvimento de uma verdade enganosa é entendido pelos autores deste modelo como a capacidade do agressor cultivar um relacionamento de amizade com sua vítima, podendo também incluir seus familiares, com o único objetivo de tirar proveito próprio em seus interesses sexuais. Normalmente o agressor é visto como alguém importante ou uma autoridade, sendo que seu comportamento é pouco questionado pela vítima e seus parentes. O isolamento pode ser apresentado tanto como um isolamento físico, em que o agressor se propõe a tomar conta da vítima, como um isolamento mental, onde o agressor se mostra para a vítima como a única pessoa em quem se pode confiar. Finalizando este ciclo tem-se a aproximação, que é definida por OLSON et al. (2007), como o contato físico inicial ou introdução verbal que ocorre antes do ato sexual.

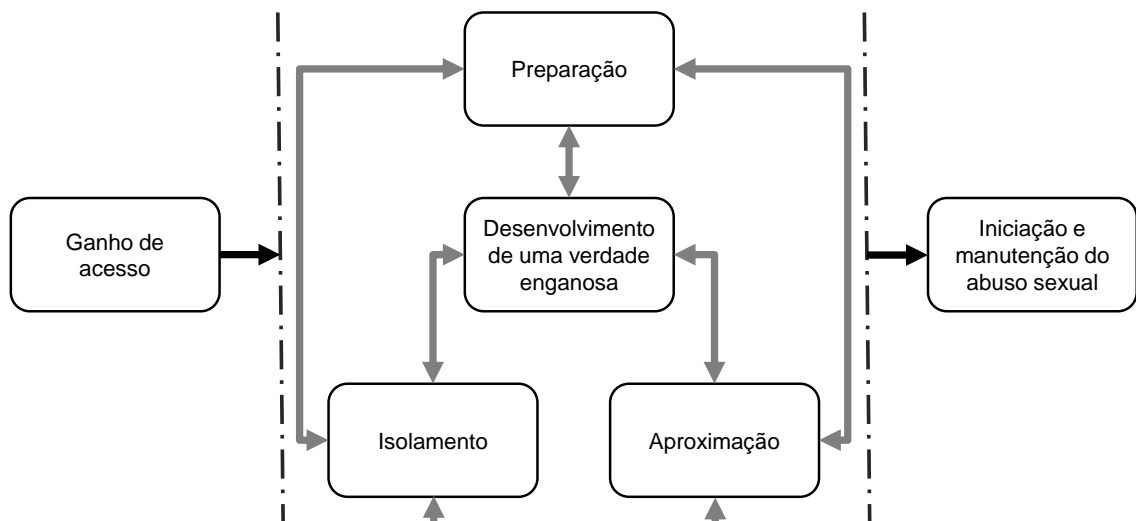


Figura 2.1 - Adaptação do Modelo de Comunicação [OLS07]

Fechando o modelo de comunicação de OLSON et al. (2007) tem-se o estágio da iniciação e manutenção do abuso sexual. A iniciação consiste em realmente praticar o ato sexual, seja este com uma penetração efetivamente, através de sexo oral ou então pela masturbação. Já para a manutenção do ato é de suma importância o sigilo da vítima. Normalmente os agressores obrigam as vítimas ao silêncio coagindo-as e ameaçando-as ou simplesmente convencendo-as de que ao relatar o abuso não serão acreditadas ou até mesmo poderão causar um desconforto familiar.

2.2.4. Estágios do aliciamento no contexto computacional

LEATHERMAN [LEA09] fez uma adaptação para o contexto virtual do modelo de comunicação proposto por OLSON et al. (2007) – Figura 2.1, o qual permite a caracterização dos estágios que um agressor desenvolve no mundo real para atrair suas vítimas com base nos estudos da Teoria de Comunicação Ludibriante (*Luring Communication Theory*) [OLS07]. Nesta adaptação, os estágios para o modelo de comunicação são descritos a seguir:

- Acesso: envolve a troca de informações pessoais tanto do agressor quanto da vítima. No aliciamento *online*, o agressor ganha acesso (*persuade*) à vítima através de fóruns de mensagens instantâneas, salas de bate-papo e *sites* de redes sociais tais como o MySpace e o Facebook. O agressor frequenta *chats* normalmente freqüentados por menores de idade;
- Desenvolvimento de uma confiança enganosa: trata-se da capacidade do agressor em cultivar relacionamentos com potenciais vítimas e possivelmente com suas famílias para benefício próprio de seus interesses sexuais. No ambiente *online* isto é dividido em quatro etapas:
 - ✓ *Informações Pessoais*: detalhes sobre a vítima e o agressor como idades, nomes, localização do computador (no quarto, porão, etc), data de aniversários, número do celular ou do telefone de casa, e fotos de si mesmos;
 - ✓ *Informações de Relacionamento*: inclui a discussão de sentimentos e atitudes em relação à manutenção, construção e dissociação de suas relações, amigos, outras pessoas importantes e membros da família;
 - ✓ *Atividades*: categoria mais compartilhada entre o agressor e a vítima, incluindo músicas, filmes, livros, esportes, *hobbies* e comidas favoritas;

- ✓ *Elogios*: envolve agressor ou vítima elogiando a aparência, as atividades e as informações pessoais com a intenção de fazer a vítima ver o agressor de uma forma positiva;
- Preparação: trata-se das estratégias sutis de comunicação que os agressores utilizam para preparar suas potenciais vítimas a aceitar a conduta sexual. Com a confiança da vítima o agressor começa a prepará-la para aceitar ofertas de contato sexual, podendo não ser explícitas, mas também como tornar a vítima menos sensível a comentários sexuais e linguagens obscenas. Há duas etapas de preparação para o aliciamento:
 - ✓ *Dessensibilização Comunicativa*: refere-se ao agressor propositadamente e frequentemente usar uma linguagem sexual vulgar na tentativa de dessensibilizar a vítima sobre seu uso. No aliciamento *online*, isto pode ser feito pelo envio de imagens pornográficas ou usando gírias sexuais;
 - ✓ *Reenquadramento*: ocorre quando criminosos sexuais se esforçam para tornar a vítima confortável com avançados experimentos sexuais por meio da *Internet*. Para este fim, a conversa sexual é apresentada de uma forma positiva e é muitas vezes referida como uma experiência de aprendizagem, um jogo para ser jogado, ou uma habilidade importante para aprender a fim de participar de relacionamentos amorosos no futuro;
- Isolamento: é essencial isolar a vítima, seja este isolamento físico ou mental. O isolamento físico é definido como organizar um tempo para ficar a sós com a vítima, já o isolamento mental é definido como aumentar a dependência da vítima com o agressor para coisas como a amizade e orientação. O agressor atinge o isolamento tendo certeza que a vítima está em salas de bate-papo sem supervisão;
- Aproximação: o agressor tenta a abordagem da vítima, sugerindo que eles se encontrem para fins sexuais. No modelo *online*, a abordagem é a etapa final, é quando o agressor pede para encontrar a vítima pessoalmente ou que eles conversem por telefone com a intenção de começar um relacionamento sexual.

Este modelo de comunicação é importante para o entendimento da complexidade do problema tratado neste trabalho, bem como, para a modelagem do modo de operação do agressor em relação à possível vítima. Cabe, portanto, descrever as abordagens encontradas na literatura para detecção de aliciamento sexual.

2.3. Abordagens para Detecção de Aliciamento Sexual

A abordagem adotada para lidar computacionalmente com a detecção de aliciamento sexual geralmente envolve o uso de filtros de palavras vinculadas a conteúdos (e.g. fotos, vídeos) com pornografia [HOW04] ou análise *online* buscando por palavras que caracterize o aliciamento [RAS08] [MIC10]. A pesquisa bibliográfica permite estabelecer a existência de três abordagens para a detecção de aliciamento sexual, são elas: a abordagem baseada em filtro de palavras, a abordagem baseada em regras e a baseada em classificadores. Cada uma destas abordagens é detalhada a seguir.

2.3.1. Baseadas em filtro de palavras

O filtro de palavras é uma técnica comumente utilizada [KON09a], porém esta abordagem é bastante limitada, pois não captura o contexto do diálogo e, portanto, está sujeita a muitos falsos positivos [HOW04] [KON09b].

A Figura 2.2 ilustra a estrutura geral do protótipo desenvolvido no projeto de NUNES et al. [NUN09] e como ele se utilizou do filtro de palavras para fazer a “Detecção de Assédio Moral em e-mails”. A partir de um documento de entrada [a], o sistema prepara os dados [b] transformando os textos em um vetor de palavras [c]. O algoritmo *N-gram* é aplicado no vetor de palavras e na base de dados *dicionario_bd* [d]. A base de dados *dicionario_bd* reúne um léxico que representa algumas palavras-chaves que caracterizam assédio moral. O algoritmo *N-gram* faz uma análise comparativa entre as palavras contidas na base de dados *dicionario_bd* e no documento de entrada, gerando os resultados [f].

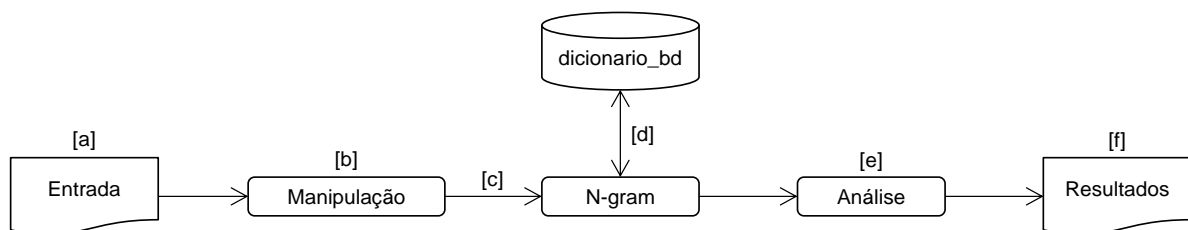


Figura 2.2 - Panorama do Protótipo para Detecção de Assédio Moral em e-mails [NUN09]

As frases ilustradas abaixo foram retiradas de NUNES et al. (2009) e representam alguns e-mails identificados como assédio pelo sistema proposto pelos autores, apresentando com destaque as palavras identificadas:

- ✓ “A empresa não é lugar para **doente**. Aqui você só **atrapalha**.”;
- ✓ “Não quero mais falar sobre esse tipo de e-mail, o próximo que escrever **besteira** será **demitido**, nem que eu tenha que ir junto”.

Neste trabalho de NUNES et al. (2009), na classificação dos 25 *e-mails*, por similaridade, utilizando o dicionário de assédio moral, os autores conseguiram uma taxa de acerto de 90,91%, com uma taxa de falso positivo igual a 9%.

2.3.2. Baseadas em regras

A abordagem baseada na criação de regras faz uma análise no contexto do diálogo [KON09b]. Esta técnica não faz uso apenas de um filtro de palavras, mas também cria um algoritmo que busca nas frases de um diálogo uma ligação entre verbos, substantivos e outros elementos da gramática para se identificar o estágio de comunicação entre o agressor e a vítima.

2.3.3. Baseadas em classificadores

Pode-se citar como outra forma de se encontrar os estágios de um diálogo a aprendizagem de máquina [MCG11]. Esta técnica auxilia a identificar quais os atributos relevantes que devem ser considerados nos algoritmos de aprendizagem não supervisionada. A dificuldade encontrada na aplicação destes métodos é dispor de bases de dados que permitam o treinamento e validação dos classificadores empregados.

2.4. Análise Textual

A seguir serão descritas as técnicas que foram empregadas para se identificar os estágios de comunicação entre um agressor e sua vítima em transcrições de conversas (enunciados) de salas de bate-papo.

2.4.1. Algoritmos de Aprendizagem de Máquina

Na década de 70 houve um aumento no uso de técnicas de computação que se utiliza de Inteligência Artificial (IA) para resolução de problemas reais [FAC11]. Nesta época, a solução de um determinado problema era feita por meio da codificação em um algoritmo do conhecimento de algum especialista da área do problema. Esses algoritmos eram conhecidos como Sistemas Baseados em Conhecimento [FAC11]. Com o crescimento dos problemas a

serem tratados computacionalmente, surgiu a necessidade do desenvolvimento de ferramentas mais sofisticadas e menos dependentes de intervenção humana. Assim, novas técnicas foram desenvolvidas para trabalharem de maneira autônoma, nas quais a partir de uma experiência passada, uma hipótese ou uma função fosse criada para resolver o problema em questão. A esse processo denominou-se Aprendizado de Máquina.

Para Mitchell [MIT97] a definição de AM é dada como “a capacidade de melhorar o desempenho na realização de alguma tarefa por meio da experiência”. Pode-se afirmar que AM é multidisciplinar, pois pode ser empregada em diferentes áreas, tais como Probabilidade e Estatística, Teoria da Informação, Neurociências e Filosofia.

Para o sucesso da Aprendizagem de Máquina conta-se com o que é denominado de aprendizado indutivo, o qual consiste em um processo que permite realizar generalizações a partir de uma base de dados [FAC11]. De um modo geral, a tarefa de aprendizado pode ser dividida em preditivas ou descritivas. No aprendizado preditivo, também conhecido como supervisionado, um conjunto de treinamento com uma classe associada é fornecido ao algoritmo de aprendizado. Basicamente cada exemplo é descrito por um vetor de características associado ao rótulo da classe. Esse algoritmo tem por objetivo construir um classificador capaz de determinar de forma correta a classe para novos exemplos que não tenham sido rotulados. Para valores de classe discretos, esse problema é conhecido como classificação. Para os valores de classe contínuos o problema é conhecido como regressão [MON03].

Já no aprendizado descritivo, ou não-supervisionado, o algoritmo de aprendizado analisa os exemplos fornecidos, que não possuem uma classe associada, e tenta agrupá-los de alguma maneira, formando agrupamentos ou associações [FAC11]. A Figura 2.3 mostra a hierarquia de aprendizado conforme os tipos de tarefas de aprendizado.

Como existem diferentes algoritmos de AM e também da área de Estatística para a aquisição automática do conhecimento, neste projeto foi dado ênfase ao *Latent Semantic Analysis* (LSA) e, ainda, ao *Support Vector Machine* (SVM), como um aprendizado supervisionado, ambos voltados, no contexto deste trabalho, à análise textual. Segue-se com a fundamentação teórica de ambos os métodos.

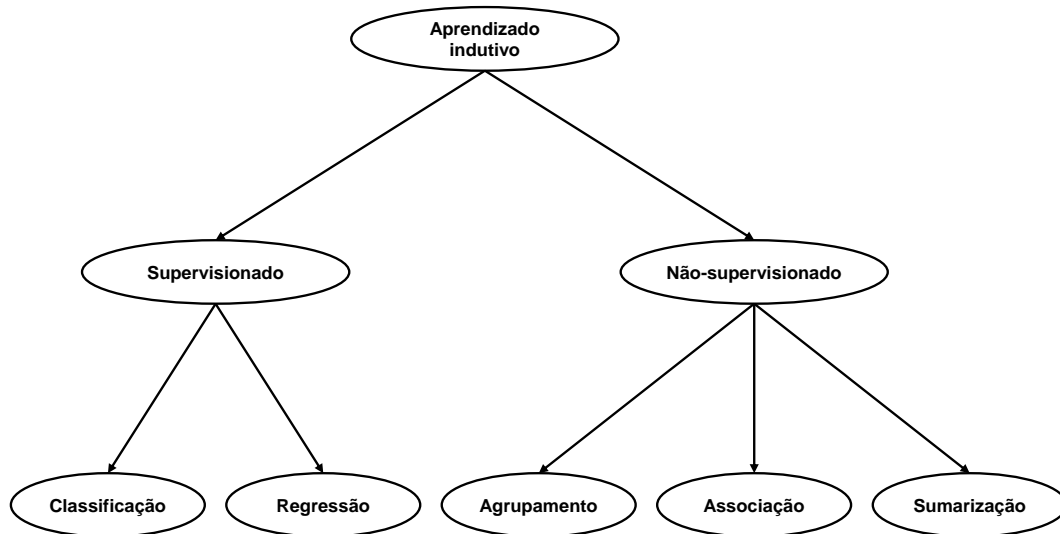


Figura 2.3 - Hierarquia de Aprendizado [FAC11]

2.4.1.1 Support Vector Machine (SVM)

O SVM foi proposto pela primeira vez em 1995 por Vladimir Vapnik e é embasado na Teoria de Aprendizado Estatístico. Trata-se de um modelo de aprendizagem indutiva ou supervisionada, em que algoritmos de aprendizagem são associados para análise dos dados e reconhecimento de padrões [LOR03]. Dentre as principais características do SVM pode-se destacar a boa capacidade de generalização [SMO99]. A capacidade de generalização é dada pela eficiência na classificação de dados que não pertencem ao conjunto de treinamento.

O SVM é um entre muitos algoritmos de AM que foram desenvolvidos originalmente para a solução de problemas de dicotomia. Ele é usado em muitas aplicações, incluindo a categorização de textos [JOA02]. Para o SVM, um conjunto de treinamento é dito linearmente separável se for possível separar as classes desse conjunto em pelo menos um hiperplano [RUS95]. No caso do SVM, para duas classes há diversos separadores lineares possíveis, porém há apenas um hiperplano ótimo, que maximiza a margem de separação entre as classes do conjunto apresentado ao classificador, como mostrado na Figura 2.4. Essa maximização tem por base os vetores de suporte, que correspondem ao elemento de cada classe que está mais próximo ao hiperplano [MAN08].

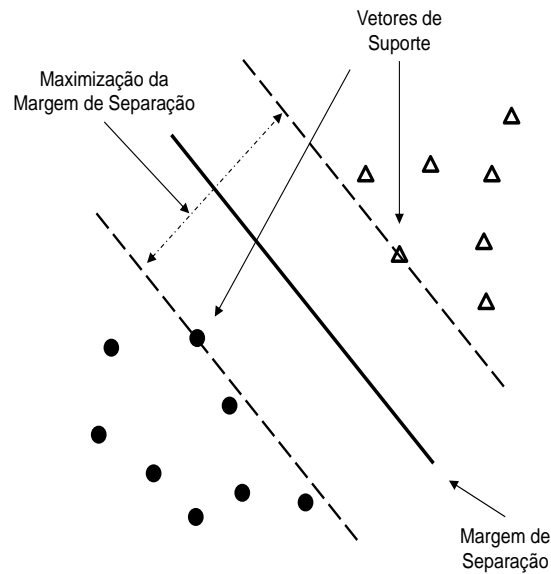


Figura 2.4 - Adaptação do Hiperplano Ótimo [MAN08]

Existem casos em que não é possível fazer a divisão dos dados de treinamento por meio de um único hiperplano, como destaca a Figura 2.5. Para esse caso, uma separação linear das classes não seria adequada.

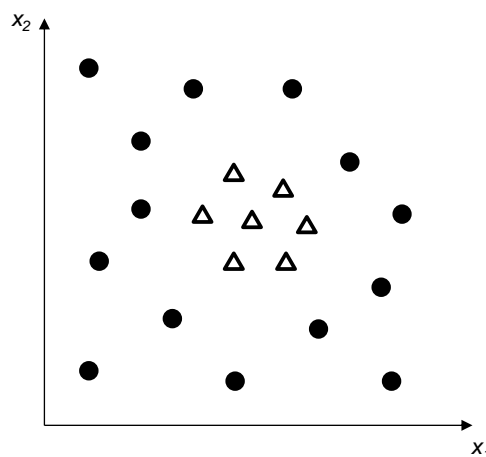


Figura 2.5 - Conjunto de Dados não Lineares. Adaptado de [LOR03]

O SVM trabalha com problemas não lineares mapeando o conjunto de treinamento de seu espaço original, tido como de entradas, para um espaço de maior dimensão, denominado espaço de características [HEA98]. Matematicamente, seja $\Phi: X \rightarrow \mathcal{H}$ um mapeamento, em que X é o espaço de entradas e \mathcal{H} denota o espaço de características. A escolha apropriada de

Φ faz com que o conjunto de treinamento mapeado em \mathcal{Y} possa ser separado por um SVM linear, conforme ilustra a Figura 2.6. Esse procedimento somente é viável se duas condições forem satisfeitas: que a transformação seja não linear e que a dimensão do espaço de características seja suficiente alta [HAY99].

Para realizar o mapeamento é necessário efetuar o cálculo de produtos escalares entre os dados no espaço de características. Esse cálculo é obtido com o uso de funções denominadas *kernels*. Um *kernel* K é uma função que recebe dois pontos x_i e x_j do espaço de entradas e computa o produto escalar desses dados no espaço de características, conforme descrito na Equação 2.1:

$$K(x_i, x_j) = \Phi(x_i) \times \Phi(x_j) \quad (2.1)$$

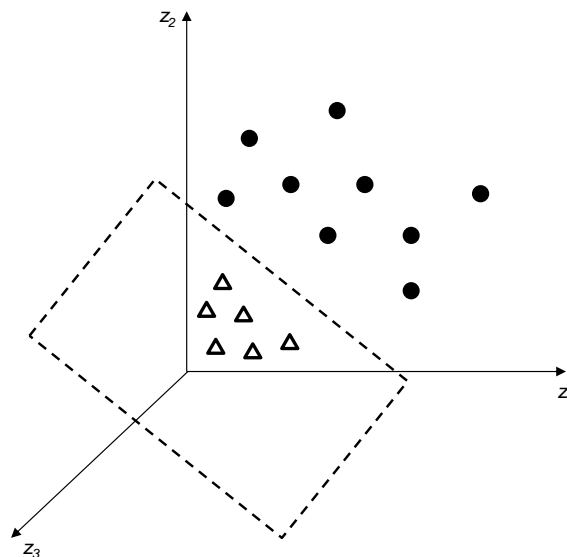


Figura 2.6 - Fronteira Linear no Espaço de Características. Adaptado de [MUL01]

Os *kernels* mais comumente empregados são os Polinomiais, os Gaussianos ou RBF (*Radial-Basis Function*) e os Sigmoidais. Cada um deles possui parâmetros que devem ser indicados pelo usuário. A pesquisa de seleção de parâmetros pode ser realizada fazendo o uso do algoritmo *grid-search*. O *grid-search* faz a busca pelos melhores parâmetros por meio da análise dos resultados obtidos na execução do algoritmo para um intervalo de parâmetros. O *kernel* RBF é utilizado como padrão no SVM, embora os outros *kernels* também possam ser utilizados [CHA11]. O *kernel* RBF possui dois parâmetros que podem ser ajustados

utilizando-se do *grid-search* que são: o custo C e o *gamma* γ . O objetivo da pesquisa de seleção dos parâmetros C e γ é identificar os melhores resultados para esses parâmetros que sejam capazes de aumentar a precisão do classificador para os dados que não pertencem ao conjunto de treinamento [CHA11].

Embora o SVM tenha sido projetado para solucionar problemas binários, ele também pode ser utilizado para problemas de classificação multiclasse. Uma das estratégias empregadas para a solução desse problema é conhecida como decomposição, técnica que faz a combinação dos classificadores gerados em subproblemas binários [FAC11]. As saídas desses classificadores binários são combinadas na obtenção da classificação final.

Algumas estratégias podem ser empregadas na decomposição de problemas multiclasse [MIL06], a saber: um-contra-todos (do inglês *one-against-all*) e a decomposição hierárquica. Nesse trabalho foi adotada a estratégia todos-contra-todos (do inglês *all-against-all*). Nesta estratégia, dadas n classes, o número de classificadores gerados para a solução do problema está na ordem de n^2 . Embora pareça lógico que o tempo total de treinamento na estratégia todos-contra-todos é muito alto, pois é necessário treinar muitos classificadores binários ao mesmo tempo, isso não se concretiza para o SVM, pois o tempo de formação de um SVM aumenta mais do que linearmente com o número de amostras de treinamento. Desta forma, como na estratégia todos-contra-todos cada um dos subproblemas binários envolve um pequeno número de amostras, se torna mais fácil de resolvê-los e mais rápido o treinamento nesta estratégia [MIL06].

2.4.1.1.1. Criação dos Vetores de Características

O modelo TF-IDF (*Term Frequency – Inverse Document Frequency*) [SAL88] é um método utilizado para a representação de textos como um vetor de características. Esse modelo reflete a importância de uma palavra para um texto na base de dados de maneira que quanto maior a frequência de uma palavra em um texto, mais representativa ela é para o conteúdo e quanto mais textos contiverem uma palavra, menos discriminante ele é. O modelo atribui um peso $W_{T,u}$ para uma palavra T em um texto u conforme as Equações 2.2 e 2.3:

$$W_{T,u} = tf_{T,u} \times idf_T \quad (2.2)$$

$$idf_T = \log \frac{N}{df_T} \quad (2.3)$$

Onde:

- $tf_{T,u}$ é a quantidade de vezes que o palavra T aparece no texto u ;
- N é a quantidade total de textos que compõem a base de entrada;
- df_T é a quantidade de textos que contém a palavra T .

Para reduzir a dimensão e os ruídos de um texto e conseqüentemente os termos de um vetor de características pode-se utilizar o Ganho de Informação (GI) [MIT97] de uma palavra. Dada uma base de entrada (S) que apresenta c classes distintas, a entropia de S é dada pela Equação 2.4:

$$Entropia(S) = \sum_{i=1}^c p_i \log_2 p_i \quad (2.4)$$

Onde:

- p_i é a proporção de dados em S que pertencem à classe i .

O Ganho de Informação para uma palavra T de uma base de entrada (S) fornece a medida da diminuição da entropia esperada ao utilizar a palavra T para fazer a partição da base de entrada.

Considerando $P(T)$ o conjunto dos valores que T pode assumir; seja v um elemento deste conjunto e seja S_v o subconjunto de S formado pelos dados em que $T = v$; a entropia que se obtém ao particionar S em função da palavra T é dada pela Equação 2.5:

$$E(T) = \sum_{v \in P(T)} \frac{|S_v|}{|S|} Entropia(S_v) \quad (2.5)$$

O Ganho de Informação de uma palavra T é definido pela Equação 2.6:

$$GI(S, T) = Entropia(S) - E(T) \quad (2.6)$$

2.4.1.2 Método *Latent Semantic Analysis* (LSA)

O presente trabalho estudou também outro método para auxiliar na identificação dos estágios de comunicação entre um agressor e sua vítima. Tal método é denominado de *Latent Semantic Analysis* (LSA, também conhecido como *Latent Semantic Indexing* - LSI), o qual

consiste em um método probabilístico para a identificação das relações entre as palavras de um texto [LAN05], sendo comumente empregado na identificação de emoções em textos [STR08] [DOS12] [MAR10].

O LSA emprega técnicas estatísticas e como resultado final apresenta uma matriz para correlacionar palavras presentes em um texto. Alguns trabalhos que empregaram o método do LSA obtiveram resultados com uma taxa de sucesso não superior a 70% [MAR10] [STR08] [DOS12].

Para executar o LSA, a primeira tarefa a ser feita é representar os textos de um documento em uma matriz denominada “termo-documento”. Nesta matriz, cada coluna representa um texto e cada linha representa uma palavra, sem repetição, existente nos textos do documento. Cada célula contém a frequência de ocorrência de cada palavra nos textos. As células são submetidas a um processo de cálculo preliminar para atribuir um valor a cada frequência de ocorrência. São atribuídos dois valores diferentes: um para a importância da palavra em cada texto e outro para o conjunto de textos analisados.

Para uma representação reduzida da matriz termo-documento é empregada a técnica denominada *Single Value Decomposition* (SVD) [LAN05]. Essa técnica enfatiza os padrões e as relações entre as palavras e/ou textos mais fortes, e descarta os ruídos e as relações e padrões mais fracos. O SVD faz a decomposição da matriz termo-documento em três novas matrizes, conforme descrito na Equação 2.7:

$$A = U \times V^T \times S \quad (2.7)$$

Sendo:

- A é a matriz termo-documento a ser decomposta;
- U é uma matriz que descreve as linhas da matriz A , ou seja, as palavras;
- V^T é a matriz transposta de V , descreve as colunas da matriz A , ou seja, os textos analisados;
- S é uma matriz diagonal que contém os “valores singulares” (singular values), que definem a relação entre os termos de U e V^T .

A matriz S permite estimar quantas dimensões devem ser usados a fim de se obter os melhores resultados. Este número de dimensões (denominado K), conforme GROSSMAN e FRIEDER [GRO04] pode ser estimado de forma arbitrária através de experimentos e

comparações. Adota-se como padrão $K = 50$, pois segundo a literatura os melhores resultados são obtidos com valores de K menores do que 100. Segundo GROSSMAN e FRIEDER (2004), resultados com K superior a 100 não demonstram melhorias significativas nos experimentos realizados.

Após determinadas as três matrizes do SVD, deve-se definir a localização de cada estágio de comunicação no mesmo espaço criado com o SVD. Para determinar esse espaço é necessário ter um léxico para cada estágio. De posse deste léxico e com o auxílio da matriz U , calcula-se um ponto médio no espaço obtido anteriormente com o SVD. Esse ponto médio traz a localização de cada estágio naquele espaço.

Finalizando a descoberta do estágio de comunicação com o auxílio do LSA é utilizada a matriz V^T e o conjunto de coordenadas dos estágios. Por meio da similaridade Cossenoidal [GAR06] é possível determinar a distância entre as matrizes geradas pelo SVD e o grupo dos estágios pré-definidos. As respostas devem ser analisadas em módulo para a descoberta do estágio de comunicação.

Assim, pode-se entender como os algoritmos LSA e SVM podem ser aplicados sobre uma base de dados de conversações (enunciados). A partir dos resultados obtidos com estes métodos, foi aplicado os Modelos Escondidos de Markov - MEM (*Hidden Markov Models* – HMM), de modo a se definir um modelo probabilístico associado aos estágios do aliciamento e, assim, apresentar o nível de susceptibilidade ao abuso para uma determinada sequencia de enunciados. Os MEMs [RAB93] caracterizam-se por ser um método de classificação estatístico e estrutural, pois leva em consideração o encadeamento de observações ao longo do tempo. Neste sentido, entende-se que os Modelos Escondidos de Markov são o método adequado para implementação da fase final da detecção do aliciamento sexual de crianças e adolescentes em conversações em salas de bate-papo.

2.5. Processos Estocásticos

Conforme Reichl [REI04] é necessário entender o conceito de processo estocástico e variáveis aleatórias para poder aplicar a Teoria das Probabilidades no mundo real. Um processo estocástico é uma sequência de variáveis aleatórias X_t , em que t pertence a um determinado conjunto temporal T . Os processos estocásticos usualmente são utilizados para descrever o funcionamento de um sistema que opera durante um determinado período de

tempo e podem ser classificados em relação ao Estado (X_t) e em relação ao tempo T da seguinte forma:

- Estado Discreto: $X_t = [0, 1, 2, \dots]$;
- Estado Contínuo: $X_t = [0, +\infty)$
- Tempo Discreto: $T = [0, 1, 2, \dots]$;
- Tempo Contínuo: $T = [t: 0 \leq t < +\infty)$.

Sendo assim, passa-se ao estudo das Cadeias de Markov, visto ser de interesse ao desenvolvimento deste projeto, uma vez que busca-se modelar com base em probabilidades os estágios de comunicação do aliciamento associado aos perfis do agressor e da vítima.

2.5.1. Cadeias de Markov

Cadeias de Markov referem-se a processos estocásticos de estado e tempo discretos, e que exibem a propriedade Markoviana representada na Equação 2.8:

$$P\{X_{t+1} = j \mid X_t = i\} = P\{X_{t+1} = j \mid X_t = i, X_{t-1} = i_1, X_{t-2} = i_2, \dots, X_0 = i_t\} \quad (2.8)$$

para $t = 0, 1, 2, \dots$ e toda sequência j, i, i_1, \dots, i_t , ou seja, espaço e tempo discretos. A Equação 2.8 pode ser dita como a probabilidade de ocorrência do estado j no tempo $t+1$ está condicionada à ocorrência do estado i no tempo t , do estado i_1 no tempo $t-1$ e assim por diante, até o estado inicial.

A Figura 2.7 traz a representação gráfica das cadeias de Markov, ou seja, o denominado diagrama de transição. Neste diagrama observam-se os seguintes elementos da cadeia:

- Estado: indica a situação do processo num certo instante, a qual é aferida por meio de uma variável aleatória;
- Estágio: o estágio está ligado ao tempo;
- Espaço de estados: é o conjunto de todas as variáveis aleatórias;
- Transição: é a evolução de um estágio para outro, podendo haver mudança de estado ou não.

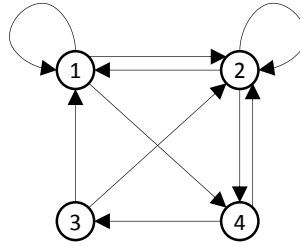


Figura 2.7 - Diagrama de transição da cadeia de Markov

As cadeias de Markov podem ser descritas por meio de uma matriz de transição. Considerando uma cadeia de Markov com N estados e p_{ij} a probabilidade de transição do estado i para o estado j , então a matriz $P = [p_{ij}]$ é denominada a matriz de transição desta cadeia de Markov (Figura 2.8).

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix}$$

Figura 2.8 - Matriz de transição de uma cadeia de Markov com 4 estados

Na matriz de transição, as probabilidades condicionais p_{ij} e a matriz estocástica devem apresentar as seguintes propriedades:

- Ser uma matriz quadrada de dimensão $N \times N$, onde N é o número finito de estados do processo;
- $0 \leq p_{ij} \leq 1$, uma vez que p_{ij} representa uma probabilidade;
- A soma de cada linha da matriz de transição deve ser igual a 1, uma vez que os estados correspondem a eventos mutuamente exclusivos num estágio qualquer.

A matriz de transição P é dita homogênea quando $P(t) = P$, ou seja, quando a matriz é estacionária, tendo as probabilidades de transição p_{ij} independentes no tempo. A matriz P também é dita regular quando todos os elementos da matriz elevados a alguma potência, $P^{(n)}$, forem diferentes de zero.

Deste modo, a transição entre estados pode ser dada diretamente como também pode passar por estágios intermediários. Quando isso ocorre, diz-se ser a probabilidade de transição

em n passos. Numa matriz de transição $P = [p_{ij}]$ a probabilidade de transição de um estado i para um estado k e do estado k para o estado j é dada pela Equação 2.9:

$$P_{ij}^{(n)} = \sum_{k=1}^N P_{ik} \times P_{kj} \quad (2.9)$$

É importante observar que a partir do estudo e aplicação das cadeias de Markov, pode-se obter a probabilidade de ocorrência de uma dada sequência de observações associadas aos estados do modelo por meio da matriz de transição, a qual representa o próprio modelo. Este tipo de cálculo se demonstra interessante ao problema tratado neste trabalho, pois se pode obter um modelo de transição associado ao modelo de comunicação do aliciamento a partir dos estágios determinados *a priori* com a aplicação dos métodos de LSA e SVM.

2.5.2. Modelos Escondidos de Markov

O Modelo Escondido de Markov (MEM) é o resultado da associação de uma cadeia de Markov com número finito de estados e um conjunto de densidades de probabilidade de evolução do sistema. Mesmo conhecendo os parâmetros do modelo não é possível determinar a evolução da cadeia de Markov para esse processo. A cadeia de Markov serve de suporte para o conjunto de observações associadas a um estado [RAB93], sendo que se necessita conhecer:

- Os parâmetros que definem o Modelo Escondido de Markov;
- As topologias;
- Os problemas básicos tratados pelo MEM.

Estes elementos seguem descritos visando permitir o entendimento dos MEMs, bem como sua complexidade.

2.5.2.1 Definição de Modelo Escondido de Markov

A definição de um MEM denotado por $\lambda = (N, M, T, A, B, \pi)$ é dada pelos seguintes parâmetros [HUA90] [RAB93]:

- N = número de estado do modelo. O conjunto de estados pode ser descrito como $S = \{s_1, s_2, \dots, s_N\}$ e S_t denota o estado da cadeia no instante t ;
- M = número de símbolos da observação. O conjunto dos símbolos podem ser descritos por $V = \{v_1, v_2, \dots, v_M\}$;

- T : comprimento da sequência de observações. Este comprimento pode ser variável e a sequência de observações pode ser descrita por $O = \{o_1, o_2, \dots, o_T\}$. O_t denota a observação instante t ;
- $A = \{a_{ij}\}$, $1 \leq i, j \leq N$ = matriz de transição entre estados. O elemento a_{ij} corresponde a probabilidade que a cadeia de Markov transite do estado s_i para o estado s_j . Essa matriz de transição deve respeitar as propriedades apresentadas na seção 2.5.1 deste Capítulo;
- $B = \{b_{jk}\}$, $1 \leq j \leq N$; $1 \leq k \leq M$ = matriz de probabilidade de observação de símbolos condicionada pelos estados da cadeia. O elemento b_{jk} representa a probabilidade de observar o símbolo v_k quando o modelo se encontra no estado s_j . Essa matriz deve respeitar as seguintes propriedades:
 - ✓ $0 \leq b_{jk} \leq 1$, uma vez que b_{jk} representa uma probabilidade;
 - ✓ A soma de cada linha da matriz de probabilidade de observação de símbolos deve ser igual a 1.
- $\pi = \{\pi_i\}$, $1 \leq i \leq N$ = conjunto de densidades de probabilidade inicial. π_i representa a probabilidade que o processo de Markov inicie no estado s_i . Esse conjunto deve respeitar as seguintes propriedades:
 - ✓ $0 \leq \pi_i \leq 1$, uma vez que π_i representa uma probabilidade;
 - ✓ A soma do conjunto de probabilidades iniciais deve ser igual a 1.

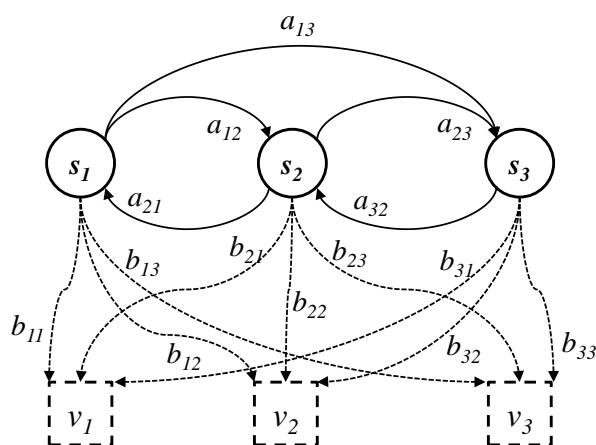


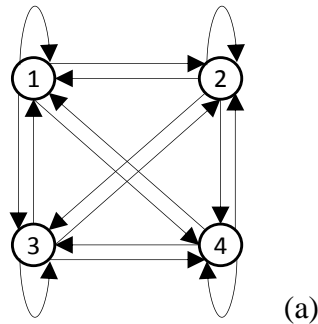
Figura 2.9 - Exemplo de parâmetros probabilísticos de um MEM: s – estados; v – símbolos de observações; a – probabilidade de transição entre estados; b – probabilidade de observação dos símbolos

O padrão apresentado na literatura [RAB89][RAB93] traz uma notação mais compacta para o MEM descrito por $\lambda = (A, B, \pi)$. A Figura 2.9 apresenta uma representação gráfica de um Modelo Escondido de Markov.

2.5.2.2 Topologias de Modelos Escondidos de Markov

Existem três tipos de topologias distintas para os MEMs [HUA90] [RAB93], a saber:

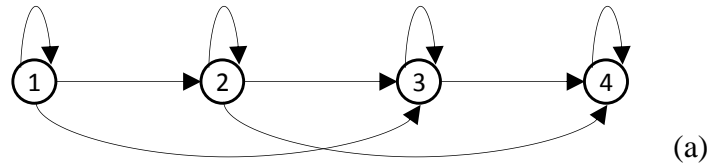
- Modelos sem restrição ou ditos ergódicos: todos os estados são alcançáveis com um número finito de passos a partir de qualquer estado s_i . A Figura 2.10a traz a representação gráfica do modelo ergódico e a Figura 2.10b sua representação matricial;



$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \quad (b)$$

Figura 2.10 - Modelo Ergódico de MEM: a) representação gráfica; b) representação matricial

- Modelos sequenciais ou esquerda-direita: com o passar do tempo, o estado permanece o mesmo ou se altera segundo uma evolução em série do modelo da esquerda para a direita, podendo alguns estados serem saltados. A representação gráfica do modelo sequencial pode ser observada na Figura 2.11a, e sua matriz de transição entre estados é triangular superior, respeitando a propriedade fundamental dos MEM com caminhamento esquerda-direita, $a_{ij} = 0, j < i$ (Figura 2.11b);



$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & 0 & a_{44} \end{bmatrix} \quad (\text{b})$$

Figura 2.11 - Modelo Sequencial de MEM: a) representação gráfica; b) representação matricial

- Modelos paralelos: vários caminhos entre estados são permitidos, porém com a mesma direção esquerda-direita. A representação gráfica deste modelo é apresentada na Figura 2.12. Já a matriz de transição deste modelo é semelhante à dos modelos sequenciais, triangular superior, pois os modelos paralelos é uma variação dos modelos sequenciais.

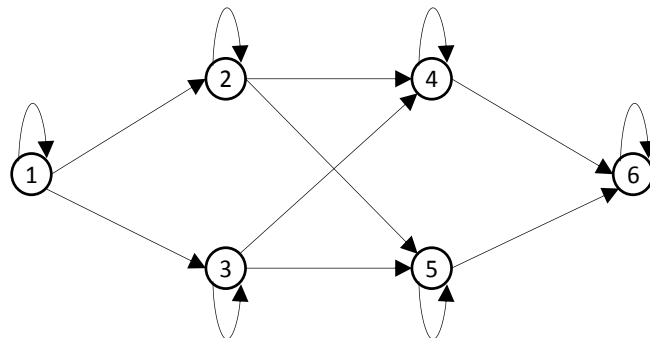


Figura 2.12 - Modelo Paralelo de MEM

2.5.2.3 Problemas Básicos de Modelos Escondidos de Markov

São três os problemas básicos encontrados em sistemas modelados por MEM [RAB93]:

- Problema da avaliação: dada uma sequência de observações $O = \{o_1, o_2, \dots, o_T\}$ e um modelo $\lambda = (A, B, \pi)$, computar eficientemente a probabilidade da sequência de observações.

- Problema de treinamento: dada uma sequência de observações $O = \{o_1, o_2, \dots, o_T\}$ e um modelo $\lambda = (A, B, \pi)$, ajustar os parâmetros do modelo para maximizar o valor de $P[O|\lambda]$.
- Problema da classificação: dada uma sequência de observações $O = \{o_1, o_2, \dots, o_T\}$ e um modelo $\lambda = (A, B, \pi)$, escolher a melhor sequência dentro do modelo capaz de gerar essas observações.

2.5.2.4 Aplicação de Modelos Escondidos de Markov na Detecção de Aliciamento Sexual de Crianças e Adolescentes

Como mencionado anteriormente, o Modelo Escondido de Markov (MEM) tem por base uma cadeia de Markov e um conjunto de densidades de probabilidade associadas aos estados desta cadeia. O aliciamento sexual de crianças e adolescentes é composto de estágios bem definidos que podem ser identificados com certo grau de precisão, por meio da aplicação dos métodos de LSA ou SVM. A seguir são mostrados como podem ser obtidos os elementos necessários para aplicação do MEM visando auxiliar na detecção de aliciamento sexual de crianças e adolescentes em conversas de sala de bate-papo.

Para definição dos estados da cadeia de Markov, utilizam-se os estágios de comunicação provenientes da classificação dos enunciados da conversação entre o agressor e a vítima. *A priori* a cadeia de Markov foi definida pelo modelo ergódico, visto que as conversações podem avançar e retroceder nos estágios de comunicação sem que haja uma regra específica para isto, muito menos que a conversação seja linear ou que a mesma avance progressivamente nos estágios possíveis para uma conversação relacionada à aliciamento. Já as probabilidades de transição entre estados são observadas por meio do uso do SVM (*Support Vector Machine*) e do método LSA (*Latent Semantic Analysis*). Portanto, foi estudada experimentalmente a melhor heurística de uso dos resultados destes métodos.

Considerando os elementos necessários para o MEM comentados anteriormente e uma nova conversa, entre o agressor e a vítima, fica caracterizado o problema da avaliação descrito como um dos “três problemas básicos para utilização de MEM” (Seção 2.5.2.3) [RAB93]. A partir de uma nova sequência de observações que caracterizam os estágios de comunicação entre um agressor e a vítima, deve-se computar eficientemente a probabilidade de ocorrência desta sequência de observações. Alcançando-se este objetivo, tem-se justificada a utilização do Modelo Escondido de Markov, visto que se pretende usar o MEM como um mecanismo de

previsão probabilística de que essa conversa venha a alcançar o último estágio de comunicação, ou seja, a aproximação. Esta previsão é útil para indicar a susceptibilidade de uma dada conversa convergir como crime de abuso sexual de crianças e adolescentes.

2.6. Avaliação de Classificadores

A avaliação de um algoritmo de AM é normalmente realizada por meio da análise do desempenho do classificador gerado pelo algoritmo na rotulação de exemplos que não foram apresentados previamente no seu treinamento [MON03]. A matriz de confusão é uma das formas utilizadas para visualizar o desempenho do algoritmo de classificação. Essa matriz traz a contagem dos registros de testes previstos correta e incorretamente pelo classificador [TAN09]. As linhas da matriz representam as classes verdadeiras e as classes preditas pelo classificador são representadas nas colunas. A diagonal principal da matriz representa os acertos do classificador, enquanto os outros elementos correspondem aos erros cometidos na predição [FAC11].

Uma série de medidas de desempenho pode ser derivada a partir da matriz de confusão [MON03]. Dentre essas medidas pode-se citar:

- Taxa de Falsos Negativos: proporção de exemplos de uma classe positiva classificados incorretamente (Equação 2.10).

$$TFN = \frac{FN}{(FN+VP)} \quad (2.10)$$

Sendo:

- ✓ FN corresponde aos falsos negativos;
- ✓ VP corresponde aos verdadeiros positivos.

- Taxa de Falsos Positivos: proporção de exemplos da classe negativa classificados incorretamente (Equação 2.11).

$$TFP = \frac{FP}{(FP+VN)} \quad (2.11)$$

Sendo:

- ✓ FP corresponde aos falsos positivos;
- ✓ VN corresponde aos verdadeiros negativos.
- *Acurácia*: proporção de exemplos positivos entre todos os exemplos (Equação 2.12).

$$Accuracy = \frac{VP+VN}{VP+VN+FP+FN} \quad (2.12)$$

- *Precisão*: proporção de exemplos positivos classificados corretamente entre todos aqueles preditos como positivos pelo classificador (Equação 2.13).

$$Precision = \frac{VP}{(VP+FP)} \quad (2.13)$$

- *Revocação*: corresponde à taxa de acerto na classe positiva (Equação 2.14).

$$Recall = \frac{VP}{(VP+FN)} \quad (2.14)$$

- *F_β-measure*: média harmônica entre a Precisão e a Revocação (Equação 2.15).

$$F_{\beta} - measure = (1 + \beta^2) \times \frac{(Precision \times Recall)}{((\beta^2 \times Precision) + Recall)} \quad (2.15)$$

A escolha de β corresponde aproximadamente ao grau de importância que se dá ao *Recall* sobre o *Precision*, ou ao custo da associação entre os falsos negativos sobre os falsos positivos. Valores maiores que 1 para β corresponde a mais peso para o *Recall*, enquanto que valores menores que 1 correspondem a um maior peso para o *Precision*.

Para a obtenção de um desempenho preditivo mais confiável pode-se utilizar a *cross-validation*, um dos principais métodos de amostragem existentes. Nesse método, o conjunto de dados é dividido em n subconjuntos de tamanhos iguais. Os exemplos de $n-1$ partições são utilizados no treinamento de um classificador, e a partição restante é utilizada para testar esse classificador. Esse processo é repetido n vezes, utilizando em cada ciclo uma partição

diferente para teste. O desempenho final do classificador é dado pela média dos desempenhos observados sobre cada subconjunto de teste [FAC11].

2.7. Considerações Finais

Neste Capítulo foram expostos os estudos realizados para a realização deste trabalho. Iniciou-se com a apresentação da diferença entre os conceitos relacionados ao abuso sexual de crianças e adolescentes e o que é o aliciamento, e sua organização em estágios. Além disto, foram apresentadas as abordagens existentes na literatura passíveis de serem aplicadas à detecção dos estágios de aliciamento, de modo a descrever as técnicas SVM (*Support Vector Machine*) e LSA (*Latent Semantic Analysis*) e que são utilizadas no desenvolvimento deste trabalho para a detecção dos estágios do aliciamento e como obter o alerta de susceptibilidade ao abuso sexual de crianças e adolescentes por meio do uso do HMM (*Hidden Markov Model*). No próximo Capítulo serão relatados os principais trabalhos encontrados na literatura que permitem compreender o que está contemplado na literatura sobre o tema deste trabalho.

Capítulo 3

Trabalhos Relacionados

A seguir são apresentadas as abordagens (filtro de palavras, regras e classificadores, agrupamento e árvores de decisão) encontradas na literatura para lidar com a problemática de detecção de aliciamento sexual em salas de bate-papo. Algumas abordagens são baseadas na distinção de estágios do aliciamento (Seção 2.2.4), já outras se limitam a distinguir transcrições de conversas que contenham características de aliciamento das demais que não apresentam essas características.

3.1. Detecção dos estágios baseada em filtro de palavras

No estudo de KONTOSTATHIS et al. [KON09a] é desenvolvido um *software* denominado *ChatCoder1* que faz a busca dos estágios de comunicação adaptados por LEATHERMAN [LEA09] em transcrições de conversas de salas de bate-papo. Neste trabalho, KONTOSTATHIS et al. [KON09a] usaram como base de dados 288 transcrições de conversas existentes no *site Perverted-Justice* [PER03]. Os autores desenvolveram um léxico de palavras e um manual de regras para definir os estágios do modelo de comunicação. O manual de regras foi criado para auxiliar os analistas das transcrições a marcarem corretamente palavras e frases em seus respectivos estágios do modelo. Para a construção do léxico foram analisadas apenas 12 das 288 transcrições de conversas, escolhidas aleatoriamente, que continham de 349 a 1500 linhas. O léxico foi desenvolvido utilizando-se de termos de aliciamento, palavras, ícones, frases e linguagem virtual para cada estágio do modelo conforme descrito pelo manual. Cabe ressaltar que as abreviações, os ícones e a

linguagem virtual foram expandidos (transcritos) para a linguagem escrita, sem alterar o contexto da frase.

O *software ChatCoder1* captura os diálogos e já classifica, por meio do uso do léxico, as frases em suas categorias. O *software* traz uma série de facilidades, dentre elas um resumo percentual de cada estágio encontrado na transcrição da conversa. Esta abordagem apresentou uma taxa de acerto de 60%.

3.2. Detecção dos estágios baseada em regras

Num segundo trabalho, KONTOSTATHIS et al. [KON09b] faz o uso de uma técnica baseada em regras para desenvolver o *software ChatCoder2*. O *ChatCoder1* era baseado na técnica popularmente denominada de “caça-palavras”. Se as palavras ou frases não aparecessem exatamente como no léxico não eram marcadas como um dos estágios do modelo de comunicação, o que dificultava o processo de análise como um todo.

As regras para o desenvolvimento do *ChatCoder2* foram criadas com base na transcrição de 15 conversas disponíveis também no *site Perverted-Justice* [PER03]. Neste estudo, as abreviações, os ícones e a linguagem virtual foram levados em conta tais como apareciam no diálogo. Também foram acrescentados valores nominais para os estágios do modelo para indicar se uma palavra ou frase estava em mais de um estágio, então seria considerada no estágio com maior valor nominal.

O algoritmo do *ChatCoder2* foi desenvolvido utilizando-se das seguintes regras e análise da composição das frases do diálogo para cada um dos estágios do modelo de comunicação:

- Acesso: toda saudação formal ou informal é marcada como acesso;
- Informações Pessoais: é verificada a existência de um substantivo combinado com um verbo de ação ou uma palavra de questionamento;
- Informações de Relacionamento: verifica-se primeiro um substantivo de relacionamento e um verbo de aproximação ou um verbo de ação;
- Atividades: são verificados os substantivos de atividades e também perguntas relacionadas a gostos, porém sem conotação sexual;

- Elogios: é verificado primeiramente se a linha apresenta um adjetivo. Existindo o adjetivo, verifica se há algum pronome de segunda pessoa ou um substantivo de informação para marcar a linha como um elogio;
- Dessensibilização Comunicativa: as linhas são marcadas como comunicação dessensibilizada se:
 - ✓ Apresentar um substantivo de comunicação dessensibilizada e um verbo de ação (*you have a hot body*);
 - ✓ Um substantivo de comunicação dessensibilizada, um verbo de comunicação dessensibilizada ou um verbo de ligação são encontrados juntos na mesma linha (*kiss your body*);
 - ✓ Aparecer um verbo de comunicação dessensibilizada com um pronome da segunda pessoa ou uma palavra de questionamento (*do you shave*);
 - ✓ Existir um adjetivo de comunicação dessensibilizada e um pronome da primeira ou segunda pessoa ou um verbo de ação (*get naked*);
- Reenquadramento: verifica-se a presença de um verbo de reenquadramento com um pronome da primeira ou segunda pessoa;
- Isolamento: uma linha é marcada como o estágio do isolamento se:
 - ✓ Um substantivo familiar aparecer com um verbo de aproximação ou palavras de questionamento;
 - ✓ Quando um adjetivo de isolamento aparecer com um pronome da segunda pessoa;
- Aproximação: verifica se na linha há um verbo de aproximação, não há substantivos de informação, há pronomes da primeira ou segunda pessoa ou um substantivo de aproximação.

Para avaliar o método desenvolvido os autores calcularam a Confiança entre Codificadores (*Intercoder Reliability*) usando o método de Holsti [HOL69] que compara as decisões compartilhadas entre dois codificadores fazendo uma análise de conteúdo. Para concluir que o método baseado em regras proposto por KONTOSTATHIS et al. [KON09b] apresentava uma melhora nos resultados se comparado ao método “caça-palavras” (Seção 3.1), os autores fizeram comparações entre dois codificadores humanos, o *ChatCoder1* e o

ChatCoder2. Com a implementação da técnica baseada em regras, os autores avaliaram ser possível obter uma melhora na confiabilidade máxima de 13,21% e na média de 5,81%.

3.3. Detecção dos estágios baseada em classificadores

Em MCGHEE et al. [MCG11] os autores fazem uso de algoritmos de aprendizagem de máquina para classificar as linhas de conversas de salas de bate-papo. Neste estudo, o modelo de comunicação adaptado por LEATHERMAN [LEA09] foi consolidado em apenas quatro estágios. Os autores consideraram os modelos anteriormente propostos muito complexos para conversações realizadas por meio de salas de bate-papo.

O novo modelo proposto por MCGHEE et al. (2011) baseia-se no fato de que os menores (crianças e adolescentes) simplesmente entram em salas de bate-papo e iniciam uma conversa com estranhos, logo o ganho de acesso é importante. Na sequência, para que o aliciamento de crianças e adolescentes aconteça deve-se levar em conta a idade do agressor e da vítima, e a aproximação que este faz tentando um encontro. Com isso, os estágios do novo modelo foram descritos da seguinte forma:

- Troca de informações pessoais: inclui informações de idade, sexo e localização. Número de amigos, namorados, gostos pessoais, sentimentos e atitudes sobre a família e relacionamentos também fazem parte deste estágio. A conversa que apresenta um aspecto de construção ou manutenção do relacionamento entre o agressor e a vítima também é marcada como informação pessoal. Os agressores utilizam destas informações para avaliar o quanto em comum eles possuem com a vítima e também para avaliar a estrutura familiar da vítima;
- Preparação: envolve o uso de termos sexuais e discussões sobre atos sexuais, eles estando ou não no contexto da conversa. A etapa de reenquadramento também é considerada aqui, onde o agressor se apresenta como voluntário a ensinar a vítima a ter certa desenvoltura em assuntos sexuais;
- Aproximação: obtenção de telefone, endereço, marcação de encontro ou apenas solicitação de segredo perante os pais ou autoridades na relação entre agressor e vítima estão neste estágio. Neste estágio também está o isolamento da vítima perante seus familiares e amigos;

- Outras linhas: linhas consideradas como não importantes, que aparecem no diálogo apenas para se manter uma comunicação ou então que são consideradas inocentes.

Para este estudo foram utilizadas 50 transcrições de conversas também disponíveis no *site Perverted-Justice* [PER03]. Estas conversas continham de 83 a 12.704 linhas e foram separadas, analisadas e classificadas por dois codificadores – apenas três conversas eram comuns a ambos os codificadores. Estas três conversas foram utilizadas para verificar a concordância de classificação dos codificadores. Após análise somente 33 transcrições foram utilizadas como base para o algoritmo de aprendizagem de máquina. Essa base foi utilizada para verificar a precisão da abordagem baseada em regras e também para o desenvolvimento e teste do modelo baseado em aprendizagem de máquina.

Como o número de estágios do modelo de comunicação foi reduzido neste trabalho, as regras para marcar cada linha do diálogo também foram adaptadas – as novas regras para identificar cada estágio são descritas abaixo:

- Troca de informações pessoais: as linhas são marcadas como informação pessoal se:
 - ✓ Contém um substantivo de aproximação, um substantivo de relacionamento e não contém um substantivo de informação pessoal;
 - ✓ Contém um substantivo de informação pessoal e também um verbo de ação ou uma palavra de questionamento;
 - ✓ Contém um substantivo de relacionamento e também um verbo de aproximação ou um verbo de ação;
 - ✓ Contém substantivo de atividade.
- Preparação: as linhas marcadas neste estágio apresentam:
 - ✓ Uma palavra de comunicação dessensibilizada;
 - ✓ Um substantivo de comunicação dessensibilizada e um verbo de ação ou de comunicação dessensibilizada;
 - ✓ Um verbo de comunicação dessensibilizada com um pronome da segunda pessoa ou uma palavra de questionamento;
 - ✓ Um adjetivo de comunicação dessensibilizada e um pronome da primeira ou segunda pessoa ou um verbo de ação;

- ✓ Um verbo de reenquadramento e um pronome da primeira ou segunda pessoa.
- Aproximação: as frases deste estágio contêm:
 - ✓ Um verbo de aproximação, não há um substantivo de informação, há um pronome da primeira ou segunda pessoa ou um substantivo de aproximação;
 - ✓ Um substantivo familiar e um verbo de aproximação ou de ação ou ainda um adjetivo de isolamento;
 - ✓ Um adjetivo de isolamento e um pronome da segunda pessoa.

O léxico utilizado em KONTOSTATHIS et al. [KON09b] foi também usado neste trabalho para desenvolver os atributos de entrada do algoritmo de aprendizagem de máquina. Apenas as linhas das conversações do agressor foram utilizadas. Estes atributos foram apresentados de duas formas ao algoritmo. Primeiramente na forma numérica e num segundo experimento na forma nominal.

Como vários experimentos foram realizados utilizando-se das 33 conversas para se comparar o método baseado em regras com o de aprendizagem de máquina, foi utilizado a acurácia (Seção 2.6) como métrica de avaliação para essa comparação. Os algoritmos de aprendizagem de máquina utilizados neste experimento foram o J48⁵, o JRip⁶ com $N = 1$ ⁷ e também com $N = 2$ e o IBk⁸ com $k = 1$ ⁹ e $k = 3$, todos existentes na ferramenta Weka [WEK12]. Foi usada também a técnica de *10-fold cross-validation* para validar os modelos. Quando as 33 conversas foram submetidas aos algoritmos individualmente, os autores afirmam que as técnicas de aprendizagem de máquina apresentaram um melhor desempenho se comparados às técnicas baseadas em regras. Mas como o objetivo dos autores neste experimento era comparar os dois métodos, as conversas foram agrupadas formando um único bloco de entrada. Quando isso foi realizado pelos autores, o desempenho dos classificadores foi muito semelhante ao da técnica baseada em regras, porém os atributos apresentados na forma numérica tiveram um desempenho melhor de cerca de 2%.

⁵ Classificador que generaliza uma árvore de decisão C4.5 com poda ou sem poda.

⁶ Classificador que implementa uma aprendizagem baseada em regras proposicionais, *Repeated Incremental Pruning to Produce Error Reduction* – RIPPER.

⁷ O peso total mínimo das instâncias em uma regra.

⁸ Classificador de “ k vizinhos mais próximos”. Pode-se selecionar um valor mais apropriado de k baseado na validação cruzada.

⁹ Número de “vizinhos” a ser utilizada.

3.4. Detecção de conversas com aliciamento

Em MORRIS [MOR13], o autor utiliza SVM para identificar as transcrições de conversas de salas de bate-papo que continham aliciamento. A experiência foi dividida em duas tarefas: a primeira em identificar conversas de aliciamento e a segunda em identificar os enunciados (do inglês *utterances*) que apresentavam aliciamento. A base de dados para esse trabalho apresentava 97.689 autores, sendo que neste caso, entende-se como autores todos os indivíduos que interagem no diálogo. Dentro desta base, apenas 142 autores eram previamente rotulados como aliciadores, pois suas transcrições de conversas foram retiradas do site *Perverted-Justice* [PER03]. Esses 142 autores formaram um conjunto C1. Mais dois conjuntos faziam parte do experimento. O conjunto C2 era composto pelos “falsos negativos”, que continham autores de conversas sobre diferentes tópicos e o conjunto C3 continha os “falsos positivos”, composto por pessoas adultas que conversavam sobre sexo, mas não caracterizado como aliciamento. Como cada conjunto era extraído de fontes diferentes, com aspectos diferentes, seria fácil descobrir as conversas de aliciamento. Para minimizar esse problema foram criadas manualmente as seguintes características:

- Número de participantes na conversa: os conjuntos C1 e C3 apresentam no máximo dois participantes. O conjunto C2 podia conter mais participantes na conversa;
- Número de conversas: No conjunto C3, por serem anônimos, os autores participavam de várias conversas ao mesmo tempo. Já no conjunto C1 eles participavam de uma conversa por vez, pois precisavam se identificar ao entrar nas salas de bate-papo;
- Número de participantes em conversas distintas.

Para treinar um modelo que identificasse as conversas de aliciamento, MORRIS [MOR13] transformou cada autor em um vetor de características. Esses vetores continham características léxicas e características comportamentais. Para as características léxicas foi utilizada a técnica “*bag-of-words*” com o TF-IDF (*Term Frequency – Inverse Document Frequency*) e também um conjunto de *unigrams* (1 palavra) e *bigrams* (2 palavras consecutivas).

Como o aliciador e a vítima apresentam uma simetria nos diálogos, algumas características não léxicas foram levadas em conta para distingui-los. Essas características

formaram o conjunto de características comportamentais, e foram fortemente relacionados como um comportamento de aliciamento. São elas:

- *NMessages*: O número total de mensagens enviadas pelo autor no conjunto e
- *NConversations*: O número total de conversas que o autor participava no conjunto.

Como o desbalanceamento na base era consideravelmente grande, outras características foram consideradas em três conjuntos diferentes para tentar melhorar o resultado da classificação, a saber:

- Iniciação:
 - ✓ *Initiations*: É o número de vezes que um autor inicia uma conversa, enviando uma mensagem de saudação, tais como "hey" ou "what's up";
 - ✓ *Initiation Rate*: É a característica '*Initiations*', porém normalizada pelo número de conversas;
 - ✓ *Questions*: É o número de vezes que o autor faz uma pergunta. Neste caso, uma pergunta é definida como qualquer mensagem que termine com um ponto de interrogação ou que tenha teor interrogatório;
 - ✓ *Question Rate*: Assim como a característica '*Initiation Rate*', é a característica '*Questions*' normalizada pelo número de mensagens;
- Atenção: É o conjunto de características que correspondem à tentativa do autor em manter uma conversa:
 - ✓ *Response Time*: É o tempo compreendido entre o envio da mensagem de um autor e o envio da resposta do outro autor e
 - ✓ *Repeat Messages*;
- Domínio da Conversa:
 - ✓ *Message Ratio*: É o número de mensagens por autor e
 - ✓ *Wordcount Ratio*: É o número de palavras por autor.

Na tarefa de classificação da conversa, MORRIS [MOR13] fez o uso do SVM com um *kernel* RBF. A classificação seguiu os seguintes passos:

- Foi criado um vetor de característica para cada autor;
- Treinou-se o SVM com o conjunto de treinamento;
- Foram rotulados os autores do conjunto de teste com o modelo criado com o conjunto de treinamento;
- Foram desenvolvidas algumas aplicações pós-classificatórias para apresentar os resultados.

Para a tarefa de classificação do enunciado, todas as conversas classificadas como de aliciamento na tarefa anterior serviram de entrada para a classificação do enunciado. Para essas conversas, cada termo do enunciado teve um peso atribuído em relação ao seu teor de aliciamento. Cada enunciado apresentava uma pontuação igual à soma dos pesos de todos os *unigram* e *bigram* do enunciado. Foram indicados como enunciados de aliciamento todos os enunciados com uma pontuação acima de um limite determinado pelo autor.

Uma lista de palavras também foi construída manualmente por MORRIS [MOR13]. Essa lista continha 122 *n-grams* que indicavam, automaticamente, um enunciado como sendo de aliciamento. A lista de palavras continha apenas termos que eram sexualmente explícitos. MORRIS [MOR13] ainda enfatizou os enunciados muito curtos (constituídos por quatro ou menos *n-grams* separados por espaços). Essa ênfase foi baseada no pressuposto de que os enunciados curtos podem suficientemente transmitir conteúdos de aliciamento.

Como resultados deste trabalho, o autor entendeu que o *Precision* não era a escolha adequada como métrica de avaliação devido ao desequilíbrio das classes e, portanto, optou por utilizar o *F-Measure*. A medida *F1-Measure* dá um peso igual ao *Precision* e ao *Recall*, e foi utilizada como métrica de avaliação por MORRIS [MOR13].

Cabe detalhar os resultados e métricas utilizadas por MORRIS [MOR13] visto ser este o trabalho mais recente e relacionado ao tema do presente trabalho. Além disto, estas análises nortearam os experimentos e conduta de análise de resultados do método proposto no presente trabalho. Assim, para a realização dos testes, MORRIS [MOR13] utilizou os conjuntos de características em separado e combinados, para verificar qual apresentaria melhores resultados. Na tarefa de classificação das conversas, foram utilizados os parâmetros de custo

C e γ pré-definidos para o LibSVM [CHA11]. Como resultado obteve-se $Precision = 0,91$, $Recall = 0,28$ e $F1-Measure = 0,43$. A grande disparidade entre o $Recall$ e o $Precision$ indicou que era necessário penalizar os erros de uma classe sobre as demais. O parâmetro Peso (w) do LibSVM então foi ajustado para 15, penalizando os falsos negativos 15 vezes mais do que falsos positivos. Com esse parâmetro ajustado, os resultados foram $Precision = 0,63$, $Recall = 0,65$ e $F1-Measure = 0,64$. MORRIS [MOR13] também fez o uso do *grid-search* para otimizar a configuração dos parâmetros C e γ , variando-os em uma escala logarítmica. Foi estabelecido um $C = 100$ e $\gamma = 0,0001$. Com essa configuração, o melhor resultado foi $Precision = 0,95$, $Recall = 0,74$ e $F1-Measure = 0,83$, considerando apenas os vetores de características construídos com as características léxicas. Na tarefa de classificação dos enunciados, o melhor resultado foi $Precision = 0,45$, $Recall = 0,19$ e $F1-Measure = 0,27$.

Outro trabalho relevante é o realizado por PARAPAR et al. [PAR12], o qual também fez uso do SVM para identificação de conversas de aliciamento e enunciados que continham aliciamento. Com uma base que continha 97.689 indivíduos, sendo 142 identificados como aliciadores, os autores selecionaram as conversas destes aliciadores e separaram seus enunciados para formar uma base única. Para aplicar o SVM, essas bases foram transformadas em vetores de características utilizando o TF-IDF e também fazendo o uso de uma análise psicolinguística com a utilização do *software Linguistic Inquiry and Word Count (LIWC)*. Para aprimorar um pouco mais o conjunto de características, algumas foram criadas baseadas no *chat*.

Para as características construídas utilizando-se o TF-IDF, PARAPAR et al. (2012) não considerou as palavras que apareciam menos de 10 vezes e também as palavras que continham mais de 20 caracteres. Também foram considerados *bigrams* e *trigrams*, porém apenas aqueles que apareciam mais de 3 vezes na conversa. No caso dos *bigrams* e *trigrams*, foram retirados aqueles que apareciam com mais de 40 caracteres. Foram testadas várias combinações com as características extraídas do TF-IDF.

Para PARAPAR et al. [PAR12], o aliciamento também pode ser descoberto fazendo o uso de características psicolinguísticas. O *software LIWC* faz a análise do texto e calcula o grau de utilização de diferentes categorias de palavras. O LIWC processa entradas textuais e produz como saída algumas variáveis, tais como as dimensões linguísticas padrões, categorias de palavras com teor psicológico (por exemplo, afeto, ansiosidade), categorias de interesse pessoal (por exemplo, trabalho, casa, lazer) etc. Ao todo são 80 dimensões diferentes

apresentadas pelo LIWC. Todos os documentos da base de dados foram processados, sem alterações, para se obter as 80 características LIWC associadas para cada indivíduo.

Além das características extraídas com o TF-IDF e o LIWC, foram também definidas 11 características adicionais que representavam alguns aspectos globais relacionados com a atividade dos indivíduos nos *chats*. Essas características incluíam o número de pessoas contatadas por um dado indivíduo, a porcentagem de conversas iniciada por um indivíduo, a porcentagem de enunciados escritos por um determinado indivíduo etc. Os autores acreditavam que essas características poderiam revelar algumas tendências relacionadas ao aliciamento.

Como o desbalanceamento entre as classes era representativo, os autores analisaram o uso de três diferentes estratégias para minimizar esse problema, dentre elas a clonagem, ou seja, a repetição dos vetores de características da classe minoritária. Essa estratégia não foi implementada, pois aumentaria consideravelmente o conjunto de treinamento. Outra estratégia analisada foi a retirada dos vetores de características da classe majoritária. Neste caso o resultado não seria efetivo, pois reduziria muito o número exemplos. A estratégia escolhida por PARAPAR et al. [PAR12] foi ajustar os pesos da classificação para penalizar o erro de classificar um exemplo positivo como sendo negativo.

Como resultado, na tarefa de identificação de conversas de aliciamento os melhores resultados foram $Precision = 0,92$, $Recall = 0,78$ e $F1-Measure = 0,84$. Esses valores foram obtidos com o vetor de características formado pela combinação TF-IDF (*unigram+bigram*) + características adicionais.

Já para a tarefa de classificação dos enunciados, todas as conversas classificadas como de aliciamento na tarefa anterior tiveram seus enunciados marcados como de aliciamento, um a um. Neste caso, foram testados os vetores de características utilizando-se os conjuntos de características separadamente, TF-IDF e LIWC. As características adicionais não foram utilizadas nesta tarefa, pois correspondem aos aspectos globais da atividade de um indivíduo durante o bate-papo, não sendo válidas para os enunciados. Os resultados desta tarefa foram $Precision = 0,03$, $Recall = 0,51$ e $F1-Measure = 0,05$ para o TF-IDF e $Precision = 0,05$, $Recall = 0,71$ e $F1-Measure = 0,09$ para o LIWC.

Ainda, sobre detecção de conversas com aliciamento, VILLATORO-TELLO et al. [VIL12] apresenta uma abordagem para a identificação de aliciamento que consiste na identificação de conversas suspeitas, *Suspicious Conversations Identification* (SCI). Nesta

pesquisa os autores utilizaram o SVM e também uma Rede Neural com uma camada escondida de 10 neurônios para comparar os resultados.

Nenhum pré-processamento foi feito, porém foi realizada uma pré-filtragem para retirar da base de 66.928 conversas as conversas que apresentavam apenas um participante, as que apresentavam menos de seis intervenções por indivíduo e as conversas que tinham longas sequências de caracteres não reconhecidos. Com essa pré-filtragem, os autores reduziram a 10% a base original.

Para detectar a SCI, os autores rotularam como conversa suspeita todas as transcrições em que pelo menos um aliciador aparecia. Com isso, obtiveram 5790 conversas não-suspeita e 798 suspeitas. O vetor de característica foi criado fazendo o uso do TF-IDF. Nesta abordagem, os resultados com o SVM foram melhores que com a Rede Neural, apresentando valores de $Precision = 0,99$ e $F1-Measure = 0,95$.

3.5. Detecção do perfil do agressor

KONTOSTATHIS et al. [KON09a] utilizou o algoritmo de *k-Means* para detectar os diferentes perfis de agressores por meio de seus diálogos. Para este experimento foram utilizadas 288 transcrições de conversas disponíveis no site *Perverved-Justice* [PER03], porém foram separadas as linhas correspondentes ao agressor. Foram criadas 288 instâncias com oito atributos normalizados (padronizados), cada um representando cada estágio de comunicação. Ao variar os valores de k , o melhor resultado foi verificado para o valor de k igual a 4, sugerindo que há quatro tipos diferentes de agressores para a base de dados utilizada.

3.6. Distinções entre agressor e vítima

KONTOSTATHIS et al. [KON09a] fez um experimento usando o classificador J48 existente na ferramenta Weka [WEK12]. O J48 construiu uma árvore de decisão C4.5¹⁰ com poda para redução de erro, no intuito de distinguir as frases do agressor e da vítima. Neste

¹⁰ Estrutura composta por uma folha, indicando uma classe; um nó de decisão, que especifica um teste a ser realizado no valor de um atributo e um galho para cada resposta possível do teste, que leva para uma nova sub-árvore ou uma folha.

experimento foram utilizadas 16 conversas aleatórias disponíveis no *site Perverted-Justice* [PER03]. Para cada conversa as frases do agressor foram separadas das frases da vítima, gerando com isso duas instâncias. Foram contabilizadas as frases que apareciam para cada um dos estágios do modelo de comunicação, gerando oito atributos para cada instância. Esse experimento resultou numa taxa de acerto de 60%. O *three-fold cross-validation* do Weka foi utilizado para validar os resultados.

VILLATORO-TELLO et al. [VIL12] também apresenta uma abordagem utilizando SVM e também Rede Neural com uma camada escondida de 10 neurônios para fazer uma separação entre a vítima e o aliciador, *Victim From Predator disclosure* (VFP), nas conversas classificadas como suspeitas, *Suspicious Conversations Identification* (SCI).

Para detectar o VFP, os autores utilizaram as conversas rotuladas como suspeitas e separaram os enunciados do aliciador e da vítima. O vetor de características foi construído com o uso do TF-IDF. Como resultado, a Rede Neural teve um melhor desempenho se comparado ao SVM, apresentando taxas de *Precision* = 0,94 e *F1-Measure* = 0,94.

3.7. Considerações Finais

Os trabalhos descritos neste Capítulo utilizam alguns dos métodos que serviram de base na implementação do método de detecção de aliciamento sexual de crianças e adolescentes em salas de bate-papo proposto neste trabalho. Os trabalhos de MORRIS [MOR13], PARAPAR et al. [PAR12] e VILLATORO-TELLO [VIL12] utilizam o classificador SVM, porém é importante ressaltar que essa abordagem foi utilizada para a descoberta do aliciamento e a classificação dos enunciados de aliciamento, e não para a descoberta dos estágios de aliciamento em transcrições de conversas, como contemplado no método proposto e desenvolvido nesta Dissertação de Mestrado. Uma comparação destes trabalhos é apresentada na Tabela 3.1.

Para a detecção dos estágios de aliciamento nos enunciados de transcrições de conversas, KONTOSTATHIS et al. [KON09a] [KON09b] propôs o uso de filtro de palavras e também de um método baseado em regras. Já MCGHEE et al. [MCG11] fez o uso de classificadores como o J48, o JRip e o IBk. Neste contexto, MCGHEE conseguiu os melhores resultados, com até 82% de *accuracy*.

Tabela 3.1 - Comparativo de trabalhos utilizando SVM

| Autores | Precision | Recall | F-Measure |
|--|-----------|--------|-----------|
| Classificação dos enunciados de aliciamento | | | |
| MORRIS | 0,45 | 0,19 | 0,27 |
| PARAPAR et al. | 0,03 | 0,51 | 0,05 |
| Descoberta do Aliciamento | | | |
| MORRIS | 0,95 | 0,74 | 0,83 |
| PARAPAR et al. | 0,92 | 0,78 | 0,84 |
| VILLATORO-TELLO et al. | 0,99 | 0,91 | 0,95 |

Deste modo, o método de detecção de aliciamento sexual de crianças e adolescentes apresentado no próximo Capítulo utiliza de algumas abordagens aqui apresentadas e descritas, além disso é dado um passo adicional no sentido de usar estes resultados para tentar predizer se a conversa se encaminha para o aliciamento, ou seja, determinar a susceptibilidade de que o aliciamento venha a se realizar em termos do *modus operandi* do aliciador e da Teoria da Comunicação Ludibriante (*Luring Communication Theory*).

Capítulo 4

Método Proposto

Este trabalho está baseado na identificação dos estágios de comunicação entre um possível agressor e sua vítima visando à detecção de aliciamento sexual de crianças e adolescentes em salas de bate-papo *online*. A análise é feita a partir de transcrições de conversas de salas de bate-papo, de modo a obter a susceptibilidade à ocorrência do abuso sexual de crianças e adolescentes determinada por meio da combinação dos estágios de comunicação.

O método desenvolvido utiliza o esquema apresentado na Figura 4.1. A partir de uma nova transcrição de conversa, não apresentada na fase de treinamento, foi realizado a ‘Preparação da Base de Conversação’ para separar os enunciados do agressor. Para a representação destas frases em vetores de características foi realizado o ‘Pré-Processamento’, seguido da ‘Extração de Características’.

Os vetores de características foram submetidos ao SVM, que apresentou como saída o valor da predição de cada vetor com sua classificação correspondente, associada a uma taxa de acurácia. Essa sequência da predição se transformou na sequência de observações necessárias ao HMM, que a partir de um modelo pré-estabelecido avalia a probabilidade de tal sequência de observação pertencer ao modelo.

A combinação da acurácia do SVM e a probabilidade do HMM são entendidas como a probabilidade DE SUSCEPTIBILIDADE AO ABUSO SEXUAL. Essa susceptibilidade é mostrada ao usuário como uma alerta com graus de intensidade leve, moderado ou severo. Deve-se destacar que esta combinação de técnicas é a contribuição mais importante deste trabalho, visto que os autores relacionados no Capítulo 3 não implementaram métodos que

possibilitem este tipo de resultado e, portanto, este tipo de análise do problema do aliciamento.

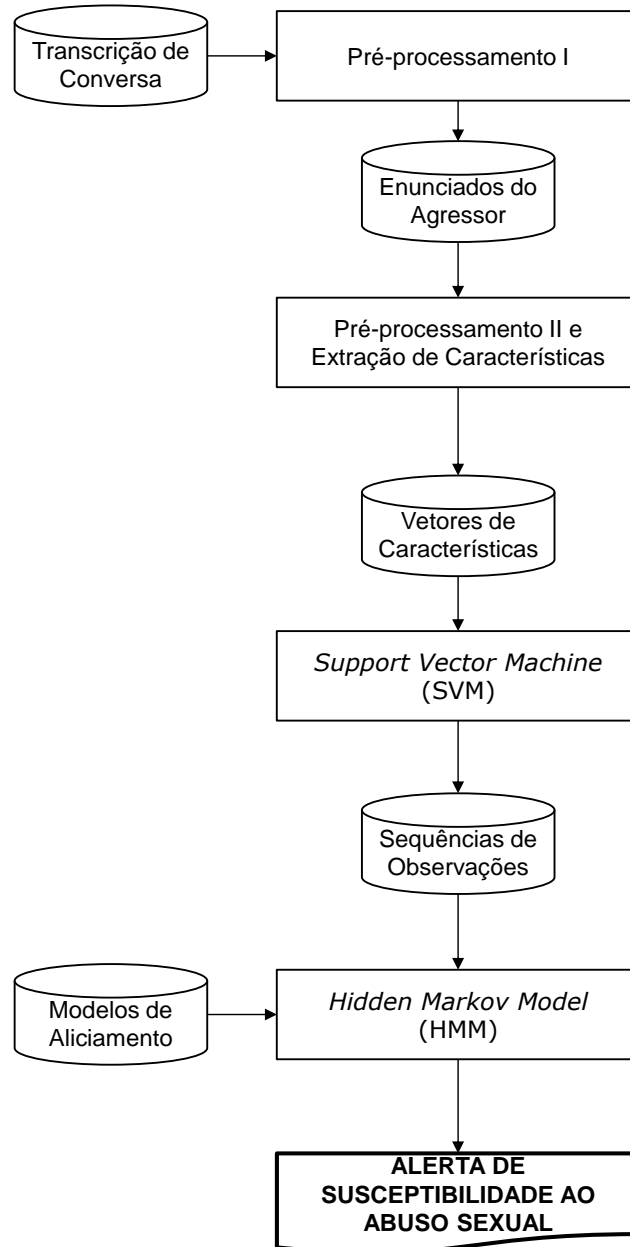


Figura 4.1 - Sequenciamento para Aplicação do Método Proposto

Para alcançar o método proposto apresentado na Figura 4.1 iniciou-se o trabalho fazendo o uso de uma técnica baseada em reconhecimento de linguagem natural (LSA) para a identificação dos estágios de comunicação, conforme mostra a Figura 4.2. Os resultados com o uso do LSA, que serão apresentados no Capítulo 5, não foram satisfatórios. Desta forma,

uma segunda abordagem foi desenvolvida, baseada num classificador híbrido (SVM e HMM), que resultou no método desenvolvido para este trabalho. Esse método pode ser observado na Figura 4.3 e se referente a etapa de treinamentos dos modelos. A seguir são detalhados os componentes de cada abordagem, mostrados na Figura 4.2 e na Figura 4.3. Esta abordagem tem também a vantagem de ser independente de língua, ou seja, obtida uma base de dados (*corpus*) em Português, pode-se utilizar o mesmo método para classificação dos estágios.

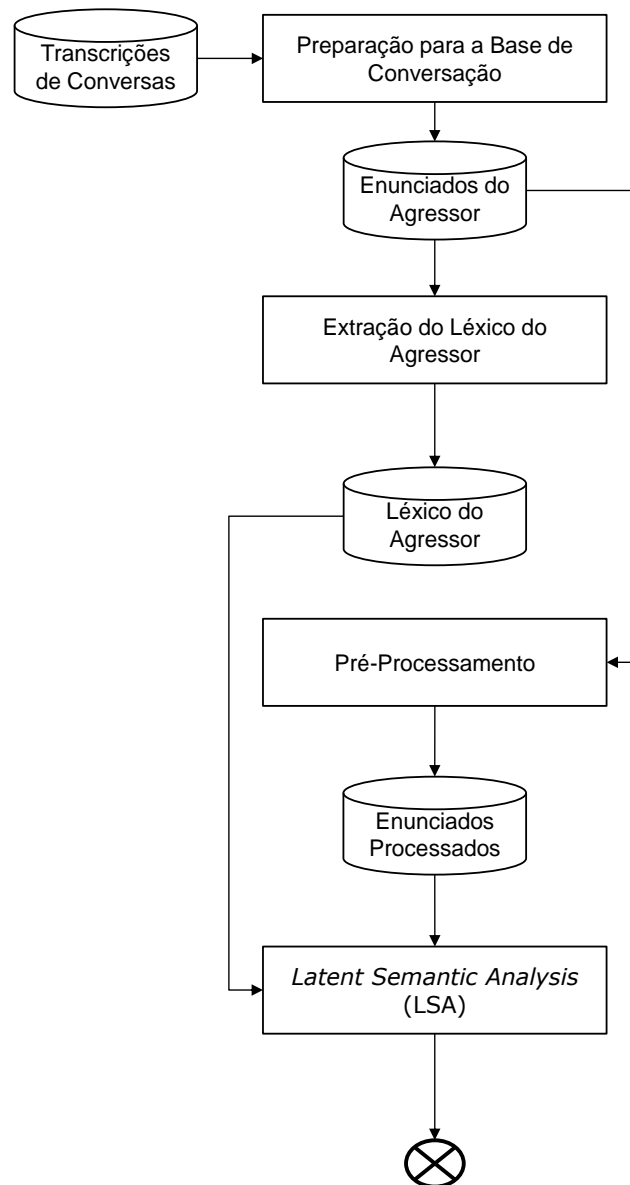


Figura 4.2 – Método estudado utilizando o LSA

4.1. Transcrição de Conversas: Base de Dados

Para que fosse possível identificar os estágios em conversas de salas de bate-papo foi necessária uma base de transcrição dessas, contendo casos reais de aliciamento sexual de crianças e adolescentes. O site *www.perverted-justice.com* [PER03] apresenta centenas de transcrições de diálogos entre agressores sexuais (*predators*) e suas vítimas. Fundado em meados de 2003 o site tem como objetivo identificar agressores sexuais que usam de meios virtuais para fazer vítimas.

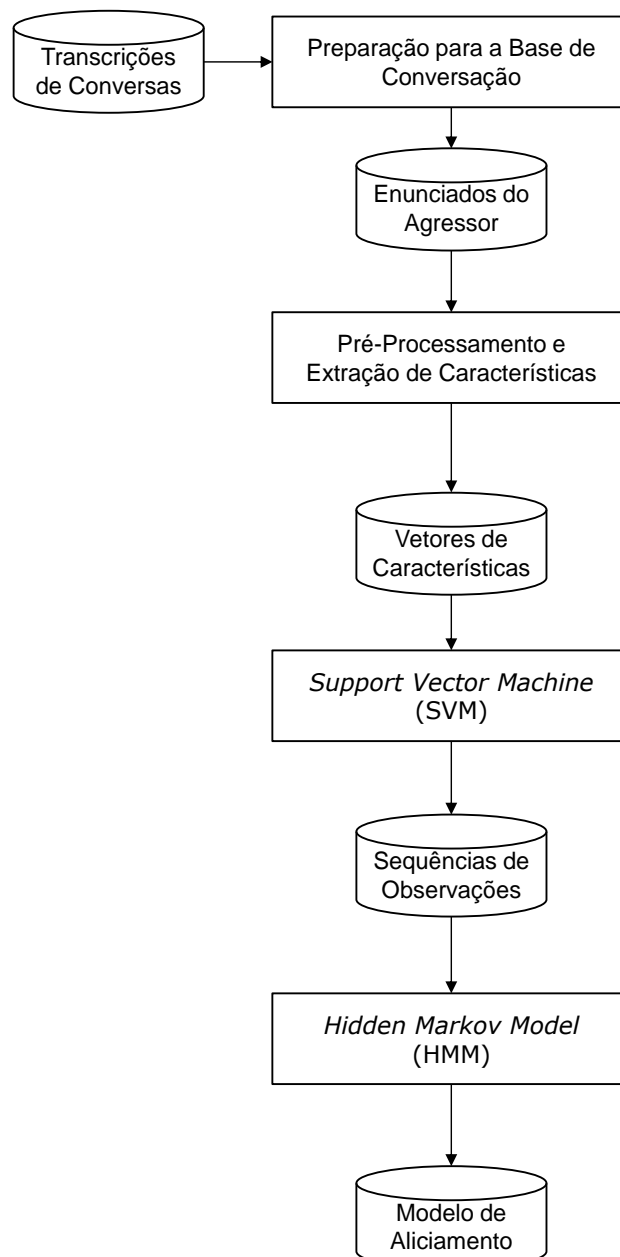


Figura 4.3 – Método desenvolvido para a fase de treinamento utilizando o SVM e HMM

Voluntários cadastrados no *Perverted-justice* (se passando por menores) respondem às abordagens feitas por agressores com o objetivo de iniciar um relacionamento sexual. Em parceria com a polícia dos EUA é “armado o flagrante”, visando confirmar o aliciamento virtual que resultará na prisão e condenação do agressor. A transcrição do diálogo na sala de bate-papo que resultou na prisão do agressor é disponibilizada no *site*. Desde junho de 2004, 550 agressores já foram condenados e as transcrições das conversas divulgadas ao público.

Algumas transcrições do *site* www.perverted-justice.com (visão geral da página inicial, Figura 4.4) formaram a base de dados deste trabalho.



Figura 4.4 - Visão geral da página inicial do *site* www.perverted-justice.com [PER03]

4.1.1. Análise das Transcrições de Conversas

Para realizar os experimentos, 40.396 linhas de enunciados de transcrições de conversas foram analisadas por três analistas. Estes analistas eram bolsistas da PUCPR (Pontifícia Universidade Católica do Paraná), alunos de PIBIC (Programa Institucional de Bolsas de Iniciação Científica). Dois deles eram estudantes do último ano de Ciência da Computação e o terceiro era estudante do último ano de Direito. Os analistas fizeram um prévio auto treinamento para uniformizar o entendimento dos estágios de comunicação definidos por LEATHERMAN [LEA09]. Tal treinamento consistiu no estudo do trabalho de LEATHERMAN (2009) e também na rotulação de poucos enunciados de transcrições de conversas. Ao surgir dúvidas sobre a rotulação era feita uma discussão, entre os próprios analistas, baseada no trabalho de LEATHERMAN (2009), para chegar a um consenso sobre o estágio de comunicação presente no enunciado analisado.

A análise das transcrições de conversas se fez necessária para associar a cada enunciado os estágios de aliciamento para o contexto virtual de aliciamento sexual de crianças e adolescentes. Os 40.396 enunciados foram observados em 37 conversas escolhidas aleatoriamente no *site Perverted-Justice* [PER03]. Uma das 37 transcrições de conversas foi separada e rotulada por todos os analistas para garantir o comum entendimento dos estágios de aliciamento; tal transcrição da conversa continha 776 enunciados. Esta conversa apresentou uma concordância de 80% na rotulação dos estágios a cada enunciado.

Para cada um dos estágios de aliciamento observado foi atribuído um valor numérico conforme ilustra a Tabela 4.1. Nos enunciados em que não foi possível observar nenhum estágio foi marcada a sigla NEO – Nenhum Estágio Observado.

Tabela 4.1 - Representação numérica para os estágios de aliciamento

| ESTÁGIO | REPRESENTAÇÃO |
|---------------------------------------|---------------|
| <i>Acesso</i> | 1 |
| <i>Informações Pessoais</i> | 2 |
| <i>Informações de Relacionamento</i> | 3 |
| <i>Atividades</i> | 4 |
| <i>Elogios</i> | 5 |
| <i>Dessensibilização Comunicativa</i> | 6 |
| <i>Reenquadramento</i> | 7 |
| <i>Isolamento</i> | 8 |
| <i>Aproximação</i> | 9 |
| <i>NEO - Nenhum Estágio Observado</i> | 10 |

4.2. Preparação da Base de Conversação

As transcrições das conversas existentes no *site Perverted-Justice* [PER03] apresentam o diálogo entre o agressor e a vítima, sendo que um fragmento de transcrição pode ser visto na Figura 4.5.

Para que as técnicas de detecção dos estágios fossem empregadas foi necessário preparar o texto das transcrições (Preparação da Base de Conversação da Figura 4.3), separando as linhas da conversa do agressor das linhas da conversa da vítima (um exemplo pode ser visto na Figura 4.6). Um algoritmo foi implementado (em linguagem Java) para a remoção de *tags* e posterior separação dos enunciados (agressor e vítima) das conversas com extensão *html* obtidas no *site Perverted-Justice*.

```

cobbler1976 (02/10/08 9:49:05 PM): what do ur panties look like?
that_polish_chick94 (02/10/08 9:49:26 PM): black bikini type
cobbler1976 (02/10/08 9:49:31 PM): sexy
that_polish_chick94 (02/10/08 9:49:35 PM): ty
cobbler1976 (02/10/08 9:50:04 PM): are u horny?
that_polish_chick94 (02/10/08 9:50:45 PM): i dunno not really
cobbler1976 (02/10/08 9:51:18 PM): i could get u horny if i was there
that_polish_chick94 (02/10/08 9:51:25 PM): yeah
cobbler1976 (02/10/08 9:51:34 PM): yep
cobbler1976 (02/10/08 9:52:21 PM): well then?
that_polish_chick94 (02/10/08 9:52:34 PM): what
cobbler1976 (02/10/08 9:52:57 PM): will u show me my cum in ur mouth?
that_polish_chick94 (02/10/08 9:53:12 PM): yeah if u wanted
cobbler1976 (02/10/08 9:53:31 PM): i could fuck u?
that_polish_chick94 (02/10/08 9:53:47 PM): yeah
cobbler1976 (02/10/08 9:54:12 PM): condom right?
that_polish_chick94 (02/10/08 9:54:28 PM): i would get preg if u didnt wouldnt i?
cobbler1976 (02/10/08 9:54:37 PM): yep
that_polish_chick94 (02/10/08 9:54:44 PM): then yeah
cobbler1976 (02/10/08 9:55:01 PM): when?
that_polish_chick94 (02/10/08 9:55:52 PM): when my mom goes outta town
that_polish_chick94 (02/10/08 9:56:00 PM): she goes sometimes
that_polish_chick94 (02/10/08 9:56:09 PM): for like at least a day
that_polish_chick94 (02/10/08 9:56:24 PM): i dunno when yet tho
cobbler1976 (02/10/08 9:56:30 PM): ok
cobbler1976 (02/10/08 9:57:03 PM): just sex or party and sex?
that_polish_chick94 (02/10/08 9:57:40 PM): i dunno both? lol
cobbler1976 (02/10/08 9:57:48 PM): sure...lol
cobbler1976 (02/10/08 9:57:54 PM): drinks?

```

Figura 4.5 – Fragmento de transcrições das conversas de aliciamento [PER03]

A partir deste ponto do trabalho somente os enunciados do agressor foram considerados. As linhas de enunciados da vítima foram desconsideradas, pois um trabalho publicado anteriormente por SANTIN et al. [SAN12] mostrou que não existe vantagem no uso dos enunciados de ambos, visto que uma conversa só progride para um aliciamento se o agressor consegue o seu objetivo. Isto é, a evolução dos estágios de aliciamento do agressor só acontece se a vítima o acompanha no desdobramento do aliciamento. Assim, optou-se por trabalhar apenas com os enunciados do agressor, pois os resultados são equivalentes a trabalhar com ambos, mas a complexidade de processamento é reduzida a metade.

4.3. Extração do Léxico do Agressor

Na etapa de identificação do estágio de comunicação por meio do LSA (*Latent Semantic Analysis*) um léxico foi utilizado para auxiliar na geração final de probabilidades que associassem a porção do texto analisado ao estágio de comunicação, usando esse léxico de palavras.

```

cobbler1976 (02/10/08 9:49:05 PM): what do ur panties look like?
that_polish_chick94 (02/10/08 9:49:26 PM): black bikini type
cobbler1976 (02/10/08 9:49:31 PM): sexy
that_polish_chick94 (02/10/08 9:49:35 PM): ty
cobbler1976 (02/10/08 9:50:04 PM): are u horny?
that_polish_chick94 (02/10/08 9:50:45 PM): i dunno not really
cobbler1976 (02/10/08 9:51:18 PM): i could get u horny if i was there
that_polish_chick94 (02/10/08 9:51:25 PM): yeah
cobbler1976 (02/10/08 9:51:34 PM): yep
cobbler1976 (02/10/08 9:52:21 PM): well then?
that_polish_chick94 (02/10/08 9:52:34 PM): what
cobbler1976 (02/10/08 9:52:57 PM): will u show me my cum in ur mouth?
that_polish_chick94 (02/10/08 9:53:12 PM): yeah if u wanted
cobbler1976 (02/10/08 9:53:31 PM): i could fuck u?
that_polish_chick94 (02/10/08 9:53:47 PM): yeah
cobbler1976 (02/10/08 9:54:12 PM): condom right?
that_polish_chick94 (02/10/08 9:54:28 PM): i would get preg if u didnt wouldnt i?
cobbler1976 (02/10/08 9:54:37 PM): yep
that_polish_chick94 (02/10/08 9:54:44 PM): then yeah
cobbler1976 (02/10/08 9:55:01 PM): when?
that_polish_chick94 (02/10/08 9:55:52 PM): when my mom goes outta town
that_polish_chick94 (02/10/08 9:56:00 PM): she goes sometimes
that_polish_chick94 (02/10/08 9:56:09 PM): for like at least a day
that_polish_chick94 (02/10/08 9:56:24 PM): i dunno when yet tho
cobbler1976 (02/10/08 9:56:30 PM): ok
cobbler1976 (02/10/08 9:57:03 PM): just sex or party and sex?
that_polish_chick94 (02/10/08 9:57:40 PM): i dunno both? lol
cobbler1976 (02/10/08 9:57:48 PM): sure...lol
cobbler1976 (02/10/08 9:57:54 PM): drinks?

```

Figura 4.6 - Separação dos enunciados do agressor e da vítima

Após a rotulação pelos analistas e separação dos enunciados, os enunciados dos agressores foram analisados manualmente novamente pelos analistas agora com o objetivo de se encontrar, em cada enunciado, as palavras de destaque que remetesse ao estágio de aliciamento, para que fosse possível formar o léxico de palavras de aliciamento que caracterizasse os estágios descritos por LEATHERMAN [LEA09]. Foram definidas 561 palavras que seguiram a distribuição apresentada na Tabela 4.2.

Tabela 4.2 - Distribuição do léxico de palavras do agressor por estágio de comunicação

| Estágio de Comunicação | Total de Palavras |
|--------------------------|-------------------|
| <i>Outros Estágios</i> | 299 |
| <i>Elogio</i> | 61 |
| <i>Dessensibilização</i> | 82 |
| <i>Reenquadramento</i> | 11 |
| <i>Isolamento</i> | 27 |
| <i>Aproximação</i> | 81 |

4.4. Pré-Processamento e Extração de Características

Para que os diálogos da sala de bate-papo fossem classificados pelas abordagens apresentadas neste trabalho, um segundo pré-processamento no texto se fez necessário (Pré-processamento da Figura 4.3). Esse pré-processamento consistiu em padronizar em caixa

baixa (minúsculas) as letras de cada enunciado, remover os caracteres especiais (sinais de pontuação, hífen e números), remover as *stopwords* (artigos, preposições, etc. da língua inglesa), além de realizar a tokenização, separação dos enunciados em palavras isoladas.

Como o texto dos enunciados das conversas de sala de bate-papo usa linguagem coloquial, uma segunda etapa neste pré-processamento foi a correção das palavras dos enunciados utilizando o método de similaridade de palavras *N-gram* (uma sequência de N letras). A correção foi feita em abreviações de palavras ou até mesmo nas escritas de maneira errônea. Assim, cada palavra foi dividida em N letras ($N = 2$) e cada *N-gram* da palavra foi pesquisada num léxico extraído do WordNet¹¹. Se a palavra existisse no léxico era considerada válida, caso contrário a mesma palavra era pesquisada numa lista de relações para encontrar uma palavra correlacionada a essa. A lista de relações foi criada manualmente pelos analistas e continha a forma correta de se escrever abreviações e palavras.

Se a palavra também não existisse nessa lista de relações, era feita uma comparação *N-gram* com todas as palavras do léxico que começassem com a mesma letra da palavra a ser verificada, e neste caso era retornada a palavra com o maior valor de similaridade. Porém, se o valor de similaridade fosse menor que 80% era feito o cálculo *N-gram* com todo o léxico, retornando a palavra com maior valor de similaridade. O percentual de 80% foi estabelecido pois com um limiar mais baixo palavras com contextos diferentes eram atribuídas àquela que se estava querendo corrigir, fazendo falsas correções. Tal pré-processamento foi implementado num algoritmo em linguagem Java.

Ao submeter dados textuais a classificadores como o SVM, além do pré-processamento apresentado acima é necessário representar os textos como um vetor de características. O modelo TF-IDF (*Term Frequency – Inverse Document Frequency*) foi utilizado para fazer essa representação. Além disso, uma redução na dimensionalidade se fez necessária. Para reduzir a dimensão e os ruídos dos enunciados foi realizada uma seleção das palavras dos enunciados para a construção dos vetores de características. Essa seleção foi baseada em três filtros, a saber:

- Quantidade de caracteres de uma palavra;
- Número mínimo de ocorrências da palavra na base de dados;
- Ganho de Informação (GI) da palavra acima de um limiar pré-estabelecido.

¹¹ Disponível em <http://wordnet.princeton.edu/>

4.5. Classificação dos Estágios utilizando o *Latent Semantic Analysis* (LSA)

O método LSA utilizado neste trabalho foi uma adaptação feita do trabalho de DOSCIATTI et al. [DOS12] para o contexto de aliciamento. Inicialmente foi implementado um algoritmo (em linguagem Java) para buscar os enunciados que apresentassem as palavras existentes no léxico do agressor, nas transcrições de conversas, para realizar o treinamento do LSA. Foram extraídos 50 enunciados para cada uma das palavras do léxico para compor o conjunto de treinamento do LSA. Se não existissem 50 enunciados para determinada palavra, era extraída da base a quantidade total de enunciados existente para essa palavra.

Os estágios de 1 a 4 (de acesso a atividades) foram considerados como um único estágio para o treinamento do LSA, denominado de “outros estágios”. Além disso, não foram utilizados para o treinamento enunciados em que nenhum estágio havia sido observado (NEO). Tal agrupamento foi feito baseado na observação de que estes estágios aparecem com maior frequência nas transcrições de conversas de salas de bate-papo e não caracterizam claramente aliciamento, podendo ser confundidos com uma conversa qualquer de salas de bate-papo.

4.6. Classificação dos Estágios utilizando o *Support Vector Machine* (SVM)

Os enunciados já transformados em vetores de características foram submetidos ao SVM usando a biblioteca LibSVM desenvolvida por CHANG e LIN [CHA11]. Essa biblioteca foi integrada ao Weka [WEK12] por apresentar algumas estatísticas de avaliação dos resultados que foram usados neste trabalho, tais como Matriz de Confusão, *Precision*, *Recall* e *F-measure*.

Primeiramente os arquivos de treinamento e teste foram submetidos ao LibSVM para que fosse possível definir, por meio do *script* `easy.py`, os parâmetros C (custo) e γ (gamma). Tais parâmetros são utilizados para aumentar a precisão do classificador para os dados que não pertencem ao conjunto de treinamento. Esses parâmetros são utilizados para fazer o treinamento e o teste dos conjuntos de dados por meio do Weka.

A função *predict* retorna a predição feita pelo classificador para cada um dos vetores de características, informando a qual classe esses vetores pertencem. A sequência desta

predição é usada como entrada para o HMM (*Hidden Markov Model*) como sequência de observações.

4.7. Identificação das Sequências de Observações com o *Hidden Markov Model* (HMM)

Assim, finalmente a saída do SVM gera a sequência de observações que é submetida ao HMM para a obtenção da susceptibilidade ao abuso sexual de crianças e adolescentes.

Na fase de treinamento, para a modelagem do aliciamento foi adaptado o algoritmo de *Baum-Welch* implementado por FREITAS [FRE01]. Foi utilizado um modelo ergódico para a representação do modelo, visto que as conversações podem avançar e retroceder nos estágios de comunicação.

Para a fase de teste, foi adaptado o algoritmo de *Viterbi* também implementado por FREITAS [FRE01]. O modelo gerado durante o treinamento foi utilizado para avaliar as novas sequências de observações, computando a probabilidade de cada sequência estar contida no modelo.

4.8. Considerações finais

Neste Capítulo foi apresentado que a partir de uma abordagem inicial com o uso do LSA para a identificação dos estágios de comunicação surgiu o método desenvolvido neste trabalho. Esse método faz o uso de um classificador híbrido usando as técnicas de SVM e HMM. Foram descritas as etapas realizadas até gerar o alerta de susceptibilidade ao abuso sexual de crianças e adolescentes. No próximo Capítulo são apresentados os experimentos realizados, bem como a análise de seus resultados.

Capítulo 5

Experimentos Realizados e Análise dos Resultados

Esse Capítulo apresenta os testes que foram realizados no desenvolvimento do método proposto como prova de conceito para avaliar a sua viabilidade, bem como, a respectiva análise e comparação dos resultados obtidos.

5.1. Identificando os Estágios de Comunicação com o *Latent Semantic Analysis* (LSA)

Para testar a identificação dos estágios de comunicação por meio do LSA foram utilizadas transcrições de conversas que estavam disponíveis no *site Perverted-Justice* (www.perverted-justice.com) [PER03]. Foi utilizado um total de 16.017 enunciados do agressor. Esses enunciados foram divididos em duas bases, a saber: uma base de treinamento com 12.229 enunciados e uma base de teste com 3.788 enunciados. Além do conjunto de enunciados do agressor e sua pré-rotulação (Tabela 5.1), foram utilizados seis léxicos com palavras relacionadas aos estágios de comunicação definidos por LEATHERMAN [LEA09] existentes numa conversa com teor de aliciamento.

Tabela 5.1 – Exemplos de enunciados

| # | <i>Enunciado</i> | <i>Pré-Rotulação</i> |
|---|-----------------------------------|----------------------|
| 1 | youre hot for 13 | dessensibilização |
| 2 | flagler im in daytona | outros estágios |
| 3 | you got big nipples | dessensibilização |
| 4 | i sent you a photo | outros estágios |
| 5 | so you will be all alone | isolamento |
| 6 | well i will pick you up somewhere | aproximação |

Os estágios de 1 a 4 (que são muito similares aos estágios de qualquer conversa em salas de bate-papo, visto envolver os estágios de acesso, informações pessoais, informações de relacionamentos e atividades e, portanto, sem conotação sexual) foram considerados como um único estágio no treinamento do LSA.

Utilizando-se da base de treinamento, o LSA foi treinado e as matrizes e dados necessários foram gerados para a identificação dos estágios de comunicação de novos enunciados. Para avaliar o resultado do teste a Tabela 5.2 foi organizada de modo que se possa apontar os resultados, por exemplo, para cada enunciado relacionado à Tabela 5.1, e mais seis colunas, uma para cada estágio de aliciamento.

Tabela 5.2 – Matriz de saída do LSA com os estágios de aliciamento encontrados nos enunciados

| # | <i>Outros Estágios</i> | <i>Elogio</i> | <i>Dessensibilização</i> | <i>Reenquadramento</i> | <i>Isolamento</i> | <i>Aproximação</i> |
|---|------------------------|---------------|--------------------------|------------------------|-------------------|--------------------|
| 1 | 1,1674 | 3,4854 | 12,7476 | 9,8617 | 1,3785 | 4,4678 |
| 2 | 4,1842 | 1,932 | 2,4493 | 0,2956 | 2,8181 | 2,7502 |
| 3 | 6,0584 | 4,1499 | 17,2653 | 11,72 | 1,094 | 12,8309 |
| 4 | 4,644 | 4,4443 | 9,0232 | 1,3741 | 38,6062 | 19,8443 |
| 5 | 15,8715 | 17,5694 | 2,0275 | 9,1086 | 18,8359 | 9,3242 |
| 6 | 0,2644 | 4,6771 | 3,5961 | 1,0305 | 22,5485 | 20,3177 |

Nas matrizes de saída do LSA, usou-se como identificação do estágio de comunicação o maior valor de relação encontrado para cada enunciado, por exemplo, como destacado na Tabela 5.2. Lê-se o resultado da seguinte forma: para cada enunciado, foi comparado o estágio identificado com o que havia sido pré-rotulado. Por exemplo, na linha 1, o enunciado havia sido pré-rotulado como dessensibilização (Tabela 5.1) e foi identificado, por meio do LSA, como sendo dessensibilização (Tabela 5.2). Fazendo-se essa comparação para a base de dados de teste, a taxa de acerto foi de aproximadamente 20%. Tal resultado foi considerado insatisfatório, visto que os trabalhos que utilizaram o mesmo método [MAR10] [STR08] [DOS12], mesmo que em outro contexto, obtiveram taxa de acerto de aproximadamente 70%.

Uma tentativa para melhorar este resultado foi a inserção de frases na base de treinamento com as palavras do léxico que caracterizavam um estágio, porém os resultados permaneceram bastante aquém do esperado. Uma das hipóteses para esse resultado bastante insatisfatório no uso desta técnica é o tamanho reduzido dos enunciados. Alguns enunciados possuem apenas uma palavra, o que dificulta a identificação do estágio de comunicação dentro do espaço criado no universo de possibilidade gerado pelas palavras inseridas nos léxicos.

A identificação do estágio gera as sequências de observação que são utilizadas como entrada no HMM para produzir o Alerta de Susceptibilidade ao Abuso Sexual. Assim, devido ao seu baixo desempenho o LSA foi descartado como método para geração das sequências de observação dos estágios em uma conversação.

No intuito de buscar uma técnica que pudesse melhorar os resultados no contexto das características extraídas das conversas de salas de bate-papo, resolveu-se investir no classificador que tem apresentado resultados compatíveis com o que é esperado na maioria das aplicações em que é aplicado, o SVM (*Support Vector Machine*).

5.2. Identificando os Estágios de Comunicação com o *Support Vector Machine* (SVM)

Para testar a identificação dos estágios de comunicação por meio do SVM também foram utilizadas transcrições de conversas que estavam disponíveis no *site Perverted-Justice* (www.perverted-justice.com) [PER03]. Essa identificação foi dividida em duas etapas, a primeira funcionando como um filtro e a segunda como um classificador propriamente dito.

Na primeira etapa foram utilizados 8.153 enunciados do agressor. Deste total, 6.259 enunciados foram utilizados para treinar um classificador SVM para a identificação de sete classes: Outros Estágios (os estágios de 1 a 4 definidos por LEATHERMAN [LEA09]), Elogio, Dessensibilização, Reenquadramento, Isolamento, Aproximação, e a classe denominada “Nenhum Estágio Observado” - NEO. O objetivo deste teste foi separar a classe NEO (encontrada em maior quantidade nas transcrições de conversa e que não caracterizam aliciamento) para que a fronteira entre as demais classes ficasse mais bem definida e o classificador apresentasse um resultado mais preciso.

Foi executado o algoritmo *grid-search* da ferramenta LibSVM [CHA11] e os melhores parâmetros capazes de aumentar a precisão do classificador foi de custo (C) igual a 8 e *gamma* (γ) igual 0,03125.

Foram utilizadas 187 características para o treinamento do modelo de aliciamento nesta primeira etapa. A seleção dessas características foi feita utilizando-se a técnica TF-IDF [SAL88], e uma redução na dimensionalidade dos vetores de características foi realizada baseada nos seguintes parâmetros:

- Quantidade de caracteres de uma palavra = 5;
- Número mínimo de ocorrências da palavra na base de treinamento = 2;
- Ganho de Informação (GI) da palavra = 0,70.

Para avaliar o modelo obtido durante o treinamento foi usada a técnica de *cross-validation* com *fold 5*. A Matriz de Confusão resultante desse primeiro teste, que funcionou como um filtro de enunciados é apresentada na Tabela 5.3 e os demais indicadores (em média) para a avaliação desse classificador foram:

- *Precisão* = 0,65;
- *Revocação* = 0,67;
- *F1-measure* = 0,62.

Tabela 5.3 - Matriz de Confusão em *Cross-validation* para o 1º Classificador

| | <i>N</i> | <i>O</i> | <i>E</i> | <i>D</i> | <i>R</i> | <i>I</i> | <i>A</i> |
|------------------------------|----------|----------|----------|----------|----------|----------|----------|
| <i>NEO (N)</i> | 3151 | 94 | 23 | 73 | 6 | 17 | 70 |
| <i>Outros (O)</i> | 774 | 172 | 8 | 38 | 1 | 18 | 31 |
| <i>Elogio (E)</i> | 135 | 5 | 84 | 16 | 0 | 1 | 1 |
| <i>Dessensibilização (D)</i> | 241 | 17 | 7 | 527 | 9 | 4 | 5 |
| <i>Reenquadramento (R)</i> | 61 | 6 | 1 | 52 | 22 | 1 | 3 |
| <i>Isolamento (I)</i> | 79 | 11 | 1 | 2 | 1 | 59 | 6 |
| <i>Aproximação (A)</i> | 237 | 18 | 3 | 4 | 0 | 3 | 161 |

Ao submeter uma base de teste contendo 1.894 enunciados ao modelo obtido no treinamento obteve-se a Matriz de Confusão apresentada na Tabela 5.4 e seus demais indicadores (em média) foram:

- *Precisão* = 0,76;
- *Revocação* = 0,78;
- *F1-measure* = 0,75.

Os resultados não foram muito satisfatórios no treinamento porque as classes se encontravam desbalanceadas, o que realmente ocorre nas transcrições de conversas, visto que alguns estágios aparecem mais do que outros. Porém observa-se que ao testar o modelo com

outra base os resultados são mais consistentes, o que mostra que o modelo está bem construído, ou seja, genérico, mas discriminante.

Tabela 5.4 - Matriz de Confusão para a Base de Teste do 1º Classificador

| | <i>N</i> | <i>O</i> | <i>E</i> | <i>D</i> | <i>R</i> | <i>I</i> | <i>A</i> |
|--------------------------|----------|----------|----------|----------|----------|----------|----------|
| <i>NEO</i> | 1200 | 22 | 9 | 17 | 0 | 6 | 13 |
| <i>Outros</i> | 198 | 61 | 0 | 6 | 0 | 4 | 8 |
| <i>Elogio</i> | 11 | 0 | 4 | 0 | 0 | 0 | 0 |
| <i>Dessensibilização</i> | 74 | 9 | 1 | 170 | 1 | 2 | 2 |
| <i>Reenquadramento</i> | 3 | 0 | 0 | 0 | 1 | 0 | 0 |
| <i>Isolamento</i> | 13 | 1 | 0 | 0 | 0 | 8 | 0 |
| <i>Aproximação</i> | 26 | 0 | 0 | 0 | 0 | 0 | 24 |

Na segunda etapa o modelo do classificador foi criado utilizando 2.825 enunciados e com 6 classes (os registros classificados como NEO foram retirados deste teste): Outros Estágios, Elogio, Dessensibilização, Reenquadramento, Isolamento e Aproximação. O algoritmo *grid-search* da ferramenta LibSVM [CHA11] foi aplicado também neste caso, sendo que os parâmetros que fornecem os melhores resultados para o classificador foram: custo (*C*) igual a 32 e *gamma* (γ) igual 0,0078125.

Nesta etapa a base de treinamento tinha 246 características e foi extraída com a técnica TF-IDF, sendo a seleção feita com os mesmos parâmetros do primeiro teste. O modelo treinado foi avaliado por *cross-validation* com *fold 5*. A Matriz de Confusão apresentada na Tabela 5.5 e os demais indicadores (em média) para este teste foram:

- *Precisão* = 0,66;
- *Revocação* = 0,66;
- *F1-measure* = 0,65.

Tabela 5.5 - Matriz de Confusão em *Cross-validation* para o 2º Classificador

| | <i>O</i> | <i>E</i> | <i>D</i> | <i>R</i> | <i>I</i> | <i>A</i> |
|--------------------------|----------|----------|----------|----------|----------|----------|
| <i>Outros</i> | 808 | 29 | 59 | 6 | 42 | 98 |
| <i>Elogio</i> | 87 | 119 | 25 | 0 | 3 | 8 |
| <i>Dessensibilização</i> | 165 | 26 | 580 | 18 | 5 | 16 |
| <i>Reenquadramento</i> | 41 | 4 | 62 | 31 | 2 | 6 |
| <i>Isolamento</i> | 62 | 3 | 3 | 1 | 78 | 12 |
| <i>Aproximação</i> | 146 | 7 | 8 | 3 | 14 | 248 |

A base de testes deste segundo modelo é originária do resultado da primeira classificação, assim restou 308 enunciados após a primeira filtragem. Quando o modelo de treinamento foi testado obteve-se a Matriz de Confusão apresentada na Tabela 5.6 e os demais indicadores (em média) foram:

- *Precisão* = 0,87;
- *Revocação* = 0,85;
- *F1-measure* = 0,85.

Tabela 5.6 - Matriz de Confusão para a Base de Teste do 2º Classificador

| | <i>O</i> | <i>E</i> | <i>D</i> | <i>R</i> | <i>I</i> | <i>A</i> |
|--------------------------|----------|----------|----------|----------|----------|----------|
| <i>Outros</i> | 64 | 0 | 9 | 0 | 4 | 8 |
| <i>Elogio</i> | 0 | 4 | 0 | 0 | 0 | 0 |
| <i>Dessensibilização</i> | 13 | 2 | 161 | 3 | 2 | 4 |
| <i>Reenquadramento</i> | 0 | 0 | 0 | 1 | 0 | 0 |
| <i>Isolamento</i> | 2 | 0 | 0 | 0 | 8 | 0 |
| <i>Aproximação</i> | 0 | 0 | 0 | 0 | 0 | 23 |

Pode-se observar com estes resultados um comportamento para teste e treinamento muito similar ao da primeira etapa, porém com resultados mais expressivos. Se comparado ao trabalho de MCGHEE et al. [MCG11] (Tabela 5.7), em que os autores também utilizaram classificadores para a identificação dos estágios de comunicação, observasse que o SVM, que até o momento não havido sido usada para a classificação dos estágios apresentou um resultado similar ao encontrado na literatura.

Tabela 5.7 - Resultados do trabalho de MCGHEE et al. [MCG11] no uso de classificadores para a identificação dos estágios de comunicação

| | IBk k = 1 | IBk k = 3 | J48 | JRip N = 1 | JRip N = 1 |
|------------------------|----------------------|----------------------|------------|-----------------------|-----------------------|
| <i>Acurácia</i> | 82,94% | 76,47% | 80,72% | 74,94% | 74,77% |

A seguir será apresentado o teste realizado para se chegar ao grau de susceptibilidade ao abuso sexual aplicando o HMM.

5.3. Criando o Modelo de Aliciamento com o *Hidden Markov Model* (HMM)

Para gerar o modelo de aliciamento por meio do HMM foram utilizadas 29 transcrições de conversas (5.532 enunciados), apenas com os enunciados do agressor, que estavam disponíveis no site *Perverted-Justice* (www.perverted-justice.com) [PER03]. Essas transcrições foram previamente rotuladas pelos analistas, como já mencionado anteriormente, e foram divididas em dois conjuntos, a saber: o primeiro conjunto era composto por 20 transcrições que foram usadas para criar o modelo; o segundo conjunto continha o restante das nove transcrições e foi usado para validar o modelo estabelecido.

Para criar e validar o modelo foram descartados todos os enunciados pré-rotulados como “NEO – Nenhum Estágio Observado”. Além disso, seguindo a linha de estudo desenvolvida com a criação dos classificadores SVM, os estágios de 1 a 4 (estágio comuns a conversações sem aliciamento) foram considerados como um único estágio, denominado “Outros Estágios”. Após essas considerações, o restante de cada transcrição de conversa deu origem às sequências de observação. O tamanho das sequências variou de 59 a 396 observações.

O algoritmo de *Baum-Welch* foi utilizado para gerar o modelo de aliciamento. O modelo criado contém 227 estados e obteve uma taxa de acerto de 96% considerando as sequências de validação. Para testar o modelo, quatro novas transcrições de conversas (620 enunciados) foram submetidas ao algoritmo de *Viterbi*, que reconheceu tais conversas com uma taxa de acerto média de 90%, conforme mostra a Tabela 5.8. O tamanho dessas sequências de observações variou de 106 a 198.

Tabela 5.8 - Taxa de Acerto do HMM para as Sequências de Teste

| Sequência | Taxa de Acerto |
|--------------|----------------|
| 1 | 0,93 |
| 2 | 0,91 |
| 3 | 0,88 |
| 4 | 0,88 |
| Média | 0,90 |

Uma segunda base de teste foi submetida ao HMM para o reconhecimento da sequência de observações. Essa base era composta de outras quatro conversas (308 enunciados), que foram classificadas pelo SVM, gerando as sequências de observações. A

taxa de acerto médio para essas conversas foi de 97%, conforme pode ser observado na Tabela 5.9.

Tabela 5.9 - Taxa de Acerto do HMM para as Sequências de Teste geradas pelo SVM

| Sequência | Taxa de Acerto |
|--------------|----------------|
| 1 | 0,97 |
| 2 | 0,98 |
| 3 | 0,94 |
| 4 | 0,98 |
| Média | 0,97 |

Os resultados mostraram que uso do HMM ergódico se mostrou bastante promissor nesta etapa de classificação, pois é a única ferramenta de aprendizado de máquina que consegue fazer classificação com vetores de entrada de tamanho diferentes, necessário para poder prever a susceptibilidade ao abuso antes mesmo que todos os estágios do aliciamento sejam identificados.

5.4. Estabelecendo os Níveis de Susceptibilidade ao Abuso Sexual de Crianças e Adolescentes

Para estabelecer os níveis de susceptibilidade foi aplicada a combinação dos resultados gerados no uso do SVM para a identificação dos estágios de comunicação e do HMM na classificação das sequências de observações dos estágios geradas pelo SVM. Como o SVM apresenta uma acurácia média de 85%, este valor é tido como base para o alerta, visto que para medir a taxa de acerto precisamente seria necessário pré-rotular os enunciados para comparar os resultados. Tal estratégia só é viável no caso de análise de novas transcrições de conversas, não sendo aplicada para o monitoramento de conversas *online*.

Assim, considerando que o HMM faz o cálculo da taxa de reconhecimento das sequências de observações com base no modelo previamente criado, a combinação das taxas do SVM e do HMM gera os diferentes níveis de susceptibilidade, conforme pode ser observado na Figura 5.1.

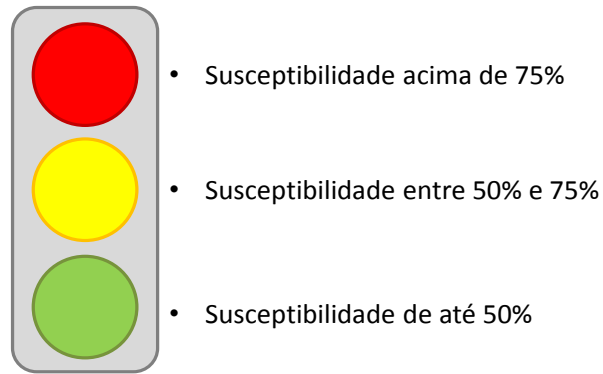


Figura 5.1 - Semáforo representando os níveis de susceptibilidade ao abuso sexual de crianças e adolescentes

Foi estabelecido o patamar de 50% para o alerta verde baseado na baixa taxa de reconhecimento do HMM para sequências de observação. O alerta amarelo foi considerado como um alerta moderado de susceptibilidade e foi estabelecido para taxas de reconhecimento, pelo HMM, de até 89% para as sequências de observação. Já o alerta vermelho é gerado a partir de um percentual de reconhecimento de 90% da sequência de observação gerada pelo SVM.

Até onde foi possível ter acesso a literatura não foi identificada nenhuma iniciativa no sentido de tentar estabelecer um indicador de susceptibilidade ao abuso sexual, visando preveni-lo antes que aconteça.

5.5. Considerações Finais

Neste Capítulo foram apresentados os resultados obtidos para cada etapa do método desenvolvido. No próximo Capítulo é realizado um fechamento deste trabalho por meio das conclusões que podem ser emitidas e, ainda, são apresentados os trabalhos futuros.

Capítulo 6

Conclusão

Este trabalho propôs um método de detecção de aliciamento sexual de crianças e adolescentes em sala de bate-papo *online*, baseado na identificação dos estágios de comunicação existentes no diálogo, tendo por fundamento o trabalho de LEATHERMAN [LEA09]. No desenvolvimento deste trabalho, após experimentos e visando o bom desempenho para aplicações de tempo real, optou-se pelo uso do SVM para a identificação dos estágios de comunicação. Foram utilizados 10.978 enunciados do agressor, que foram previamente rotulados, para treinar e testar o SVM, que apresentou uma taxa de acerto na classificação dos estágios de 85%, equivalente ao que é relatado na literatura, usando outras técnicas. Adicionalmente a classificação dos estágios do aliciamento, usou-se o HMM para reconhecer as sequências de estágios identificados pelo SVM com uma taxa de acerto de 97%.

Como uma inovação em relação à literatura, neste trabalho integrou-se os resultados apresentados com o uso do SVM e do HMM para tornar possível estabelecer o nível de susceptibilidade ao abuso sexual de crianças e adolescentes, antes mesmo que todos os estágios do aliciamento tenham sido detectados. O HMM trabalha bem com sequências de entrada de diferentes tamanhos e deste modo permite inferir se um estágio identificado tem probabilidade de pertencer ao modelo obtido no HMM que levará ao aliciamento. Com isso, três níveis de susceptibilidade foram propostos para o alerta: verde, com taxas de até 50%; amarelo, com taxas entre 50% e 75%; e vermelho, para taxas superiores a 75%.

Entende-se que detectando o aliciamento precocemente, por meio do alerta de susceptibilidade resultante do trabalho, pode-se prevenir o abuso sexual. Além disso, o método desenvolvido provê recursos de caráter auto-preventivo. Deve-se lembrar de que

quando o assunto é aliciamento e abuso sexual de crianças e adolescentes, ações de prevenção evitarão as consequências decorrentes do abuso sexual.

Atualmente, os mecanismos de proteção para controle paterno (*Parental control*) na navegação na *Internet* em geral se restringem a filtragem de palavras e *sites* considerados perigosos, mostrando-se como um recurso bastante limitado e em geral pouco conhecido.

Em geral pais, crianças e adolescente não conhecem a abordagem que o aliciador usa com suas vítimas, mas a implementação do método como uma ferramenta é dotada deste modelo de comportamento, típico de ação do aliciador, auxiliando a possível vítima a se proteger quando a situação de risco de aliciamento for identificada pelo *software*. A aplicação do método desenvolvido permite que a infraestrutura computacional oferecida às crianças e adolescentes esteja disponível de forma segura, ao menos na troca de mensagens instantâneas.

Este trabalho teve como resultados parciais até o momento uma publicação no Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais (SBSeg 2012) e a escrita de um pedido de patente do processo de detecção de aliciamento de crianças e adolescentes em mensagens instantâneas. Além disto, o presente trabalho foi avaliado e validado pela Dra. Leila Paiva, Coordenadora do Programa Nacional de Enfrentamento da Violência Sexual contra Crianças e Adolescentes.

Como trabalhos futuros pode-se destacar a adaptação do método desenvolvido para as características da língua portuguesa ou espanhola. Além disso, uma ferramenta poderá ser desenvolvida com recursos para capturar o contexto da conversa e o seu desencadeamento para concluir que determinada palavra pode se caracterizar como indutor de assédio sexual.

Até o momento, não foi identificada nenhuma ferramenta de mercado que atue no sentido de auxiliar os responsáveis pela criança ou adolescente, ou eles próprios, na detecção da possibilidade de aliciamento e abuso sexual. Assim, o desenvolvimento deste produto é uma possibilidade de trabalho futuro.

Referências

- [CFM06] Conselho Federal de Medicina. Visão atual do abuso sexual na infância e adolescência. Junho, 2006. Disponível em <
http://portal.cfm.org.br/index.php?option=com_content&view=article&id=20278&catid=46:artigos&Itemid=18>. Acesso em 19 de junho de 2013.
- [CHA11] Chang, C. C.; Lin, C. J.. *LIBSVM -- A Library for Support Vector Machines*. ACM Transactions on Intelligent Systems and Technology, 2011. Vol. 2.
- [DOS12] Dosciatti, M. M.; Martinazzo, B.; Paraiso, E. C.. *Identifying Emotions in Short Texts for Brazilian Portuguese*. IV Internacional Workshop on Web and Text Intelligence, Curitiba, 2012.
- [FAC11] Faceli, K; Lorena, A. C.; Gama, J.; Carvalho, A. C. P. L. F.. *Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina*. LTC – Livros Técnicos e Científicos Editora Ltda, Rio de Janeiro, 2011.
- [FRE01] Freitas, C. O. A. *O Uso de Modelos Escondidos de Markov para Reconhecimento de Palavras Manuscritas*. Tese de Doutorado apresentada ao Programa de Pós Graduação Em Informática. Pontifícia Universidade Católica do Paraná, 2001.
- [GAR06] Garcia, E.. *Cosine Similarity and Term Weight Tutorial*. Disponível em <
<http://www.miislita.com/information-retrieval-tutorial/cosine-similarity-tutorial.html>>. Acesso em janeiro de 2012.
- [GAU03] Gauer, G. J. C.; Cataldo Neto, A.; Furtado, N. R.. *Psiquiatria - Para Estudantes De Medicina*. Rio Grande do Sul: EDIPUCRS - PUC RS, 2003.
- [GRO04] Grossman, D. A.; Frieder, O.. *Information retrieval: Algorithms and Heuristics*. Second Edition. Springer, The Netherlands, 2004.
- [HAY99] Haykin, S.. *Neural Networks – A Comprehensive Foundation*. Prentice-Hall, 1999.

- [HEA98] Hearst, M. A.; Schölkopf, B.; Dumais, S.; Osumi, E.; Platt, J.. *Trends & Controversies – Support Vector Machines*. IEEE Intelligent Systems, 1998. p. 18-28.
- [HOL69] Holsti, O.R.. *Content Analysis for the Social Sciences and Humanities*. Addison-Wesley, 1969.
- [HOW04] Ho, W. H.; Watters, P. A.. *Statistical and Structural Approaches to Filtering Internet Pornography*. IEEE International Conference on Systems, Man and Cybernetics, 2004. p. 4792-4798.
- [HUA90] Huang, X.D.; Ariki, Y.; Jack, M.A.. *Hidden Markov Models For Speech Recognition*. Edinburgh Information Technology Series, Edinburgh, 1993. Edinburgh University Press, 276 p.
- [JOA02] Joachims, T.. *Learning to Classify Text using Support Vector Machines*. Cornell University - Department of Computer Science, Kluwer Academic Publishers, 2002.
- [KON09a] Kontostathis, A.; Edwards, L.; Leatherman, A.. *ChatCoder: Toward the Tracking and Categorization of Internet Predators*. Proceedings of 2009 Text Mining Workshop – Society for Industrial and Applied Mathematics International Conference on Data Mining, May 2009.
- [KON09b] Kontostathis, A.; Edwards, L.; Bayzick, J.; McGhee, I.; Leatherman, A.; Moore, K.. *Comparison of Rule-based to Human Analysis of Chat Logs*. International Workshop on Mining Social Media, November 2009.
- [LAN05] Landauer, T. K.; Foltz, P. W.; Laham, D.. *Introduction to Latent Semantic Analysis*. Discourse Processes 25, 2005. p. 259-284.
- [LEA09] Leatherman, A.. *Luring language and virtual victims: Coding cyber-predators online communicative behavior*. Technical Report. Ursinus College, 2009.
- [LOR03] Lorena, A. C.; Carvalho, A. C. P. L. F. de. *Introdução às Máquinas de Vetores Suporte (Support Vector Machines)*. Technical Report nº 192. Instituto de Ciências Matemáticas e de Computação da USP, 2003.

- [MAN08] Manning, C. D.; Raghavan, P.; Schütze, H.. *Introduction to Information Retrieval*. Cambridge University Press, 2008.
- [MAR10] Martinazzo, B.; Paraiso, E. C.. *Identificação de Emoções em Notícias Curtas*. CLEI - Conferência Latino-americana de Informática, 2010. Vol. 1, p. 1-10.
- [MCG11] McGhee, I.; Bayzick, J.; Kontostathis, A.; Edwards, L.; McBride, A.; Jakubowski, E.. *Learning to Identify Internet Sexual Predation*. International Journal of Electronic Commerce, 2011. Vol. 15, p. 103-122.
- [MIC10] Michalopoulos, D.; Mavridis, I.. *Towards Risk Based Prevention of Grooming Attacks*. International Conference on Security and Cryptography, July 26-28, 2010. p. 1-4.
- [MIL06] Milgram, J.; Cheriet, M.; Sabourin, R.. “*One Against One or One Against All: Which one is better for handwriting recognition with SVMs?*”. Tenth International Workshop on Frontiers in Handwriting Recognition, Outubro 2006.
- [MIT97] Mitchell, T. *Machine Learning*. McGraw-Hill, 1997.
- [MON03] Monard, M. C.; Baranauskas, J. A.. *Conceitos sobre Aprendizagem de Máquina. Sistemas Inteligentes - Fundamentos e Aplicações*, 2003. Cap. 4, Vol. 1, p. 89-114.
- [MOR13] Morris, C.. *Identifying Online Sexual Predators by SVM Classification with Lexical and Behavioral Features*. Tese de Mestrado, Departamento de Ciência da Computação, Universidade de Toronto. Janeiro, 2013.
- [MUL01] Müller, K. R.; Mika, S.; Rätsch, G.; Tsuda, K.; Schölkopf, B.. *An Introduction to Kernel-Based Learning Algorithms*. IEEE Transactions on Neural Networks, March 2001. Vol.12, p. 181-201.
- [NCM84] National Center for Missing and Exploited Children. Disponível em <http://www.missingkids.com/en_US/documents/CyberTiplineFactSheet.pdf>. Acesso em 17 de fevereiro de 2012.
- [NUN09] Nunes, A. V.; Freitas, C. O. A.; Paraiso, E.C.. *Detecção de Assédio Moral em e-mails*. I Student Workshop on Information and Human Language Technology - 7th

Brazilian Symposium in Information and Human Language Technology, Julho 20-24, 2009. V. 1, p. 01-05.

- [OLS07] Olson, L. N.; Daggs, J. L.; Ellevold, B. L.; Rogers, T. K. K.. *Entrapping the Innocent: Toward a Theory of Child Sexual Predators' Luring Communication*. *Communication Theory*, 2007. Vol. 17, p. 231-251.
- [PAR12] Parapar, J.; Losada, D. E.; Barreiro, A.. *A learning-based approach for the identification of sexual predators in chat logs*. Conference and Labs of the Evaluation Forum, PAN Workshop – Author Identification. Roma, 2012.
- [PER03] *Perverted Justice Foundation Incorporated*. Disponível em < www.perverted-justice.com >. Acesso em janeiro de 2012.
- [RAB93] Rabiner, L.R.; Juang, B-H.. *Fundamentals of speech recognition*. Prentice Hall Inc., Englewood Clifss, New Jersey, 1993.
- [RAB89] Rabiner, L. R.. *A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition*. *Proceedings of the IEEE*, February, 1989. Vol. 77, No. 2, p.257-286.
- [RAS08] Rashid, A.. *The Talk of Crime*. *Investigative Practice Journal*, July 10, 2008. p. 28-29.
- [REI04] Reichl, L. E.. *A Modern Course in Statistical Physics*. WILEY-VCH, second edition, 2004.
- [RUS95] Russell, S.; Norvig, P.. *Artificial Intelligence: A Modern Approach*. Prentice Hall, 1995.
- [SAL88] Salton, G.; Buckley, C.. *Term-weighting approaches in automatic text retrieval*. *Information Processing and Management – Cornell University*, Ithaca, 1988.
- [SAN12] Santin, P. L. L.; Freitas, C. O. A.; Paraiso, E. C.; Santin, A. O.. *Modelagem de Aliciamento de Menores em Mensagens Instantâneas de Texto*. *Simpósio Brasileiro de Segurança da Informação e de Sistemas Computacionais*, Curitiba, 2012, p. 288-301.

- [SMO99] Smola, A. J.; Barlett, P.; Schölkopf, B.; Schuurmans, D.. *Introduction to Large Margin Classifiers*. MIT Press, 1999.
- [STR08] Strapparava, C.; Mihalcea, R.. *Learning to Identify Emotions in Text*. ACM Symposium on Applied Computing, March 16-20, 2008. p. 1556-1560.
- [TAN09] Tan, P.; Steinbach, M.; Kumar, V.. *Introdução ao DATA MINING Mineração de Dados*. Rio de Janeiro: Editora Ciência Moderna Ltda., 2009.
- [TRI10] Trindade, J.; Breier, R.. *Pedofilia: aspectos psicológicos e penais*. Porto Alegre: Livraria do Advogado Editora, 2010. Ed. 2, p. 100.
- [VIL12] Villatoro-Tello, E.; Juárez-González, A.; Escalante, H. J.; Montes-y-Gómez, M.; Villaseñor-Pineda, L.. *A Two-step Approach for Effective Detection of Misbehaving Users in Chats*. Conference and Labs of the Evaluation Forum, PAN Workshop – Author Identification. Roma, 2012.
- [WEK12] *WEKA: Data Mining Software in Java*. The University of Waikato. Disponível em < <http://www.cs.waikato.ac.nz/ml/weka/> >. Acesso em janeiro de 2012.

Anexo I

Parecer da Dra. Leila Paiva, Coordenadora do Programa Nacional de Enfrentamento da Violência Sexual contra Crianças e Adolescentes, sobre o presente projeto.

Parecer: Modelagem de Aliciamento de Menores em Mensagens Instantâneas de Texto

“direitos sexuais são elemento fundamental dos direitos humanos e incluem o direito à liberdade e à autonomia e o exercício responsável da sexualidade” Plataforma de Ação de Beijing (1995)

1. Introdução.

A proposta de **Modelagem de Aliciamento de Menores em Mensagens Instantâneas de Texto** se apresenta como mais uma ferramenta de enfrentamento da violência sexual cometida contra crianças e adolescentes. A partir do pressuposto de que alguns indivíduos reconhecem na rede mundial de computadores uma forte aliada para suas estratégias de aliciamento de crianças e adolescentes para fins sexuais, por conhecerem a tendência comportamental destas de usar com frequência as salas de bate-papo e outros mecanismos de comunicação instantânea para estabelecimento de relacionamentos, desenhou-se a proposta de elaboração de um sistema para identificação de potenciais atividades de aliciamento sexual de crianças e adolescentes, com a proposta de se configurar como mais uma ferramenta de prevenção a esse tipo de violência.

O presente parecer pretende disponibilizar uma análise da proposta, a partir da perspectiva da garantia de direitos de crianças e adolescentes, quanto: 1. Aos pressupostos conceituais da proposta (embasamento teórico e conceitual); 2. À adequação da proposta às estratégias nacionais e internacionais de enfrentamento da violência sexual contra crianças e adolescentes positivadas nos documentos referenciais sobre o tema, especialmente, a Convenção sobre os direitos da Criança (CDC)¹, O Estatuto da Criança e do Adolescente (ECA), o Documento Final do III Congresso Mundial de Enfrentamento da Exploração Sexual de Crianças e Adolescentes e o Plano

¹ Convenção sobre os Direitos da Criança da ONU, promulgada no Brasil por meio do DECRETO No 99.710, de 21 de novembro de 1990.

Nacional de Enfrentamento da Violência Sexual contra Crianças e Adolescentes; 3. Às possibilidades de utilização da ferramenta como estratégia de enfrentamento da violência sexual contra crianças e adolescentes.

Preliminarmente cabe ponderar sobre a nomenclatura utilizada para identificar a referida ferramenta: “Modelagem de Aliciamento de Menores em Mensagens Instantâneas de Texto”. Sugere-se a utilização do termo “criança e adolescente” em substituição ao termo “menor” para referir-se a pessoas com menos de 18 anos, em adequação à legislação especializada brasileira, o ECA, Lei nº 8069/90.

2. Pressupostos Conceituais da Proposta de Modelagem de Aliciamento de Menores em Mensagens Instantâneas de Texto.

Para efeitos da análise pretendida neste parecer, será utilizada a categorização de direitos da CDC. A CDC, a Constituição Federal e o ECA afirmam ao longo de seus dispositivos quatro grupos de direitos da criança e do adolescente: direito à vida, ao desenvolvimento, à proteção e à participação. Estes direitos não excluem quaisquer outros dispostos em tratados internacionais de direitos humanos.

Faz-se necessário afirmar que o enfrentamento da violência sexual se dá no campo dos direitos humanos, especialmente quando se trata de uma ferramenta voltada para prevenção e também para a responsabilização. A afirmação do direito à liberdade e ao desenvolvimento sexual saudável deve embasar toda e qualquer estratégia de enfrentamento da violência sexual.

De acordo com Wanderlino Neto, “Sendo a sexualidade o ponto fontal da vida humana, o reconhecimento e a garantia da sua liberdade é uma das tarefas magnas do direito. A regulação emancipatória do exercício dessa liberdade sexual passa a ser a responsabilidade maior do Estado Democrático de Direito, dirimindo conflitos de interesses possíveis, visando o asseguramento do prazer pessoal do cidadão e da ordem social. A sexualidade humana pressupõe liberdade, diversidade, respeito e tolerância. E a livre expressão dessa sexualidade deve ser reconhecida e garantida como um direito fundamental, indisponível”.

Assim, salvo melhor entendimento, a **proposta de Modelagem de Aliciamento de Menores em Mensagens Instantâneas de Texto** deve se embasar no enfrentamento,

com foco na prevenção da violência sexual contra crianças e adolescentes no Brasil, entendida esta **como todo ato de qualquer natureza, atentatório ao direito humano e ao desenvolvimento sexual saudável da criança e do adolescente, praticado por agente (adulto ou não) em situação de poder e de desenvolvimento sexual desigual em relação à vítima.**

Esta opção conceitual impõe considerar que: a violência sexual é uma violação aos direitos humanos de crianças/adolescentes; crianças e adolescentes têm direito ao desenvolvimento protegido de sua sexualidade; que a violência sexual se expressa de forma física e/ou psicológica; e que a violência sexual é praticada por alguém numa situação de poder e desenvolvimento sexual desigual em relação a vítima (aqui considerada criança/adolescente), razão pela qual a violência sexual contra crianças e adolescentes deve ser considerada como conduta ilegal, cujos autores diretos e indiretos devem ser responsabilizados de acordo com o ordenamento jurídico pátrio.

Conclui-se, portanto, que **a proposta de Modelagem de Aliciamento de Menores em Mensagens Instantâneas de Texto** deve se constituir em mais um instrumento garantidor dos direitos de crianças/adolescentes de ter sua sexualidade protegida e promovida em observância à sua condição peculiar de pessoa em desenvolvimento (Art.6º do ECA). Partindo do pressuposto que esse direito somente pode ser exercido por pessoas em situação especial de desenvolvimento, a partir de um comportamento proativo da sociedade e do Estado para a garantia de seu exercício. Sendo assim, entende-se que o enfrentamento à violação do direito ao desenvolvimento sexual da criança e do adolescente deve aglutinar, necessariamente, os mais diferentes atores em torno de estratégias comuns de promoção/proteção dos direitos humanos, para a realização da dignidade de crianças e adolescentes.

3. A Adequação da Proposta de Modelagem de Aliciamento de Menores em Mensagens Instantâneas de Texto às Estratégias Nacionais e Internacionais de Enfrentamento da Violência Sexual contra Crianças e Adolescentes.

3.1 A proposta de Modelagem de Aliciamento de Menores em Mensagens Instantâneas de Texto nos cenários de enfrentamento da violência sexual contra crianças e adolescentes.

Atendendo à diretriz de se constituir como uma ferramenta que possa fortalecer a atuação do sistema de garantia de direitos, a **proposta de Modelagem de Aliciamento de Menores em Mensagens Instantâneas de Texto** além de ter princípios balizadores dentro do contexto de violação dos direitos humanos deve se incorporar às diretrizes instituidoras do Sistema de Garantia de Direitos da Criança e do Adolescente no Brasil, previsto no Estatuto da Criança e do Adolescente e regulamentado pela resolução nº 113 do CONANDA (Conselho Nacional dos Direitos da Criança e do Adolescente) que dispõem, respectivamente:

Lei 8069/90. Estatuto da Criança e do Adolescente: Art. 86. A política de atendimento dos direitos da criança e do adolescente far-se-á através de um conjunto articulado de ações governamentais e não governamentais, da União, dos estados, do Distrito Federal e dos municípios.

*Resolução 113 CONANDA: Art. 1º - O Sistema de Garantia dos Direitos da Criança e do Adolescente constitui-se na articulação e integração das instâncias públicas governamentais e **da sociedade civil**, na aplicação de instrumentos normativos e no funcionamento dos mecanismos de promoção, **defesa** e controle para a efetivação dos direitos humanos da criança e do adolescente, nos níveis Federal, Estadual, Distrital e Municipal (*GRIFOS NOSSOS*).*

É necessário ainda estabelecer o projeto como uma resposta aos apelos internacionais, especialmente, tratando-se de prevenção a crimes transacionais, instituídos por congressos internacionais sobre o tema, senão vejamos:

A Agenda de Estocolmo (1996) e seu Plano de Ação:

c) Desenvolver ou reforçar e aplicar medidas legais, políticas e programas nacionais para proteger as crianças vítimas da exploração sexual e comercial diante da eventualidade de que sejam castigadas como criminosas e garantir que estas tenham

pleno acesso a pessoas e serviços de apoio em todos os setores, particularmente no âmbito legal, social e de saúde;

5. a) Adotar um enfoque não punitivo para as crianças vítimas da exploração sexual e comercial em consonância com os direitos da criança, tendo especial cuidado para que os processos judiciais não agravem o trauma vivenciado pela criança e para que a resposta do sistema esteja acompanhada de medidas de assistência legal, quando necessário, e judiciais para as crianças vítimas;

5. f) Adotar não somente sanções legais contra os praticantes de crimes sexuais contra as crianças, mas também medidas psicológicas e médico-sociais para produzir mudanças de comportamento nos agressores.

O Protocolo facultativo à Convenção sobre os Direitos da Criança referente à Venda de Crianças, à Prostituição Infantil e à Pornografia Infantil de 2000:

Art. 8º - 1. Os Estados Partes adotarão as medidas apropriadas para proteger os direitos e interesses de crianças vítimas das práticas proibidas pelo presente Protocolo em todos os estágios do processo judicial crimina,(...).

O III congresso Mundial contra a Exploração sexual de Crianças e Adolescentes em seu documento final dispõe:

- Há necessidade de se melhorar os filtros de material pornográfico na Internet e de desenvolver linhas seguras (safelines) para denunciar estas práticas;*
- Imagens e materiais relacionados a abusos contra crianças devem ser proibidos por lei em todos os países: criação, distribuição e visualização desse tipo de imagens na Internet;*
- Treinamento e advocacy são necessários, e devem ser desenvolvidos em cooperação com assistentes sociais e na construção de uma relação de confiança com a indústria das Tecnologias de Informação e Comunicação (TIC);*

- *Unidades nacionais especializadas precisam ser treinadas para prosseguir com as investigações e para a correta utilização do banco de dados de imagem da Interpol;*
- *O acesso a linhas prioritárias de denúncia (hotlines) deve ser priorizado;*
- *Os procedimentos para queixas / denúncias devem ser simplificados;*
- *De acordo com a Convenção, os governos devem adotar os mesmos padrões legais básicos que incluem como crimes, todas as formas de exploração sexual de crianças e adolescentes, como a pornografia e a prostituição, com disposições específicas contra a pornografia infantil, mesmo quando já existe uma lei contra essa prática;*
- *No que diz respeito à pornografia infantil, indivíduos devem ser processados pela simples posse de material;*
- *O Plano de Ação deve ser elaborado e executado imediatamente;*

3.2. Os Conceitos da Violência Sexual.

Entende-se ainda que a **proposta de Modelagem de Aliciamento de Menores em Mensagens Instantâneas de Texto** deve adotar os conceitos assumidos pela rede de enfrentamento da violência sexual contra crianças e adolescentes e presentes no Plano Nacional de Enfrentamento da Violência sexual contra Crianças e Adolescentes.

Para melhor entendimento, é importante conceituar as chamadas expressões da violência, como forma de facilitar a compreensão e o desenho das diversas metodologias de enfrentamento do fenômeno. Assim, passa-se a elencar alguns conceitos, os quais são de imprescindível domínio para a constituição de uma ferramenta para o enfrentamento da violência sexual contra crianças e adolescentes.

- **Violência Física** – Uso da força física de forma intencional, não-acidental, por um agente agressor adulto (ou mais velho que a criança ou o adolescente). Geralmente, esses agentes são os próprios pais ou responsáveis, que muitas vezes machucam a

criança ou adolescente sem a intenção de fazê-lo. A violência física pode deixar marcas evidentes e, em casos extremos, até causar a morte (Fonte: Guia Escolar, 2004).

- **Violência Psicológica** – Conjunto de atitudes, palavras e ações para envergonhar, censurar e pressionar a criança de modo permanente. Ela ocorre quando xingamos, rejeitamos, isolamos, aterrorizamos, exigimos demais das crianças e dos adolescentes, ou, mesmo, os utilizamos para atender a necessidades dos adultos (Fonte: Guia Escolar, 2004).
- **Violência Sexual** – É uma violação dos direitos sexuais, porque abusa do corpo e da sexualidade, seja pela força ou outra forma de coerção, ao envolver crianças e adolescentes em atividades sexuais impróprias para a sua idade cronológica, ou para seu desenvolvimento psicosssexual. Trata-se de toda ação na qual uma pessoa, em situação de poder, obriga outra à realização de práticas sexuais, por meio da força física, da influência psicológica (intimidação, aliciamento, sedução) ou do uso de arma ou droga.

Assim, a **proposta de Modelagem de Aliciamento de Menores em Mensagens Instantâneas de Texto** deverá considerar os conceitos das diversas formas de desrespeito ao direito de crianças e adolescentes ao exercício da sua sexualidade, consideradas como expressões da violência sexual.

O Plano Nacional de Enfrentamento da Violência Sexual contra Crianças e Adolescentes conceitua violência sexual como uma violação de direitos, que se subdivide em **exploração sexual** e **abuso sexual**.

3.2.1. Abuso Sexual.

O **abuso sexual** se caracteriza pela utilização do corpo de uma criança ou adolescente para a prática de qualquer ato de natureza sexual. Nesse tipo de violência não há qualquer intuito de lucro, qualquer relação de compra ou troca. No abuso sexual, o agressor visa unicamente satisfazer seus desejos por meio da violência sexual. Uma característica que costuma compor a violência é a relação de confiança entre o agressor e a vítima, ainda que momentânea e enganosa, e geralmente é praticada por alguém que participa do mesmo convívio. Isso não necessariamente significa que seja convívio familiar, podendo ser comunitário.

O abuso sexual pode expressar-se de duas formas: **intrafamiliar** e **extrafamiliar**:

- a. O abuso sexual intrafamiliar é assim considerado quando a agressão ocorre dentro da família, ou seja, a vítima e o agressor possuem alguma relação de parentesco. Aqui é

importante considerar o contexto familiar ampliado, já que a diferença estabelecida sob o aspecto conceitual objetivou apenas diferenciar as estratégias e metodologias de prevenção, proteção e responsabilização. Assim, quando o agressor compõe a chamada família ampliada ou possui vínculos afetivos familiares, o abuso deve ser caracterizada como intrafamiliar.

- b.** O abuso sexual extrafamiliar se dá quando não há vínculo de parentesco entre o agressor e a criança ou adolescente.

3.2.2. Exploração Sexual.

A **exploração sexual** caracteriza-se pela utilização sexual de crianças e adolescentes com a intenção do lucro, seja financeiro ou de qualquer outra espécie. Nesse caso, pode haver a participação de um agente entre a criança ou adolescente e o usuário ou cliente. É por isso que se diz que a criança ou adolescente foi explorada, e nunca prostituída, pois ela é vítima de um sistema de exploração de sua sexualidade.

A partir de uma proposição do Instituto Interamericano del Niño, la Niña y Adolescentes, da OEA, em 1998 passou-se a classificar a **exploração sexual** em quatro modalidades²: a pornografia, o turismo com fins sexuais, a prostituição convencional e o tráfico para fim **sexual**.

Posteriormente, a partir das discussões do III Congresso Mundial Contra a Exploração Sexual de Crianças e Adolescentes, esses conceitos foram atualizados para:

- exploração sexual no contexto de prostituição
- tráfico para fins de exploração sexual
- exploração sexual no contexto do turismo
- pornografia infantil (ver item d, adiante)

a. Exploração sexual no contexto de prostituição.

É a expressão mais próxima do contexto do comércio da exploração sexual de crianças e adolescentes. É muito comum aparecerem adultos como intermediários nessa forma de exploração sexual, rede de aliciadores, agenciadores, facilitadores, pessoas que lucram com a exploração sexual. Porém, esse tipo de exploração sexual pode ocorrer sem intermediários. Ainda que a princípio possa parecer uma atividade autônoma, como no caso de crianças ou adolescentes que oferecem seus corpos nas ruas, caracteriza-se

² Em <http://www.recrianacional.org.br/glossary/exploracao-sexual>. Acesso em 14/02/2012.

como exploração, já que o usuário pagará pela utilização de seu corpo com dinheiro ou outros elementos de troca.

Os artigos 244-A e 244-B do Estatuto da Criança e do Adolescente (ECA), incluídos pelas leis n. 9.975, de 23/06/2000, e n. 12.015, de 07/08/2009, tratam da exploração sexual com maior especificidade.

Art. 244-A. Submeter criança ou adolescente, como tais definidos no caput do art. 2º desta Lei, à prostituição ou à exploração sexual: (Incluído pela Lei n. 9.975, de 23.6.2000)

Pena – reclusão de quatro a dez anos, e multa.

§ 1º- Incorrem nas mesmas penas o proprietário, o gerente ou o responsável pelo local em que se verifique a submissão de criança ou adolescente às práticas referidas no caput deste artigo. (Incluído pela Lei n. 9.975, de 23/06/2000)

§ 2º- Constitui efeito obrigatório da condenação a cassação da licença de localização e de funcionamento do estabelecimento. (Incluído pela Lei n. 9.975, de 23/06/2000)

Art. 244-B. Corromper ou facilitar a corrupção de menor de 18 (dezoito) anos, com ele praticando infração penal ou induzindo-o a praticá-la: (Incluído pela Lei n. 12.015, de 07/08/2009)

Pena – reclusão, de 1 (um) a 4 (quatro) anos. (Incluído pela Lei n. 12.015, de 07/08/2009)

§ 1º- Incorre nas penas previstas no caput deste artigo quem pratica as condutas ali tipificadas utilizando-se de quaisquer meios eletrônicos, inclusive salas de bate-papo da internet. (Incluído pela Lei n. 12.015, de 07/08/2009).

§ 2º- As penas previstas no caput deste artigo são aumentadas de um terço no caso de a infração cometida ou induzida estar incluída no rol do art. 1º da Lei n. 8.072, de 25 de julho de 1990. (Incluído pela Lei n. 12.015, de 07/08/2009)

A partir da alteração trazida pela Lei n. 12.015, de 7 de agosto de 2009, o Código Penal (CP) foi acrescido de um artigo específico que define o crime de exploração sexual contra crianças e adolescentes³.

³ **Favorecimento da prostituição ou outra forma de exploração sexual de vulnerável**

Art. 218-B. Submeter, induzir ou atrair à prostituição ou outra forma de exploração sexual alguém **menor de 18 (dezoito) anos** ou que, por enfermidade ou deficiência mental, não tem o necessário discernimento para a prática do ato, facilitá-la, impedir ou dificultar que a abandone:

Pena - reclusão, de 4 (quatro) a 10 (dez) anos.

§ 1º Se o crime é praticado **com o fim de obter vantagem econômica**, aplica-se também multa.

§ 2º Incorre nas mesmas penas:

I - **quem pratica conjunção carnal ou outro ato libidinoso com alguém menor de 18 (dezoito) e maior de 14 (catorze) anos na situação descrita no caput deste artigo;**

II - o proprietário, o gerente ou o responsável pelo local em que se verifiquem as práticas referidas no **caput** deste artigo.

O comércio sexual em todas as suas modalidades (bares, motéis, hotéis, boates) desenvolve diferentes ações encontradas nos artigos do CP e do ECA. Os autores desses crimes podem submeter, induzir ou atrair à prostituição ou outra forma de exploração sexual alguém menor de 18 (dezoito) anos. Importante salientar que no caso de pessoas maiores de dezoito anos, o exercício da prostituição não constitui crime, pelo menos no Brasil.

b. Tráfico para fins de exploração sexual.

É a promoção ou facilitação da entrada, saída ou deslocamento no território nacional, ou para outro país, de crianças e adolescentes com o objetivo de exercerem a prostituição ou outra forma de exploração sexual.

O Protocolo de Palermo⁴ unificou o conceito de tráfico de pessoas (art. 3º, “a”), deixando claro que o consentimento dado pela vítima de tráfico é irrelevante, e o simples fato de recrutar, transportar, transferir ou alojar a criança para fins de exploração sexual será considerado “tráfico de pessoas” (art. 3º, “b”).

Os bens jurídicos que se quer proteger nesse crime são a liberdade e a dignidade sexual, podendo cometer o delito qualquer pessoa (sujeito ativo), podendo a vítima também ser qualquer ser humano (sujeito passivo). O crime é consumado com a prática efetiva de uma destas três condutas: promover, intermediar e facilitar.

Por meio da Lei n. 11.106/2005 e da Lei n. 12.015/2009, o art. 231 do CP foi alterado. O crime “tráfico de mulheres” passou a ser “tráfico de pessoas”, o que trouxe a possibilidade de o homem também ser vítima (sujeito passivo) do delito. Ao lado dessa alteração, foi inserida a figura do “tráfico interno” (art. 231-A). Com a alteração de 2009, o capítulo V do título VI do CP, em que está inserido o artigo 231, mudou a denominação para

“DO LENOCÍNIO E DO TRÁFICO DE PESSOA PARA FIM DE PROSTITUIÇÃO OU OUTRA FORMA DE EXPLORAÇÃO SEXUAL”⁵.

§ 3º Na hipótese do inciso II do § 2º, constitui efeito obrigatório da condenação a cassação da licença de localização e de funcionamento do estabelecimento.

⁴ *Protocolo Adicional à Convenção das Nações Unidas Contra a Criminalidade Organizada Transnacional relativo à Prevenção, Repressão e Punição do Tráfico de Pessoas, em especial Mulheres e Crianças*
5 *Tráfico internacional de pessoa para fim de exploração sexual*

Art. 231. Promover ou facilitar a entrada, no território nacional, de alguém que nele venha a exercer a prostituição ou outra forma de exploração sexual, ou a saída de alguém que vá exercê-la no estrangeiro.

Pena – reclusão, de 3 (três) a 8 (oito) anos.

§ 1º Incorre na mesma pena aquele que agenciar, aliciar ou comprar a pessoa traficada, assim como, tendo conhecimento dessa condição, transportá-la, transferi-la ou alojá-la.

§ 2º A pena é aumentada da metade se:

I – a vítima é menor de 18 (dezoito) anos;

As condutas para tipificação do delito de tráfico foram expandidas, pela legislação atual, para “promover ou facilitar a entrada, agenciar, aliciar ou comprar a pessoa traficada e, ainda, tendo conhecimento de que a pessoa está na condição de tráfico, transportá-la, transferi-la ou alojá-la”.

Para um país continental como o Brasil, a alteração efetuada em 2005 foi fundamental, porquanto enquadrou o tráfico interno como delito⁶.

Antes de 2005, somente o tráfico internacional era previsto, o que constituía uma lacuna na legislação penal brasileira. O movimento de proteção dos direitos de crianças e adolescentes foi muito importante para o reconhecimento do tráfico interno. A Pesquisa Nacional sobre o Tráfico de Mulheres, Crianças e Adolescentes (Pestraf), realizada em 2002, que mapeou 241 rotas de tráfico de pessoas para fins de exploração sexual, incluindo 78 interestaduais e 32 intermunicipais, alertou para a necessidade de tipificação dessa forma de violência. Assim como no artigo anterior, o tráfico interno deve ser para o exercício da prostituição ou outra forma de exploração.

Em tais casos, cabe ao Ministério Público a responsabilidade pela instauração da ação penal.

Outros tipos penais também caracterizam o crime de tráfico de crianças e adolescentes, no CP e no ECA, como no artigo 245 do CP, que trata da entrega de filho menor a pessoa inidônea, e os artigos 238 e 239 do ECA, que tratam respectivamente de “prometer ou efetivar a entrega de filho ou pupilo a terceiro, mediante paga ou

II – a vítima, por enfermidade ou deficiência mental, não tem o necessário discernimento para a prática do ato;

III – se o agente é ascendente, padrasto, madrasta, irmão, enteado, cônjuge, companheiro, tutor ou curador, preceptor ou empregador da vítima, ou se assumiu, por lei ou outra forma, obrigação de cuidado, proteção ou vigilância; ou

IV – há emprego de violência, grave ameaça ou fraude.

§ 3º Se o crime é cometido com o fim de obter vantagem econômica, aplica-se também multa.

⁶ Tráfico interno de pessoa para fim de exploração sexual

Art. 231-A. Promover ou facilitar o deslocamento de alguém dentro do território nacional para o exercício da prostituição ou outra forma de exploração sexual:

Pena – reclusão, de 2 (dois) a 6 (seis) anos.

§ 1º Incorre na mesma pena aquele que agenciar, aliciar, vender ou comprar a pessoa traficada, assim como, tendo conhecimento dessa condição, transportá-la, transferi-la ou alojá-la.

§ 2º A pena é aumentada da metade se:

I – a vítima é menor de 18 (dezoito) anos;

II – a vítima, por enfermidade ou deficiência mental, não tem o necessário discernimento para a prática do ato;

III – se o agente é ascendente, padrasto, madrasta, irmão, enteado, cônjuge, companheiro, tutor ou curador, preceptor ou empregador da vítima, ou se assumiu, por lei ou outra forma, obrigação de cuidado, proteção ou vigilância; ou

IV – há emprego de violência, grave ameaça ou fraude.

§ 3º Se o crime é cometido com o fim de obter vantagem econômica, aplica-se também multa.

recompensa” e “promover ou auxiliar a efetivação de ato destinado ao envio de criança ou adolescente para o exterior com inobservância das formalidades legais ou com o fito de obter lucro”. Diferenciam-se, porém, dos artigos 231 e 231-A do CP, já que estes abordam especificamente o tráfico para fins de prostituição.

c. Exploração sexual no contexto do turismo.

É a exploração sexual de crianças e adolescentes por visitantes de países estrangeiros ou turistas do próprio país, podendo ocorrer com o envolvimento, cumplicidade ou omissão de estabelecimentos comerciais de diversos tipos.

A conceituação desse tipo de exploração é uma tarefa complexa, pois muitos discursos o confundem com abuso de crianças e adolescentes ou tráfico internacional de pessoas. “Em certos discursos, o estrangeiro, pelo mero contato sexual com algum brasileiro ou brasileira, é qualificado de “turista sexual”, mesmo quando ausente qualquer violência sexual⁷”.

A exploração sexual no contexto do turismo, enquanto violação aos direitos sexuais, não pode ser entendida como a vinda de pessoas estrangeiras ao território brasileiro com o objetivo de se relacionarem sexualmente com brasileiros ou brasileiras, mas, sim, como a vinda dessas pessoas visando à realização de alguma conduta que constitua violência sexual. Em outras palavras, a relação sexual entre um estrangeiro e uma brasileira, se não houver agressão ao exercício de sua sexualidade, não poderá configurar exploração sexual. É importante salientar a noção de que a exploração sexual no contexto do turismo “parece ser definido no campo legal-jurídico brasileiro de forma diferente [do senso comum], como algo muito mais específico: a violação por estrangeiros das leis brasileiras que regulam o comportamento sexual⁸”.

A exploração sexual no contexto do turismo não é em si uma expressão da violência sexual, pois ele é apenas uma qualificação dada a certas práticas quando o agressor é estrangeiro e vem ao país visando à violência, sozinho ou agenciado por uma empresa do ramo, ou, mesmo não a desejando anteriormente, por meio de sua conduta acaba por agredir o exercício sexual de alguma pessoa no país.

⁷ SILVA, Ana Paula da; BLANCHETTE, Thaddeus. “Nossa Senhora da Help”: sexo, turismo e deslocamento transnacional em Copacabana, Cadernos Pagu, Campinas, n. 25, p. 253. 2005.

⁸ SILVA, Ana Paula da; BLANCHETTE, Thaddeus. “Nossa Senhora da Help”: sexo, turismo e deslocamento transnacional em Copacabana, Cadernos Pagu, Campinas, n. 25, 2005. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-83332005000200010&lng=pt&nrm=iso>. Acesso em: 14.02.2012.

d. Pornografia infantil.

Pornografia infantil é a expressão da exploração sexual que se caracteriza por qualquer representação, por qualquer meio, de uma criança envolvida em atividades sexuais explícitas reais ou simuladas, ou qualquer representação dos órgãos sexuais de uma criança para fins primordialmente sexuais, de acordo com o Decreto n. 5.007, de 8 de março de 2004, que promulga o Protocolo Facultativo à Convenção sobre os Direitos da Criança referente à venda de crianças, à prostituição infantil e à pornografia infantil.

A legislação brasileira vigente tipifica como crime: Apresentar, produzir, vender, fornecer, divulgar ou publicar, por qualquer meio de comunicação, inclusive rede mundial de computadores ou Internet, fotografias ou imagens com pornografia ou cenas de sexo explícito envolvendo criança ou adolescente.

3.3. Novas formas de violência sexual.

O abuso sexual e a exploração sexual também podem ocorrer via internet. Várias práticas têm sido caracterizadas como tal, ou muitas vezes iniciam um processo de abuso ou exploração. Algumas já vêm ocorrendo com maior frequência, tornando-se, portanto, imperativo conhecê-las.

O *sexting*⁹ é uma expressão de abuso sexual recente, no qual adolescentes, jovens ou adultos usam celulares, *e-mail*, salas de bate-papo, comunicadores instantâneos e *sites* de relacionamento, para enviar fotos sensuais, mensagens de texto eróticas ou com convites sexuais para conhecidos. Algumas vezes essa prática, ao cair na rede, pode configurar abuso sexual ou exploração sexual, a depender da forma como será utilizada.

O *sexcasting* consiste na troca de mensagens sexuais em serviços de conversas instantâneas.

O *sextosión* se configura a partir do “sexting”. É a prática de chantagens com fotografias ou vídeos da criança ou adolescente sem roupa ou em relações íntimas que foram compartilhados por “sexting” com fins de exploração sexual¹⁰;

O *grooming*, caracterizado pela ação de um adulto ao se aproximar de crianças ou adolescentes via internet, por meio de *chats* ou redes sociais, com o objetivo de praticar abuso sexual ou exploração sexual.

⁹ Sexting – Neologismo originado da união de duas palavras em inglês: “sex” (sexo) com “texting” (envio de mensagens).

¹⁰ Luz María Velázquez Reyes, Boletim da Universidade Nacional Autónoma do México (Unam), do Instituto Superior de Ciências da Educação do Estado do México. 11/2011.

4. Possibilidades de utilização da ferramenta como estratégia de enfrentamento da violência sexual contra crianças e adolescentes.

Além da adequação às normativas e opções conceituais nacionais e internacionais, já elencadas, para apresentar a proposta em análise como uma ferramenta de enfrentamento da violência sexual contra crianças e adolescentes, é importante que alguns elementos sejam positivados na eleição dos critérios de funcionamento da **proposta de Modelagem de Aliciamento de Menores em Mensagens Instantâneas de Texto**. Assim, passa-se a eleger alguns considerados de extrema importância, porém não exaustivos:

- Ausência de construção de perfis - Para o atendimento ao propósito de enfrentar a violência sexual sem correr o risco de atentar contra os princípios de direitos humanos, é importante não propiciar qualquer tipo de caracterização dos (as) usuários(as), entendendo-se aqui tanto vítimas como autores. Quaisquer tentativas de caracterização ou de criação de perfis de vítimas ou autores(as) “ideais” poderia correr o risco contribuir para a estigmatização de determinadas pessoas e/ou grupos sociais.
- Enfrentamento ao abuso e à **exploração sexual** - A partir da leitura do material enviado para análise¹¹ pode-se depreender que o foco eleito até o momento para o monitoramento é o abuso sexual, todavia os crimes de exploração sexual facilitados pelo uso inadequado da internet têm aumentado consideravelmente. Assim, impõe considerar essa também como uma estratégia a ser experimentada para o enfrentamento da violência sexual na modalidade exploração, notadamente por meio da pornografia.
- Possibilidade do exercício de Autoproteção – A ferramenta tal como descrita pode significar uma importante estratégia de autoproteção de crianças e adolescentes, considerando que seria possível estabelecer etapas de risco de fácil visualização por parte das possíveis vítimas a fim que estas recebessem alertas do sistema sobre potenciais violações de direitos. Neste caso é fundamental atentar para a linguagem utilizada garantindo o fácil entendimento por parte de crianças e adolescentes, além de viabilizar no processo de validação, o debate com grupos de crianças, adolescentes e suas famílias.

¹¹ Modelagem de Aliciamento de Menores em Mensagens instantâneas de Texto - Priscila L. L. Santin, Cinthia O. A. Freitas, Emerson C. Paraíso, Altair O. Santin. Pontifícia Universidade Católica do Paraná (PUCPR) – Escola Politécnica – Programa de Pós-Graduação em Informática (PPGIa) – Curitiba – PR – Brasil

- Natureza Preventiva – O grande diferencial da ferramenta, tal como proposta, seria estabelecer o caráter? como característica principal e justificativa de sua atuação. Importante salientar que esta definição implicaria na opção de privilegiar a defesa dos direitos de crianças e adolescentes mesmo que por vezes esta opção signifique abrir mão da responsabilização de possíveis autores(as) de violência. O grande diferencial de qualquer ação de enfrentamento da violência sexual contra crianças e adolescentes é a possibilidade de evitar o dano às vítimas, de evitar a violência. Essa é a possibilidade real da ferramenta em análise, caso faça a opção metodológica de privilegiar o caráter preventivo.
- Inserção no fluxo de defesa do Sistema de Garantia de Direitos – O processo de definição metodológica de atuação pressupõe a construção coletiva de fluxo de funcionamento com os demais atores do Sistema de Garantia de Direitos. A proposta de contribuir para o aprimoramento de atuação no processo de defesa dos direitos humanos de crianças e adolescentes impõe a definição de um processo de apresentação, validação e pactuação da ferramenta junto aos conselhos de direitos e demais órgãos que compõem o sistema no eixo de defesa (Conselhos Tutelares, Ministério Público, Defensorias Públicas, Segurança Pública, Poder Judiciário, Centros de Defesa).
- Proteção dos direitos e Produção de prova – A **proposta de Modelagem de Aliciamento de Menores em Mensagens Instantâneas de Texto** tem a oportunidade de constituir referência no enfrentamento da violência sexual contra crianças e adolescentes com foco na prevenção. O foco de atuação não impede sua contribuição com os processos de responsabilização dos(as) possíveis autores(as) da violência. Para tanto, é fundamental construir com a Polícia Federal uma metodologia de produção de provas que garanta prioridade à proteção da criança/adolescente vitimizado e não a responsabilização. A violência sexual é um crime, como tal é necessária e imprescindível a responsabilização deste crime, porém esta não deve ser feita à custa do bem-estar da criança/adolescente. Assim, é fundamental o estabelecimento prévio de metodologias que consigam equilibrar estes objetivos dando prioridade absoluta à proteção de crianças e adolescentes preconizada na Constituição Federal.

5. Conclusão.

A proposta em análise de desenvolvimento de **Modelagem de Aliciamento de Menores em Mensagens Instantâneas de Texto** para o enfrentamento da violência sexual contra crianças e adolescentes, facilitada pela utilização criminosa da rede

mundial de computadores em salas de diálogos, pode se configurar em uma importante ferramenta de autoproteção de crianças e adolescentes e de inserção dos responsáveis no diálogo sobre os riscos da utilização da internet.

O sucesso deste processo certamente será diretamente proporcional à legitimidade conquistada no decorrer de sua construção. Portanto, é fundamental o envolvimento do sistema de garantia de direitos na validação e pactuação do fluxo para utilização da ferramenta.

Cabe ressaltar o caráter inovador do compromisso da Pontifícia Universidade Católica do Paraná em apoiar estudos voltados para a garantia dos direitos de crianças e adolescentes e sugerir que os resultados e aprendizagens desse processo multidisciplinar, que certamente foi necessário para o projeto, sejam compartilhados com outras universidades, como forma de disseminação.

A referida proposta, a princípio, da forma como descrita no material disponibilizado para análise, vem ao encontro das necessárias ações propugnadas nos diferentes instrumentos normativos nacionais e internacionais de enfrentamento da violência sexual contra crianças e adolescentes, já citados neste parecer. Vislumbra-se, portanto, a possibilidade de constituir-se em uma real e necessária estratégia de proteção dos direitos humanos de meninas e meninos brasileiros(as). Para tanto, será condição para seu êxito, a interlocução com o poder público no âmbito do executivo, responsável pela implementação de políticas públicas para a infância e adolescência, e com o Sistema de Garantia de Direitos, por meio dos representantes dos seus três eixos - promoção, defesa e controle.

É o parecer.

Brasília, 2 de junho de 2013

Leila Paiva
OAB/CE 9515