Classificação de Risco de Suicídio Utilizando Análise de Linguagem Natural

Nayron M. Almeida[1], José Flávio G. Barros[1], Breno C. da Silva[1], Francisco das Chagas de S. S.[1], João Pedro G. Feitosa[1], Gerson James M. F. Guimarães[1], Luís Fernando M. [1]

[1] Instituto Federal de Educação, Ciência e Tecnologia do Maranhão (IFMA) – Campus Caxias

{nayronmorais, fchagas.sousa45, joaopedro7000, gerson.mfg}@gmail.com,
{flavio.barros, breno.caetano, luis.maia}@ifma.edu.br

Abstract. *In recent decades, the high rates of suicide worldwide have been drawing the attention of government agencies and society in general. Strategies for treatment and prevention have been formulated, however, there is a great difficulty in detecting people in situations of risk. In this context, the present work proposes a method for suicide risk classifier using Natural Language Processing (NLP), which seeks to identify suicidal intentions in text messages. The classifier is a Naive Bayes learning algorithm that delivers 70.45% accuracy, 54.2% acceptance rate for suicidal messages and 95% for non-suicidal.*

Resumo. *Os altos índices de suicídio em todo mundo têm chamado bastante atenção dos órgãos governamentais, de saúde e da sociedade em geral nas últimas décadas. Estratégias voltadas ao tratamento e prevenção vêm sendo formuladas, porém, existe uma grande dificuldade na detecção de pessoas em situações de risco. Neste contexto, o presente trabalho propõe um método para classificação de risco suicida utilizando Processamento de Linguagem Natural (NLP), que busca identificar intenções suicidas em mensagens textuais. Para tanto, como algoritmo de aprendizagem foi utilizado Naive Bayes que apresentou acurácia de 70,45%, e, taxa de acerto de 54,2% para mensagens suicidas e 95% para as não suicidas.*

1. Introdução

Nas últimas décadas o comportamento suicida vem tomando grande impulso no que se refere aos índices em todo o mundo. Segundo pesquisa realizada pela Organização Mundial de Saúde (2016), o suicídio no ano de 2012, estava ocupando segundo lugar no ranking mundial de causas de morte, e a previsão para 2020 é que o número de mortes por ano possa chegar até 1,6 milhão.

No Brasil, no ano de 2006, o Ministério da Saúde por meio da portaria Nº 1.876, de 14 de agosto de 2006, instituiu as Diretrizes Nacionais para Prevenção do Suicídio, no qual devem ser implantadas em todas as unidades federadas, respeitadas as competências das três esferas de gestão [Ministério da Saúde 2006]. Por meio desta portaria foi dado o alicerce para a formulação de estratégias para a prevenção do suicídio a nível nacional. No entanto, isso não foi suficiente para que houvesse redução nos números de casos de tentativas e de mortes por suicídio.

A Organização Mundial de Saúde (2006) adverte que o desafio chave de tal prevenção consiste em identificar as pessoas que estão em risco e que a ele são

vulneráveis, entender as circunstâncias que influenciam o seu comportamento autodestrutivo e estruturar intervenções eficazes. De forma que uma melhor detecção na comunidade, o encaminhamento para especialistas e a gestão do comportamento suicida são passos importantes na prevenção do suicídio.

Nesse contexto, o presente trabalho busca identificar intenções que remetem ao suicídio em mensagens textuais utilizando o algoritmo de aprendizagem de máquina Naive Bayes, que apesar de ser o mais simples de todos os modelos de classificação apresenta excelentes resultados na maioria de suas aplicações [Mccallum e Nigam 1998] [Lewis 1998]. A identificação se dá por meio da extração de atributos característicos do grupo de mensagens suicidas pertencente a uma base de treinamento, para posterior classificação de mensagens desconhecidas.

2. Trabalhos Relacionados

Recentemente Hettige et. al (2017) buscaram identificar pacientes que se encontram em alto risco suicida. Em sua pesquisa foram utilizados 345 participantes diagnosticados com esquizofrenia, e, o risco suicida de cada um foi identificado utilizando a Columbia Suicide Severity Rating Scale (C-SSRS) e Beck Suicide Ideation Scale (BSS). Para a classificação de risco suicida foram utilizados quatro algoritmos de aprendizagem: regression regularized, random forest, elastic net e support vector machine.

Para o treinamento dos modelos de classificação, foram utilizadas variáveis clínicas e socioculturais, ambas coletadas por meio de questionários construídos com base nas escalas citadas anteriormente. Como resultado, foi obtido aproximadamente 67% de acurácia para o modelo de regressão logística, 66% para o Random Forest, 67% para o SVM e 65% para o Elastic Net.

Outro trabalho voltado a mesma problemática deste trabalho é o de Pak et. al (2012), que apresenta o sistema LIMSI e realiza a extração de emoções a partir de textos em linguagem natural. Este trabalho descreve o sistema LIMSI participando da segunda etapa do desafio da i2b2/VA 2011 em Processamento de Linguagem Natural para Dados Clínicos, cujo desafio era detectar opiniões descritas em anotações suicidas. Neste caso, as sentenças deveriam ser etiquetadas em apenas uma ou várias das seguintes categorias: instrução, informação, desesperança, delito, culpa, raiva, tristeza, medo, abuso, amor, gratidão, esperança, felicidade-paz, orgulho e perdão. O Corpus de treinamento foi disponibilizado pela organização do evento, sendo que já estava todo anotado manualmente.

Para a classificação, os autores utilizaram o algoritmo de aprendizagem de máquina (SVM) combinado com Transdutor. Para a extração de atributos para a classificação das mensagens foram utilizados N-Grams taggers, dicionários de sentimentos e atributos definidos manualmente através da análise do Corpus. Valores como a acurácia do algoritmo não foram construídos devido a abordagem da aprendizagem, no qual, metade do processo de classificação se deu por meio de avaliação de estados finitos. Como resultado do uso combinado dos dois métodos de classificação foi obtido precisão de 53,8%.

Spasić et. al (2012), assim como Pak et. al (2012), participaram do desafio da i2b2/VA 2011. O trabalho de Spasić et. al (2012) se assemelha mais ainda a este trabalho, pois o algoritmo Naive Bayes também foi utilizado como método de classificação. Os atributos foram extraídos das mensagens suicidas por meio de análise

linguística, combinando análise léxica e semântica. Os resultados obtidos tiveram maior destaque que os de Pak, alcançando 55% de precisão.

3. Metodologia

3.1. Base de Treinamento

O início do desenvolvimento se deu por meio da construção da base de treinamento, a qual foi construída com mensagens escritas relacionadas a sentimentos suicidas e não suicidas. Para tanto, esta etapa foi desenvolvida em três subfases: Busca, Seleção e Caracterização.

Primeira subfase – Busca: Em sítios eletrônicos de acesso público em que ocorrem um grande fluxo de troca de informações entre pessoas, grupos e organizações como redes sociais, blogs e sites de autoajuda foram realizadas buscas por postagens de mensagens escritas que possuíssem apelo emocional. Também foram feitas buscas em trabalhos acadêmicos da área psiquiátrica.

Segunda subfase – Seleção: Como critério para adição à base de dados do classificador, esta deveria relacionar o autor a ele mesmo ou terceiros, desempenhando significado de desejo (s), decepção (ões) ou gratidão. Estes critérios foram escolhidos com base nos fatores de risco delineados por Botega e Werlang (2006), onde o primeiro significado (desejo) refere-se a sonhos que desejara realizar, posição de outras pessoas diante dela, emprego ou cargos que almejava ter. O segundo (decepção) está em sua maioria relacionado a decepções ocasionadas pelo autor a outras pessoas, mas também inclui traições de parceiros em relações afetivas. Por fim, o terceiro (gratidão) está relacionado a doenças crônicas e/ou físicas, no qual o autor não suportava observar o estado físico/mental de pessoas próximas diante de seu estado.

Terceira subfase – Caracterização: Depois de selecionadas, as mensagens eram atribuídas a duas classes: Possível Risco Suicida (Explícito ou Implícito) e Baixo Risco Suicida. A primeira classe caracteriza-se por mensagens que apresentam indícios de risco suicida, podendo ainda ser explícito, sem a necessidade de uma abstração mais elevada ou implícito, que dependendo do contexto a qual a mensagem está inserida pode apresentar outros sentidos. E a última classe caracteriza mensagens que não possuem inclinação suicida, podendo apresentar significados relacionados a tipos relacionados da etapa de Seleção, no entanto, não ao suicídio. Vale ressaltar que as classes foram determinadas por senso comum, com isso temos que mensagens relacionadas a classe Possível Risco Suicida (Explícito) podem ser associadas a intenções suicidas de forma direta por quantidade expressiva de indivíduos, mesmo sem terem conhecimento básico da psiquiatria, o mesmo se repete as demais classes.

3.2. Método Proposto

O método foi desenvolvido utilizando a linguagem de programação *Python* em sua versão 3.6.3, os módulos de tokenização, classificação e corpus da biblioteca de processamento de linguagem natural NLTK, na versão 3.2, e uma base de treinamento, que foi construída conforme descrito na seção anterior.

Inicialmente foi realizada a remoções de ruídos, isto é, a normalização das mensagens de forma manual para a forma coloquial da língua portuguesa, foram feitas correções ortográficas, essas correções incluíram principalmente a substituição de abreviações de palavras da linguagem utilizada na internet por sua representação formal na língua portuguesa. Como exemplo podemos citar a palavra “você”, abreviada para

“vc”, a palavra “porque”, cuja abreviação é “pq”, dentre muitas outras comumente abreviadas em textos na internet.

Por seguinte, temos que dada a alta complexidade na construção de relações semânticas em uma frase ou conjuntos de frases e também a forma como o algoritmo *Naive Bayes* realiza a aprendizagem, se fez necessário uma análise da base de treinamento em busca dos atributos mais relevantes nas mensagens de cada classe alvo. Para tanto, foram selecionadas as classes de palavras que mais caracterizavam cada conjunto. Essa avaliação se deu primeiramente com algoritmo de frequência pertencente a biblioteca NLTK e posteriormente com etiquetador *N-Gram Tagging* treinado com o Corpus *Mac Morpho*, este processo foi realizado de forma semelhante ao trabalho de Phuc e Phung (2007), realizando adaptações para a língua portuguesa. As classes no qual as palavras foram etiquetadas são as existentes no Corpus, conforme a Figura 1.

CLASSE GRAMATICAL	ETIQUETA
ADJETIVO	ADJ
ADVÉRBIO CONECTIVO SUBORDINATIVO	ADV-KS
ADVÉRBIO RELATIVO SUBORDINATIVO	ADV-KS-REL
ARTIGO (def. ou indef.)	ART
CONJUNÇÃO COORDENATIVA	KC
CONJUNÇÃO SUBORDINATIVA	KS
INTERJEIÇÃO	IN
SUBSTANTIVO	N
SUBSTANTIVO PRÓPRIO	NPROP
NUMERAL	NUM
PARTICÍPIO	PCP
PALAVRA DENOTATIVA	PDEN
PREPOSIÇÃO	PREP
PRONOME ADJETIVO	PROADJ
PRONOME CONECTIVO SUBORDINATIVO	PRO-KS
PRONOME PESSOAL	PROPESS
PRONOME RELATIVO CONECTIVO SUBORDINATIVO	PRO-KS-REL
PRONOME SUBSTANTIVO	PROSUB
VERBO	V
VERBO AUXILIAR	VAUX
SIMBOLO DE MOEDA CORRENTE	CUR

Figura 1. Classes gramaticais e respectivas etiquetas do Corpus MacMorpho [Barbosa et. al 2017].

Como último processo temos o treinamento do classificador. Neste caso foi utilizada a implementação clássica do *Naive Bayes*, pertence a biblioteca NLTK, este modelo de aprendizagem foi escolhido devido sua simplicidade, desempenho e alta aplicabilidade em tarefas de NLP. Este modelo recebe um conjunto de atributos e um rótulo (classe) para o referido conjunto. Neste método, os atributos são o conjunto de palavras da saída do processo anterior referente aos exemplos de cada classe. Os rótulos são as classes alvo, POSSÍVEL RISCO SUICIDA e BAIXO RISCO SUICIDA. Assim, na Tabela 1 é demonstrada a quantidade de mensagens inserida para o treinamento do algoritmo.

Tabela 1. Quantidade de mensagens utilizadas para o treinamento do algoritmo.

Possível Risco Suicida	45 mensagens
Baixo Risco Suicida	35 mensagens

É notório na Tabela 1, a quantidade do conjunto de treinamento, onde é necessário ressaltar que as mensagens são bem diversificadas no que diz respeito ao tamanho, a menor possui cerca de 14 palavras e a maior cerca de 1012.

4. Resultados e Discussões

Na fase de análise manual foi observado quais as características mais relevantes do conjunto de treinamento. Como resultado foram obtidas 12 classes gramaticais (conforme classes do Corpus *Mac Morpho*) das palavras que possuem maior nível de importância nas mensagens, sendo elas: Pronome Pessoal, Pronome Conectivo

Subordinativo, Advérbio, Advérbio Conectivo Subordinativo, Advérbio Relativo Subordinativo, Conjunção Coordenativa, Conjunção Subordinativa, Verbo, Verbo Auxiliar, Palavra Denotativa, Nome e Nome Próprio.

Na última etapa relativa aos testes de acurácia do algoritmo, foram obtidos resultados relevantes dado a complexidade do processo de construção semântica e o método de aprendizagem do *Naive Bayes*, que desconsidera a relação semântica entre as palavras. Foram testados um conjunto de 44 mensagens não pertencentes a base de treinamento, sendo 24 referentes a classe de POSSÍVEL RISCO SUICIDA (11 implícita, 13 explícita) e 20 a BAIXO RISCO SUICIDA.

O algoritmo apresentou taxa de acerto para a classe POSSÍVEL RISCO suicida equivalente a 54,2 %, e para a classe de BAIXO RISCO SUICIDA de 95%, demonstrando eficiência regular para mensagens pertencentes a primeira classe e desempenho superior para a segunda. Além disso, o algoritmo apresentou acurácia de 70,45%.

O desempenho regular para a primeira classe é justificável. O método obteve êxito na classificação de todas as 13 mensagens que faziam referência ao ato suicida de forma explícita, e apresentou baixo desempenho para a subclasse implícita. Devido as mensagens pertencentes à subclasse implícita necessitar de alto nível de abstração para entendimento do significado, tornando-se difícil até mesmo para humanos. Assim, avaliando somente a presença de palavras, como faz o *Naive bayes*, dificilmente serão alcançados níveis de abstração elevados, sendo somente possível por meio de análises pragmáticas.

5. Conclusão

O suicídio é um grave problema de saúde pública, de origem multifatorial e que tem crescido de forma alarmante entre a comunidade jovem - mas não se restringindo a estes - em todo o mundo nas últimas décadas. Diversos foram os mecanismos criados pelos países com maior acometimento de suicídio e órgãos não governamentais com intuito de amenizar e até mesmo o extinguir. No entanto, não é uma tarefa fácil, devido principalmente a etapa de identificação das pessoas que estão em risco e que a ele são vulneráveis. Isso se justifica na necessidade do entendimento das circunstâncias que influenciam o seu comportamento autodestrutivo, para que se possa estruturar intervenções eficazes.

Diante deste contexto, este trabalho buscou desenvolver um classificador que possa ser utilizado como meio auxiliar para a identificação de pessoas que estão em potencial risco suicida. Este trabalho contribui principalmente para com a comunidade, sendo parte de uma solução voltada a saúde pública e para com o meio científico, com novas formas de aplicações das tecnologias de Inteligência Artificial.

O método desenvolvido nesta pesquisa apresentou resultados relevantes dada a quantidade de amostras de cada classe para o treinamento do algoritmo, como também o próprio algoritmo. Apresentado deficiências na classificação de mensagens que requerem abstração elevada para construção do significado. Com relação a estas deficiências, propõe-se em trabalhos futuros realizar análise pragmática por meio de histórico de mensagens de cada pessoa, fazendo com que se tenha resultados mais precisos por meio do relacionamento destas.

Ainda como trabalhos futuros, pretende-se implementar um aplicativo móvel para monitoramento da conversação, principalmente para crianças e adolescentes. O

monitoramento ocorrerá nos aplicativos de troca de mensagens. Isso se faz relevante para que os pais possam observar as intenções de seus filhos, dado que o público juvenil está entre um dos mais acometidos pelo suicídio.

Referências

Barbosa, Jardeson Leandro Nascimento; Vieira, João Paulo Albuquerque; Santos, Roney Lira de Sales; Junior, Gilvan Veras Magalhães; MUNIZ, Mariana dos Santos; Moura, Raimundo Santos. Introdução ao Processamento de Linguagem Natural usando Python. In: III ESCOLA REGIONAL DE INFORMÁTICA DO PIAUÍ. 1. Anais... SBC, 2017, v. 1. p. 336-360.

Botega, Neury José; WERLANG, Blanca Susana Guevara; CAIS, Carlos Filinto da Silva; MACEDO, Mônica Medeiros Kother. Prevenção do comportamento suicida. Psico, v. 37, n. 3, p. 5, 2006.

Hettige, Nuwan C. et. al. Classification of suicide attempters in schizophrenia using sociocultural and clinical features: A machine learning approach. General Hospital Psychiatry, v. 47, p. 20-28, 2017.

Lewis, David D. Naive (Bayes) at forty: The independence assumption in information retrieval. In: European conference on machine learning. Springer, Berlin, Heidelberg. p. 4-15, 1998.

Mccallum, Andrew; Nigam, Kamal. A comparison of event models for naive bayes text classification. In: AAAI-98 workshop on learning for text categorization. p. 41-48, 1998.

Ministério da Saúde. Portaria Nº 1.876, de 14 de agosto de 2006. Institui Diretrizes Nacionais para Prevenção do Suicídio, a ser implantadas em todas as unidades federadas, respeitadas as competências das três esferas de gestão.

Organização Mundial da Saúde. Departamento de Saúde Mental e de Abuso de Substância. Prevenção do Suicídio um Recurso para Conselheiros. Genebra, 2006. Disponível em: <http://www.who.int/mental_health/media/counsellors_portuguese.pdf>. Acesso em: 01 jan. 2018.

Organização Mundial da Saúde. Mental health: Suicide data. 2016. Disponível em: http://www.who.int/mental_health/prevention/suicide/suicideprevent/en/. Acesso em: 01 jan. 2018.

Pak, Alexander; Bernhard, Delphine; Paroubek, Patrickand; Grouin, Cyril. A combined Approach to emotion Detection in suicide notes. Biomedical informatics insights, v. 5, n. Suppl 1, p. 105-114, 2012.

Phuc, Do; Phung, Nguyen Thi Kim. Using Naïve Bayes model and natural language processing for classifying messages on online forum. In: Research, Innovation and Vision for the Future, 2007 IEEE International Conference on. IEEE, 2007. p. 247-252.

Spasić, Irena; Burnap, Pete; Greenwood, Mark; Arribas-Ayllon, Michael. A naïve bayes approach to classifying topics in suicide notes. Biomedical informatics insights, v. 5, n. Suppl 1, p. 87-97, 2012.