

Identificando Sinais de Comportamento Depressivo em Redes Sociais

Rodolpho da Silva Nascimento¹, Pedro Parreira¹,
Gabriel Nascimento dos Santos¹, Gustavo Paiva Guedes¹

¹CEFET/RJ - Centro Federal de Educação Tecnológica Celso Suckow da Fonseca
Av. Maracanã, 229 - Rio de Janeiro - RJ - Brasil.

{rodolpho.nascimento, pedro.cruz}@eic.cefet-rj.br,
gustavo.guedes@cefet-rj.br

Abstract. *There are over 300 million individuals suffering from depression in the world. Among the many consequences of this disorder, the most catastrophic is suicide. In this aspect, several individuals manifest in social networks the depressive signs before reaching the end of the suicide. In this scenario, the present study aims to construct a data set of words related to depression for the Portuguese language. Preliminary experiments indicate that the dataset created helps to classify emotions in texts of a Brazilian social network.*

Resumo. *Existem mais de 300 milhões de indivíduos que sofrem de depressão no mundo. Dentre as diversas consequências provenientes desse transtorno, a mais catastrófica é o suicídio. Nesse aspecto, diversos indivíduos manifestam em redes sociais os sinais depressivos antes de chegar ao extremo do suicídio. Nesse cenário, o presente estudo tem objetiva construir um conjunto de dados de palavras relacionadas à depressão para a língua portuguesa. Experimentos preliminares indicam que o conjunto de dados criado auxilia na classificação de emoções em textos de uma rede social brasileira.*

1. Introdução

A Organização Mundial de Saúde (OMS) destaca que a cada quatro pessoas no mundo, uma é afetada por algum tipo de problema mental em determinada fase de sua vida [Organization 2001]. Ainda segundo o estudo da OMS, a depressão é a principal causa de incapacidade global e ocupa o quarto lugar de destaque entre as dez principais doenças. Com isso, a OMS evidencia a relevância de um lançamento de campanhas por todo o mundo, objetivando prevenir sintomas da depressão ou consequências a ela associadas, como suicídio, epilepsia e esquizofrenia [WHO 2015].

Estima-se que mais de 300 milhões de indivíduos sofrem de depressão no mundo, tendo havido um aumento de 18% entre os anos de 2005 e 2015 [WHO 2017]. Dentre outras consequências negativas provenientes da depressão, a mais catastrófica é o suicídio, que é a segunda maior causa de morte entre os adolescentes [WHO 2018]. Em alguns casos de suicídio, as vítimas apresentaram, previamente, sinais depressivos em conteúdos postados nas redes sociais. Dentre eles, podemos destacar o caso de uma estudante de Rio Branco no Acre, que momentos antes de seu suicídio postou o seguinte texto *Já viram*

*alguém morrer ao vivo?*¹. Também houve o caso de um jovem de Itaparica na Bahia, que antes de cometer o suicídio, publicou uma carta de despedida no Facebook².

Nesse cenário é interessante destacar que metade dos brasileiros utilizam a internet (i.e., 48%) e desses, 92% utilizam alguma rede social [SECOM 2015]. Dessa maneira, diversos esforços vêm sendo empreendidos para evidenciar sintomas depressivos em redes sociais [Goh and Huang 2009]. Frente a isso, é importante considerar a necessidade de intervenções nos casos em que esses sintomas sejam detectados [Moreno et al. 2011].

Dado o panorama apresentando nos parágrafos precedentes, o presente trabalho tem dois objetivos principais: (i) o primeiro corresponde à criação de um conjunto de dados de palavras relacionadas à depressão para o português do Brasil, visto que não foram encontrados semelhantes na literatura; (ii) o segundo objetivo refere-se à realização de experimentos iniciais com base no conjunto de dados criado, objetivando melhorar a detecção de textos com emoção negativa. Esse objetivo se fundamenta no estudo proposto em [Shen et al. 2013], que considera a “identificação de manifestação de emoções negativas” como primeiro passo para a detecção de potenciais candidatos à depressão.

O restante desse artigo está organizado da seguinte forma: a Seção 2 descreve trabalhos relacionados ao contexto de detecção de depressão em redes sociais; a Seção 3 discute os conjuntos de dados utilizados no presente estudo; a Seção 4 detalha a criação do conjunto de dados de depressão; a Seção 5 apresenta a metodologia utilizada para avaliar o conjunto de dados criado; a Seção 6 descreve os resultados alcançados no presente estudo e, por fim, a Seção 7 conclui e apresenta alguns cenários de trabalhos futuros.

2. Trabalhos relacionados

[Nguyen et al. 2014] apresentaram um trabalho no qual investigaram comunidades de interesse em depressão e estudaram seus fatores diferenciais para outras comunidades de outros temas. Três aspectos foram examinados: afeto, processos psicolinguísticos e tópicos dentro de cada conteúdo postado. Os autores utilizaram o conjunto de dados LiveJournal³ para efetuar análise de sentimentos. Os resultados alcançados indicaram alta prevalência de palavras com características depressivas (e.g., *ódio*, *suicídio*, *dor*) nas comunidades de interesse em depressão em comparação com comunidades de outros temas.

[Nambisan et al. 2015] coletaram dados do Twitter para identificar *posts* relacionados à depressão. Este diagnóstico constituiu na formação de um conjunto de palavras ou frases, tais como “eu estou sofrendo E depressão”, “eu E medicamentos depressão”, “eu E terapia E depressão”, dentre outras. Esse conjunto de palavras serviu como filtro de seleção dos *tweets* e assim, foi criada uma amostra com os termos desses *posts*. Seguidamente, elaboraram uma segunda amostra em que foram coletados *posts* aleatoriamente. Os resultados deste estudo indicaram que os usuários deprimidos do Twitter exibem as mesmas características comportamentais quando estão conectados (postam entradas) e quando não estão, especialmente em tópicos como sono, dor e pensamentos suicidas. O

¹<https://g1.globo.com/ac/acre/noticia/estudante-transmite-suicidio-ao-vivo-em-rede-social-ja-viram-alguem-morrer.ghtml>

²<http://www.itaberabanoticias.com.br/plantao/jovem-evangelico-comete-suicidio-apos-deixar-carta-de-despedida-no-facebook>

³<https://www.livejournal.com/>

estudo também indicou que o Twitter é utilizado por indivíduos com depressão para expressarem pensamentos e refletirem sobre os sintomas e problemas que os incomodam.

Os trabalhos supracitados envolveram a elaboração de léxicos, que foram utilizados para análise de palavras que continham características depressivas, no entanto, todos foram criados para a língua inglesa. Este estudo objetiva desenvolver um léxico com características semelhantes porém, com base em palavras do português do Brasil.

3. Conjuntos de dados existentes

ANew-Br: O ANEW-br [Haag Kristensen et al. 2011] é um conjunto de dados composto por 1.046 palavras em português do Brasil. Cada uma dessas palavras é associada a valores de valência que correspondem o intervalo entre 1.16 a 8.80. Valores superiores a 5 indicam valência positiva. Analogamente, valores inferiores a 5 indicam valência negativa. Do total de palavras, 466 possuem valência negativa, que formam o conjunto de palavras relevante para o presente estudo.

LIWC: O Linguistic Inquiry and Word Count (LIWC) é um sistema proposto por James Pennebaker [Rude et al. 2004]. O LIWC classifica palavras em uma ou mais categorias psicolinguísticas com auxílio de um dicionário interno. A versão brasileira do dicionário do LIWC [Balage Filho et al. 2013], aqui denominada LIWC2007-pt-br, contém 127.149 palavras e 64 categorias psicolinguísticas. É importante destacar que, nesse estudo são utilizadas apenas as categorias *negemo*, *posemo* e as palavras associadas a essas categorias.

MQD: O Meu Querido Diário (MQD) é uma rede social brasileira criada em 2009 e tem o objetivo de funcionar como um diário online, em que os usuários podem descrever sobre o dia-a-dia. É interessante destacar que no momento em que escreve suas entradas, os usuários podem classificá-las com 1 entre as 6 emoções básicas propostas por Paul Ekman [Ekman 1992], essas são, *felicidade*, *tristeza*, *raiva*, *medo*, *nojo*, *surpresa*. O MQD possui atualmente 69.452 entradas. Dentre essas entradas, 32.244 estão associadas com a emoção *felicidade* e 26.921 estão associadas com a emoção *tristeza*. As entradas com a emoção *felicidade* e *tristeza* formam o conjunto de dados MQD-FT, que compreende 59.165 entradas.

4. Depress-pt-br

O conjunto de dados de depressão proposto neste artigo, denominado `Depress-pt-br`, foi construído em 3 fases, conforme ilustra a Figura 1. A primeira fase consistiu na aquisição de textos com palavras de cunho depressivo. Essa aquisição foi efetuada com base na coleta de dados provenientes do site Yahoo Respostas⁴, conforme adotado em [De Choudhury et al. 2013] para a língua inglesa.

Devido a limitações técnicas do Yahoo Respostas, a coleta de dados restringiu-se a 1.000 resultados por *string* de busca. Desta maneira, as *strings* de busca utilizadas na coleta foram *depressão*, *depressão+suicídio* e *depressão+socorro*, respectivamente. Essa tarefa resultou em um total de 3.000 documentos únicos. Para cada resultado, foram extraídos o (i) título resumido da pergunta; (ii) o detalhe da pergunta (nem sempre informado) e (iii) a melhor resposta para a pergunta, sendo essa indicada

⁴<http://br.answer.yahoo.com>

pelos próprios usuários do site. Essa coleta resultou em um conjunto de documentos de texto contendo termos relacionados à depressão.

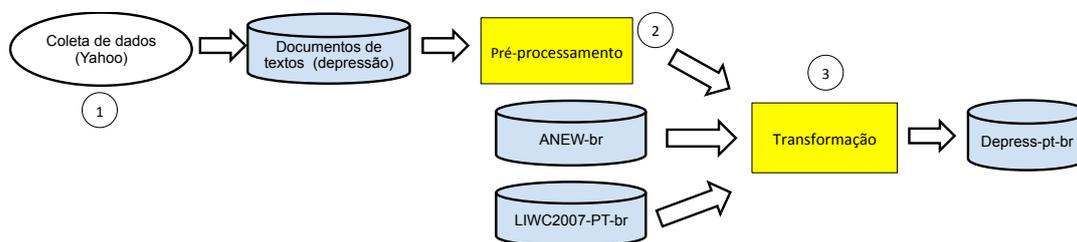


Figura 1. Fases para a construção do conjunto de dados Depress-pt-br.

A segunda fase consistiu no pré-processamento de todos os documentos de textos com termos relacionados à depressão. Foram efetuadas as tarefas de *tokenização* e *case folding* (conversão dos *tokens* para letra minúscula). É importante ressaltar que as pontuações foram removidas.

Por fim, a terceira fase consistiu em unificar o resultado da saída da fase 2 (pré-processamento) com os conjuntos de dados ANEW-br e LIWC2007-pt-br. Inicialmente, foram selecionadas todas as palavras do ANEW-br que possuem valência negativa, resultando em 466 palavras. Esse procedimento foi adotado com base na premissa de que indivíduos com depressão tendem a utilizar mais palavras com valência negativa [Rude et al. 2004].

Das 466 palavras com valência negativa do ANEW-br, foram sub-selecionadas as que estavam contidas no resultado do pré-processamento da fase 2, o que resultou em 335 palavras. Para finalizar a terceira fase, foram selecionadas todas as palavras do LIWC2007-pt-br para inclusão no Depress-pt-br, além das categorias *posemo* e *negemo*. Após essa inclusão, o Depress-pt-br continha 127.149 palavras e 2 categorias. Em seguida, foi incluída a categoria *depress* para representar as palavras relacionadas à depressão. Por fim, a categoria *depress* do Depress-pt-br foi associada a cada uma das 335 palavras sub-selecionadas do ANEW-br⁵.

5. Metodologia

A metodologia proposta neste estudo consiste em utilizar o conjunto de dados descrito na Seção 4 (i.e., Depress-pt-br) para auxiliar na detecção de entradas positivas e negativas em um conjunto de dados proveniente do MQD. Esse conjunto de dados, aqui denominado de MQD-FT, consiste em 32.244 entradas com a emoção *felicidade* e 26.921 entradas com a emoção *tristeza*.

Inicialmente, os textos de cada entrada e tiveram todos os caracteres convertidos para letra minúscula. Em seguida, os textos foram transformados em vetores de contagem de palavras. Cada dimensão do vetor correspondente diz respeito a uma das 3 dimensões presentes no Depress-pt-br (i.e., *posemo*, *negemo* e *depress*). A Tabela 1 ilustra um exemplo contendo duas entradas. Pode-se notar, por exemplo, que a entrada representada por e_1 possui 43, 52 e 14 palavras associadas à categoria de emoções positivas, emoções negativas e depressão, respectivamente.

⁵A lista com as 335 palavras se encontra em <https://github.com/LaCAfe/Depress-pt-br>.

Entrada (e)	<i>posemo</i>	<i>negemo</i>	<i>depress</i>
e_1	43	52	14
e_2	50	11	1

Tabela 1. Representação das entradas em vetores de contagem de palavras.

Em seguida, esses vetores foram submetidos ao algoritmo J48 para que fosse realizada a tarefa de classificação em duas classes no MQD-FT: *felicidade* e *tristeza*. Foi adotado o procedimento de validação cruzada com 10 partições aleatórias (*10-fold cross-validation*). Os resultados foram analisados em termos de F1 e número de entradas classificadas corretamente. O algoritmo J48 se encontra na plataforma Weka⁶.

6. Resultados

Os resultados deste estudo são exibidos com a avaliação dos atributos do conjunto de dados *Depress-pt-br* na tarefa de classificação de emoções no conjunto de dados MQD-FT. A Tabela 2 apresenta o resultado encontrado. Os atributos foram avaliados separadamente e combinados.

Atributo(s)	F1	Class. corret.
<i>negemo</i>	0,676	39.988
<i>posemo</i>	0,552	32.638
<i>depress</i>	0,611	36.153
<i>negemo</i> + <i>posemo</i>	0.679	40.242
<i>negemo</i> + <i>depress</i>	0,681	40.287
<i>posemo</i> + <i>depress</i>	0,611	36.379
<i>negemo</i> + <i>posemo</i> + <i>depress</i>	0,681	40.444

Tabela 2. Resultados da classificação obtidos com a aplicação do algoritmo J48.

Na análise dos atributos separadamente, o atributo *negemo* apresenta o melhor F1 ao ser realizada a tarefa de classificação, ou seja, 0,676. O segundo melhor atributo foi o *depress* que obteve F1 igual a 0,611. A combinação de dois atributos indica que os atributos *negemo* e *depress* alcançam o melhor F1 (0,681), ou seja, o mesmo valor alcançado pela combinação dos três atributos em questão. No entanto, a combinação dos 3 atributos possui maior número de acertos (40.444).

7. Conclusão

O presente estudo objetivou a criação de um conjunto de dados de palavras com conteúdo depressivo para o português do Brasil, denominado *Depress-pt-br*. Em seguida, utilizou esse conjunto para efetuar a classificação de emoções em textos das classes *felicidade* e *tristeza* com base em entradas de uma rede social brasileira. Os resultados indicam que o *Depress-pt-br* foi capaz de melhorar os resultados na tarefa de classificação, quando comparada a utilização dos atributos *negemo* e *posemo* do LIWC.

No conhecimento dos autores, não há trabalhos anteriores que criaram um conjunto de dados de palavras com conteúdo depressivo em português do Brasil. Dessa maneira, esse conjunto pode servir de *baseline* para novos estudos que objetivem a detecção

⁶<https://www.cs.waikato.ac.nz/ml/weka/>

de depressão em nessa língua. Como trabalhos futuros, serão adicionadas novas palavras ao conjunto de dados e será avaliada a atribuição de pesos às palavras.

Referências

- Balage Filho, P. P., Pardo, T. A. S., and Aluísio, S. M. (2013). An evaluation of the brazilian portuguese liwc dictionary for sentiment analysis. In *Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology*.
- De Choudhury, M., Gamon, M., Counts, S., and Horvitz, E. (2013). Predicting depression via social media. *ICWSM*, 13:1–10.
- Ekman, P. (1992). An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200.
- Goh, T.-T. and Huang, Y.-P. (2009). Monitoring youth depression risk in web 2.0. *VINE*, 39(3):192–202.
- Haag Kristensen, C., Azevedo Gomes, C. F., Reuwsaat Justo, A., and Vieira, K. (2011). Normas brasileiras para o affective norms for english words. *Trends in Psychiatry and Psychotherapy*, 33(3).
- Moreno, M. A., Jelenchick, L. A., Egan, K. G., Cox, E., Young, H., Gannon, K. E., and Becker, T. (2011). Feeling bad on facebook: Depression disclosures by college students on a social networking site. *Depression and anxiety*, 28(6):447–455.
- Nambisan, P., Luo, Z., Kapoor, A., Patrick, T. B., and Cisler, R. A. (2015). Social media, big data, and public health informatics: Ruminating behavior of depression revealed through twitter. *2015 48th Hawaii International Conference on System Sciences*.
- Nguyen, T., Phung, D., Dao, B., Venkatesh, S., and Berk, M. (2014). Affective and content analysis of online depression communities. *IEEE Transactions on Affective Computing*, 5(3):217–226.
- Organization, W. H. (2001). *The World Health Report 2001: Mental health: new understanding, new hope*. World Health Organization.
- Rude, S., Gortner, E.-M., and Pennebaker, J. (2004). Language use of depressed and depression-vulnerable college students. *Cognition & Emotion*, 18(8):1121–1133.
- SECOM (2015). Pesquisa brasileira de mídia 2015. url: <http://www.secom.gov.br/atuacao/pesquisa/lista-de-pesquisas-quantitativas-e-qualitativas-de-contratos-atuais/pesquisa-brasileira-de-midia-pbm-2015.pdf/view>.
- Shen, Y.-C., Kuo, T.-T., Yeh, I.-N., Chen, T.-T., and Lin, S.-D. (2013). Exploiting temporal information in a two-stage classification framework for content-based depression detection. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 276–288. Springer.
- WHO (2015). Comprehensive mental health action plan 2013–2020. url: http://www.who.int/mental_health/action_plan_2013/en/.
- WHO (2017). "depression: let's talk" says who, as depression tops list of causes of ill health. url: <http://www.who.int/mediacentre/news/releases/2017/world-health-day/en/>.
- WHO (2018). Depression. url: <http://www.who.int/mediacentre/factsheets/fs369/en/>.