Modelos de Previsão do Tom Emocional de Usuários em Comunidades de Saúde Mental no Reddit

Bárbara Silveira, Ana Paula Couto da Silva, Fabricio Murai

[1]Departamento de Ciência da Computação
Universidade Federal de Minas Gerais (UFMG) – Belo Horizonte, MG – Brazil

{barbarasilveira, ana.coutosilva, murai}@dcc.ufmg.br

Abstract. *The rise in the number of people afflicted by mental health problems has placed these disorders among the main public health problems worldwide. As a result, user activity in communities related to mental health in online social networks has spiked. Here we analyze how interactions through posts and comments influence the emotional state of users of mental health communities on Reddit, investigating whether seeking help on these networks results in changes in the feelings expressed by users over time. Our results show that the state of users changes over time, and that in 68% of cases the change is positive. We also propose predictive models to capture the variation of the emotional tone.*

Resumo. *O crescimento do número de pessoas atingidas por problema de saúde mental colocou tais distúrbios entre os principais problemas de saúde pública em todo o mundo. Como resultado, aumentou-se a procura por comunidades sobre saúde mental em redes sociais online. Neste artigo, analisamos como as interações através de posts e comentários influenciam no estado emocional de usuários de comunidades de saúde mental no Reddit, verificando se a busca por auxílio nestas redes resulta em mudanças nos sentimentos expressos pelos usuários ao longo do tempo. Nossos resultados mostram que o estado dos usuários muda ao longo do tempo, e que em 68% dos casos a mudança é positiva. Também propomos modelos preditivos para capturar a variação do tom emocional.*

1. Introdução

Dados da Organização Mundial de Saúde (OMS) mostram que a cada 40 segundos uma pessoa morre devido ao suicídio, sendo esta a segunda causa de morte entre jovens entre 15 e 29 anos[1]. Diversos fatores podem resultar nesta situação extrema, dentre eles, transtornos mentais, como a ansiedade, a depressão e a bipolaridade. Para citar algumas estatísticas, a ansiedade atinge 264 milhões de indivíduos no mundo. O transtorno bipolar[2] afeta cerca de 60 milhões de pessoas e, muitas vezes, este problema pode ser confundido com um caso de depressão ou de ansiedade, levando a um tratamento inadequado. Apesar desses números preocupantes, muitas das pessoas que sofrem de distúrbios mentais não recebem tratamento. Segundo a OMS, a cada 4 pessoas, 3 não recebem qualquer tipo de

[1]https://www.who.int/docs/default-source/mental-health/suicide/live-life-brochure.pdf?sfvrsn=6ea28a12_2$download = true$

[2]https://www.who.int/en/news-room/fact-sheets/detail/mental-disorders

tratamento e 45% da população mundial vive em um país com menos de um psiquiatra para cada 100 mil habitantes[3].

Muitas pessoas que precisam de apoio acabam não sendo tratadas devido a ausência de recursos para apoiá-la ou pelo estigma social associado aos transtornos mentais [Barney et al. 2006], o que pode gerar consequências irrevisíveis. Assim, novas ferramentas e tratamentos para auxiliar pessoas que passam por problemas relacionados à saúde mental estão sendo propostos nos últimos anos. Em especial, podemos citar as redes sociais, as quais são facilmente acessadas a baixo custo pela maioria das pessoas. O trabalho realizado por [De Choudhury 2013] mostra que o compartilhamento online de pensamentos, sentimentos e experiências com outras pessoas contribui com a melhoria do bem-estar dos indivíduos.

Neste contexto, o presente artigo analisa como as interações através de *posts* e comentários influenciam o estado emocional dos usuários de comunidades de saúde mental no Reddit[4], representado pelo tom emocional de suas mensagens. O Reddit é um site de fóruns com características de redes sociais: é composto por comunidades (subreddits), onde os usuários compartilham suas experiências e dúvidas sobre os mais diversos assuntos. Nesta rede, um usuário inicia uma *thread* ao publicar um *post*. Este *post* pode ser respondido por comentários de outros usuários ou mesmo daquele que escreveu o *post*. Comentários, por sua vez, podem ser respondidos por outros comentários. Dados de 2019 mostram que o Reddit é constituído por 430 milhões de usuários ativos por mês. Ao todo são 199 milhões de *posts*, 1,7 bilhões de comentários[5].

Nosso trabalho tem dois objetivos principais: (1) compreender e medir como redes sociais online são utilizadas para apoiar na melhoria do tom emocional de usuários com problemas de saúde mental e (2) definir modelos que permitam acompanhar a evolução do tom emocional do usuário. A partir dos objetivos listados, nossas contribuições são:

1. Análise do Tom Emocional dos Usuários. Tom emocional (TE) é o sentimento (combinação do positivo e negativo) extraído através de um texto. Sabendo que existem usuários que procuram estas comunidades em busca de ajuda e outros que estão dispostos a ajudar, analisamos a variação do estado emocional de um usuário dentro de uma *thread*, verificando se os outros usuários influenciam no TE do usuário que iniciou a discussão. Como principais resultados, observamos que, em geral: existe uma diferença positiva entre o TE do último comentário e o do *post*; o TE do último comentário do autor da *thread* é maior que o do *post* que a iniciou e superior à média do TE dos comentários da árvore de discussão.
2. Modelos para Previsão do Tom Emocional dos Usuários. A partir das análises do TE dos usuários, propomos modelos para prever a evolução do TE dos participantes destas comunidades. Utilizamos o *Multilayer Perceptron* que consiste em uma rede neural com pelo menos três camadas. Estes modelos capturam com boa acurácia (MSE 0.66) a variação do estado emocional dos usuários (que assume valores entre -2 e 2). Uma possível aplicação dos modelos é auxiliar intervenções promovidas por profissionais da área de saúde em redes sociais.

[3] `https://www.who.int/mental_health/evidence/atlas/interactive_infographic_2015.pdf`

[4] `https://www.reddit.com/`

[5] https://redditblog.com/2019/12/04/reddits-2019-year-in-review/

Este artigo está organizado da seguinte forma: a Seção 2 descreve os principais trabalhos relacionados; a Seção 3 detalha os dados e métodos utilizados; os resultados da análise do tom emocional e dos modelos são apresentados na Seção 4; as implicações deste trabalho e trabalhos futuros são discutidos na Seção 5.

2. Trabalhos Relacionados

Com o passar dos anos, asn Redes Sociais Online (RSO) passaram a permitir o compartilhamento de vários tipos de mídia, resultando em um grande número de trabalhos na literatura que utilizam dados oriundos destas redes para auxiliar o comportamento dos usuários em diferentes aspectos da vida cotidiana. Por exemplo, podemos citar o estudo realizado por [Cunha et al. 2017] que propõe um *framework* para relevar relações de causalidade entre interações de usuários no subreddit "loseit" e perda de peso.

Além da saúde física, alguns trabalhos na literatura utilizam RSOs como ferramenta para entender problemas relacionados à saúde mental. Os autores em [Blair and Abdullah 2018] utilizam dados do Instagram para identificar e analisar os desafios enfrentados pelos usuários que divulgam seus problemas de saúde mental nesta rede. O trabalho mostra que a mídia social pode ter um impacto positivo para esses usuários, uma vez que permite que eles construam suas próprias comunidades. O trabalho feito por [Islam et al.] utiliza o Facebook como objeto de estudo para detecção de depressão através de técnicas de aprendizado de máquina. Em [Gruda and Hasan 2019, Sahota and Sankar 2019, Silveira et al. 2018], os autores abordam a detecção da ansiedade, bipolaridade, depressão e suicídio em diferentes redes sociais e procuram entender o comportamento dos usuários presentes nessas redes sociais como forma de propor políticas que contribuam para a diminuição do volume de pessoas afetadas por estes problemas. Considerando o Reddit, os autores em [De Choudhury and De 2014] utilizam dados coletados desta rede para analisar os posts e comentários dos usuários, investigando como o grau de desinibição nos comentários e *posts* feitos por usuários anônimos se difere daqueles feitos pelos usuários que se identificam.

A maior parte dos trabalhos listados anteriormente tem como foco analisar o comportamento das pessoas que sofrem de doenças relacionadas à saúde mental, visando auxiliar a elaboração de políticas de saúde pública, desenvolver aplicações que auxiliem estas pessoas, entre outras medidas. Diferentemente dos trabalhos listados, neste artigo analisamos o tom emocional de usuários de quatro comunidades online relacionadas a transtornos mentais: depressão, suicídio, ansiedade e bipolaridade, com o principal objetivo de mensurar o impacto que as interações nessas comunidades geram no estado de seus usuários. Nosso estudo analisa o usuário desde o momento em que publica um *post* até o momento em que realiza seu último comentário dentro de cada *thread*, permitindo a análise da evolução de seu tom emocional e de suas interações dentro da comunidade. A partir da análise que realizamos, propomos modelos que podem ser utilizados para prever tom emocional dos usuários e que podem ser futuramente utilizados para auxiliar intervenções promovidas por profissionais da área de saúde em redes sociais.

3. Metodologia

Nesta seção, apresentamos: (i) a metodologia para analisar como as interações através de *posts* e comentários influenciam o estado emocional de comunidades de saúde mental

e (ii) os conceitos e modelos para prever a variação do tom emocional dos usuários que participam destas comunidades.

Conjunto de Dados. Existem mais de 25 subreddits que focam na discussão de transtornos mentais. Os quatro com o maior número de *posts* e comentários [Gkotsis et al. 2016] são[6]: Depression (/r/depression), SuicideWatch (/r/suicide), Anxiety (/r/Anxiety) e Bipolar (/r/bipolar). Neste trabalho, utilizamos todos os *posts* e comentários compartilhados nestas comunidades entre Janeiro de 2011 e Dezembro de 2017[7].

A Tabela 1 apresenta estatísticas básicas agregadas sobre o período selecionado. Consideramos todos *posts* e comentários, inclusive aqueles que foram removidos, deletados ou estavam vazios. Nas estatísticas apresentadas, desconsideramos os *posts* e comentários realizados por usuários deletados (i.e., usuários que foram excluídos do Reddit), porque não conseguiríamos distinguir os autores destas publicações. Podemos notar que a comunidade Bipolar é a que realiza, proporcionalmente, mais *posts* e comentários. Além disso, a média dos comentários por usuário nesta comunidade é cerca de três vezes maior que nas outras três. Isto é um indício que os usuários nesta comunidade são mais engajados durante as discussões.

	Depression	**SuicideWatch**	**Anxiety**	**Bipolar**	**Total**
Usuários Únicos	333.624	162.363	121.945	35.081	569.122
Posts	468.507	169.541	142.168	73.254	853.470
Comentários	2.128.991	1.024.171	685.867	607.262	4.446.291
Posts/usuário	1,40	1,04	1,17	2,09	1,51
Comentários/usuário	6,38	6,31	5,62	17,31	7,81
Comentários/post	4,54	6,04	4,82	8,29	5,21

Tabela 1. Estatísticas básicas de cada comunidade.

Análise Textual. Tom emocional é o sentimento (combinação do positivo e negativo) extraído através de um texto. Assim, para mensurar o estado de um usuário, analisamos o texto dos *posts* e comentários coletados utilizando a ferramenta de análise de sentimentos Vader [Hutto and Gilbert 2014]. Após a execução do algoritmo, o texto analisado é classificado, proporcionalmente, nas seguintes categorias léxicas: positiva, negativa ou neutra. Adicionalmente, a ferramenta calcula a composição dos três valores léxicos normalizados entre -1 (extremo negativo) e 1 (extremo positivo), denominado *compound*, a que se refere por tom emocional.

Análise do Tom Emocional. A análise do tom emocional e a quantificação da sua variação ao longo do tempo são realizadas considerando as árvores de discussão extraídas das interações entre os usuários dos subreddits considerados. Definimos como árvore de discussão uma *thread* inciada pelo *post* de um usuário, seguido por um ou mais comentários. Para analisarmos a variação do tom emocional em cada árvore de discussão isoladamente, consideramos apenas sequências de comentários em que: (i) exista pelo menos uma interação com outro usuário e (ii) o intervalo entre duas interações consecutivas seja menor ou igual a 24 horas.

A Figura 1 ilustra a metodologia para seleção das árvores de discussão. O eixo

[6]Não consideramos o subreddit Opiates (/r/opiates) por estar mais ligado à dependência de remédios.

[7]Os dados foram recuperados em: `http://files.pushshift.io/reddit/`

horizontal corresponde ao tempo. O usuário em questão realiza atividades (posts e/ou comentários) nas *threads* 1 a 5 durante o intervalo ilustrado. As atividades realizadas por este usuário são marcadas por *ticks* mais grossos, identificados por P no caso de *post*, ou C_i no caso do i-ésimo comentário que ele faz na *thread*. Atividades de outros usuários são marcadas pelos *ticks* mais finos. No exemplo, a *thread* 4 não será selecionada para análise porque entre o *post* e seu primeiro comentário na *thread*, ele interagiu na *thread* 2. Por razão análoga, o comentário c_2 na *thread* 2 será desconsiderado. Dado que a ordem em que as *threads* são avaliadas impacta no conjunto final de árvores de discussão, as *threads* são ordenadas de forma crescente segundo a data de criação do *post*. Após aplicar o critério de seleção apresentado, permanecemos com 87.906 árvores no Depression (27,78% do total), 41.087 no SuicideWatch (37,81% do total), 33.821 no Anxiety (32,64% do total) e 19.267 no Bipolar (33,17% do total).

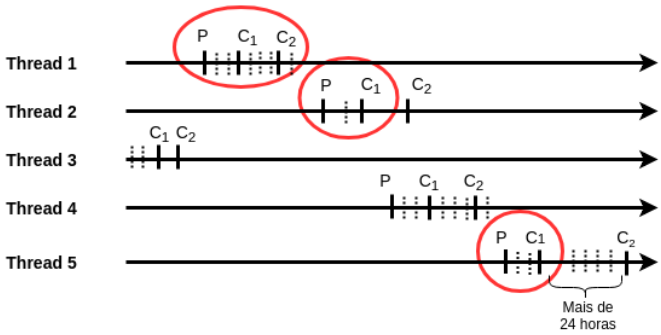


Figura 1. Apenas os trechos das árvores de discussão circulados foram considerados. O usuário A inciou as *threads* 1, 2, 4 e 5. Não consideramos a *thread* 4, pois o usuário ainda realizou interação na *thread* 2.

Tarefa de Previsão. A tarefa de previsão do tom emocional de um usuário é definida a seguir. Consideramos como entradas dos modelos de previsão: o *post* inicial p, com tom emocional TE(p) realizado por um usuário u, e o conjunto de comentários da mesma *thread* ($c_1, c_2, ..., c_{last}$), onde o comentário c_{last}, com tom emocional TE(c_{last}), é o último comentário realizado pelo usuário u. O resultado do modelo é uma previsão da variação do tom emocional, TE(c_{last}) - TE(p), do usuário u. A partir deste resultado, será possível verificar se houve alguma mudança do sentimento do usuário após as interações realizadas com os demais usuários da comunidade em questão.

Representação dos Posts e Comentários. Para desenvolver a tarefa proposta, é preciso representar posts e comentários através de atributos. Neste trabalho, utilizamos a técnica de *Word Embeddings* (WE), também conhecida como *words vectors* ou *word representations*, que representam palavras como vetores em um espaço latente. Nesse espaço, vetores similares representam palavras com o significado ou função semelhantes.

Para a criação dos *words-embeddings* usamos todos os *posts* e comentários de 2010 a 2017 de cada subreddit. No pré-processamento, utilizamos a ferramenta ekphrasis[8], a qual realiza normalização das palavras e correção ortográfica. Aplicamos a normalização no e-mail, telefones e horas. Também removemos as pontuações, caracteres especiais e excesso de espaço. Para remover as *stopwords* utilizamos o *Natural Language Toolkit* (NLTK).

[8]https://github.com/cbaziotis/ekphrasis

Na geração dos *Words Embeddings* (WE) utilizamos o pacote Gensim do Python. Os parâmetros para execução do modelo foram: arquitetura *SkipGram*; contexto igual à 5; considerando todas as palavras (que tenham aparecido ao menos uma vez no corpus); e dimensão dos vetores 300. O WE retorna um vetor para cada palavra do *post*/comentário, treinado com o corpus que possuímos. Para agregar os *embeddings* associados a cada publicação, tomamos a média dos *embeddings* das palavras. Portanto, cada publicação é representada por um vetor composto pela média dos *embeddings* de todas as palavras.

Modelos de Previsão. Neste trabalho, utilizamos o modelo de rede neural Multilayer Perceptron (MLP) para prever o tom emocional dos usuários das comunidades analisadas. Escolhemos esse modelo por capturar possíveis relações não-lineares entre o TE e pares de features (o que não seria possível com uma regressão linear), sem o elevado número de parâmetros de uma arquitetura mais profunda. Utilizamos 3 camadas ocultas, uma com 128 neurônios, as outras duas com 256. Foram criadas dois tipos de redes neurais: (i) uma apenas com perceptrons e (ii) outra com camadas de regularização. Para a regularização utilizamos a camada de $Dropout$ com os valores de 0.3, 0.5 e 0.7. Depois dos testes, permanecemos com o valor de 0.5. Consideramos 5 variações do conjunto de atributos de entrada para cada uma das redes apresentadas, discutidas a seguir. É importante ressaltar que em todos modelos, desconsideramos o último comentário do autor do *post* na árvore de discussão ao calcular o vetor de entradas para a rede (i.e., as *features*).

- Modelo 1 - Agregação de comentários (dimensão da entrada = 300): a entrada é um vetor de dimensão 300 que é o *embedding* médio de todos os comentários da árvore de discussão.
- Modelo 2 - Agregação de *Post* e Comentários (dimensão da entrada = 300): a entrada é um vetor de dimensão 300 que é a média do *embedding* do *post* com o *embedding* médio de todos comentários da árvore de discussão.
- Modelo 3 - *Post* e agregação de Comentários (dimensão da entrada = 600): a entrada é composta por dois vetores concatenados, sendo um vetor de dimensão 300 que o *embedding* médio de todos comentários da árvore de discussão, e outro vetor de dimensão 300 que é o *embedding* do *post*.
- Modelo 4 - *Post*, agregação de comentários do autor e agregação de outros comentários (dimensão da entrada = 900): a entrada é composta por três vetores, sendo um vetor de dimensão 300 que é o *embedding* médio de todos comentários (excluindo aqueles do usuário que realizou o *post*), outro vetor de dimensão 300 que é o *embedding* médio dos comentários do autor do *post*, e um vetor de dimensão 300 que é o *embedding* do *post*.
- Modelo 5 - Agregação de *post* e comentários do autor e agregação de outros comentários (dimensão da entrada = 600): a entrada é composta por dois vetores, sendo um vetor de dimensão 300 que é o *embedding* médio de todos comentários (excluindo aqueles do usuário que realizou o *post*); outro vetor de dimensão 300 igual a média entre o *embedding* do *post* e o *embedding* médio dos comentários do autor do *post*.

A avaliação da acurácia dos diferentes modelos é realizada a partir da métrica de MSE (*Mean Squared Error*).

4. Resultados

Nesta Seção apresentamos os resultados da análise do tom emocional dos usuários e dos modelos de predição propostos.

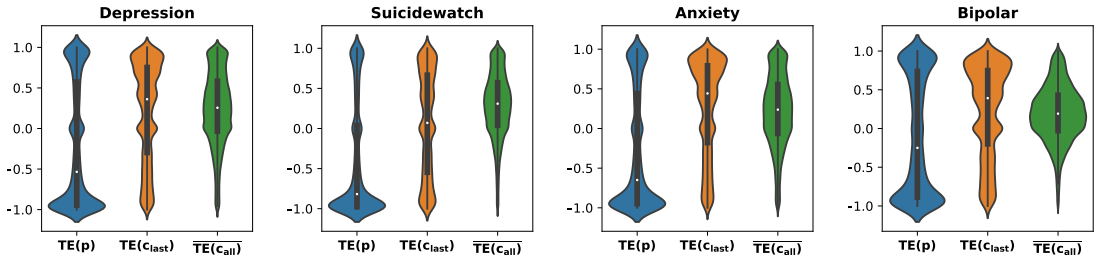
4.1. Análise do Tom Emocional

Primeiramente, introduzimos a notação que será utilizada nesta análise. Seja TE(p) o tom emocional de um *post* p; TE(c_{last}) o tom emocional do comentário feito pelo autor de uma *thread* naquela mesma *thread*; e TE(c_{all}) comentários dos outros usuários, excluindo o autor do *post*.

Características TE Árvores de Discussão. A Figura 2 apresenta a distribuição de, TE(p), TE(c_{last}) e $\overline{\text{TE}(c_{all})}$, i.e, TE do *post*, do último comentário do autor do *post* e a média de TE dos comentários da árvore de discussão, sem incluir c_{last}. O gráfico apresentado é do tipo violino, semelhante ao boxplot, mas cuja diferença é a possibilidade de se analisar a densidade dos valores: quanto mais largo o formato, maior é a concentração de pontos naquela região. O ponto branco indica a mediana. A barra mais grossa é o intervalo interquartil.

É possível observar que, em todos subreddits, o 1º quartil e a mediana referentes ao *post* estão bem abaixo das respectivas medianas dos comentários dos últimos comentários do autor e dos comentários dos outros usuários. Além disso, o 3º quartil referente ao *post* também se encontra abaixo daquele referente à média dos comentários, exceto no subreddit Bipolar. Em geral, os comentários tendem a ser mais positivos que os *posts*, exceto para o subreddit Bipolar, que apresenta também um grande volume de *posts* com o tom emocional mais positivo. Note que, em geral, há maior concentração de *posts* na região de TE negativo. Por outro lado, no caso do último comentário feito pelo autor da *thread*, há maior concentração na região de TE positivo.

Além disso, a mediana do TE médio dos comentários (i.e., de $\overline{\text{TE}(c_{all})}$) é menor que o TE do último comentário do autor do *post*, exceto no subreddit SuicideWatch. Esta análise sugere que as interações do autor de um *post* com outros usuários, ao receber comentários com TE positivo, podem contribuir para o aumento do seu TE.

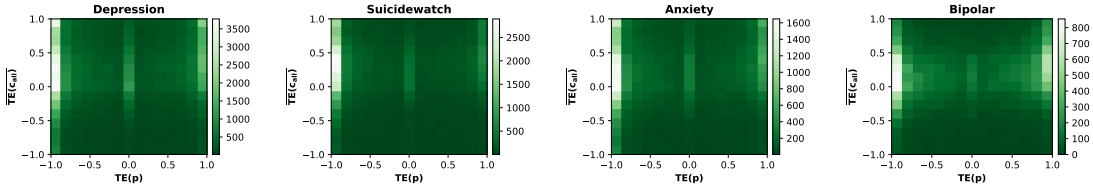


(a) Depression (b) Suicide (c) Anxiety (d) Bipolar

Figura 2. Tom Emocional.

Relação TE do Post com os Comentários. Para entender como os usuários que participam de uma árvore de discussão interagem com o autor do *post*, apresentamos a Figura 3, que mostra um mapa de calor do número de árvores de discussão, agrupadas de acordo com o TE do *post* (eixo x) e com o TE médio dos comentários da árvore de discussão,

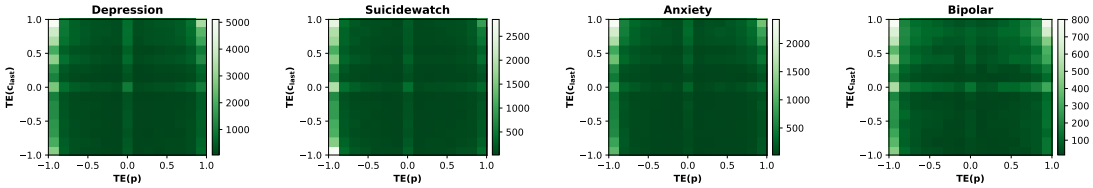
excluindo-se o *post* inicial. É possível observar que quando o *post* tem TE extremamente negativo (-1), a média dos comentários dos outros usuários na árvore de discussão tende a ser mais positiva (valores acima 0).



(a) Depression (b) Suicide (c) Anxiety (d) Bipolar

Figura 3. Tom Emocional dos Posts x Tom Emocional Médio dos Comentários.

Relação TE do Post com o Último Comentário. Averiguamos também se ocorre mudança do TE entre o momento em que um usuário escreve um *post* e o momento em que faz seu último comentário na árvore de discussão, utilizando um mapa de calor semelhante ao anterior, mostrado na Figura 4. Note que, autores de *posts* extremamente negativos (-1) tendem a escrever comentários mais positivos ao final da *thread*. Esta variação sugere que os comentários realizados por outros usuários podem ajudar os usuários que se encontram em situações difíceis.

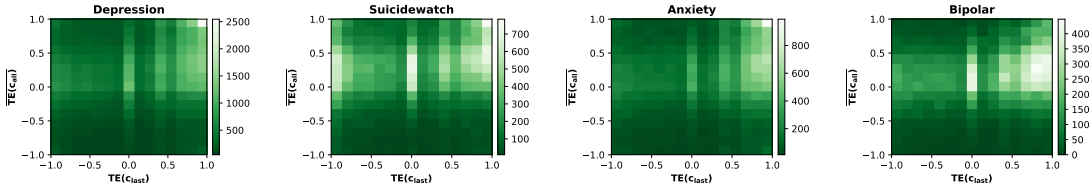


(a) Depression (b) Suicide (c) Anxiety (d) Bipolar

Figura 4. Tom Emocional dos Posts x Tom Emocional do Último Comentário.

Relação TE do Último Comentário com os demais Comentários. O passo seguinte foi investigar se o TE do último comentário do autor tem relação com a média do TE dos comentários da árvore de discussão (excluindo-se o *post* inicial), calculando-se o mapa de calor mostrado na Figura 5. Note que, em todos subreddits, um TE positivo no último comentário do autor do *post* está relacionado a um TE positivo médio entre os comentários *thread*. No SuicideWatch também ocorre uma grande quantidade de comentários negativos por parte dos autores da *thread*, mesmo quando estão relacionados a uma árvore de discussão com TE positivo. Isto pode ser um indício de que existem usuários nesta comunidade que a melhoria do tom emocional, ou seja, o aumento da positividade nos textos das suas atividades, é mais difícil de ocorrer.

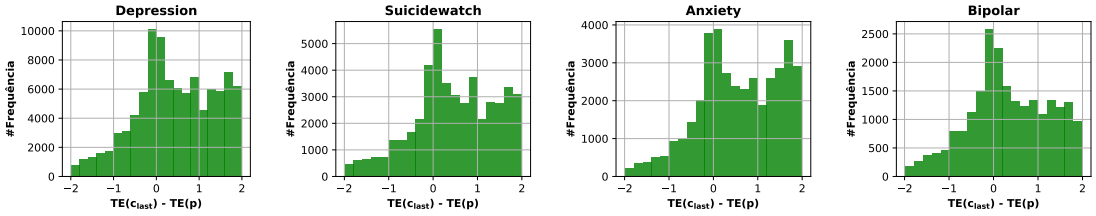
Relação TE dos Comentários com Diferença do TE do último comentário e o Post. Como observado anteriormente, o TE dos comentários das árvores de discussões estão positivamente correlacionados com o TE do último comentário. Consequentemente, podem estar relacionados com uma variação positiva, considerando-se a diferença entre o TE do *post* e o TE do último comentário. Para investigar melhor estas relações, analisamos a diferença, para cada árvore de discussão, entre o tom emocional do último comentário do autor e do *post*. A Figura 6 apresenta a distribuição desta diferença. Observe que ocorre



(a) Depression (b) Suicide (c) Anxiety (d) Bipolar

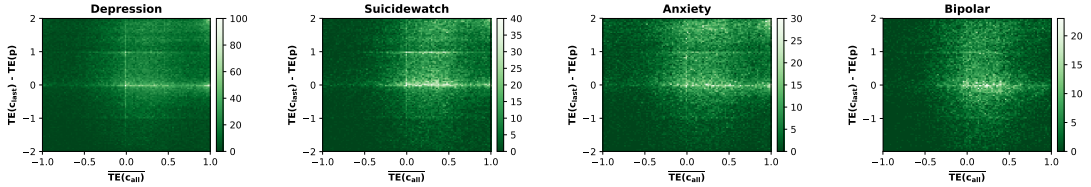
Figura 5. Média do Tom Emocional dos Comentários x Tom Emocional do último comentário

maior concentração nos valores acima de zero, mostrando que o usuário, na maioria das vezes, "melhora" ao final da *thread*, em relação ao TE. Analisamos então a relação entre a média do TE dos comentários e a variação do TE observada pelo autor do *post*. Na Figura 7, notamos que quando a média é positiva, existe uma diferença positiva entre o TE do último comentário e o *post*, a qual pode ser observada no canto superior direito. Note que quando a média é negativa, não existe um padrão claro para a diferença do TE.



(a) Depression (b) Suicide (c) Anxiety (d) Bipolar

Figura 6. Histograma da Diferença do Tom Emocional entre comentário e post



(a) Depression (b) Suicide (c) Anxiety (d) Bipolar

Figura 7. Média do Tom Emocional dos Comentários x Diferença do tom emocional entre último comentário e post.

Resumindo, as análises realizadas fornecem um conjunto de observações relevantes a respeito da dinâmica do tom emocional de um usuário ao participar em uma árvore de discussão. Investigamos vários aspectos do TE dentro da árvore de discussão: nos *posts*, nos comentários, no último comentário do autor do *post* e na diferença do TE. Um dos resultados mais importantes é a existência da correlação entre o último comentário do usuário que realizou a postagem com os comentários dos demais usuários na árvore de discussão. Além disso, analisamos que um comentário influencia o TE do próximo comentário, um indício de que o aspecto temporal pode impactar na evolução do comportamento do usuário dentro da comunidade. A seguir, avaliamos a acurácia do modelo para prever a diferença do TE do usuário ao longo tempo, dentro da árvore de discussão.

4.2. Previsão do Tom Emocional

Nesta seção apresentamos os resultados da acurácia dos modelos de previsão propostos.

Baselines. Definidos três baselines para avaliar o quão efetivos são os modelos propostos. Seja D a variável aleatória que representa a variação do TE sofrida pelo autor de uma árvore de discussão escolhida uniformemente ao acaso, definida por $D = \text{TE}(c_{last}) - \text{TE}(p)$. Sejam ainda D_i a variação do TE associada à *thread* i e $\hat{D}_i$ a respectiva previsão, retornada por um dado modelo. Considerando o *Mean Squared Error* (MSE) como medida de erro, podemos avaliar os modelos utilizando a equação

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(D_i - \hat{D}_i)^2, \tag{1}$$

onde as variáveis $D_1, \ldots, D_n$ representam a variação de TE associada às *threads* do conjunto de teste. Para calcular as métricas de erro, utilizamos *cross-validation* com 5 *folds*.

Os resultados dos modelos propostos neste artigo são comparados com os resultados obtidos pelos seguintes baselines:

- Baseline 1: O primeiro baseline é o modelo que sempre retorna 0 para a variação do TE, ou seja, que assume que o usuário manteve seu TE na árvore de discussão. Neste caso, o MSE do Baseline 1 é dado por: $\text{MSE}(B_1) = E[D^2]$.
- Baseline 2: O segundo baseline é o modelo que sempre retorna a variação média do TE, $\bar{D} = \frac{1}{n}\sum_{i=1}^{n} D_i$. Neste caso, o erro de previsão é calculado por: $\text{MSE}(B_2) = E[(D - \bar{D})^2]$, ou seja, é a variância de D.
- Baseline 3: O terceiro baseline utilizado foi uma regressão linear considerando como atributos o total de comentários e a média do TE da árvore de discussão (excluindo-se àqueles de quem fez o *post*) para prever a diferença do TE. O erro correspondente é calculado através da Eq. 1.

A Tabela 2 apresenta o MSE obtido para cada baseline avaliado em cada comunidade. Observe que os MSE's do baseline 3 são menores que os MSE's dos baselines 1 e 2, mas que a diferença em relação ao modelo 2 é pequena, indicando a dificuldade de prever D a partir de $\overline{\text{TE}(c_{all})}$. Assim, nossos modelos terão o objetivo apresentar resultados mais precisos que os resultados do baseline 3.

	Mean Squared Error (MSE)		
Subreddit	**Baseline 1**	**Baseline 2**	**Baseline 3**
Depression	1,023	0,854	0,848
SuicideWatch	1,029	0,843	0,839
Anxiety	1,062	0,818	0,813
Bipolar	0,964	0,819	0,811

Tabela 2. MSE's dos Baselines para cada comunidade.

Avaliação. Testamos cada modelo com e sem regularização, a Tabela 3 apresenta os resultados das 10 configurações (5 modelos $\times$ uso/não-uso de regularização) para a comunidade Bipolar. Observe que os resultados dos modelos com e sem regularização são semelhantes. Entretanto, utilizar a regularização é importante para evitar *overfitting*. Assim escolhemos, dentre os modelos com regularização, o modelo 4, por resultar no menor MSE dentre os demais, sendo inclusive bem menor que o do baseline 3. Em suma, executamos todos os modelos para a comunidade Bipolar e quando encontramos o melhor modelo, aplicamos para as demais comunidades. Note que este modelo é o que mais

distingue dentre os tipos de publicação, uma vez que existe uma dimensão para os comentários dos usuários que não realizaram o *post*, outra para os comentários de quem fez o *post* e outra para o *post*. Isto mostra que separar os textos em diferentes vetores de entrada é importante para a previsão da diferença do TE.

A única diferença do modelo 5 para modelo 4 é que, no modelo 5, o vetor do *embedding* médio dos comentários do autor do *post* e o vetor dos *embedding* do seu *post* são agregados em um único vetor, que possui características textuais do mesmo autor. Esta diferença ocasiona um erro 8.8% maior do modelo 5. Apesar desses textos terem sido realizados pelo mesmo autor, é importante distinguir entre *post* e comentários do autor para melhor prever a variação do TE. Isto ocorre possivelmente devido à mudança no discurso dos comentários do autor do *post*, influenciado pelos demais usuários.

Aplicamos o modelo 4 às demais comunidades, Depression, SuicideWatch e Anxiety. A Tabela 4 apresenta os resultados. A maior diferença percentual no MSE entre comunidades é de apenas 1,02% (entre as comunidades Bipolar e Depression), mostrando que os resultados para as quatro comunidades são similares. Em comparação com o Baseline 3, o modelo 4 obteve melhores resultados de acurácia para todas as comunidades: na comunidade Depression, a diminuição do MSE é de 20,79%, no SuicideWatch 20,04%, no Anxiety 18,02% e no Bipolar 18,02%. Uma possível explicação para o melhor desempenho do modelo 4 na tarefa de previsão proposta, está na separação entre o *post*, os comentários do autor da *thread* e os comentários dos outros usuários, isolando diferentes tipos de tons emocionais inerentes às atividades consideradas: por exemplo, o *post* tem o tom mais negativo e normalmente é um pedido de ajuda, os comentários dos outros usuários costumam ter o tom mais positivo e tem o papel de oferecer suporte e conselhos, já os comentários do autor do *post* podem refletir um processo de melhora.

	Sem Regularização		Com regularização	
	Media MSE	DP MSE	Media MSE	DP MSE
Modelo 1	0,8186	0,0236	0,8061	0,0192
Modelo 2	0,784	0,0182	0,7844	0,0186
Modelo 3	0,671	0,0084	0,6961	0,0446
Modelo 4	0,6781	0,0205	0,6651	0,0191
Modelo 5	0,7174	0,0121	0,7295	0,0203

Tabela 3. Resultados MSE das variações dos Modelos MLP's para a comunidade Bipolar. Baseado em 5 folds em termos de média e desvio padrão.

Subreddit	Media MSE	DP MSE
Depression	0,6719	0,0225
SuicideWatch	0,6707	0,02568
Anxiety	0,6665	0,0264
Bipolar	0,6651	0,0191

Tabela 4. Resultados do MSE Modelo MLP. Baseado em 5 *folds* em termos de média e desvio padrão.

Em suma, os modelos propostos neste trabalho são eficazes para capturar a variação do TE do usuário em uma árvore de discussão. Esses modelos podem ser utilizados no monitoramento de conversas, por exemplo, em redes sociais de suporte a usuários com problemas de saúde mental. Ao identificar que o usuário não está melhorando seu tom emocional, especialistas podem intervir para evitar situações extremas.

5. Conclusão

As análises apresentadas neste trabalho mostram que as interações entre os usuários de RSOs relacionadas aos transtornos de saúde mental auxiliam na melhoria das suas condições de saúde, uma vez que o TE dos usuários tendem a sofrer variações positivas

ao longo das interações realizadas. Com base nesta análise, propomos modelos que preveem com boa acurácia a variação do TE dos usuários participantes destas comunidades.

Como os diversos trabalhos que utilizam dados de redes sociais online, nosso trabalho possui como principal limitação a impossibilidade de verificar se uma variação positiva no TE está correlacionada com uma melhora real no estado emocional do usuário, já que não temos um contato direto com usuário para averiguação. No entanto, nossos modelos podem ser aplicados para monitorar a evolução do tom emocional de tais usuários durante as discussões realizadas nestas comunidades: caso o TE não esteja evoluindo positivamente, intervenções podem ser realizadas por profissionais da área, evitando situações extremas (por exemplo, o suicídio).

Referências

[Barney et al. 2006] Barney, L. J., Griffiths, K. M., Jorm, A. F., and Christensen, H. (2006). Stigma about depression and its impact on help-seeking intentions. *Australian & New Zealand Journal of Psychiatry*, 40(1):51–54.

[Blair and Abdullah 2018] Blair, J. and Abdullah, S. (2018). Supporting constructive mental health discourse in social media. In *PervasiveHealth '18*.

[Cunha et al. 2017] Cunha, T., Weber, I., and Pappa, G. (2017). A warm welcome matters!: The link between social feedback and weight loss in/r/loseit. In *WWW'17 Companion*, pages 1063–1072. International World Wide Web Conferences Steering Committee.

[De Choudhury 2013] De Choudhury, M. (2013). Role of social media in tackling challenges in mental health. In *Proceedings of the 2nd international workshop on Socially-aware multimedia*, pages 49–52. ACM.

[De Choudhury and De 2014] De Choudhury, M. and De, S. (2014). Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *ICWSM*.

[Gkotsis et al. 2016] Gkotsis, G., Oellrich, A., Hubbard, T., Dobson, R., Liakata, M., Velupillai, S., and Dutta, R. (2016). The language of mental health problems in social media. In *Proceedings of the Third Workshop on Computational Lingusitics and Clinical Psychology*.

[Gruda and Hasan 2019] Gruda, D. and Hasan, S. (2019). Feeling anxious? perceiving anxiety in tweets using machine learning. *Computers in Human Behavior*, 98:245–255.

[Hutto and Gilbert 2014] Hutto, C. J. and Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *ICWSM*.

[Islam et al.] Islam, M. R., Kabir, M. A., Ahmed, A., Kamal, A. R. M., Wang, H., and Ulhaq, A. Depression detection from social network data using machine learning techniques. *Health Information Science and Systems*.

[Sahota and Sankar 2019] Sahota, P. K. and Sankar, P. L. (2019). Bipolar disorder, genetic risk, and reproductive decision-making: A qualitative study of social media discussion boards. *Qualitative health research*.

[Silveira et al. 2018] Silveira, B., da Silva, A. P. C., and Murai, F. (2018). Análise de comunidades de suporte a transtornos de saúde mental do reddit. In *Anais do VII Brazilian Workshop on Social Network Analysis and Mining*. SBC.