

Um estudo sobre o aprimoramento de tuítes com base na substituição de adjetivos

Juliano M. Pasa¹, Lucas L. de Oliveira², Sérgio L. S. Mergen³

¹Centro de Tecnologia – Universidade Federal de Santa Maria (UFSM)
97105-900 – Santa Maria – RS – Brasil

²Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brazil

³Departamento de Linguagens e Sistemas de Computação
Universidade Federal de Santa Maria (UFSM)

{jdpasa,mergen}@inf.ufsm.br, lloliveira@inf.ufrgs.br

Abstract. *Nowadays, social networks like Twitter are used not just for entertainment, but as a work tool, for political, artistic or purely commercial purposes. Many of the messages posted are created carefully, in order to gain loyalty and expand the base of followers. In this context, this work proposes the re-writing of tweets in order to increase their engagement. The proposed rewrite replaces adjectives with synonyms considered more appropriate, taking into account the adjectives that are commonly associated with tweets that have a high engagement rate. Experimental results using a machine learning algorithm as the evaluator demonstrate that the rewrite process potentially generates better tweets.*

Resumo. *Hoje em dia, redes sociais como o Twitter são usadas não apenas para entretenimento, mas como uma ferramenta de trabalho, seja de cunho político, artístico ou puramente comercial. Muitas das mensagens publicadas são concebidas com cuidado, de modo a fidelizar e ampliar a base de seguidores. Nesse contexto, este trabalho propõe a reescrita de tuítes visando aumentar o seu engajamento. A reescrita proposta substitui adjetivos por sinônimos considerados mais adequados, levando em consideração os adjetivos que são comumente associados à tuítes com alta taxa de engajamento. Os resultados experimentais usando um algoritmo de aprendizado de máquina como avaliador demonstram que o processo de reescrita potencialmente gera tuítes melhores.*

1. Introdução

Conforme demonstrado pela plataforma de análise estatística, Statista¹, o Twitter ainda está entre as redes sociais mais utilizadas no mundo, com cerca de 353 milhões de usuários, além de estar entre aquelas com o maior engajamento em países como os Estados Unidos. Por ser uma fonte rica em texto livre, muitas bases foram criadas a partir de dados coletados do Twitter, que serviram como ponto de partida para diversos estudos e análises de diferentes áreas da computação, mas principalmente, para o processamento

¹Statista: <https://www.statista.com/topics/737/twitter/>

de texto e linguagem natural. Uma grande parcela destes trabalhos tem como objetivo principal a análise de sentimentos e a predição da popularidade de uma publicação, ou suas chances de ser retuitada. Estes tópicos vão ao encontro com uma das preocupações de usuários dessa rede, sejam eles criadores de conteúdo ou empresas, que é obter um maior engajamento em suas publicações.

Com o avanço de tecnologias como as redes neurais e *word embeddings*, não especificamente no contexto do Twitter, muitas ferramentas já foram desenvolvidas para processar e auxiliar no processo de escrita. Um claro exemplo disso é a funcionalidade Smart Compose [Chen et al. 2019], que foi adicionada ao Gmail para tornar a escrita de e-mails mais rápida, sugerindo novas palavras.

Neste contexto, este trabalho tem por objetivo o desenvolvimento de um modelo capaz de processar e reescrever tuítes, substituindo adjetivos por sinônimos que resultem em mensagens com maiores chances de engajamento. Para isso, é usada uma base pontuada de adjetivos, na qual a pontuação de cada palavra está associada à taxa de engajamento de tuítes que contenham essa palavra.

O restante deste artigo está estruturado nas seguintes seções. A seção 2 apresenta os trabalhos relacionados. A seção 3 apresenta a proposta, juntamente com a arquitetura de preparação e transformação dos dados. A seção 4 apresenta os experimentos realizados, bem como o processo de validação dos resultados. A seção 5 apresenta as considerações finais.

2. Trabalhos Relacionados

Desde sua criação, em 2012², a API do Twitter permite a interação com recursos e a coleta de dados de sua plataforma. Sendo uma fonte de dados tão rica e disponível para coleta, muitos trabalhos já foram desenvolvidos tendo os tuítes como base. Dentre estes trabalhos, destacam-se aqui aqueles que analisam características extraídas dos tuítes, como também aqueles que fazem o processamento de seus textos. Muitos deles têm, como parte de seus objetivos, a análise e identificação de fatores impactantes no conteúdo das mensagens propagadas.

No que se refere à análise de características, [Suh et al. 2010] e [Oliveira and Mergen 2018] avaliaram um conjunto de atributos extraídos das mensagens, como a utilização de URLs, *hashtags*, tamanho das mensagens e polaridade, para analisar o impacto sobre os indicadores de engajamento, sendo eles a quantidade de retuítes, curtidas e seguidores do autor. Os resultados mostraram que a utilização de determinados recursos como *hashtags* e URLs são fatores muito significativos e que influenciam no engajamento de uma publicação.

Outros trabalhos utilizando o Twitter como base, mas com diferentes objetivos, podem ser mencionados aqui. [Duan et al. 2010] propôs um novo método de ranqueamento de tuítes, considerando a relevância de seu conteúdo, além de características como o uso de URLs. [Benevenuto et al. 2010] coletou uma base com mais de um bilhão de tuítes e desenvolveu um modelo para classificar quando estas mensagens são spam ou não. Em [Naveed et al. 2011] foi feita uma análise sobre o que pode tornar um tuíte interessante e propôs um modelo para prever as chances de uma mensagem ser retuitada.

²Twitter API: <https://developer.twitter.com/en/docs/twitter-api>

Ainda neste contexto, em [Zhang et al. 2016] foi desenvolvida uma rede neural profunda para realizar este mesmo tipo de predição de retuíte. Um fator em comum entre eles é a utilização de algoritmos de aprendizado de máquina na realização de seus experimentos. Dentre as principais técnicas utilizadas estão: SVM, (*Support Vector Machine*); árvores de decisão; Naive Bayes; Regressão logística; PCA (*Principal Components Analysis*) e redes neurais.

Já no contexto de processamento de textos, mais especificamente NLP (*Natural Language Processing*), existem uma série de subdomínios, como reconhecimento de entidades nomeadas e reconhecimento de fala. Dentro deste campo, [Derczynski et al. 2013] realiza uma análise detalhada de métodos de *POS tagging* (*Part-of-speech tagging*) aplicados a textos do Twitter, o que impõe maior dificuldade a este processo, por serem textos que contém ruídos, erros ortográficos e estilo próprio de escrita. [Jain et al. 2019] propõe um arquitetura de redes neurais profundas, utilizando *WordNet* and *ConceptNet* na representação semântica das palavras para detectar mensagens de spam.

Não focando apenas do contexto do Twitter, outros autores direcionaram suas pesquisas para a geração e/ou reescrita de textos. [Shang et al. 2015] criou uma rede neural treinada sobre textos coletados de um serviço de microblog, semelhante ao Twitter, para geração de respostas em textos curtos, para conversação, capaz de gerar frases gramaticalmente corretas. [Xu et al. 2019] propõe uma ferramenta para auxiliar na reescrita de textos e tarefas de geração de linguagem natural. [Goldfarb-Tarrant et al. 2019] propôs um sistema interativo entre humano e computador capaz de auxiliar na escrita de histórias.

Esses trabalhos mostram que os campos de pesquisa abrangendo NLP e dados do Twitter são muito amplos e ainda existem outros trabalhos relevantes nessa área. Diferentemente dos trabalhos mencionados, o presente artigo busca analisar a modificação no uso das palavras empregadas em textos de tuítes, correlacionando-as com o engajamento obtido por eles.

3. Proposta

A proposta deste estudo é o aprimoramento de tuítes com substituição de palavras por outras que possam gerar um maior engajamento. As substituições são restritas às palavras que têm função de adjetivação. A hipótese a ser validada é que essas palavras (adjetivos) tenham um potencial maior para sensibilizar os leitores, e sua escolha criteriosa possa ajudar no impulsionamento de uma publicação.

Define-se como *taxa de engajamento* a métrica utilizada para medir o alcance, ou popularidade, de uma publicação a partir das interações dos usuários com este conteúdo. Conforme mostra [Pillat and Pillat 2017], por convenção, utiliza-se a Equação 1 – atualizada para as interações permitidas no Twitter – para calcular a taxa de engajamento de uma publicação, da qual obtém-se um número contínuo que varia de 0 a infinito.

$$E(x) = \frac{\text{curtidas} + \text{retuítes} + \text{respostas}}{\text{seguidores}} * 100 \quad (1)$$

Para este trabalho, devido a limitações na extração dos dados e restrições da API do Twitter, a quantidade de respostas para cada tuíte não pode ser obtido e, portanto, será ignorada.

A Figura 1 exibe a arquitetura responsável pelo processamento de tuítes. De modo geral, a etapa de preparação é responsável por gerar um modelo de engajamento, formado por uma lista de adjetivos associados com um valor de pontuação. Já a etapa de transformação utiliza esse modelo para realizar a reescrita de um tuíte. Todos os módulos foram desenvolvidos em Python e as próximas seções irão detalhá-los.

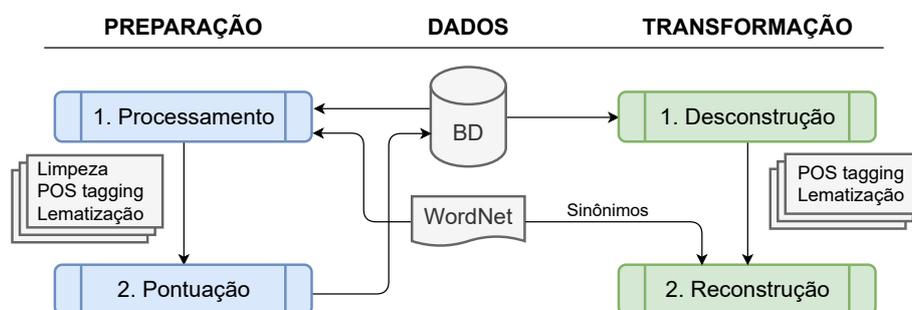


Figura 1. Arquitetura de processamento

3.1. Preparação dos dados

A etapa de preparação dos dados utiliza tuítes previamente armazenados em um repositório para criar o modelo de engajamento. O processo de preparação é dividido em duas etapas: i) processamento do texto, e ii) pontuação.

3.1.1. Processamento do texto

Em um primeiro momento, os tuítes passam por uma etapa de processamento, cujo objetivo é identificar no texto as partes que irão compor o modelo. Um processo inicial de limpeza é encarregado de retirar todos os sinais de pontuação, padronizar o alfabeto (mantendo apenas símbolos presentes na tabela ASCII) e deixar todas as letras restantes em minúsculo.

Em seguida, de cada tuíte, são mantidos no modelo apenas os termos que têm função de adjetivação. Esses termos são identificados a partir de um módulo de *POS tagging*. *Part-of-speech tagging* (marcação de partes do discurso) é uma técnica utilizada para identificar as classes gramaticais de cada palavra presente em uma frase baseada em seu contexto.

Por fim, é realizada a lematização das palavras. Este processo reduz as palavras à sua forma de origem. A lematização é importante pois ela reduz a quantidade de palavras que precisam ser controladas e cria um modelo mais coeso, no qual palavras que carregam o mesmo significado e derivem do mesmo *lema* são tratadas de modo consonante. Isso é especialmente relevante durante a etapa de pontuação, como será apresentado a seguir.

Para realizar a lematização, foi utilizada a biblioteca *NLTK*³ (Natural Language Toolkit). Com essa biblioteca é possível utilizar a *WordNet*⁴, uma base de dados lexical da língua inglesa que armazena mais de 100 mil *synsets* – conjuntos de palavras que apresentam significados próximos.

³*NLTK*: <https://www.nltk.org/>

⁴*WordNet*: <https://www.nltk.org/howto/wordnet.html>

3.1.2. Pontuação

Este módulo tem como finalidade criar o modelo que será usado para a reescrita de tuítes. O modelo é formado por um conjunto de adjetivos, sendo que cada adjetivo está associado a uma pontuação. A pontuação é um valor numérico (não necessariamente normalizado) que indica a importância inferida do adjetivo para alavancar uma publicação.

A pontuação das palavras é determinada por uma função que leva em consideração a taxa de engajamento de cada tuíte em que a palavra ocorra. As seguintes funções foram propostas:

- **Média simples:** soma a taxa de todas as ocorrências da palavra e divide pela frequência total da palavra;
- **Média harmônica:** soma o inverso da taxa de todas as ocorrências da palavra e divide a frequência total da palavra pela soma;
- **Somatório:** soma a taxa de todas as ocorrências da palavra.

3.2. Transformação

A etapa de transformação recebe um tuíte original e devolve um tuíte reescrito, em que os adjetivos são substituídos por outros que mantenham o sentido mas que possuam maior pontuação. O processo de transformação é dividido em duas etapas: i) desconstrução do texto, e ii) reconstrução do texto.

3.2.1. Desconstrução do texto

Esta etapa aplica ao texto de entrada os mesmos passos do processamento realizados anteriormente, já descritos na Seção 3.1.1, em que os adjetivos do tuíte são identificados e lematizados. No entanto, o tuíte original é preservado após a etapa de *POS Tagging*, na qual cada palavra é rotulada com a sua classe gramatical. A presença do tuíte original rotulado serve como contexto para que a substituição do adjetivo ocorra na posição correta.

3.2.2. Reconstrução do texto

Nesta etapa do processo, para realizar a reconstrução do texto, primeiramente, é necessário fazer a busca pelas possíveis palavras que substituirão cada adjetivo do tuíte original, considerando suas respectivas pontuações. A busca utiliza o WordNet para identificar os sinônimos a partir do *synset* de uma palavra. As palavras encontradas formam a lista de candidatos para a substituição.

Uma versão simplificada do processo realizado nesta etapa é apresentado em pseudo código descrito no Algoritmo 1. Para cada palavra dessa lista deve ser encontrado o seu item correspondente no modelo e recuperada a sua pontuação. Palavras que não existam no modelo recebem uma pontuação nula. É importante ressaltar que cada sinônimo também passa pelo processo de lematização, para que seja encontrado seu correspondente na base. Finalmente, o sinônimo com a maior pontuação é escolhido para a substituição.

Algorithm 1 Pseudoalgoritmo para reescrita dos tuítes

```
1: function REALIZATROCA(todosTuites)
2:   for each tuite in todosTuites do
3:     todosSinonimos ← encontraSinonimos(tuite)
4:     for each sinDoAdjetivo in todosSinonimos do
5:       melhorAdj ← sinDoAdjetivo[maiorPontuacao]
6:       texto[adjCorrespondente] ← melhorAdj
7:
8: function ENCONTRASINONIMOS(tuite)
9:   sinonimos ← []
10:  for each classe, palavra in tuite['classes'], tuite['texto'] do
11:    if (classe == adjetivo) then
12:      lemas ← []
13:      for each synset in synsets(palavra) do
14:        for each lema in synset.lemma do
15:          lemas.novo(lema)
16:      sinonimos.novo(baseDeSinonimos(palavra, lemas))
17:  return sinonimos
```

4. Experimentos

O objetivo dos experimentos é comparar os tuítes originais com os tuítes que foram modificados, de modo a identificar quais deles têm potencial de obter um maior engajamento. Como não existe um oráculo humano capaz de fornecer uma resposta definitiva, utilizamos um algoritmo de aprendizado de máquina como avaliador. O processo completo de avaliação segue os processos demonstrados na Figura 2.

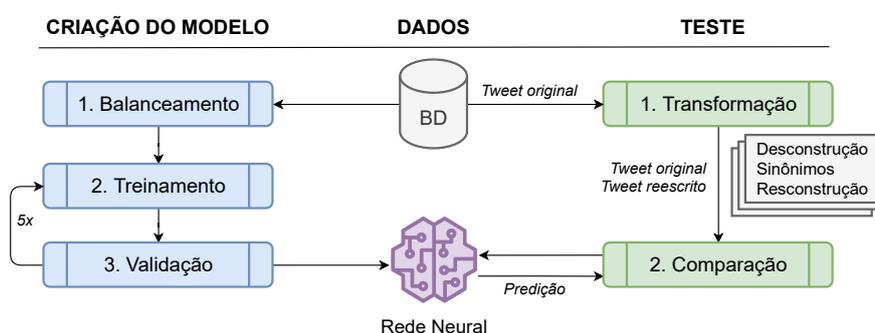


Figura 2. Processo de validação

A primeira etapa corresponde à criação do modelo a ser usado como oráculo utilizando uma base de tuítes para treiná-lo. Antes do treinamento, a base passa por um balanceamento e, após o treinamento, o modelo é validado de modo a verificar a sua eficácia. A seção 4.2 detalha como ocorre o balanceamento e a validação. A segunda etapa engloba todo o processo de reescrita do tuíte, já descrito na Seção 3.2. Uma vez obtido o tuíte reescrito, o modelo é usado para avaliar se a nova versão é mais adequada do que a versão original.

4.1. Base de dados utilizada

A base utilizada contém no total 434.238 tuítes, que datam entre 2008 e 2021, distribuídos entre 100 contas de usuários. Foram utilizadas contas de usuários influentes, como Barack Obama, Rihanna, Jimmy Fallon, NASA, entre outros, e inclusive de usuários que não participam mais do Twitter, como Donald Trump. Foram escolhidas contas que tem como principal idioma a língua inglesa, devido a maior quantidade de recursos disponíveis.

Além disso, a escrita dos adjetivos em inglês possui uma variação morfológica menor em comparação com o português por exemplo.

Os dados foram coletados usando os recursos de *streaming* já empregados em [Oliveira and Mergen 2018] e também explorando recursos que permitem acompanhar a *timeline* dos usuários selecionados. A quantidade de tuítes coletados de cada usuário varia, bem como a data do tuíte mais antigo, dependendo do nível de atividade de cada usuário. Por exemplo, a conta Reuters Top News é a mais ativa, tendo enviado 7.313 tuítes dentro do período utilizado para a coleta.

Após uma filtragem, que desconsidera os tuítes que não sejam em inglês, mensagens que sejam retuítes e também que não possuam todas os atributos para o cálculo de engajamento, restaram um total de 262.468 tuítes únicos.

4.2. Definição do Modelo

O algoritmo de aprendizado de máquina a ser usado como oráculo deve receber um tuíte e classificá-lo como popular ou não popular. A saída do algoritmo deve ser um escore entre 0 e 1, sendo que quanto mais próximo de 1, maior a confiança de que o tuíte seja popular. Com esse cenário de avaliação configurado, dados os tuítes original e reescrito, aquele que obtiver o maior escore é considerado mais popular. Ainda, a diferença percentual de melhora entre os escores indica o quanto um tuíte é considerado mais popular que o outro.

Foram testados três modelos de inteligência artificial que são comumente utilizados em tarefas de classificação de texto: LSTM (*Long short-term memory*), GAP1D (*Global Average Pooling 1D*) e SGD (*Stochastic Gradient Descent*). Para a criação do modelo, a base de dados foi balanceada entre tuítes populares e não populares, sendo utilizado 80% e 20% dos dados para treinamento e teste da rede, respectivamente. Dentro da etapa de treino, os dados foram divididos novamente na mesma proporção para treinamento e validação do treino e, por fim, a rede analisa-os cinco vezes.

Cada modelo foi avaliado com dois pontos de corte para a classificação da popularidade dos tuítes. Usando a taxa de engajamento, caso ela esteja acima de um determinado valor, o tuíte será considerado popular, caso contrário, não popular. Os pontos foram: i) mediana das taxas de engajamento, que foi 0,5%, e ii) taxa de engajamento de 1%, comumente adotada para considerar tuítes como populares. No primeiro caso, foi utilizada a totalidade dos tuítes, 262.468. Já para o segundo caso, foi necessário um passo de *undersampling*, descartando os tuítes não populares excedentes, para que ambas as classes permaneçam balanceadas, restando 206.814.

Ao fim dos testes, o modelo que obteve os melhores resultados foi o GAP1D⁵, utilizando o segundo método de balanceamento, com uma precisão de 65,24%. Dado a sua eficácia superior, este foi o modelo escolhido para ser o oráculo da avaliação.

4.3. Resultados

Nesta seção, são apresentados os resultados obtidos com experimentos realizados, utilizando o processo apresentado na Figura 2 e as diferentes funções de pontuação apresentadas na Seção 3.1.2. Para cada função de pontuação, foram submetidos o total de tuítes

⁵Criado utilizando o TensorFlow, com os seguintes hiper-parâmetros modificados: *textVectorization_max_tokens* = 30000; *textVectorization_output_sequence_length* = 60; *embedding_output_dim* = 16;

que possuíam adjetivos (176.323), mas a quantidade que passou por alterações é diferente para cada uma das funções, pois as substituições irão variar conforme as pontuações individuais das palavras.

Para fins de ilustração, alguns exemplos de adjetivos e possíveis palavras para substituição, juntamente de suas respectivas pontuações (usando média simples), são apresentados na Tabela 1. Considerando estes exemplos, o adjetivo “*angry*” seria substituído por “*crazy*”, enquanto que “*advanced*” não seria substituído, por já possuir a maior pontuação.

Tabela 1. Exemplos de adjetivos e possíveis palavras para alteração

Adjetivo	Sinônimos: Pontuação
angry: 0.07	(crazy: 0.08), (furious: 0.03), (savage: 0.06), (violent: 0.03)
advanced: 0.06	(modern: 0.03), (new: 0.03)
attempt: 0.007	(search: 0.01), (seek: 0.03), (try: 0.06)
individual: 0.03	(mortal: 0.07), (somebody, 0.08): (soul: 0.04)

Na Tabela 2, está distribuída a porcentagem de tuítes que foram alterados pelo modelo. Para os casos em que os tuítes não sofreram alterações, o modelo identificou que as palavras utilizadas já são aquelas com a maior pontuação possível. Dentre as funções, a média simples foi a que fez o maior número de modificações nos tuítes.

Tabela 2. Percentual de tuítes alterados com cada uma das funções

Tuítes	Média simples	Média harmônica	Somatório
Alterados	63,51%	50,97%	43,90%
Não alterados	36,49%	49,03%	56,10%

Observando os resultados dentre os tuítes que sofreram modificações, a Tabela 3 mostra o percentual daqueles que foram alterados positivamente, negativamente e aqueles que seu score foi mantido, em comparação com sua versão original. Pode-se observar que a média simples foi a função que obteve os melhores resultados, com a maior porcentagem de tuítes que foram favorecidos pelas alterações e a menor taxa de tuítes que foram prejudicados. Para as três funções, o percentual de tuítes em que o score foi mantido foi muito próximo, com uma variação inferior a 1% entre eles.

Tabela 3. Resultados da alteração de tuítes para cada função de pontuação

Resultado	Média simples	Média harmônica	Somatório
Positivo	62,66%	51,22%	57,54%
Negativo	24,89%	36,52%	30,08%
Mantido	12,45%	12,26%	12,37%
Positivo + Mantido	75,11%	63,48%	69,91%

Somado o total de casos positivos com os total daqueles que foram mantidos, pode-se dizer que 75,11% dos tuítes alterados atingiram sua melhor forma – de acordo com o processo de validação proposto.

Detalhando os resultados de cada função, na Figura 3 apresenta-se a distribuição dos tuítes conforme as porcentagens de ganho ou perda. Pode-se observar que para todos

os casos, a maior concentração de tuítes está presente na faixa com até 20% de ganho. Além disso, no caso da média simples, 37% dos tuítes tiveram um ganho acima desses 20%, enquanto que apenas 7% tiveram uma perda acima disso.

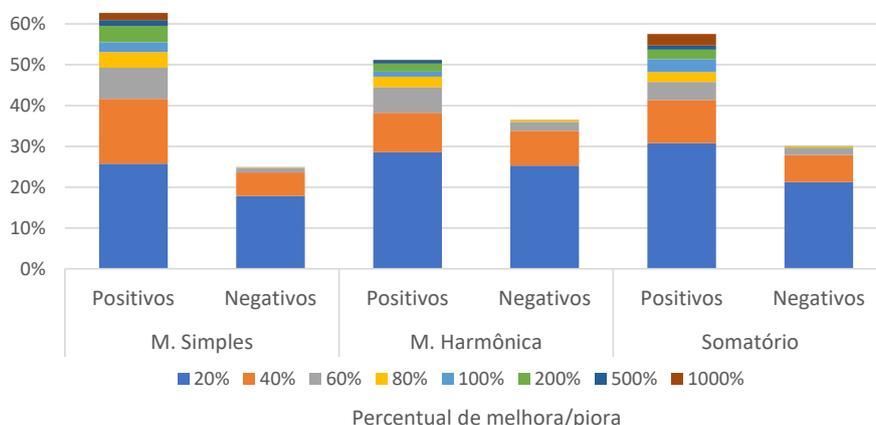


Figura 3. Distribuição dos tuítes por porcentagem de melhora ou piora em relação ao texto original, para as diferentes funções de pontuação utilizadas

Convém destacar, como uma das limitações deste trabalho, que durante a substituição dos adjetivos, o processo de ajuste de concordância ainda não é realizado. Como exemplo hipotético, o texto “*he is angrier*” se tornaria “*he is crazy*”. Isso acontece pois somente as formas lematizadas das palavras são armazenadas na base.

5. Considerações Finais

Atualmente o Twitter, assim como a maioria das redes sociais, é utilizado como ferramenta de trabalho, seja para assuntos políticos, artísticos ou comerciais, muitas vezes influenciando inclusive no mercado de investimentos. Por vezes, as mensagens veiculadas pela rede devem ser elaboradas com cuidado, pensando especialmente no seu público, com o objetivo de ampliar seu alcance.

Neste contexto, este trabalho realizou um estudo sobre os textos veiculados em publicações do Twitter, realizando alterações em seu corpo, especificamente em adjetivos, com o intuito de verificar a capacidade de aumentar o engajamento nestas publicações. As substituições levam em consideração adjetivos associados à tuítes com uma alta taxa de engajamento. Os resultados obtidos durante os experimentos, usando um algoritmo de aprendizado de máquina como avaliador, demonstraram que apesar de prejudicar uma porcentagem dos tuítes, o processo de reescrita potencialmente gera tuítes melhores. Ao final de todo processo, o saldo de tuítes que obtiveram sua melhor forma foi positivo.

Como trabalhos futuros, pretende-se realizar a adaptação das palavras que são substituídas para que respeitem a concordância da frase, sem que o texto perca o sentido. Além disso, também podem ser realizadas análises específicas considerando as contas de usuários, com o objetivo de obter modelos especializados, considerando a audiência de cada conta, para isso seria necessário expandir a base, coletando um maior número de tuítes por usuário. Já no que se refere ao processo de validação, pretende-se evoluir o modelo de predição para aumentar sua precisão e, com isso, a confiança dos resultados.

Referências

- Benevenuto, F., Magno, G., Rodrigues, T., and Almeida, V. (2010). Detecting spammers on twitter. In *Collaboration, electronic messaging, anti-abuse and spam conference (CEAS)*, volume 6, page 12.
- Chen, M. X., Lee, B. N., Bansal, G., Cao, Y., Zhang, S., Lu, J., Tsay, J., Wang, Y., Dai, A. M., Chen, Z., et al. (2019). Gmail smart compose: Real-time assisted writing. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2287–2295.
- Derczynski, L., Ritter, A., Clark, S., and Bontcheva, K. (2013). Twitter part-of-speech tagging for all: Overcoming sparse and noisy data. In *Proceedings of the international conference recent advances in natural language processing ranlp 2013*, pages 198–206.
- Duan, Y., Jiang, L., Qin, T., Zhou, M., and Shum, H.-Y. (2010). An empirical study on learning to rank of tweets. In *Proceedings of the 23rd International Conference on Computational Linguistics, COLING '10*, pages 295–303, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Goldfarb-Tarrant, S., Feng, H., and Peng, N. (2019). Plan, write, and revise: an interactive system for open-domain story generation. *arXiv preprint arXiv:1904.02357*.
- Jain, G., Sharma, M., and Agarwal, B. (2019). Spam detection in social media using convolutional and long short term memory neural network. *Annals of Mathematics and Artificial Intelligence*, 85(1):21–44.
- Naveed, N., Gottron, T., Kunegis, J., and Alhadi, A. C. (2011). Bad news travel fast: A content-based analysis of interestingness on twitter. In *Proceedings of the 3rd International Web Science Conference, WebSci '11*, pages 8:1–8:7, New York, NY, USA. ACM.
- Oliveira, L. L. and Mergen, S. L. S. (2018). Análise da popularidade de tuítes com base em características extraídas de seu conteúdo. *Escola Regional de Banco de Dados (ERBD)*, 14(1/2018).
- Pillat, V. G. and Pillat, V. G. (2017). Comparação entre duas fórmulas utilizadas para o cálculo da taxa de engajamento utilizando como base a porcentagem de visualizações e o total de fãs. *Revista Brasileira de Pesquisas de Marketing*.
- Shang, L., Lu, Z., and Li, H. (2015). Neural responding machine for short-text conversation. *arXiv preprint arXiv:1503.02364*.
- Suh, B., Hong, L., Pirolli, P., and Chi, E. H. (2010). Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. In *2010 IEEE Second International Conference on Social Computing*, pages 177–184. IEEE.
- Xu, Q., Xu, C., and Qu, L. (2019). Alter: Auxiliary text rewriting tool for natural language generation. *arXiv preprint arXiv:1909.06564*.
- Zhang, Q., Gong, Y., Wu, J., Huang, H., and Huang, X. (2016). Retweet prediction with attention-based deep neural network. In *Proceedings of the 25th ACM international on conference on information and knowledge management*, pages 75–84.