

Análise de sentimentos: duas possíveis aplicações

Matias Giuliano Gutierrez Benitez¹, Rui Jorge Tramontin Júnior².

¹ Acadêmico (a) do Curso de Ciência da Computação – CCT – Voluntário PROVIC

² Orientador, Departamento de Ciência da Computação – CCT – rui.tramontin@udesc.br

Palavras-chave: Análise de Sentimentos. Mineração de Opinião. Processamento de Linguagem Natural. Língua Portuguesa.

O Processamento de Linguagem Natural (PLN) é uma vertente da Inteligência Artificial que ajuda computadores a entender, interpretar e manipular a linguagem humana. O PLN resulta de diversas disciplinas, incluindo Ciência da Computação e Linguística Computacional, que buscam preencher a lacuna entre a comunicação humana e o entendimento dos computadores. Dentro do PLN existe uma subárea chamada Análise de Sentimentos (AS) em ela se estudam as opiniões, atitudes e emoções das pessoas em relação a uma entidade utilizando o poder computacional. Esta entidade pode ser representada por indivíduos, eventos ou tópicos.

O objetivo da AS é encontrar opiniões, identificar o sentimento que estas opiniões expressam e, por último, classificar sua polaridade. Existem três níveis de classificação, o primeiro é em *nível de sentença*, que visa classificar o sentimento expresso em uma frase. O segundo é em *nível de documento*, que tem como objetivo classificar um conjunto de frases e dar para elas uma polaridade positiva ou negativa. Em último lugar, há o *nível de aspecto*, que utiliza o conceito de entidades para fazer a classificação. Esse nível identifica uma entidade dentro do texto e procura por opiniões referentes a tal entidade.

Existem também variados métodos de classificação, entre os mais conhecidos podemos encontrar os seguintes grupos: (1) Abordagem de aprendizado de máquina: esta é uma das mais usadas na área, dentro do grupo podemos achar as seguintes técnicas: *Naïve Bayes*, *Maximum Entropy*, *Support Vector Machines*, Árvores de decisão, entre outras; (2) Abordagem baseadas em léxico: este grupo se divide em três subgrupos: baseadas em *corpus*, baseadas em dicionários e baseadas em estatística.

Há diversas aplicações possíveis para a AS, como por exemplo, análise de comentários na rede social *Twitter* para predizer o resultado das eleições, identificar sentimentos em músicas para aplicá-los em sistemas de recomendação e análise da opinião pública sobre um determinado produto para o desenvolvimento de estratégias de *marketing*. Em resumo, onde existe uma opinião dirigida a algum tema em específico, é possível aplicar a AS. Levando isso em consideração, no projeto estamos planejando duas possíveis aplicações: a primeira é fazer um estudo comparativo das técnicas de aprendizado de máquina, mas focadas totalmente ao português do Brasil. Existem outros artigos que já fizeram estudos comparativos de certas técnicas, mas o comum entre eles é que trabalham com o texto em inglês, ou seja, primeiro traduzem o texto para logo aplicar as técnicas, fato que termina afetando a precisão do resultado, pois existem palavras e frases no português que, ao serem traduzidas para outra língua perdem o sentido ou possuem mais de um significado diferente. Esta primeira ideia surgiu da falta de conteúdo que existe na área de AS em português. Nesse sentido, procuramos ampliar a pesquisa na área e dar uma contribuição para futuros trabalhos que envolvam nosso idioma.

A segunda possível aplicação seria fazer uma AS das opiniões em *Twitter* referentes a distintas criptomoedas e estudar se elas afetam no valor que estas possuem no mercado. Como é observado hoje em dia, as moedas digitais são uma tendência, e o número de usuários que as utilizam cresce diariamente. Elas são conhecidas por serem bastante voláteis quando falamos do preço devido a que não existe um órgão governamental que as controla e dependem totalmente da oferta, demanda e tecnologia que elas oferecem. O objetivo final desta aplicação seria analisar se existe uma relação entre os sentimentos gerados pelos *Tweets* e a variação no preço final das criptomoedas.