

Análise de Sentimentos: Interpretando Opiniões nas Avaliações de Serviços *Online* com Léxicos de Sentimentos

Mateus Nepomuceno Ladeira¹, Rui Jorge Tramontin Júnior².

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

³ Orientador, Departamento de Ciência da Computação – CCT – rui.tramontin@udesc.br

Palavras-chave: Processamento de Linguagem Natural. Análise de Sentimentos. Abordagem Lexical. Recurso Léxico. Serviços *Online*.

Com o crescente desenvolvimento de novas tecnologias, as tarefas do dia a dia têm se tornado cada vez mais simples. Nesse contexto, aplicações *mobile* permitem que a sociedade tenha acesso aos mais diversos serviços *online*. Como consequência, ocorre um constante aumento do número de opiniões disponíveis na internet, tornando-se fundamental o desenvolvimento de aplicações que as analisem de forma automatizada. Nesse sentido, diversos estudos sobre Análise de Sentimentos, ramo do processamento de linguagem natural, estão em desenvolvimento. Esse campo de estudos busca verificar a ocorrência de opiniões em textos, e, quando houver, classificá-las de acordo com sua polaridade. Quando não há opinião no escrito observado, ele é classificado como objetivo. Caso contrário, ele é subjetivo. Após essa primeira verificação, constatando-se a subjetividade, outra análise deve ser feita, verificando se a opinião expressa pelo conteúdo é positiva, negativa ou neutra. Vale destacar que essa análise pode ser realizada em três diferentes níveis: documento, sentença e aspecto. O primeiro consiste em um agrupamento de frases, das quais, cada uma pode expressar opiniões distintas. O segundo define um grupo de palavras, que também possuem uma polaridade. O último compreende a análise de uma característica presente no texto, que definirá a polaridade do conjunto como um todo.

Nesse contexto, existem diferentes métodos para analisar-se o sentimento de um texto, dentre os quais, dois se destacam: classificação por meio de um recurso léxico e análise com aprendizado de máquina. Um léxico de sentimentos é uma das ferramentas mais utilizadas para mineração de opinião. Ele compreende um conjunto de palavras que possuem uma polaridade atribuída. Nesse cenário, existem diferentes ferramentas disponíveis para essa categoria de análise na língua portuguesa e no inglês. Diversos métodos não-supervisionados são desenvolvidos com os quatro recursos disponíveis para o português: *OpLexicon*, *SentiLex*, *Brazilian Portuguese Linguistic Inquiry and Word Count (LIWC)* e *Onto.PT*. Já entre os recursos da língua inglesa, o *SentiWordNet* ocupa posição de destaque pela grande variedade de palavras classificadas, já que dispõe de 117.374 entradas vindas dos *synsets* do WordNet 3.0. Há também estudos desenvolvidos para língua portuguesa que utilizam o *SentiWordNet*, traduzindo-o de forma automática através da API do Google Tradutor. Uma vez que a escolha do léxico de sentimentos adequado define a qualidade da aplicação desenvolvida, é comum o teste de diferentes recursos em uma mesma aplicação, comparando os resultados obtidos para encontrar a maior precisão.

Tendo em vista os excelentes resultados obtidos na classificação de documentos a partir da escolha de um léxico de sentimentos, este estudo visa desenvolver uma aplicação que analise as avaliações de serviços *online*. Nesse sentido, serão comparados os resultados obtidos pelo algoritmo utilizando diferentes recursos léxicos da língua portuguesa. Para a obtenção das avaliações, será utilizada a API disponibilizada pelo provedor do serviço, que permite a retirada do comentário e do número de estrelas dado pelo consumidor.

Após a extração dos comentários, é necessário realizar o pré-processamento dos dados obtidos. Sendo assim, ocorre o tratamento do texto, permitindo que o algoritmo o percorra facilmente para localizar as classes gramaticais e então iniciar a classificação. Funções comuns dessa etapa são: remoção de *stop words*, palavras que não agregam valor à polaridade do documento; remoção de caracteres especiais, números e *tags* HTML (*Hypertext Markup Language*).

Ao concluir-se a fase de pré-processamento dos textos, inicia-se o processo de classificação. Ao elaborar-se uma lógica de classificação, é importante levar em conta as classes gramaticais presentes no documento e como elas interferem na formação de opiniões. Como claro exemplo disso, tem-se os advérbios de negação, como a palavra “não”. Observando-se a frase “Esta marca é boa!”, nota-se que ela é subjetiva, com polaridade positiva. Entretanto, adicionando-se a negação, tem-se “Esta marca não é boa!”. Ao analisar-se novamente essa sentença, percebe-se que a inclusão da negação inverte a polaridade, resultando em uma nova sentença também subjetiva, porém, negativa. Ademais, no trecho analisado, o vocábulo responsável por caracterizar opinião ao texto foi o adjetivo “bom”. Isso torna evidente que os adjetivos são a principal classe a ser observada na mineração de opinião utilizando léxico semântico, e deve ser estudado como a presença de diferentes advérbios intensifica, diminui ou inverte a polaridade do conteúdo.

Portanto, fica evidente que esse estudo visa o desenvolvimento de um classificador de opinião utilizando um léxico de sentimentos para a análise de textos sobre serviços *online*, contribuindo no entendimento de como o uso de advérbios interfere na polaridade de documentos, predominantemente definida por adjetivos. Como primeira opção, será verificada a possibilidade de análise dos comentários disponíveis no *iFood*. Caso o uso desses dados seja inviável, será buscada outra fonte de textos, por exemplo, postagens em blogs ou comentários no *TripAdvisor*. Ademais, serão calculadas as métricas de desempenho padrão de *Acurácia*, *F-measure* e *Entropia* para cada recurso léxico utilizado, constatando-se o melhor resultado.