

Comparação de algoritmos de aprendizado de máquina na classificação e análise de sentimentos

Proposta de Trabalho

Gabriel Erzinger Dousseau - 33592

Orientador: Isabela Neves Drummond

Instituto de Matemática e Computação - IMC

Universidade Federal de Itajubá - UNIFEI

Início: 26/02/2018 Término: 30/11/2018

Este documento é uma *proposta de trabalho* para o Trabalho Final de Graduação - TFG no curso de *Ciência da Computação* da Universidade Federal de Itajubá e deve ser entregue, preenchido e assinado pelo aluno e pelo orientador(a) ao coordenador(a) de TFG.

Problema

Com o crescimento do tamanho de dados gerado diariamente devido às redes sociais, sites de publicações de artigos, jornais e outras diversas formas de exposição de conhecimento e opinião, novas tecnologias de bancos de dados e processamento de dados surgiram para comportar o aumento na quantidade de informação gerada. Consequentemente, o volume de dados aumenta rapidamente e, para extrairmos informações úteis destes dados são necessárias as técnicas de processamento.

Dessa forma surge a Análise de Sentimentos, um campo de estudos que tem como objetivo

exatamente extrair informações importantes de grandes conjuntos de dados, analisando opiniões, sentimentos, avaliações, críticas, atitudes e emoções pessoais em relações a entidades como produtos, serviços, indivíduos, problemas, eventos, tópicos e seus atributos. (5)

A análise de sentimentos é um campo notório de pesquisa uma vez que, quando aplicada com sucesso, permite o processamento de um enorme volume de textos escritos em *Lingua Natural* (como o Português ou o Inglês por exemplo) para classificar e diferenciar essas informações, sendo alvo principal de aplicações como:

- Análise de entusiasmo eleitoral em redes sociais
- Análise da reputação de uma marca no meio online
- Análise de sentimentos em relação a publicações pessoais
- Análise futura de tendências em investimentos ou ações

Na esfera conceitual apresentada, o papel deste trabalho se concentra em *projetar, analisar, estudar e comparar* as diversas técnicas aplicadas aos problemas de análise de sentimentos.

No campo das técnicas mais utilizadas, temos as léxicas e as de aprendizado de máquina (*Machine Learning*) em conjunto com mecanismos de *Deep Learning*. Aprendizado de máquina é o campo de estudos que tem como escopo a construção de programas que se aperfeiçoem com a experiência, desta forma, *aprendendo*. (6). Neste campo, *deep learning* é uma solução que permite o aprendizado de máquina de forma que a mesma entenda o mundo em forma de conceitos hierárquicos de relações que podem ser representadas num grafo profundo, onde cada camada compõem um *nível de conhecimento* (3, 4) e o conjunto de todas camadas represente o aprendizado, constituindo assim, o que é chamado de aprendizado profundo e que tem sido

usado em outros estudos de classificação de sentimentos. (2)

Palavras-chave: Análise de Sentimentos, Aprendizado de Máquina, Aprendizado Profundo, SVM, Naive-Bayes, Processamento de Linguagem Natural

Objetivos

O objetivo deste trabalho é analisar e estudar os principais algoritmos aplicados na análise de sentimentos, com uma ênfase maior nos algoritmos de aprendizado de máquina como o *SVM* (*Support Vector Machine*) e o *Naive-Bayes Classifier*, que se destacam na maior parte dos artigos e trabalhos científicos como os melhores algoritmos de *machine learning* deste campo de estudo e compará-los entre si e com outros algoritmos pertinentes à matéria como algoritmos de *deep learning*.

Como contribuição científica, o projeto tem como escopo o desenvolvimento de uma maior compreensão das técnicas de análise de sentimentos, processamento de linguagem natural, aprendizado de máquina e aprendizado profundo. Dessa forma, é parte do objetivo do projeto a execução destes algoritmos, assim como a análise e o estudo dos resultados obtidos em cada execução, no que diz respeito a precisão, tempo de execução, memória e outros parâmetros de comparação.

Planejamento

O projeto tem como prazo concreto os dois períodos acadêmicos do ano de 2018 (dois mil e dezoito). Dessa forma as fases da pesquisa serão realizadas neste período de tempo, tendo início no mês de março.

De forma geral, o desenvolvimento do trabalho se dá pela completude das seguintes etapas:

- Decidir, junto a orientadora do trabalho, quais algoritmos estudar (além dos já citados)
- Fundamentação teórica e estudo em análise de sentimentos, aprendizagem de máquina, aprendizagem profunda e os algoritmos que serão comparados
- Definição dos conjuntos de dados que serão usados para os testes
- Normalização dos conjuntos de dados e visualização dos mesmos (análise estatística, distribuição de dados, etc)
- Executar os algoritmos utilizando bibliotecas científicas e coletando os parâmetros de execução
- Apresentar os resultados obtidos na etapa anterior
- Estudar, analisar e elaborar conclusões a respeito dos resultados.

É planejado que as etapas do projeto sejam desenvolvidas de forma igualmente distribuída entre o prazo do trabalho.

Referências

1. Ronen Feldman. Techniques and applications for sentiment analysis. *Commun. ACM*, 56(4):82–89, 2013.
2. Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Domain adaptation for large-scale sentiment classification: A deep learning approach. In *Proceedings of the 28th International Conference on International Conference on Machine Learning, ICML'11*, pages 513–520, USA, 2011. Omnipress.

3. Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.
<http://www.deeplearningbook.org>.
4. Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, May 2015.
5. Bing Liu. *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers, 2012.
6. Thomas M. Mitchell. *Machine Learning*. McGraw-Hill, Inc., New York, NY, USA, 1 edition, 1997.

É importante notar que, durante o desenvolvimento do trabalho, novas referências poderão ser adicionadas, na medida que forem necessárias.

Assinatura do aluno:

Assinatura da orientadora: