

Utilizando Deep Learning na Classificação Multi Rótulo de Textos

Mestrado

Zacarias Curi Filho, Emerson Cabrera Paraiso

Contexto: O crescente avanço no uso das redes sociais possibilitou que cada vez mais usuários expressassem suas ideias e opiniões sobre diversos temas. Devido ao grande interesse em analisar essas informações diversas técnicas foram estudadas para tratar este problema, tendo o Deep Learning se destacado em muitos domínios onde uma grande quantidade de dados são utilizados. Uma das técnicas de Deep Learning que obteve grande sucesso nos últimos anos foi o Long short-term memory (LSTM). Essa técnica possui diversas aplicações, possuindo a análise de sentimentos em textos como uma das mais comuns. Um estudo com as publicações nessa área nos permitiu identificar que a maioria das bases utilizadas são classificadas com polaridade ou polaridade associada com graus de intensidade. Embora alguns trabalhos tenham utilizado bases multi classe estas permitiam apenas um rótulo por instancia, como as bases formadas por *reviews* de produtos, onde ao escrever sua opinião o usuário seleciona uma quantidade de estrelas. Uma abordagem diferente, como a utilização de multi rótulos, pode ser realizada. Nessa abordagem são considerados um grupo de rótulos para cada instância, como no caso onde diversos usuários expressam sua opinião sobre determinado tema utilizando *emoticons*. Embora essa abordagem possa ser aplicada em diversos domínios nenhum trabalho utilizando Deep Learning em uma base multi rótulos de textos foi encontrado na literatura. **Objetivo Geral:** Esta pesquisa tem como objetivo desenvolver um método utilizando Deep Learning para a classificação de textos multi rótulos. **Questão de Pesquisa:** A utilização de técnicas de Deep Learning na classificação multi rótulo de textos permite melhores resultados que os algoritmos rasos? **Método de Pesquisa:** Para alcançar o objetivo, será realizada a construção de um novo corpus com textos multi rótulos. Será realizada a modificação do algoritmo LSTM para possibilitar a sua utilização na classificação multi rótulos. **Resultados Preliminares:** Duas bases foram construídas com dados de um site de notícias. As bases, nos idiomas português e inglês, foram anotadas utilizando a reação dos usuários. Uma base não supervisionada composta de notícias do site G1.com foi construída para a utilização na operação de embedding do método. Diversos testes foram realizados com diferentes bases multi classes para a escolha do melhor algoritmo.

Palavras-chave: Deep Learning, LSTM, Multi Label.