



APRENDIZADO DE MÁQUINA EM FLUXO CONTÍNUOS DE DADOS NÃO-ESTRUTURADOS: ANALISANDO SENTIMENTOS EM TWEETS

Marcelo Campos Silva, Rodrigo Coelho Barros (orientador)

Faculdade de Ciência da Computação, PUCRS.

Resumo

O Aprendizado de Máquina é uma subárea da inteligência Artificial dedicada ao desenvolvimento de algoritmos e técnicas que permitam ao computador aprender de forma automática. Com a ascensão de redes sociais como Facebook e Twitter, usuários utilizam para conversar e compartilhar mensagens, gerando diariamente uma grande quantidade de dados. Os usuários além de explorar as redes para comunicação, mas também como um espaço para compartilhar opiniões, reflexões e discussões. Desta forma, redes sociais se transformaram em valiosas fontes de conteúdo subjetivo, despertando o interesse de pesquisadores e diversos segmentos da sociedade, pela enorme quantidade de informação útil. Diversos trabalhos têm utilizado o Aprendizado de Máquina para detectar a opinião do público online em mídias sociais com relação à determinada temática. Definida como uma tarefa de classificação em Aprendizado de Máquina, Análise de Sentimento tem por objetivo atribuir categorias a textos escritos em linguagem natural, de acordo com a polaridade que esses textos expressam (positiva ou negativa). No entanto, apenas uma pequena parcela de trabalhos tem investigado a Análise de Sentimento sob a ótica do fluxo contínuo de dados, aonde estes chegam contínua e infinitamente, gerando uma demanda por ambientes dinâmicos que atendam as restrições e limitações de recursos computacionais. Este trabalho tem por objetivo realizar Análise de Sentimento em *tweets* sob o viés da área de fluxo contínuo de dados. Para tanto, em um primeiro momento foram utilizados dados coletados através de uma API gratuita do Twitter, através de um protocolo de coleta de dados e posterior armazenamento em banco de dados. Em um segundo momento, foi realizado um levantamento bibliográfico de técnicas de preparação de dados capazes de reduzir o impacto do informalismo das redes sociais, característica especialmente marcante dos *tweets*. Em sua fase corrente, estão sendo realizados

experimentos simulando fluxos contínuos de *tweets* e posterior aplicação de algoritmos de aprendizado, visando à obtenção de modelos de classificação de alta capacidade preditiva para a tarefa de Análise de Sentimento de *tweets*.

Palavras-chave

Aprendizado de Máquina; Análise de Sentimento; Redes sociais.