

# APLICANDO MINERAÇÃO DE PADRÕES DE SEQUENCIAS NA ANALISE DE SENTIMENTOS

Mestrado

Willian Augusto Dias dos Reis, Emerson Cabrera Paraiso

As emoções fazem parte da vida do ser humano e podem ser expressas e percebidas por meio de gestos e expressões, da fala e por meio da escrita. Existe uma área de pesquisa que busca inserir o tratamento das emoções no ambiente computacional: a Computação Afetiva. Entre as pesquisas realizadas em Computação Afetiva encontra-se a identificação automática de emoções em textos, conhecida como Análise de Sentimentos. As pesquisas em Análise de Sentimentos se iniciaram, e têm crescido, em decorrência da grande quantidade de dados que são disponibilizadas diariamente na internet. Dentre os tipos de dados, os sequenciais têm suas características e que são importantes para vários tipos de aplicações. Como um subcampo da mineração de dados, a mineração de sequência obtém conhecimento através da análise temporal dos dados. Existem vários exemplos de aplicações nesta área como transações financeiras, mudanças globais do clima, sequencias de DNA, etc. Com base nesse contexto, este trabalho apresenta um método de identificação de emoções em textos, que permite identificar as emoções básicas em textos, além de identificar também textos neutros (sem emoção), utilizando mineração de sequencias. O método utiliza uma abordagem de Mineração de Padrões de Sequencias onde é possível identificar padrões de sequencias de sentimentos na escrita e aplica-los para corrigir problemas de classificação. O método pode ser aplicado em duas abordagens, sendo que, para as duas abordagens se extrai padrões através das sequencias de sentimentos no documento. Uma sequência para este método é a classificação de cada sentença dentro de um documento, analisados como um vetor de emoções. Através de um alinhamento entre as sequencias, utilizando algoritmos como Needleman-Wusch, os padrões são encontrados. Na primeira abordagem se identifica erros entre a base anotada e a primeira classificação realizada por algum método multiclasse e aplica-se os padrões nestes pontos para correção. Na segunda abordagem, aplica-se os padrões encontrados em toda a extensão do documento a ser classificado. O método foi avaliado por meio de um corpus de contos infantis, de língua inglesa, onde se percebeu a existência de padrões de escrita entre os documentos. Ao ser avaliado com o corpus de contos, o método obteve uma taxa de acerto de 83,4% ao identificar as seis emoções básicas, além de neutro, em textos. Outras pesquisas com o mesmo corpus obtiveram 57,41% e 70,33% de acerto.

**Palavras-Chave:** Analise Sentimentos, Sequence DataMining, Classificação, Padrão Sequencia