Edital PRPq 04/2010

BITIB – Fapemig

Projeto: Análise em Tempo Real de Sentimentos na Web

Adriano Alonso Veloso
Departamento de Ciência da Computação
Universidade Federal de Minas Gerais

Projeto apresentado em resposta ao Edital PRPq 04/2010.

Belo Horizonte, MG

1 Caracterização do Problema

A Web mudou drasticamente a maneira como expressamos nossas opiniões. Todos os dias, pessoas disponibilizam grandes volumes de conteúdo opinativo sobre os mais diversos eventos, produtos, e pessoas [4], em canais sociais como *Twitter*[1] e *Facebook*[2]. Monitorar e dar sentido a tal avalanche de opiniões é uma tarefa desafiadora, denominada **Análise de Sentimentos** [1], e pode levar pessoas e empresas à descoberta de conhecimento altamente estratégico [5, 3]. Tal tipo de análise pode, por exemplo, ajudar as empresas a compreender o que seus clientes pensam sobre seus produtos ou serviços[3], ou mesmo facilitar o entendimento do que seus clientes pensam sobre seus concorrentes.

No entanto, são vários os desafios que permeiam a efetividade da análise de sentimento em canais sociais como o *Twitter*. Primeiramente, 140 caracteres não fornecem espaço suficiente para explicações elaboradas, e os usuários devem ir direto ao assunto principal, e expor suas opiniões através de poucas palavras. Outro desafio é a utilização de gírias e outras formas de expressão, como sarcasmo e ironia, o que dificulta a análise das opiniões (expressões sarcásticas correspondem a cerca de 90% do conteúdo postado no *Twitter*). Por último destacamos a utilização de meta-dados, como *hashtags*, que podem conter significado útil para a análise de sentimentos, mas que precisam ser encontrados e processados automaticamente.

A maioria dos algoritmos para a análise de sentimentos são baseados em técnicas de processamento de linguagem natural [6], e utilizam como entrada um vocabulário controlado. Tal vocabulário é comparado com o texto contento conteúdo opinativo. A partir dessa comparação, o algoritmo fornece predições acerca das opini oes implícitas no texto. Tal estratégia não é adequada em domínios de aplicação como o *Twitter*, uma vez que os usuários podem entrar com texto livre, que não necessariamente obedece um vocabulário controlado.

Os algoritmos a serem propostos serão capazes de oferecer altos índices de efetividade, mesmo perante todos os desafios mencionados. Eles serão baseados em técnicas de aprendizado de máquina [2], que utilizam como entrada um "conjunto de treinamento"que contém exemplos de textos com diferentes gradações de opinião acerca de um assunto ou entidade (i.e., exemplos de opiniões com conotação negativa e positiva). A partir desses exemplos, o algoritmo é capaz de construir uma função que é utilizada para predizer a conotação de um grande volume de opiniões.

A elaboração e o desenvolvimento destes algoritmos nos possibilitarão realizar análises mais ricas e detalhadas do grande volume de opiniões disponíveis em canais sociais na Web.

[1] `http://twitter.com`
[2] http://facebook.com
[3] Recentemente, a empresa Apple realizou uma análise de sentimento sobre um de seus novos produtos (`http://techcrunch.com/2010/04/06/ipad-sentiment-analysis`)

2 Objetivos e Metas

O objetivo deste projeto é desenvolver algoritmos para analisar, em tempo real, o grande volume de texto opinativo disponível em canais sociais da Web, como o *Twitter*. As metas do projeto estão relacionadas com a formação de recursos humanos qualificados, a produção e a publicação de resultados de pesquisa, e a disseminação de conhecimento para a sociedade. De forma mais específica temos como metas gerais:

1. Publicar 1 artigo em conferência nacional.
2. Publicar 1 artigo em conferência internacional.
3. Desenvolver algoritmos para a análise de opiniões, em tempo real, com base em canais como o *Twitter*.

3 Metodologia

A execução do projeto ora proposto seguirá a metodologia apresentada a seguir:

1. Levantamento de técnicas existentes: Este levantamento consiste de avaliar trabalhos técnicos e acadêmicos pertinentes à tarefa de análise de sentimentos, além de localizar, obter e avaliar ferramentas que possam ajudar parcialmente na tarefa.
2. Projeto: O projeto descreve o que será implementado, quais as condições de execução, a natureza da saída e os procedimentos de validação dos resultados.
3. Implementação: A implementação especificada deve ser feita primando pela clareza e eficiência do código, assim como preceitos tradicionais de engenharia de software, como manutenibilidade, modularidade e robustez. A implementação também compreende o teste dos códigos gerados.
4. Execução: Uma vez que os algoritmos estejam implementados e validados, realiza-se a execução atentando para os mecanismos de controle pertinentes.
5. Avaliação: A cada novo conjunto de resultados obtido, será realizada uma avaliação da sua qualidade e efetividade, de forma a subsidiar novos tipos de análises.

Finalmente, é importante mencionar que em nossos experimentos vamos utilizar bases de dados reais, coletadas do *Twitter*, e utilizando mecanismos de validação e comparação já existentes e bem conhecidos.

4 Resultados e Impactos Esperados

1. Proposição de novos algoritmos para a análise de sentimentos e opiniões em canais sociais como o *Twitter*, que sejam capazes de prover resultados em tempo real.

2. Avaliação dos algoritmos propostos utilizando eventos reais, que traduzam algum impacto para a sociedade Brasileira, como as Eleições 2010 ou o Campeonato Brasileiro de Futebol.

5 Recursos Necessários

Será necessária uma bolsa de Iniciação Científica, e recursos de custeio no valor de R$1.000,00 que serão usados para a aquisição de softwares e de passagens. Os equipamentos de laboratório que são necessários para os experimentos, e as bolsas associadas a outros alunos envolvidos indiretamente com este projeto, entrarão como contrapartida.

Referências

[1] Michelle de Haaf. Sentiment analysis, hard but worth it! *Customer Think*, 2010 (`http://www.customerthink.com/blog/sentiment_analysis_hard_but_worth_it`).

[2] Tom Mitchell. *Machine Learning*. McGraw Hill, 1997.

[3] Maria Ogneva. How companies can use sentiment analysis to improve their business. *Business*, 2010 (`http://mashable.com/2010/04/19/sentiment-analysis`).

[4] Adriano Veloso and Wagner Meira Jr. Efficient on-demand opinion mining. In *Brazilian Symp. on Databases*, pages 332–346, 2007.

[5] Alex Wright. Mining the web for feelings, not for facts. *New York Times*, 2009 (`http://www.nytimes.com/2009/08/24/technology/internet/24emotion.html?_r=2`).

[6] Tae Yano and Noah A. Smith. What is worthy of comment? content and comment volume in political blogs. In *International AAAI Conference on Weblogs and Social Media*, pages 318–326, 2010.