

Mineração de dados e análise de opinião em redes sociais – Um estudo de caso sobre as Olimpíadas 2012 utilizando o *Twitter*

Luan S. Oliveira¹, Guilherme O. Campos¹, Rodrigo S. da Silva²

¹Bacharelado em Sistemas de Informação – Universidade Federal de Viçosa – Campus Rio Paranaíba – Rio Paranaíba, MG – Brazil

²Departamento de Sistemas de Informação
Universidade Federal de Viçosa – Campus Rio Paranaíba – Rio Paranaíba, MG – Brazil

{luan.oliveira,guilherme.o.campos}@ufv.br, smarzaro@ufv.br

Abstract. *Currently there is a great advance in technology, especially in the social networks area, which allows us to access information recently released in a fascinating speed and comprehensive audience. Inserted in the business, the data mining area are gradually being used in the media, aiming to find groups of people who have an ideal standard of shopping for a particular company. Here in the following article a case study related to data mining over the social network Twitter.*

Resumo. *Atualmente há um grande avanço da tecnologia, principalmente na área de redes sociais, que nos permite acessar informações recém-divulgadas em uma velocidade fascinante e um público-alvo abrangente. Inserido no âmbito empresarial, a área de mineração de dados vêm, aos poucos, sendo utilizada nos meios de comunicação social, com finalidade de encontrar grupos de pessoas que possuam um padrão ideal de compras para uma determinada empresa. Apresentamos no artigo a seguir um estudo de caso relacionado à mineração de dados em cima da rede social Twitter.*

1. Introdução

Um dos fatores mais importantes para o sucesso de uma empresa é conhecer seu público alvo. Isso porque ao se conhecer o cliente é possível direcionar campanhas de *marketing* específicas e adaptar os serviços prestados à vontade do mesmo. No entanto, avaliar a vontade e opinião dos clientes em relação aos produtos e serviços oferecidos pela empresa não é tarefa simples. Uma das formas utilizadas pelas empresas para coletar tais informações é a distribuição de questionários e realização de pesquisas de opinião, que visam possibilitar uma análise da aceitação do produto oferecido bem como possíveis questões a serem mudadas.

Contudo o uso de questionários em pesquisas é, muitas vezes, ineficiente. Isso porque esta forma de coleta de dados possui um alcance relativamente reduzido, chegando a apenas 1% em questionários de filantropia, como citado por Courage e Baxter (2005), o que acaba prejudicando a precisão dos dados coletados. Outro motivo é que tais técnicas possuem um custo consideravelmente alto de aplicação. Esse custo ainda é maior quando tais questionários e pesquisas não são automatizados, isto é, quando são necessárias pessoas para suas aplicações e tabulação.

Já quando se trata de redes sociais, os dados são dispostos pelos usuários por vontade própria. Além de garantir um amplo alcance, evita problemas como influência nas respostas por parte de aplicadores, como o que acontece com questionários. No entanto, apesar dos dados serem distribuídos livremente em redes sociais, a captura e análise automática destes é complicada. Isso porque existem várias dificuldades envolvidas na extração de opinião de textos livres, já que a tarefa de quantificar opiniões contidas em textos, muitas vezes subjetivos, utilizando computadores é bastante complexa.

Tendo este cenário em mente surge a necessidade de uma forma alternativa de coleta de dados, que não necessite da procura do cliente e nem de seu tempo dedicado. Uma das alternativas que vem sendo estudadas atualmente é o uso dos dados coletados a partir de redes sociais para geração de informações a respeito de opiniões das pessoas. As redes sociais estão inseridas no cotidiano das pessoas, alcançando um grande número de possíveis clientes. Por tudo isso, propõe-se com este trabalho a coleta de dados de uma rede social em especial e o uso destes para obtenção de informações, através da utilização de técnicas de mineração de dados.

2. Metodologia

Basicamente, o trabalho será dividido em cinco partes, como mostra a Figura 1.

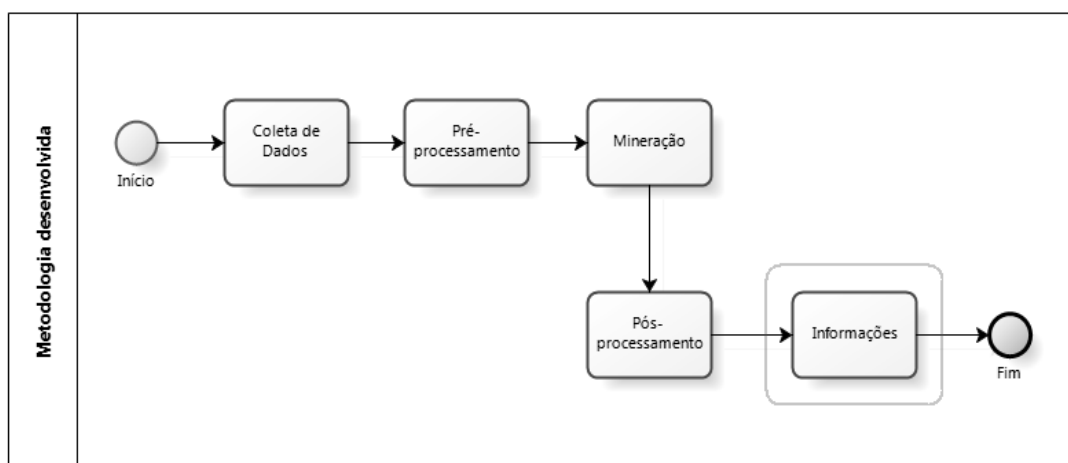


Figura 1. Metodologia de desenvolvimento do trabalho

A primeira etapa é a coleta dos dados que será realizada por um programa desenvolvido em Java, utilizando API do *Twitter* chamada de *twitter4j*. Essa coleta será iniciada no dia 14 de julho, duas semanas antes de começar as Olimpíadas, e terá fim dia 19 de agosto, uma semana após o término da mesma. Na segunda etapa, é realizado um pré-processamento em cima destes dados, para deixá-los prontos para aplicação dos algoritmos de *text mining*. Na terceira etapa, os dados serão submetidos a diferentes algoritmos de agrupamento. Padrões de filtragem, formas de visualização e interpretação dos padrões serão realizados após o agrupamento. Por fim, as informações obtidas são analisadas e colocadas à mostra como forma de análise de *marketing*, gerando informações que responderão perguntas do tipo: qual foi o atleta que atraiu a maior atenção do público em determinado esporte? Qual esporte foi o mais comentado na rede social? Existem relações entre as pessoas que comentam sobre dois esportes distintos?

3. Redes Sociais

De acordo com Regina Maria Marteleto (2001), redes sociais são “[...] *um conjunto de participantes autônomos, unindo ideias e recursos em torno de valores e interesses compartilhados*”. Aplicando-se esta ideia às redes sociais *on-line*, percebe-se que tais grupos acabam por se dividir naturalmente de acordo com interesses e/ou características em comum com os demais membros do grupo. No entanto, tais características e interesses em comum são evidentes, muitas vezes, apenas para os membros pertencentes àquele determinado grupo. Este fato torna-se um problema quando se deseja saber o que os membros de um conjunto de pessoas pensam sobre determinado assunto, já que esta informação se dispersa em torno de várias atividades realizadas pelas pessoas que compõe tal grupo. Além disso, é comum a existência de vários conjuntos de pessoas que possuem pensamentos semelhantes, mas não estão diretamente conectados pela rede social. Neste ponto torna-se necessário o uso de técnicas que permitam aproximar estes grupos semelhantes e classificar seus membros, bem como suas linhas de pensamento e opiniões.

4. Estudo de Caso Proposto – Coleta de Dados do Twitter Relacionado às Olimpíadas

Considerando-se o cenário descrito, este trabalho tem como objetivo específico fazer uso dos dados produzidos no *microblog twitter* para obter informações relacionadas às olimpíadas de 2012;

O *twitter* é um serviço de *micro-blogging* que permite que seus usuários postem mensagens de até 140 caracteres para que sejam visualizadas por outras pessoas. Além disso, também funciona como rede social, permitindo que os usuários sigam outros usuários, para que possam ver as mensagens que são postadas por este. Também são permitidas trocas de mensagens privadas, citações de usuários nas mensagens, dentre outros serviços (Matioli, S. L. et al. 2010).

Dado amplo uso desta rede e a grande diversidade de seus usuários, surge uma série de oportunidades a serem exploradas. Um dos pontos mais interessantes e com maior potencial neste tipo de serviço é a possibilidade de apresentação de anúncios de marketing e campanhas publicitárias para os usuários. Prova disso é o relatório de Michael Stelzner (2009) que indica que 88% dos 900 empresários entrevistados já empregavam as redes sociais para fazer marketing e 81% acreditava que esse uso trazia benefícios para as suas marcas.

Contudo, mesmo que campanhas publicitárias genéricas possuam bons retornos e sejam amplamente utilizadas, a possibilidade de oferecer campanhas personalizadas sobre produtos específicos para determinados usuários é algo que aumentaria as chances de sucesso desta campanha. Não faz sentido oferecer produtos relacionados a esportes para alguém que fala mal ou não gosta desse tipo de atividade. Também não seria interessante oferecer produtos para a prática de natação para um usuário que tem como atividade o futebol. Tendo isto em vista, fica claro que a escolha do produto a ser oferecido tem influência direta nas chances de atrair a atenção de um possível cliente.

Mas como saber o que oferecer a cada usuário? Como saber os interesses de milhões de usuários de diferentes lugares? Isso, claramente, não é uma tarefa manual e exige um tratamento mais cuidadoso. Uma alternativa para solucionar este problema é o

uso de técnicas de mineração de dados e descoberta de opinião, que se encaixam em um campo em grande crescimento nos últimos anos.

Assim, fazendo-se uso das informações geradas pelo *twitter* e de técnicas de mineração de texto e descoberta de opinião, pretende-se obter informações que possibilitem: classificação de usuários dentro de grupos de interesses, análise de opiniões e aceitação de certos assuntos.

5. Resultados Esperados

Os autores da pesquisa entraram em contato com algumas empresas conhecidas por patrocinarem atletas, para obter informações sobre as características relevantes ao se analisar a viabilidade de um patrocínio. Uma das respostas foi da Petrobrás, que indicou como um dos principais aspectos o mérito intrínseco e a relevância para a cultura brasileira, além dos resultados do atleta no esporte que pratica. Isso mostra a importância da popularidade do atleta, e sua conseqüente visibilidade. Tendo isso em vista este trabalho tem como resultados esperados análises de viabilidade de atletas, juntamente com uma classificação de opinião sobre estes, em opiniões positivas e negativas, o que poderia ajudar a empresas como a Petrobrás a verificar a popularidade de atletas. Desta forma o presente trabalho pode ser continuado de várias maneiras e os dados obtidos por ele também podem servir de entrada para outros projetos.

6. Referências

- Cortês, S. da C., Porcaro, R. M., Lifschitz, S. “*Mineração de Dados – Funcionalidades, Técnicas e Abordagens*”. PUC-RioInf.MCC, Rio de Janeiro, MG, 2002.
- Courage, C. and Baxter, K. “*Understanding your users: a practical guide to user requirements, methods, tools, and techniques*”. San Francisco, CA: Morgan Kaufmann Publishers, 2005.
- Marteletto, R. M. “*Análise de redes sociais – aplicação nos estudos de transferência da informação*”. Programa de Pós-Graduação em Ciência da Informação, volume 30, paginas 71-81, UFRJ, Rio de Janeiro, RJ, 2001.
- Matioli, S. L. “*Protótipo Para Mineração de Opinião Em Redes Sociais: Estudo de Casos Selecionados Usando o Twitter*”. Lavras, MG, 2010.
- Meira, S. R. de L., Costa, R. A., Jucá, P. M., Silva, E. M. “*Sistemas Colaborativos*”. Elsevier, paginas 53-64, 2011.
- Pang, B. and Lee, L. “*Opinion Mining and Sentiment Analysis*”. Foundations and Trends. Volume 2, 2008.
- Pimentel, M., Gerosa, M. A., Fuks, H. “*Sistemas de comunicação para colaboração*”. Elsevier, paginas 65-93, 2011.
- Stelzner, M. “*Social Media Marketing Industry Report*”. 2009.