

Twitter, análise de sentimento e desenvolvimento de produtos: Quanto os usuários estão expressando suas opiniões?

Twitter, sentiment analysis and product development: How much users are expressing their opinions?

Leandro Matioli Santos

Universidade Federal de Lavras (UFLA), Brasil
matioli.leandro@gmail.com

Ahmed Ali Abdalla Esmin

Professor do Departamento de Ciência da Computação da Universidade Federal de Lavras (UFLA), Brasil
aesmin@gmail.com

André Luiz Zambalde

Professor do Departamento de Ciência da Computação da Universidade Federal de Lavras (UFLA), Brasil
zamba@ufla.br

Frank Mendes Nobre

Universidade Federal de Minas Gerais (UFMG), Brasil
franknobre@gmail.com

Resumo

Este trabalho visa realizar uma investigação sobre o Twitter, analisando-o como um meio de contato direto com os clientes de empresas. Através do método de aprendizado de máquina conhecido como SVM, pretende-se estimar quanto os usuários estão expressando suas opiniões sobre um certo produto. O sistema operacional Windows 7® foi escolhido como estudo de caso e uma coleta de mensagens foi

Abstract

This paper aims to perform an investigation about how Twitter can work as a way to make direct contact with products' clients. Using a machine learning method called SVM, it intends to estimate how much Twitter users are expressing their opinions about a certain product. The operating system called Windows 7® was chosen as a study case and a message gathering from Twitter was done. The SVM was trained by a set of

realizada no Twitter. Treinou-se o SVM com um conjunto de mensagens classificadas manualmente como neutras ou opinativas. Um teste realizado da mesma maneira também foi feito, resultando numa acurácia de 80%. Por fim, as mensagens restantes foram submetidas ao classificador já treinado e separadas em duas classes distintas (opinativas e neutras). Concluiu-se que a maior parte do que é postado no Twitter não demonstra a opinião do usuário, sendo necessária uma quantidade considerável de mensagens para fazer a pesquisa com apenas aquelas que contém sentimentos relacionados ao produto em questão.

Palavras-chave: Análise de Sentimento; Twitter; Inovação; Aprendizado de Máquina. **Keywords:** *Sentiment Analysis; Twitter; Innovation; Machine Learning.*

1. Introdução

Com o uso significativo da web e o número crescente de redes sociais disponíveis, empresas tem buscado novas formas de obterem retorno sobre seus produtos. A pesquisa de opinião com questionário ou entrevista para saber o que as pessoas acharam de determinado lançamento, seus pontos fortes e fracos, possíveis melhorias, etc. já não é o único meio de obter informações do cliente, pois cada usuário pode expressar sua opinião ou sentimento acerca de um produto espontaneamente na web.

Essa opinião informal ou formal, dependendo de onde foi postada, é de extrema importância, pois refletirá o sentimento sem censura do usuário. Através de uma entrevista direta com a mesma pessoa, esta poderia se sentir intimidada e não responderia com a franqueza que geralmente é encontrada em opiniões postadas pela web.

Críticas, comentários e outras formas de análise de um produto estão disseminadas por toda a web em vários tipos de sites, variando de blogs pessoais a sites de compras. Uma fonte ainda não muito explorada nesse contexto é o Twitter.

Twitter é um serviço de micro-blogging que permite a postagem de mensagens (atualizações) de até 140 caracteres para que outras pessoas as visualizem. Atuando

também como uma rede social, permite que um usuário siga outros usuários e receba em sua página inicial ou aparelhos móveis as atualizações dessas pessoas na ordem em que foram postadas. A troca de mensagens privadas entre usuários cadastrados também é possível, assim como respostas públicas direcionadas.

Lançado em 2006, ainda é uma tecnologia recente e em expansão. Operando inicialmente apenas como um serviço de postagens, teve várias funções incorporadas à medida que os usuários criavam alternativas para suas necessidades como, por exemplo, a comunicação entre dois usuários através da citação da pessoa no início da mensagem com o símbolo @ antes do nome, indicando que aquela mensagem era direcionada àquele usuário.

De acordo com um estudo realizado pelo Global Language Monitor, como visto em Lea (2009), a palavra Twitter foi eleita a palavra mais usada em 2009, superando Obama e H1N1(segundo e terceiro lugares, respectivamente). Com tamanha repercussão, torna-se instigante um estudo relacionado às mensagens postadas no Twitter, considerando-as como reflexo das mentes dos consumidores.

Essas mensagens postadas por um usuário recebem o nome de tweets na terminologia padrão do Twitter. Neste artigo, serão utilizados os termos mensagem e tweet como palavras de mesmo significado.

Análise de sentimento ou mineração de opinião é um ramo da mineração de textos preocupado em classificar textos não por tópicos, e sim pelo sentimento ou opinião contida em determinado documento. Geralmente associado à classificação binária entre sentimentos positivos e negativos, o termo é usado de uma forma mais abrangente para significar "o tratamento computacional de opinião, sentimento e subjetividade em textos" (Pang e Lee, 2008: 10). Possuindo significado semelhante e uso conjunto na literatura, será utilizado nesse artigo o termo mineração de opinião como sinônimo de análise de sentimento, assim como feito em Pang e Lee (2008).

Nessa pesquisa, foram coletadas mensagens relacionadas ao novo sistema operacional da Microsoft®, o Windows 7®, pois o mesmo foi lançado recentemente para o público (22 de outubro de 2009) e os usuários do Twitter estariam provavelmente comentando sobre o produto.

Buscou-se mensurar o quanto os usuários estão expressando suas opiniões sobre o novo sistema operacional através da separação de mensagens neutras e mensagens opinativas utilizando métodos de análise de sentimento e mineração de opinião.

É interessante notar que no site do Windows 7[®] foi colocado um quadro contendo mensagens postadas no Twitter elogiando ou contendo mensagens neutras como: “win7 up and running finaly ! now taking a break.” ou “testing windows7 on new netbook.”. No geral, só aparecem mensagens positivas como: “Using Windows 7. So far, so good.” ou “Finally upgraded both our PCs to Win7. Love it! Shutdown/start times much better.”. Como pode-se imaginar, essas mensagens foram selecionadas de maneira a demonstrar a boa receptividade do Windows 7[®] e não servem como indicação de que não se fala mal do novo sistema operacional da Microsoft[®] no Twitter.

Sendo usado por uma grande empresa como veículo de propaganda, uma pesquisa sobre a viabilidade de analisar como está a receptividade de um produto no Twitter se torna bastante importante do ponto de vista mercadológico.

2. Metodologia

O presente trabalho constitui um estudo exploratório realizado utilizando práticas computacionais de mineração de dados na web e aprendizagem de máquina, tendo como unidade-caso, textos postados no Twitter sobre o Windows 7[®]. Através dessas práticas, buscou-se investigar se o usuário apresenta ou não opiniões sobre o produto da Microsoft[®].

Optou-se pelo uso do método supervisionado de aprendizagem de máquina conhecido como SVM (Support Vector Machine) para realizar a classificação binária desejada (mensagens neutras e mensagens contendo opiniões relacionadas ao Windows 7[®]). Foi utilizada a implementação fornecida em <http://svmlight.joachims.org/>.

Basicamente, o método consiste em representar cada documento por um ponto ou vetor em um espaço t-dimensional e traçar um hiperplano que separe da melhor forma possível as duas classes de documentos em questão, i.e., deve-se maximizar a distância do hiperplano e dos elementos (documentos) de borda das classes, que são os chamados support vectors.

Utilizando a Twitter Streaming API, pôde-se coletar os tweets praticamente em tempo real através de uma conexão persistente. Foram salvas 60923 mensagens. A coleta foi realizada entre os dias 21 e 30 de dezembro de 2009 em horários aleatórios e com durações variadas, de modo a não tendenciar a amostra coletada.

Como a Twitter Streaming API não permite a filtragem por palavras-chave contendo espaço, optou-se por realizar uma pesquisa mais abrangente com a palavra-chave “windows”, o que resultou na coleta de mensagens não relacionadas ao software da Microsoft® (e.g. "Through all windows, I see only infinity."). Além desse problema inicial, tem-se também mensagens em várias línguas, uma vez que não é possível especificar o idioma das mensagens retornadas. Sendo assim, tem-se um corpus (coleção de textos) inicial bastante ruidoso.

Com esses dois problemas iniciais em vista, optou-se primeiro por eliminar as mensagens que não estavam relacionadas ao Windows 7®. Foi implementada uma aplicação para verificar cada tweet e descartar aqueles que não eram de interesse. Esse primeiro filtro retornou 3156 tweets que continham a palavra “windows 7”. Para obter apenas mensagens em determinada língua, utilizou-se primeiramente a API do site <http://langid.net>, que é baseada na Google Ajax API, para identificar o idioma de cada mensagem. Sabendo-se em qual língua cada tweet estava, realizou-se mais um filtro para separar apenas as mensagens em inglês, visto que esse idioma representava a maior parte das mensagens coletadas (aproximadamente 60%). As demais mensagens estavam em várias línguas, como português, espanhol, italiano, francês, etc.. Por fim, restaram 1885 tweets.

Tem-se então um corpus contendo apenas mensagens sobre o assunto e no idioma desejado. Selecionou-se aleatoriamente 357 tweets para serem classificados de acordo com a interpretação do leitor em questão. Segundo Kim e Gilbert (2009), uma das dificuldades para definir uma boa base de treinamento é a falta de consenso entre os avaliadores das mensagens, ocorrendo classificações diferentes para um mesmo tweet dependendo do modo como cada um o interpretou. Como para a realização desse artigo apenas uma pessoa realizou tal tarefa, tais incongruências não ocorreram, mas também não é garantida a correteza da classificação, uma vez que textos estão sujeitos à interpretação de cada indivíduo.

Definida a base de treinamento, removeu-se a pontuação e palavras especiais do Twitter (e.g. @usuario, que indica que é uma mensagem citando aquela pessoa ou #assunto, que indica que o tweet trata daquele assunto especificado). Links de páginas da web também foram removidos. Optou-se pela não remoção de stopwords (palavras que não acrescentam informações significativas para o sentimento global da mensagem), pois as mensagens do Twitter são pequenas devido à própria natureza restritiva do serviço e não representariam um grande problema para serem processadas.

Por fim, representou-se cada tweet como um vetor de características, sendo tal representação conhecida na literatura como Modelo de Espaço Vetorial, onde “documentos são representados como pontos em um espaço Euclidiano t-dimensional em que cada dimensão corresponde a um token do léxico” (Carrilho Junior, 2007: 48). Neste caso, cada palavra (ou token) foi representada pelo índice dessa palavra no dicionário computado para o corpus em questão e o peso dela foi calculado de acordo com o tf-idf, sigla para term frequency-inverse document frequency. Ver Manning, Raghavan e Schütze (2009) para maiores detalhes.

Tal representação também é conhecida como saco de palavras (do inglês, bag of words), em que “um documento é visto como um container de tokens, aonde a ordem e a ligação entre os tokens não tem nenhum valor para o sistema” (Carrilho Junior, 2007: 48). Embora conhecidamente falha para identificar certas inflexões da língua escrita, “a codificação bag of words vem apresentando bons resultados na literatura, justificando a sua abordagem puramente estatística” (Carrilho Junior, 2007: 48).

Além das 357 mensagens classificadas, sorteou-se para uma base de teste mais 150 mensagens, que também foram classificadas manualmente de modo a verificar a qualidade do classificador. Com um acerto de 80%, verificou-se que o classificador estava apto para realizar a classificação das outras mensagens automaticamente.

Ainda faltando 1376 tweets para serem classificados, os mesmos foram convertidos para representação vetorial e submetidos ao classificador já treinado.

3. Resultados

Com o classificador treinado e já submetido ao teste para validá-lo, realizou-se a classificação dos 1376 tweets restantes. Foram identificadas 108 mensagens como contendo opiniões de usuários e o restante como mensagens neutras. Esta situação encontra-se representada na Figura 1. O eixo X representa o número dos tweets, variando de 1 à 1376 e o eixo Y representa o valor dado como resultado pelo SVM. Valores acima de 0 representam a classe de tweets contendo opiniões e sentimentos relacionados ao Windows 7[®] postadas pelos usuários e valores abaixo de 0 representam a classe daqueles que continham informações neutras sobre o assunto em questão.

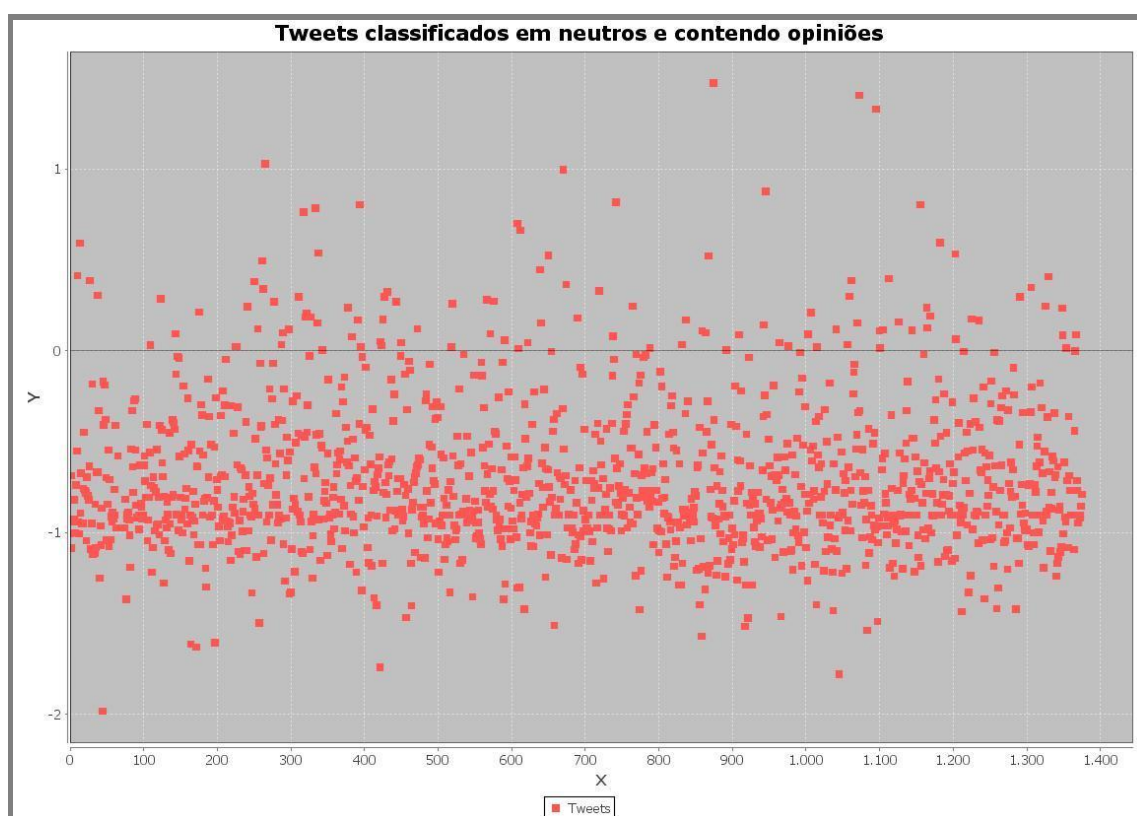


Figura 1: Gráfico mostrando a saída do classificador SVM.

Para uma melhor visualização da disposição dos tweets classificados, está representada na Figura 2 a mesma distribuição entre neutros e contendo opiniões.

Na base de teste, das 150 mensagens submetidas, 29 foram classificadas erroneamente. Ao verificar de quais tipos eram essas mensagens, contabilizou-se que 25 eram mensagens que continham opiniões e que tinham sido classificadas de maneira errada como neutras. Elas correspondem à 86% do total desses tweets que o classificador errou.

Como o classificador supostamente errará em 20% das mensagens, de acordo com o resultado obtido pela base de teste, tem-se 275 tweets classificados erroneamente do total de 1376. Considerando que 86% desses 275 tweets contém carga emocional relacionada ao Windows 7®, obtém-se 237 mensagens contendo opiniões além dos 107 identificados. Porém, houveram 38 mensagens neutras classificadas erroneamente como contendo opinião, segundo estimativa calculada, então deve-se subtrair esses 38 tweets dos 107 identificados inicialmente, resultando em 70 mensagens contendo opinião. Para chegar à distribuição final, soma-se as 237 mensagens com as 70 já conhecidas.

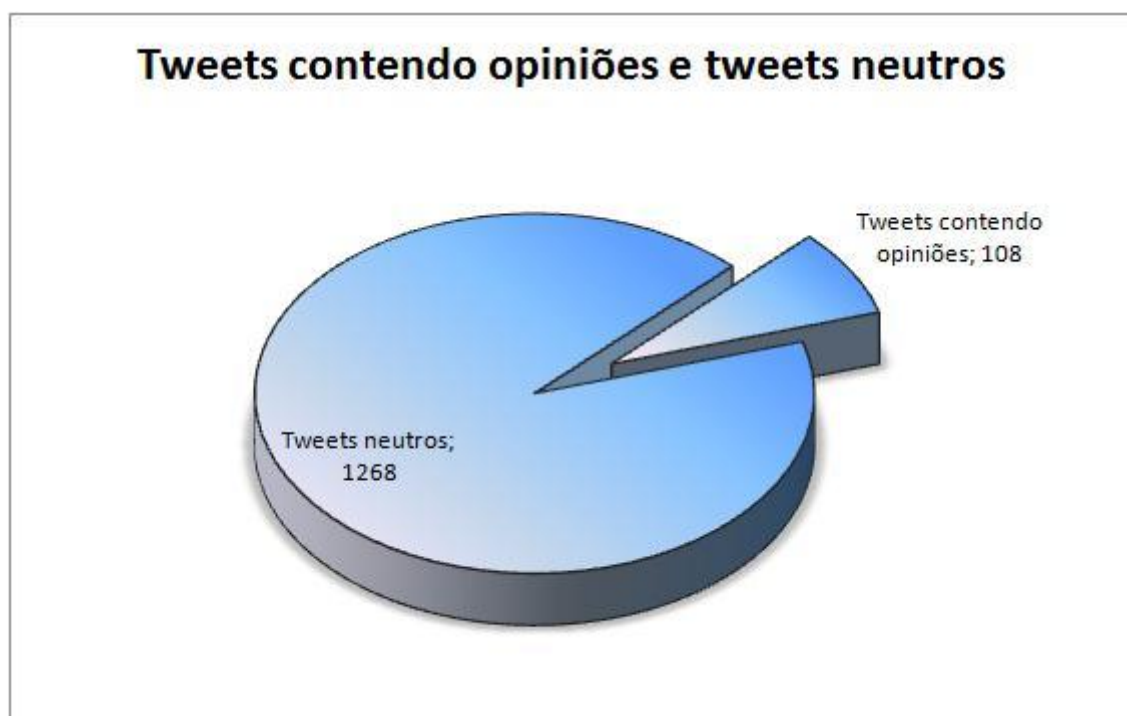


Figura 2: Gráfico mostrando a quantidade de tweets neutros e tweets contendo opiniões de acordo com a classificação realizada.

Por fim, tem-se o que equivaleria num total de 22% de mensagens contendo opiniões. O gráfico da Figura 3 mostra como ficaria a distribuição final nessa amostra coletada.



Figura 3: Distribuição real conforme suposição realizada.

Analisando as mensagens classificadas manualmente para treinar o classificador, percebe-se que a quantidade de tweets que contém opiniões de usuários é o equivalente à 26% dessas mensagens, mostrando que o classificador está realizando uma análise correta, sendo que temos valores tão próximos nessas duas distribuições (a amostra selecionada randomicamente para interpretação humana e a amostra coletada para ser classificada automaticamente pelo SVM).

4. Conclusão

Sendo uma rede social recente, não se sabe como será o uso do Twitter num futuro próximo. Muitas pessoas ainda desconhecem o serviço ou não o utilizam por não saberem exatamente seu conceito e função. Segundo estudo feito pelo site Pear Analytics (2009), a maior parte do que é postado no Twitter são informações inúteis, mas como várias pessoas já perceberam, existem também muitas informações importantes, bastando apenas saber filtrá-las.

Em análise feita nesse artigo, pôde-se perceber que para se extrair uma opinião sobre determinado produto, é necessário uma grande quantidade de tweets coletados, pois o que

é comentado, em sua maioria, são informações neutras sobre o assunto em questão. Conclui-se que o SVM possibilitou uma separação eficiente entre mensagens neutras e as que continham sentimento. Infelizmente, com a base de textos coletada, não foi possível realizar a classificação entre positivos e negativos após a separação das mensagens neutras, pois dispunha-se de poucas mensagens para treinar novamente o classificador para realizar outra classificação binária (dessa vez, positivos e negativos).

No entanto, o escopo desse artigo visava apenas mensurar o quanto os usuários estavam expressando seus sentimentos sobre determinado produto no Twitter e, para tal, a eliminação das mensagens neutras foi suficiente para o alcance da meta estipulada.

Tendo em vista o gráfico mostrado na Figura 4, percebe-se a grande quantidade de tweets não relevantes para uma análise de sentimento ou mineração de opinião no Twitter. Considerando que a distribuição entre mensagens neutras e contendo opiniões se mantenha na faixa dos 24%, deve-se ponderar o quão interessante é fazer uma análise da receptividade de determinado produto no Twitter. Como dito anteriormente, somente coletando uma grande quantidade de mensagens sobre o assunto é que se poderá realizar a tarefa de analisar o que os usuários estão opinando.

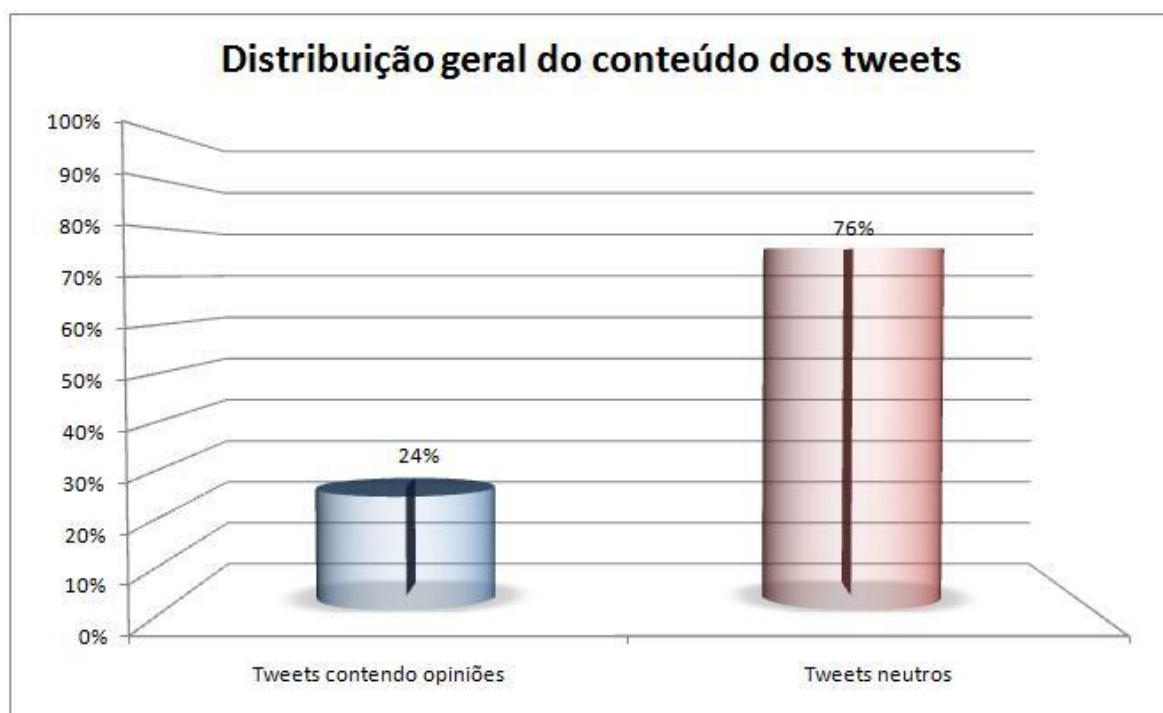


Figura 4: Porcentagem de tweets que contém opinião e aqueles que são mensagens neutras.

Essa distribuição necessita ser comprovada através de mais estudos sobre o que os usuários estão expressando no Twitter. Trabalhos futuros envolveriam a definição de novos produtos para coletar tweets relacionados e nova análise da distribuição entre mensagens neutras e contendo opiniões para comprovar ou não que esta distribuição se mantém, em geral, nessa faixa.

É importante lembrar também que a coleta feita no Twitter é realizada em tempo real à medida que as pessoas vão postando suas atualizações. Torna-se interessante então uma análise de sentimento no período de lançamento de um certo produto, visto que as pessoas tendem a comentar sobre assuntos atuais no Twitter e não sobre algo que já estão usando há algum tempo.

Com sua limitação no tamanho das mensagens, pode-se questionar a validade do Twitter como fonte de informação a respeito do que a pessoa pensa sobre determinado produto ou marca, mas pode ser justamente nesse ponto que está a sua qualidade. Sendo forçada a postar uma mensagem pequena, a pessoa simplesmente escreverá sua opinião direta a respeito do produto, sem se perder em pontos fortes e fracos ou divagações acerca do assunto.

5. Referências bibliográficas

CARRILHO JUNIOR, João Ribeiro; PASSOS, Emmanuel Piseces Lopes (Orientador). (2007). *Desenvolvimento de uma Metodologia para Mineração de Textos*. Dissertação de Mestrado – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro. Rio de Janeiro, 2007, 98p.

JOACHIMS, T. (1999). *Making large-Scale SVM Learning Practical*. *Advances in Kernel Methods - Support Vector Learning*. B. Schölkopf and C. Burges and A. Smola (ed.), MIT-Press, 1999.

KIM, Elsa; GILBERT, Sam. (2009). *Detecting Sadness in 140 Characters: Sentiment Analysis and Mourning Michael Jackson on Twitter*. Web Ecology Project. Aug. 2009.

LEA, Richard. (2009). *'Twitter' declared top word of 2009*. The Guardian, Reino Unido, 30 nov. 2009. Disponível em: <http://www.guardian.co.uk/books/2009/nov/30/Twitter-declared-top-word-of-2009>. Acesso em: 27 dez. 2009.

- MANNING, Christopher D.; RAGHAVAN, Prabakhar; SCHÜTZE; Hinrich. (2009). *Term frequency and weighting*, in An Introduction to Information Retrieval. England: Cambridge University Press, 2009. p 117-120. Disponível em: <http://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf>. Acesso em: 29 dez. 2009
- PANG, Bo; LEE, Lilian. (2008). *Opinion Mining and Sentiment Analysis*. Foundations and Trends in Information Retrieval 2(1-2), pp. 1–135, June 2008.
- PEAR ANALYTICS. (2009). *Twitter Study*. ago. 2009. Disponível em: <http://www.pearanalytics.com/wp-content/uploads/2009/08/Twitter-Study-August-2009.pdf>. Acesso em: 22 dez. 2009.