

# **Mineração de Opinião / Análise de Sentimentos**

Carlos Augusto S. Rodrigues  
Leonardo Lino Vieira  
Leonardo Malagoli  
Nikolas Timmermann

## Introdução

É evidente o crescimento da quantidade de informação disponível na Web nos últimos anos. Além do conteúdo disponibilizado por empresas de notícias, os usuários passaram a compartilhar na Web seus conhecimentos, críticas e opiniões em blogs pessoais, sites de relacionamentos, microblogs, dentre outros meios. Com este aumento da participação dos usuários nos conteúdos da Web, surgiu também uma nova área de pesquisa relacionada ao Processamento de Linguagem Natural, chamada de Análise de Sentimentos ou Mineração de Opinião.

Para os cientistas da computação, essa montanha crescente de dados oferece um vislumbre fascinante da consciência coletiva dos usuários de internet.

## Definição

A Análise de Sentimentos visa identificar o sentimento que os usuários apresentam a respeito de alguma entidade de interesse (um produto específico, uma empresa, um lugar, uma pessoa, dentre outros) baseado no conteúdo disponível na Web. O objetivo principal é permitir que um usuário obtenha um relatório contendo o que as pessoas andam dizendo sobre algum item sem precisar encontrar e ler todas as opiniões e notícias a respeito.

Cada vez mais pessoas e principalmente empresas, estão interessadas em observar as opiniões de um grupo de pessoas sobre temas que lhe interessam. Um caso comum é uma empresa que se interessa em medir a aceitação de um novo produto, monitorando as opiniões de um grupo em relação a ele. A análise de sentimento é um tipo de mineração de dados que possibilita tal demanda.

Utiliza-se ontologia para a realização da mineração, além de graus numéricos para indicar o quanto uma palavra ou expressão indica o sentimento. Não é analisado o que o conteúdo dos textos expressa, mas sim a emoção predominante no mesmo (positivo ou negativo dentre outros).

## Desafios

- Textos com erros e sentenças sintaticamente mal formadas (o que é bastante comum nos blogs e redes sociais) dificultam a busca e classificação dos mesmos;
- Não é trivial distinguir se um texto é opinião ou fato, e principalmente identificar em um fato se existem opiniões embutidas;
- Textos podem conter sarcasmos e ironias, que são difíceis de serem identificados e podem impactar os resultados;
- Um texto pode referenciar mais de um item de interesse (pode citar iPhone e iPod) com opiniões diferentes sobre os itens, o que pode confundir a classificação;
- Uso de pronomes para referenciar itens pode dificultar a identificação de sentenças que mencionam o item de interesse;
- Uso de termos informais da internet, como por exemplo “blz”, “fds” e “:.)”, devem ser considerados no vocabulário;

- Propaganda disfarçada, em que blogueiros recebem dinheiro para falar bem de alguma empresa ou produto pode impactar os resultados.

## Aplicações práticas

Um minerador de opiniões pode ser aplicado, por exemplo, nos seguintes cenários:

**Sentimento sobre um produto:** Esta análise pode ser usada por uma empresa para determinar a aceitação de um certo produto lançado no mercado, contribuindo para alterações estratégicas nas áreas de marketings. Ainda é possível verificar o sentimento sobre um produto que foi anunciado mas ainda não foi lançado, podendo ajudar em algumas alterações durante o desenvolvimento.

**Empresas na bolsa de valores:** Este cenário é usado para analisar empresas que contem ações na bolsa de valores. O sentimento sobre determinada empresa pode ser obtida através da opinião de especialistas e sites de notícias. O analisador ajuda na tomada de decisão sobre compras e vendas de ações, mas não deve ser um fator decisivo.

**Análise sobre pessoas:** O minerador de opinião pode ser usado para se obter o sentimento sobre determinadas pessoas, especialmente daqueles que aparecem muito na mídia. É possível saber, por exemplo, a aceitação ou rejeição de um determinado político durante a época de eleições, dando alternativa para que o político altere sua estratégia de campanha.

Os dados para uma análise podem ser obtidas de diversas fontes na internet, antigamente restrito a blogs e fóruns, hoje se expande as redes sociais. Algumas das fontes de minerações:

**Usando mecanismos de busca:** É possível usar mecanismos de busca como Google ou Bing para obtenção de opiniões na internet, principalmente em fóruns e blogs, mas podendo estender a qualquer site. Este método de obtenção de opinião é variável, conforme a ferramenta de busca altera seus resultados, e não é muito acurado, sendo que páginas favorecidas pela mecanismo de busca podem alterar o resultado. O site [whatdoestheinternetthink.net](http://whatdoestheinternetthink.net) possui um algoritmo para saber a opinião da internet sobre determinado assunto, utilizando a ferramenta Bing.

**Twitter:** O twitter é um ótimo local para se obter opinião sobre determinado assunto, já que é uma rede social focada em opiniões de usuários. Ele possui mais de 250 milhões de tweets diários (2011), e os tweets são, em sua maioria, públicos, fornecendo uma base rica em informação e de fácil acesso. O site [sentiment140 \(http://www.sentiment140.com\)](http://www.sentiment140.com) obtém informações de tweets recentes sobre determinado assunto e forma um gráfico analisando as opiniões dos usuários.

**Facebook:** O facebook possui uma grande quantidade de usuários (1 bilhão de usuários ativos em outubro de 2012). Os usuários podem demonstrar suas opiniões através de mensagens, compartilhamentos e “likes”. A obtenção destes dados pode não ser tão trivial em comparação ao twitter, já que não é possível obter estas informações de forma simples, podendo ser necessário a instalação de aplicativos. O Facebook possui uma ferramenta em suas fan pages chamada “Facebook Insight”, ele não funciona exatamente como um minerador de opiniões, mas possui informações sobre a página, incluindo feedbacks negativos.

## **Etapas da mineração de opinião**

O processo de mineração de opinião consiste basicamente em 3 etapas. Como segue:

- Coleta de conteúdo;
- Classificação;
- Sumarização dos resultados;

### **Coleta de conteúdo**

Etapa na qual é feita uma busca em fontes diversas, tais como artigos em sites, comentários, anúncios, documentos dentre outras. É importante a utilização de técnicas avançadas de busca, visando identificar se o conteúdo encontrado, trata-se de uma opinião ou um fato. Fatos por si só devem ser descartados, porém opiniões expressas em fatos, devem ser mantidas.

A tarefa de busca do conteúdo e classificação em fato ou opinião não é uma atividade trivial, ainda hoje existe uma grande demanda por pesquisas na área, e várias delas encontram-se em andamento.

### **Classificação**

A classificação é etapa mais importante do processo, e nesta etapa que medimos a polaridade ou orientação da opinião. Esta etapa pretende determinar se um opinião é positiva, neutra ou ainda negativa.

Na classificação ou análise de sentimentos são as palavras opinativas que tem maior importância, através delas, é possível determinar a polaridade expressa pelo autor. Exemplos de palavras opinativas: bom, legal, ótimo, ruim, péssimo etc.

Para alcançar o objetivo da classificação, existem diversas técnicas, dentre as quais:

- Aprendizagem de máquina;
- Seleção de palavras;
- análise sintática.

### **Aprendizagem de máquina**

A aprendizagem de máquina utiliza algoritmos já conhecidos e consolidados, tal como redes neurais.

Nessa técnica é necessário um conjunto de conteúdo previamente classificado, que será utilizado como base do treinamento do modelo. A partir do modelo já treinado, novos conteúdos são classificados.

A desvantagem dessa técnica, está no uso de um conjunto de conteúdos previamente classificados e que para diferentes tópicos, diferentes modelos, e conseqüentemente diferentes conjuntos de treinamentos devem ser adotados. Por exemplo, um modelo de classificação de celular não é adequado para classificar opiniões referentes à lugares.

### **Seleção de palavras**

Abordagem na qual um determinado conteúdo é avaliado com base no resultado da avaliação de suas palavras, tendo como base um conjunto de palavras previamente classificadas. Nesta técnica existe um dicionário de palavras positivas, negativas e neutras, por exemplo, bom, legal, ótimo são palavras positivas, já péssimo, ruim, horrível são palavras que expressão um sentimento negativo.

A vantagem desta técnica está na sua simplicidade, e o fato de não ser necessário classificar previamente um conjunto de conteúdo para treinamento, no entanto, o conjunto de palavras deve ser cuidadosamente selecionado, tratando ainda a ambigüidade de uma mesma palavra em um mesmo tópico. Por exemplo, a palavra grande, no contexto de avaliação de um celular, pode representar tanto uma opinião positiva, quanto uma opinião negativa. O dicionário selecionado deve ser ajustado para cada tópico

### **Análise sintática.**

Essa etapa visa analisar sintaticamente o conteúdo, identificando adjetivos e ou advérbios que podem indicar a polaridade do conteúdo. Como vantagem desta técnica, ela também não precisa de um conjunto de dados previamente classificados, porém , é fundamental para o sucesso desta técnica, um classificador sintático eficiente.

Estas três técnicas podem ainda serem utilizadas em conjunto, formando um modelo híbrido, e ajudando a aumentar a acurácia da classificação.

## **Sumarização**

Etapa focada na apresentação dos resultados, que pode ser de forma textual, ou gráfica. A apresentação dos resultados de forma textual não é uma tarefa trivial, e é por si só um tópico de estudos. Ela envolve geração de linguagem natural, e apesar das recentes pesquisas, ainda é um tópico em aberto. A forma mais simples de sumarizar os resultados, e a forma gráfica, onde os resultados são sumarizados em totais e dados estatísticos, facilitando sua apresentação.

## **Bibliografia**

**PROTÓTIPO PARA MINERAÇÃO DE OPINIÃO EM REDES SOCIAIS: ESTUDO DE CASOS SELECIONADOS USANDO O TWITTER , LEANDRO MATIOLI SANTOS , 2010.**

<http://blog.vettalabs.com/2009/10/26/mineracao-de-opiniao/>

<http://tecnologia.terra.com.br/noticias/0..OI3939723-EI12884.00-Analise+de+sentimentos+e+novo+campo+na+web.html>

<http://webholic.com.br/2010/01/13/analise-de-sentimento/>