

Uma Ferramenta para Análise de Sentimentos em Redes Sociais Utilizando o SenticNet

Yuri Malheiros^{1,2}, George Lima²

¹Departamento de Ciências Exatas – Universidade Federal da Paraíba (UFPB)
Rio Tinto – PB – Brasil

²Centro de Informática - Universidade Federal de Pernambuco (UFPE)
Recife – PE – Brasil

yuri@dce.ufpb.br, georgelimal@gmail.com

Abstract. *Social networks are now a global communication medium with a huge number of users. The Facebook, for example, has more than 1 billion active users per month, and these users share a myriad of information, including opinions about many subjects. This big quantity of public information accessible has attracted the attention of companies and institutions that want to know better what people on the Internet are thinking about them. In this paper, we present a sentiment analysis tool for social network messages using the SenticNet knowledge base. The early tests to evaluate the tool have shown that our approach properly classifies the sentiment in messages, and it supports the large quantity and speed of the sharings in a social network.*

Resumo. *As redes sociais se tornaram um meio global de comunicação com um número muito grande de usuários, o Facebook, por exemplo, já possui mais de 1 bilhão de usuários ativos por mês, no qual são compartilhadas diversos tipos de informações, entre elas opiniões sobre os mais variados assuntos. Tal riqueza de informações acessíveis publicamente tem atraído a atenção de empresas e instituições que desejam conhecer melhor o que as pessoas na Internet estão pensando sobre elas. Neste trabalho, é apresentada uma ferramenta de análise de sentimentos em mensagens compartilhadas em redes sociais usando o SenticNet como base de conhecimento. Os primeiros testes realizados para avaliação da ferramenta já começam a mostrar que a abordagem consegue classificar mensagens adequadamente e suporta a grande quantidade e velocidade com que elas são enviadas para as redes sociais.*

1. Introdução

O uso maciço das redes sociais pelos usuário da Internet criou um ambiente digital onde um número muito grande de pessoas consegue se expressar através do compartilhamento de fotos, vídeos, textos e muitos outros tipos de dados suportados pela Web. Para se ter uma ideia da importância das redes sociais e dos dados compartilhados nelas, o Facebook¹ hoje possui mais de 1 bilhão [Facebook 2012] de usuários ativos e ameaça os sites de buscas, que sempre tiveram um papel chave na Web, como site mais popular da Internet [Harvey 2010].

¹<http://facebook.com>

Essa avalanche de dados gerada pelos usuários das redes sociais é preciosa e muitas vezes traz informações que não são percebidas com facilidade. Por exemplo, é possível extrair automaticamente de mensagens textuais sobre que assunto elas se referem, que idioma estão as mensagens e também que sentimento elas carregam: felicidade, tristeza, excitação, reclamação, etc. Este último tipo de informação é tratado especificamente pela área de análise de sentimentos.

A detecção de sentimentos através de computadores tem ganhado muita atenção nos últimos anos, tanto nas universidades quanto nas empresas. Um dos motivos de tal interesse é justamente o grande aumento do conteúdo gerado pelas pessoas na Web, principalmente quando elas estão expressando opinião sobre determinados assuntos, serviços e produtos.

Conseguir extrair o que os usuários das redes sociais estão sentindo em relação a algo pode ter importância vital para empresas e instituições em geral. Pois, dessa forma, é possível saber a todo momento, o que milhares, ou milhões de pessoas estão pensando, sentindo e expressando sobre assuntos de importância para quem estiver monitorando os usuários. Através de técnicas de análise de sentimentos aplicadas às rede sociais podem ser realizadas pesquisas de opinião com as seguintes características:

Grande escala: por causa do grande número de usuários nas redes sociais, que passa dos bilhões, pode-se capturar a opinião de muitas pessoas dos mais diferentes lugares sem muito esforço.

Não invasiva: os usuários não precisam saber que a pesquisa de opinião está sendo realizada. Já que eles compartilham opiniões naturalmente na Internet, nenhum trabalho extra precisa ser exigido para quem estiver tendo sua opinião analisada.

Rápida: com técnicas eficientes é possível em alguns segundos ou minutos classificar que tipo de sentimentos os usuários estão tendo em relação a um determinado assunto.

Autêntica: ninguém precisará pedir ou forçar outra pessoa a responder um questionário ou algo parecido, os sistemas de análise de sentimentos recuperam e analisam mensagens que já são compartilhadas espontaneamente pelas pessoas, o que pode trazer uma autenticidade maior para uma pesquisa de opinião.

Barata: com uma ferramenta que faça a análise de sentimentos em redes sociais em funcionamento, o custo de analisar milhares de mensagens de milhares de pessoas é muito baixo, principalmente se comparado com pesquisas de opinião tradicionais.

Automática: todo o trabalho de procurar mensagens e analisar cada uma é feito por um computador, fazendo com que a interferência humana seja quase desnecessária.

Nesta pesquisa, deseja-se criar e avaliar uma ferramenta que extraia sentimentos em mensagens textuais compartilhadas em redes sociais usando o SenticNet [Cambria et al. 2010], uma base de dados para análise de sentimentos e mineração de opinião construída por meio da *sentic computing* [Cambria and Hussain 2012]. Dessa forma, será possível entender o que muitos usuários nas redes sociais estão sentindo sobre determinados assuntos em tempo real.

O restante do artigo está estruturado da seguinte forma. Na seção 2 são descritos alguns trabalhos relacionados da área, a ferramenta proposta é descrita em mais detalhes na seção 3 e resultados preliminares são apresentados na seção 4. Por fim, conclusões e trabalhos futuros são discutidos na seção 5.

2. Trabalhos Relacionados

Atualmente existem diversos trabalhos na área de análise de sentimentos, entre eles vários voltados a analisar conteúdo criado pelos usuários da Internet.

No trabalho de [Grassi et al. 2011] foi criado um *website* no qual é possível assistir e navegar entre 100 vídeos retirados do Youtube². Entretanto, o *website* vai além da navegação que estamos acostumados, pois, usando técnicas da inteligência artificial e da web semântica para classificar os vídeos, ele torna possível fazer buscas de acordo com os sentimentos associados aos vídeos. Por exemplo, o usuário pode listar todos vídeos que causam medo ou vídeos engraçados.

A ferramenta SPOONS do Netflix [Augustine et al. 2012], é um exemplo prático da utilidade da análise de sentimentos em redes sociais. O SPOONS usa mensagens compartilhadas pelos usuários do Twitter³ (tweets) para detectar falhas de disponibilidade do Netflix. Várias técnicas são utilizadas para tentar detectar falhas no serviço através dos tweets, entre elas o SPOONS classifica as mensagens em categorias, por exemplo, feliz, neutro e reclamação, para saber se muitos usuários estão reportando problemas na rede social. Isto é possível através do uso de classificadores que são treinados com tweets que foram rotulados em categorias manualmente por especialistas.

O trabalho de [Davidov et al. 2010] usou uma ideia simples e inteligente para treinar um classificador de sentimentos em tweets. Ao invés de usar especialistas para rotular manualmente as mensagens do conjunto de treinamento, foram utilizadas 50 *hashtags* e 15 *emoticons*, estruturas que aparecem naturalmente nos tweets, como os possíveis rótulos. Dessa forma, o trabalho de classificar as mensagens do conjunto de treinamento passou a ser automático, possibilitando gerar um conjunto muito maior em um tempo muito menor.

Uma abordagem um pouco diferente foi proposta por [Hassan et al. 2010], neste trabalho não foram classificados quais sentimentos estão relacionados a uma mensagem textual, o foco dele foi avaliar a interação entre usuários através de mensagens de texto, para saber se um usuário tem uma atitude positiva ou negativa em relação aos outros usuários.

3. Ferramenta de Análise de Sentimentos

A ferramenta em desenvolvimento para análise de sentimentos em mensagens de redes sociais está dividida em três componentes principais: o componente de aquisição, o componente de classificação e a aplicação. A organização dos componentes e como eles interagem entre si pode ser visualizada na Figura 1.

O componente de aquisição é responsável por se conectar às redes sociais e recuperar mensagens compartilhadas pelos usuários. Atualmente este componente se conecta com o Facebook, rede social mais popular da atualidade, e realiza uma busca por mensagens públicas de acordo com palavras-chave fornecidas. Dessa forma, é possível, através deste componente, listar as mensagens do Facebook que estejam relacionadas com um determinado assunto, produto, evento, entre outros.

²<http://youtube.com>

³<http://twitter.com>

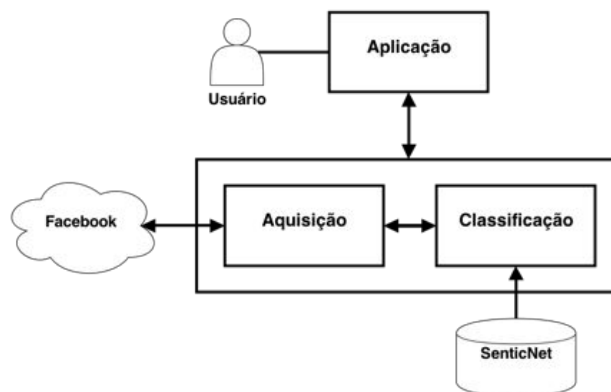


Figura 1. Arquitetura da ferramenta

Após a obtenção das mensagens, elas são passadas para o componente de classificação, para que ele calcule a polaridade de cada uma. Os valores das polaridades variam entre -1 e 1, onde valores abaixo de zero representam sentimentos negativos, sendo -1 o mais negativo possível e valores acima de zero representam sentimentos positivos, sendo 1 o mais positivo possível.

Para classificar as mensagens é utilizado o SenticNet como base de conhecimento. Ele consiste em um arquivo RDF [Lassila et al. 1998] que traz a polaridade de várias palavras e expressões da língua inglesa. Assim, para calcular a polaridade de uma mensagem, a ferramenta procura por cada palavra ou conjunto de palavras da mensagem no SenticNet, soma os valores das polaridades de todas as ocorrências encontradas e divide pela quantidade de ocorrências encontradas. Ou seja, a média das polaridades das ocorrências é o valor final da polaridade da mensagem.

Por exemplo, para a frase “Cool! It is friday night” a ferramenta encontra a ocorrência de “cool” no SenticNet que possui polaridade +0,383 e também a expressão “friday night” que possui polaridade +0,728. Portanto, a polaridade da frase é $(0,383+0,728)/2 = 0,5555$.

A classificação das mensagens usando o SenticNet é muito veloz, analisar milhares de mensagens pode levar apenas poucos segundos. Esta é uma importante vantagem para a ferramenta, pois, com o grande volume de mensagens compartilhadas em uma rede social, seria inviável classificá-las em tempo real usando uma técnica muito custosa computacionalmente.

O componente de aplicação ainda se encontra em fase de planejamento, mas já são conhecidas as principais funcionalidades que ele precisará ter. Através da aplicação o usuário poderá fazer buscas por palavras-chave para que sejam recuperadas as mensagens desejadas e em seguida a ferramenta exibirá a classificação de cada uma delas e uma polaridade global de todas as frases. Este componente utilizará a infraestrutura que é provida pelos dois primeiros componentes.

A ferramenta em funcionamento pleno trará para os seus usuários a capacidade de medir o que as pessoas ao redor de todo mundo que usam redes sociais estão sentindo sobre determinados assuntos, basta que seja feita uma procura que reflita o assunto

desejado.

4. Resultados Preliminares

Nesta seção serão apresentados alguns resultados preliminares do analisador de sentimentos implementado, assim como algumas discussões.

Para os primeiros testes foram obtidas mensagens do Facebook que possuísem a palavra “war”. O termo em inglês foi escolhido, pois o SenticNet possui apenas classificações para conceitos em inglês. Era esperado que uma procura por “war”, palavra que remete à fatos negativos, retornasse muitas frases com polaridade negativa, mas não foi o que aconteceu. A média das polaridades das frases recuperadas foi positiva, 0,2425.

Alguns exemplos de mensagens positivas com a palavra “war” foram: “*Watching War Horse. What a great movie!*”, que possui polaridade 0,356. E “*lol my dad is wanting me to join him with watching war horse movie..just started soo his way of telling me to come and watch is by narrating every single detail of whats happening xD*”, que possui polaridade 0,328.

Ao buscar mensagens com a palavra “war” o sistema acabou analisando várias frases sobre o filme de Steven Spielberg, War Horse, que foi nomeado a 6 Oscars. Os diversos comentários positivos sobre o filme influenciaram bastante para levar o valor da média das polaridades a um valor positivo.

Mesmo com a média positiva, ainda existiram as esperadas mensagens negativas, por exemplo: “*i hate my family in the morning..my wife bark out, the kids crying..am i in the war zone..??*”, que possui polaridade -0,287. E “*We continue to wage war on evil men, instead of on what makes men evil...!!!*”, que possui polaridade -0,0609.

O tempo levado para classificação das mensagens foi medido, pois um dos requisitos do sistema é conseguir analisar mensagens em um tempo muito próximo do real. Todo o processo levou 16,06 segundos para classificar 500 mensagens, sendo 15,2 segundos para se conectar e recuperar as mensagens do Facebook e 0,86 segundos para fazer a análise de sentimentos.

5. Conclusões e Trabalhos Futuros

O trabalho em andamento já apresenta bons resultados para as funcionalidades requeridas inicialmente. O SenticNet classifica coerentemente muitas mensagens, apesar de conter um número limitado de conceitos, e a base de conhecimento proporciona uma velocidade muito grande para a ferramenta, pois o conhecimento para classificação das frases já está todo embutido no RDF.

Assim, a infraestrutura já construída será útil para que uma futura aplicação consiga trazer para usuários finais a possibilidade de entender em tempo real o que as pessoas estão sentindo sobre determinados assuntos nas redes sociais.

O uso do SenticNet traz uma limitação em relação ao idioma usado, pois a base de conhecimento foi criada com expressões em inglês. Assim, mensagens em outras línguas não podem ser classificadas diretamente utilizando o SenticNet. Uma possível solução para este problema seria utilizar um tradutor nas mensagens compartilhadas antes

de serem classificadas. Outra limitação está relacionada ao conhecimento contido no RDF. O arquivo contém muitos conceitos, mas faltam muitos outros e sempre faltarão. Com isso, algumas mensagens, por mais que tenham sentimento nelas, podem ficar com classificação neutra na ferramenta.

Existem muitas possibilidades de evolução do trabalho atual, além do término da aplicação, que tornará as funcionalidades já desenvolvidas disponíveis de forma amigável para os usuários, e das melhorias que podem ser realizadas para superar as dificuldades apontadas anteriormente. Primeiramente será feita uma avaliação da classificação de sentimentos das mensagens pela ferramenta e da utilidade dos resultados apresentados para os usuários. Em seguida, o próximo passo será a integração com mais redes sociais. Apesar do Facebook ser a rede social mais popular, outras redes sociais são bastante conhecidas e utilizadas como o Twitter e Google+⁴. Ainda podemos ir adiante e adicionar redes sociais mais específicas como LinkedIn⁵, Quora⁶ e Pinterest⁷.

Referências

- Augustine, E., Cushing, C., Dekhtyar, A., McEntee, K., Paterson, K., and Tognetti, M. (2012). Outage detection via real-time social stream analysis: leveraging the power of online complaints. In *Proceedings of the 21st international conference companion on World Wide Web, WWW '12 Companion*, pages 13–22, New York, NY, USA. ACM.
- Cambria, E. and Hussain, A. (2012). *Sentic Computing: Techniques, Tools, and Applications*. Springer Netherlands.
- Cambria, E., Speer, R., Havasi, C., and Hussain, A. (2010). Senticnet: A publicly available semantic resource for opinion mining.
- Davidov, D., Tsur, O., and Rappoport, A. (2010). Enhanced sentiment learning using twitter hashtags and smileys. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters, COLING '10*, pages 241–249, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Facebook (2012). Facebook reports third quarter 2012 results.
- Grassi, M., Cambria, E., Hussain, A., and Piazza, F. (2011). Sentic web: A new paradigm for managing social media affective information. *Cognitive Computation*, 3(3):480–489.
- Harvey, M. (2010). Facebook ousts google in us popularity.
- Hassan, A., Qazvinian, V., and Radev, D. (2010). What's with the attitude?: identifying sentences with attitude in online discussions. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, EMNLP '10*, pages 1245–1255, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Lassila, O., Swick, R. R., Wide, W., and Consortium, W. (1998). Resource description framework (rdf) model and syntax specification.

⁴<http://plus.google.com>

⁵<http://linkedin.com>

⁶<http://quora.com>

⁷<http://pinterest.com>