



Universidade Federal de Pernambuco

Graduação em Ciência da Computação

2012.2

Análise de Sentimentos e Mineração de Links em uma Rede de Co-ocorrência de Hashtags

Proposta de Trabalho de Graduação

Aluno: Arthur Cavalcanti Alem (aca3@cin.ufpe.br)

Orientador: Ricardo Bastos Cavalcante Prudêncio (rbcp@cin.ufpe.br)

Recife, Janeiro de 2013

Sumário

1. Contexto	3
2. Objetivos.....	4
3. Cronograma	5
4. Referências Bibliográficas.....	6
5. Possíveis Avaliadores.....	7
6. Assinaturas	8

1. Contexto

Muitos conjuntos de dados hoje são mais bem representados como uma coleção de objetos interligados entre si, como é o caso de redes sociais, onde o atributo amizade pode representar um link entre duas pessoas distintas. Mineração de Links diz respeito a técnicas de mineração de dados que explicitamente levam em conta estas ligações ao construir modelos descritivos ou preditivos destes dados, e abrange tarefas como detecção de grupos, predição de links, e classificação coletiva, entre outras [1].

Em plataformas como o Twitter, os usuários tendem a expressar livremente e publicamente os seus sentimentos, o que cria um meio ideal de se capturar as opiniões comuns sobre diversos tópicos, como empresas, produtos ou celebridades [2]. Esta informação, no entanto, acaba por ser muito extensa, inviabilizando que se analise na íntegra todas as opiniões expressadas, e tornando atrativo o uso de ferramentas automáticas capazes de extrair o sentimento geral contido nos dados. O Twitter possui ainda a característica de os usuários frequentemente “etiquetarem” as próprias mensagens com *hashtags*, palavras precedidas do símbolo “#”, para deixar ainda mais explícito o seu sentimento e/ou o tópico ao qual se referem.

Na tentativa de identificar a orientação da opinião de usuários, geralmente em positiva ou negativa, várias técnicas de Análise de Sentimento já foram propostas, a maioria delas baseadas em classificação de textos [3]. Estudos sugerem, no entanto, que a classificação automática de dados pode ser consideravelmente aprimorada ao se considerar a vizinhança destes em uma estrutura de grafo [4]. É irrefutável que links entre os dados representam dicas semânticas de alta qualidade, que são totalmente desconsideradas diante de um classificador de texto puro, porém explorar esta informação de links não é trivial, pois a mesma costuma apresentar muitos ruídos [5]. Já foi amplamente observado também, que várias redes, incluindo as redes sociais, tendem a favorecer desproporcionalmente a ligação entre dados que possuem atributos ou características similares, um conceito denominado Assortatividade [6].

Este Trabalho de Graduação pretende explorar o uso de hashtags no Twitter como uma nova abordagem para se classificar tópicos em relação à opinião dos usuários, uma vez que a análise baseada apenas em classificação de textos tende a não ser muito confiável devido à alta sensibilidade à qualidade dos dados de treinamento [2].

2. Objetivos

Este trabalho tem como objetivo principal avaliar os conceitos de Mineração de Opiniões e Assortatividade sendo aplicados em uma rede real montada a partir da ocorrência de hashtags em mensagens no Twitter, onde cada hashtag representa um tópico ou sentimento.

Utilizando técnicas de classificação coletiva, os tópicos serão classificados automaticamente como relacionados a percepções positivas ou negativas dos usuários do Twitter em um determinado período de tempo, baseado principalmente nos outros tópicos aos quais estão frequentemente atrelados, ao invés de apenas na análise de sentimento dos textos das próprias mensagens.

3. Cronograma

Atividade	Dezembro		Janeiro		Fevereiro		Março	
Revisão Bibliográfica								
Estudo das ferramentas auxiliares								
Implementação								
Elaboração do relatório								
Preparação da apresentação								

4. Referências Bibliográficas

- [1] L. Getoor and C. P. Diehl, “Link mining: a survey,” *ACM SIGKDD Explorations Newsletter*, vol. 7, no. 2, pp. 3–12, 2005.
- [2] X. Wang, F. Wei, X. Liu, M. Zhou, and M. Zhang, “Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach,” in *Proceedings of the 20th ACM Conference on Information and Knowledge Management*, pp. 1031–1040, 2011.
- [3] J. Rabelo, R. Prudêncio and F. Barros, “Collective Classification for Sentiment Analysis in Social Networks”, In *Proceedings of the 24th IEEE International Conference on Tools with Artificial Intelligence*, 2012.
- [4] R. Angelova and G. Weikum, “Graph-based text classification: learn from your neighbors”. In *ACM SIGIR*, pp. 485–492, 2006.
- [5] S. Chakrabarti, B. Dom, and P. Indyk, “Enhanced hypertext categorization using hyperlinks,” in *Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data*, pp. 307–318, 1998.
- [6] J. Bollen, B. Gonçalves, G. Ruan, and H. Mao, “Happiness is assortative in online social networks,” *Artificial Life*, vol. 17, no. 3, pp. 237–251, 2011.
- [7] X. Meng, F. Wei, X. Liu, M. Zhou, S. Li, and H. Wang, “Entity-Centric Topic-Oriented Opinion Summarization in Twitter”. *KDD’12*, 2012.
- [8] S. A. Macskassy and F. Provost, “Classification in networked data: A toolkit and a univariate case study,” *Journal of Machine Learning Research*, vol. 8, pp. 935–983, 2007.
- [9] M. Conover, B. Gonçalves, J. Ratkiewicz, A. Flammini, and F. Menczer, “Predicting the political alignment of twitter users,” in *Proceedings of 3rd IEEE Conference on Social Computing (SocialCom)*, 2011.

5. Possíveis Avaliadores

- Flávia de Almeida Barros (fab@cin.ufpe.br)
- Patrícia Cabral de Azevedo Restelli Tedesco (pcart@cin.ufpe.br)

6. Assinaturas

Ricardo Bastos Cavalcante Prudêncio
Orientador

Arthur Cavalcanti Alem
Aluno