

Caracterização Automática de Notícias no Contexto do Mercado de Ações Brasileiro

PROPOSTA DE TRABALHO

Fulano Beltrano da Silva

fbeltrano@email.edu

Curso de Sistemas de Informação

Centro Universitário Ritter Dos Reis - UNIRITTER

1. Relevância do trabalho

Muitos investidores atuam no mercado de ações com o objetivo de realizar investimentos rentáveis e uma previsão confiável do comportamento do mercado pode ser útil neste contexto [2]. Contudo, esta não é uma tarefa fácil e ainda não há um método amplamente aceito para prever o movimento dos preços das ações [6]. Os desafios envolvem a compreensão da dinâmica do mercado, onde parâmetros mudam constantemente e muitas vezes se quer estão (são) bem definidos(conhecidos). Por esta razão, a análise do mercado de ações tem atraído o interesse da comunidade científica.

A *análise técnica* e a *análise fundamentalista* [3] são duas das principais vertentes no contexto de prever o comportamento do mercado financeiro. Basicamente, a análise técnica emprega dados históricos de preços no processo de prever o comportamento futuro destes preços, enquanto que a análise fundamentalista baseia-se em fatores econômicos que afetam o ramo de atividade das empresas, como indicadores econômicos e quebras de produção. Neste contexto, eventos/notícias podem ter grande impacto no comportamento dos preços das ações no mercado e, por esta razão, um dos pré-requisitos de um analista de mercado é estar a par de notícias associadas a ramos de atividades de uma carteira de investimentos.

Com base no grande e crescente volume de informações que são gerados diariamente nos canais de notícias, se torna impraticável para um analista monitorar uma grande quantidade de eventos relativos ao mercado de capitais, sobretudo quando este processo deve ocorrer na velocidade das oscilações de preços dos ativos financeiros. Assim, para embasar a tomada de decisão sobre títulos transacionados em bolsa de valores, se torna útil o processamento computacional prévio destas informações de forma a entregar para o profissional especialista somente as informações relevantes ao contexto de interesse. Neste sentido, técnicas de mineração de dados textuais têm mostrado resultados relevantes na tarefa de automatizar a análise de notícias ¹ e seu impacto em mercados internacionais [6, 4].

No contexto discutido acima são raros os trabalhos que envolvem o mercado nacional e sua suscetibilidade à notícias/eventos. Quando o foco é o mercado de ações brasileiro, a literatura apresenta métodos que exploram o histórico de preços das ações do ponto de vista de técnicas de análise de séries temporais [1], porém não verificou-se trabalhos que visam automatizar a busca por padrões de conteúdo em notícias (texto) como forma de entender o comportamento do mercado de ações.

2. Objetivos

2.1. Objetivo geral

Este trabalho propõe um estudo sobre as potencialidades de métodos clássicos de mineração de texto focado na caracterização automática do conteúdo de notícias que impactam no mercado de ações brasileiro.

2.2. Objetivos específicos

A abordagem testada neste trabalho visa avaliar o desempenho de uma abordagem *bag-of-words* [5] para propósitos de descoberta de padrões noticiosos e seus reflexos no mercado de ações brasileiro. Com isso busca-se avaliar as

¹<http://www.thestocksonar.com/>

potencialidades de uma abordagem clássica e difundida para o problema de mineração de textos, para então, em um trabalho futuro, implementar a classificação de notícias em conjunto com técnicas de seleção de feições mais elaboradas, como aquelas baseadas em extração de tópicos [4]. Deste modo, entre os objetivos específicos deste trabalho pode-se destacar:

- Avaliar a existência de padrões em notícias que versam sobre o mercado financeiro brasileiro em contextos de alta e baixa de preços.
- Investigar a influência de falsos positivos em dados de treinamento para a classificação supervisionada, uma vez que assume-se que todas as notícias que ocorrem em janelas temporais caracterizadas pela queda de preços são, de fato, notícias que levam a queda destes e vice-versa.
- Prover um estudo inicial sobre a bolsa brasileira no sentido de responder, de forma automatizada, perguntas como *"O quão similar as notícias de hoje são em relação àquelas do passado?"*, provendo suporte a decisão para analistas que buscam medir o impacto de determinados eventos sobre um conjunto de ações.

3. Solução proposta

Este trabalho visa avaliar o desempenho de uma abordagem bag-of-words [5] para propósitos de descoberta de padrões noticiosos e seus reflexos no mercado de ações brasileiro. Especificamente, busca-se avaliar as potencialidades de uma abordagem clássica e difundida de mineração de textos para a caracterização/classificação de textos noticiosos em contextos de alta e baixa de preços no mercado de ações brasileiro. Neste sentido, avalia-se as potencialidades da abordagem proposta para realizar previsões de curto e médio prazo sobre o comportamento do mercado acionário brasileiro.

Os experimentos serão conduzidos sobre uma base textual histórica que contém mais de 5000 notícias sobre o mercado do petróleo e de ações no Brasil. Os textos provêm das principais fontes de notícias brasileiras que publicam o seu conteúdo na internet, como Boletim Reuters, Correio Braziliense, Isto é Dinheiro, O Estado de SP, O Globo, Folha de SP e Valor Econômico. As cotações dos ativos serão obtidas do site da Bovespa, que mantém para consulta pública até dez anos de valores de abertura e fechamento de todos os papéis negociados pela entidade.

4. Cronograma de desenvolvimento

A Tabela 1 apresenta o cronograma de desenvolvimento do trabalho conforme a numeração das atividades abaixo:

1. Levantamento de dados históricos (documentos e cotações);
2. Pré-processamento dos dados históricos;
3. Estudo sobre as principais técnicas de mineração de textos;
4. Levantamento sobre abordagens de mineração de textos voltadas ao mercado de ações;
5. Testes preliminares com bibliotecas de mineração de dados;
6. Elaboração do artigo;
7. Apresentação do trabalho.

Referências

- [1] E. D. Amoroso. Um estudo de caso sobre mineração de dados como instrumento de aprendizado para o investidor do mercado de ações. Dissertação de mestrado, Universidade Católica de Brasília, Brasília, Fevereiro 2009. 1
- [2] J. Bollen, H. Mao, and X. Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1 – 8, 2011. 1
- [3] G. A. Deschatre and A. Majer. *Aprenda a investir com sucesso em ações: análise técnica e fundamentalista*. Ciência Moderna, 2006. 1
- [4] A. Mahajan, L. Dey, and S. M. Haque. Mining financial news for major events and their impacts on the market. In *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, pages 423–426, 2008. 1, 2

Tabela 1. Cronograma de atividades.

Atividade	7/16	8/16	9/16	10/16	11/16	12/16
1	x	x				
2		x	x	x		
3	x	x	x			
4	x	x	x	x		
5				x	x	
6			x	x	x	
7						x

- [5] C. Manning, P. Raghavan, and H. Schütze. *Introduction to information retrieval*. Cambridge University Press, New York, 2008. 1, 2
- [6] R. P. Schumaker and H. Chen. A discrete stock price prediction engine based on financial news. *IEEE Computer*, 43:51–56, 2010. 1