Paulo de Tarso Gomide Castro Silva

A System for Stock Market Forecasting and Simulation

DISSERTAÇÃO DE MESTRADO

DEPARTAMENTO DE INFORMÁTICA

Postgraduate Program in Informatics

Rio de Janeiro
March 2011

PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO DE JANEIRO

Paulo de Tarso Gomide Castro Silva

A System for Stock Market Forecasting and Simulation

Dissertação de Mestrado

Dissertation presented to the Postgraduate Program in Informatics of the Departamento de Informática, PUC-Rio as partial fulfillment of the requirements for the degree of Mestre em Informática

Advisor: Prof. Ruy Luiz Milidiú

Rio de Janeiro
March 2011

Paulo de Tarso Gomide Castro Silva

**A System for Stock Market
Forecasting and Simulation**

Dissertation presented to the Postgraduate Program
in Informatics of the Departamento de Informática,
PUC-Rio as partial fulfillment of the requirements for
the degree of Mestre em Informática

Prof. Ruy Luiz Milidiú
Advisor
Departamento de Informática — PUC–Rio

Prof. Carlos José Pereira de Lucena
Departamento de Informática – PUC-Rio

Prof. Eduardo Sany Laber
Departamento de Informática – PUC-Rio

Prof. José Eugenio Leal
Coordenador Setorial do Centro Técnico Científico —
PUC–Rio

Rio de Janeiro, March 25th 2011

Paulo de Tarso Gomide Castro Silva

Graduated in Computer Science by the Universidade Federal de Minas Gerais. His research is focused on Machine Learning and Stock Market Forecasting and Simulation.

To my parents, my brothers and Camilla.

Acknowledgements

To God, for everything.

To my parents, my brothers and my beloved Camilla, for their love, understanding and continuous support.

To the SI^2 team, especially to my eternal partners Leonardo Conegundes, Gabriel Lana and Mateus Lana, for their friendship and the endless talks that always lead me to the best solutions.

To my coaches Daniel Fleischman and Eduardo Cardoso, my teammates Caio Valentim, Daniel Marques, Guilherme de Napoli and Pedro Veras, as well as all my programming contests fellows, for the numerous problems solved, or not, making the beauty of computation even more explicit.

To my *WhileTrue* friends, for having accepted me as a friend, the infinite email threads and the long, and sometimes productive, discussions.

To my friends from Itaúna and Belo Horizonte, my roommates Michel Quintana and Gabriel Senra, as well as my neighbors Allan Rocha, Fernando del Carpio and Julio Daniel, for their indispensable friendship and company.

To my Professor and first Advisor, Marcus Poggi, for his fellowship and help in making important decisions.

To my UFMG Professors, Antônio Loureiro, Geraldo Robson Mateus and Rodolfo Resende, for making this master's degree possible.

To my Advisor, Ruy Milidiú, for trusting my work and encouraging me during my dissertation.

To Carlos Crestana, Eraldo Fernandes, Leandro Alvim, and the LEARN team, for their talks, help and suggestions.

To PUC-Rio, CAPES and CNPq, for their financial support.

Abstract

Gomide, Paulo; Milidiú, Ruy (Advisor). **A System for Stock Market Forecasting and Simulation**. Rio de Janeiro, 2011. 63p. MSc. Dissertation — Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

The interest of both investors and researchers in stock market behavior forecasting has increased throughout the recent years. Despite the wide number of publications examining this problem, accurately predicting future stock trends and developing business strategies capable of turning good predictions into profits are still great challenges. This is partly due to the nonlinearity and noise inherent to the stock market data source, and partly because benchmarking systems to assess the forecasting quality are not publicly available. Here, we perform time series forecasting aiming to guide the investor both into Pairs Trading and buy and sell operations. Furthermore, we explore two different forecasting periodicities. First, an interday forecast, which considers only daily data and whose goal is predict values referring to the current day. And second, the intraday approach, which aims to predict values referring to each trading hour of the current day and also takes advantage of the intraday data already known at prediction time. In both forecasting schemes, we use three regression tools as predictor algorithms, which are: Partial Least Squares Regression, Support Vector Regression and Artificial Neural Networks. We also propose a trading system as a better way to assess the forecasting quality. In the experiments, we examine assets of the most traded companies in the BM&FBOVESPA Stock Exchange, the world's third largest and official Brazilian Stock Exchange. The results for the three predictors are presented and compared to four benchmarks, as well as to the optimal solution. The difference in the forecasting quality, when considering either the forecasting error metrics or the trading system metrics, is remarkable. If we consider just the mean absolute percentage error, the proposed predictors do not show a significant superiority. Nevertheless, when considering the trading system evaluation, it shows really outstanding results. The yield in some cases amounts to an annual return on investment of more than 300%.

Keywords

Resumo

Gomide, Paulo; Milidiú, Ruy. **Um Sistema para Predição e Simulação do Mercado de Capitais**. Rio de Janeiro, 2011. 63p. Dissertação de Mestrado — Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Nos últimos anos, vem crescendo o interesse acerca da predição do comportamento do mercado de capitais, tanto por parte dos investidores quanto dos pesquisadores. Apesar do grande número de publicações tratando esse problema, predizer com eficiência futuras tendências e desenvolver estratégias de negociação capazes de traduzir boas predições em lucros são ainda grandes desafios. A dificuldade em realizar tais tarefas se deve tanto à não linearidade e grande volume de ruídos presentes nos dados do mercado, quanto à falta de sistemas que possam avaliar com propriedade a qualidade das predições realizadas. Nesse trabalho, são realizadas predições de séries temporais visando auxiliar o investidor tanto em operações de compra e venda, como em *Pairs Trading*. Além disso, as predições são feitas considerando duas diferentes periodicidades. Uma predição *interday*, que considera apenas dados diários e tem como objetivo a predição de valores referentes ao presente dia. E uma predição *intraday*, que visa predizer valores referentes a cada hora de negociação do dia atual e para isso considera também os dados *intraday* conhecidos até o momento que se deseja prever. Para ambas as tarefas propostas, foram testadas três ferramentas de predição, quais sejam, Regressão por Mínimos Quadrados Parciais, Regressão por Vetores de Suporte e Redes Neurais Artificiais. Com o intuito de melhor avaliar a qualidade das predições realizadas, é proposto ainda um *trading system*. Os testes foram realizados considerando ativos das companhias mais negociadas da BM&FBOVESPA, a bolsa de valores oficial do Brasil e terceira maior do mundo. Os resultados dos três preditores são apresentados e comparados a quatro *benchmarks*, bem como com a solução ótima. A diferença na qualidade de predição, considerando o erro de predição ou as métricas do *trading system*, são notáveis. Se quando analisado apenas o Erro Percentual Absoluto Médio os preditores propostos não mostram uma melhora significativa, quando as métricas do *trading system* são consideradas eles apresentam um resultado bem superior. O retorno anual do investimento em alguns casos atinge valor superior a 300%.

Palavras–chave

Aprendizado de Máquina; Predição de Séries Temporais; Regressão por Mínimos Quadrados Parciais; Regressão por Vetores de Suporte; Redes Neurais Artificiais; Mercado de Capitais; Trading System.

Contents

List of Figures

PUC-Rio - Certificação Digital Nº 0912857/CA

List of Tables

The belief of the scientist-dreamer has triumphed and will always triumph over the vulgar opportunism of the ambitious scientist without philosophical belief! Kelimet-Oul-Iah!

Malba Tahan, *Júlio César de Mello e Souza.*