

Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce Anastasia Serebryannikova
Název práce Predicting Stock Market Trends from News Articles
Rok odevzdání 2018
Studijní program Informatika **Studijní obor** Matematická lingvistika

Autor posudku Vladislav Kuboň
Pracoviště ÚFAL MFF UK

Role Vedoucí

Text posudku:

This diploma thesis attempts to solve a very difficult task - a prediction of stock market trends based upon a collection of news articles. The prediction concentrates on overall trend of the stock exchange as a whole, it does not attempt to predict the trends for individual shares. Even in such a general form, it is still a very challenging endeavor which has resisted several previous attempts to solve it (as it is documented in a rich bibliography of the thesis).

The core of the thesis consists of six main chapters, which sufficiently cover all steps of the investigation of the problem, starting with a short theoretical overview of it from an economical perspective, and ending with a discussion of main results achieved in an impressive number of experiments.

A very interesting material can be found already in the second chapter, in which the author describes previous work done in this field. The author did here a very good job, she has found out several articles on this topic and described their main achievements. The assessment of these articles is critical, the author successfully tried to discover some weak points of those articles, which gave her an inspiration how to proceed in her own experiments.

The data used for the experiments are described in the third chapter. In order to be able to compare her own results with the ones reported previously, the author has decided to use the same dataset. It consists mainly from financial news produced by Bloomberg and Reuters over the period of 7 years (2006-2013). The author describes the structure of data in detail, from several points of view. It is obvious that she tries to understand what kind of information may be hidden in the dataset and how to exploit these findings in the experiments.

The experiments are described in the two subsequent chapters, which describe the initial basic model and its extensions. Also these two chapters are written with a great care for detail. In her experiments, the author tries to discover the best observation window, i.e. the time interval when the news have a greatest impact on future stock market prices, as well as the usefulness of either closing or opening stock market prices. She also devotes a great deal of attention to the treatment of Monday data, which are obviously different than the data for the other days of the week (because Mondays follows after Saturday and Sunday, i.e. the two days without any activity on the stock market).

The last two chapters describe the results of all experiments. Both of them contain not only an overview of the results achieved, but also a thorough discussion of the reasons why the results are in some respect surprising. The author does not hide the fact that she actually did not

reached any substantial improvement on testing data (although the results on the development dataset looked promising), instead, she provides a good explanation of the possible reasons for this result. Taking into account a wide variety of experiments which the author has performed and described in the previous chapters of the thesis, it is very hard to say what actually went wrong. The whole thesis leads to one logical conclusion – the whole task has been probably too ambitious. It seems that even the papers mentioned in the second chapter, which reported certain improvements of the prediction, may have obtained those positive results only for very specific settings, which were not transferable to different datasets. In the light of this discovery it is necessary to judge the work described in the thesis positively for its completeness and thoroughness of the experiments, despite the lack of score improvements on testing data.

From the formal point of view, the thesis standard is also high, the text is easy to read, the language quality is high and the thesis meets all formal requirements for a diploma thesis. For these reasons I wholeheartedly recommend it for the defense.

Práci doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

Pokud práci navrhuje na zvláštní ocenění (cena děkana apod.), prosím uveďte zde stručné zdůvodnění (vzniklé publikace, významnost tématu, inovativnost práce apod.).

Datum 3.9.2018

Podpis