STUDY OF PRICE FORECASTING EFFECTIVENESS BASED ON NEWS ARTICLES

Polovyi V.O., Orekhov S.V., Malyhon G.V., Vincent Agalab
National Technical University «Kharkiv Polytechnic Institute», Kharkiv

Usually less successful technical analysis [1] tries to predict future prices based on past prices, whereas fundamental analysis tries to base predictions on factors in the real economy (e.g. inflation, trading volume, organizational changes in the company, demand for products or services offered by the company). As financial textual data (news articles) became available on the web, a new source of indicators appeared, which potentially could contain useful information for fundamental analysis. The objective of this project is to analyze and extract such information, and derive numerical indicators from financial text [2].

The first step in our research is to select a method for extracting keywords and creating specific keyword dictionaries. Second step is to select prediction method and its precision. We evaluate the outcome of our positive or negative prediction method using a keyword dictionary based on the overwhelming number of positive or negative keywords in this news item.

Our goal is to develop method and software prototype to predict stock price movements using news articles and investigate efficiency compared to conventional methods.

Common approaches to extracting keywords include manually assigning keywords based on the content of the article and the judgment of the authors. It takes a lot of time and effort, and can also be inaccurate in terms of choosing the right keywords. With the advent of natural language processing (NLP), keyword removal has turned into effective and efficient.

ARIMA method is used to predict selected values. The first step in applying ARIMA methodology is to check for stationary. "Stationary" implies that the series remains at a fairly constant level over time. If a trend exists, as in most economic or business applications, then your data is NOT stationary. The data should also show a constant variance in its fluctuations over time. This is easily seen with a series that is heavily seasonal and growing at a faster rate. In such a case, the ups and downs in the seasonality will become more dramatic over time. Without these stationary conditions being met, many of the calculations associated with the process cannot be computed.

References:
1. Lavrenko, V., Schmill, M., Lawrie, D., Ogilvie, P., Jensen, D., & Allan, J. (2000) Mining of Concurrent Text and Time Series. KDD-2000: Workshop on Text Mining.
2. Schumaker, Rob & Chen, Hsiu-chin. (2010). A Discrete Stock Price Prediction Engine Based on Financial News. IEEE Computer. 43. 51-56. 10.1109/MC.2010.2.