

Prediction and Analysis for Stock Market using Web-Mining

Rachana Pathak

Dept. of Computer Science & Engineering,
N.K.Orchid College of Engineering,
Solapur, Maharashtra, India

V.V. Bag

Dept. of Computer Science & Engineering,
N.K.Orchid College of Engineering,
Solapur, Maharashtra, India

ABSTRACT

The goal is to design a recommendation system so that we can provide investors with the latest updates and to encourage maximum investors to participate in online trading without the need for profound technical expertise.

Investors in Stock Market needs expert advice and are strongly influenced by what people in the throng think about. Stock reviews from online groups or troop are valuable to take investment decisions. In order to support investors for daily updates, there is a need to develop a system that allows investors to include the crowd's recommendations in making investment decisions and using it further to manage a portfolio.

Keywords

Recommendation System, Stock Market, Crawler, technical index, review index

1. INTRODUCTION

People are often panicked to invest in stock market. Proper recommendations are also needed if the investor has absolutely no knowledge about stock market. Even though various sites provide with different recommendations, still people doubt to invest. Along with recommendations, use of technical indicators will produce remarkable results and provide proper guideline to investors.

2. LITERATURE SURVEY

2.1 Existing System

Opinion mining also known as sentimental analysis refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials. Sentimental analysis is process of determining attitude of an individual towards any product or topic. Sentimental Analysis is about extracting the opinions or sentiments when given a piece of text. It involves various techniques that are part of Natural language Processing and Data mining. It provides great source of unstructured information especially opinions that may be useful to others, like companies and their competitors and other consumers. The World Wide Web is growing day-by-day not only in size but also in the types of services and contents provided. Individual users are participating more actively and are generating vast amount of new data. Internet provides excellent platforms for business and commercial purposes. The electronic shopping platforms allow consumers to make intelligent comparison and purchasing decision on consumer products. In addition to comparing product specifications as described on electronic catalogue for better purchasing decision, consumers also search for consumer reviews to get the best products that fulfil their requirements. Any organization needs to conduct surveys and collect

reviews, in order to improve their product quality. There are number of websites which deals with product reviews. All these reviews are nothing but the opinions of people all over the world about different products. Opinion mining is nothing but to analysing the reviews. A consumer review is the review written by the product owner or the user who has used that product and uses the services. Bing Liu, Mining Hu and Jun shengCheng [1] proposed an analysis system with a visual component to compare consumer opinions of different products. With a glance of its visualization, the user can clearly see the strengths and weakness of each product and what people think about that specific product. Mining Hu and Bing Liu [2] studied the problem of opinion summarization or customer reviews of products sold online. The task is to identify the features of the products that customers have expressed opinions and according to their views that they appear in the reviews. For each feature, identify how many customer reviews have positive or negative opinions. The specific reviews that express these opinions are attached to the feature. This helps the customers in browsing.

2.2 Problem Formulation

There are many technical indicators available in market but these technical indicators show only the graphical representation. Technical indicators are helpful in deciding companies to invest.

2.3 Conclusion from Literature Survey

We can derive from all available sources and currently available technologies that there are many applications available for sentimental analysis.

Sentimental Analysis is used at various phases and is useful for prediction based approaches.

With the help of sentimental analysis we can predict in stock markets. The daily updates required for the constant resulting of predictions.

There is an instant need for enhancing the current tools used for predictions.

For, more and more convenience to the investors we find they follow regular investors. Designing a system that makes easier predictions on stocks where expert advices are also indulged.

Statistical methods include various techniques such as the root mean square error, the mean absolute error and the mean squared prediction error, statistical indicators like the autocorrelation, the correlation coefficient, the mean absolute deviation, the squared correlation and the standard deviation. However, these methods give only average prediction in handling complex data.

Necessity of the proposed system: The proposed system is innovation of new system which provides aggregation of expert reviews and technical indicators. These reviews are important so that we can decide where the exact investment is to be done.

3. METHODOLOGY

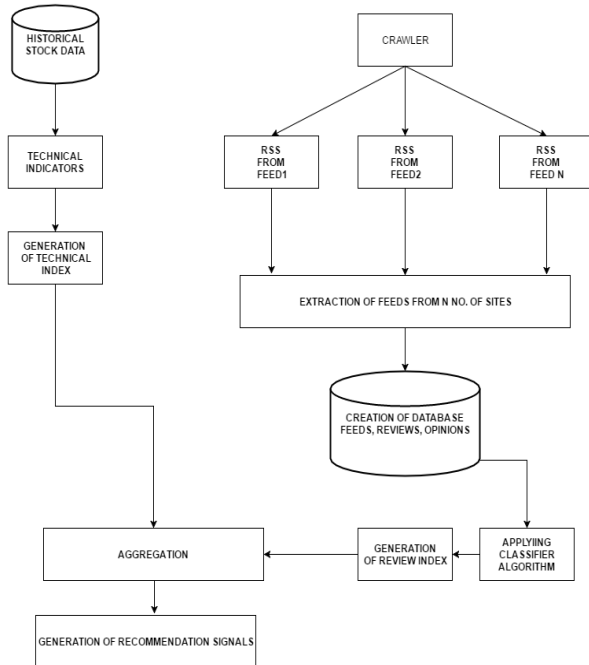


Fig Proposed System Architecture

3.1 Historical Stock Data

Historical stock data is the pre-existing data derived from analysis and study of previously occurring events. The requirement of these kinds of data is essential to analyse future trends.

3.2 Technical Indicators

Technical Indicator is a mathematical calculation based on historic price, volume or open interest information that aims to forecast financial market directions. These are used for predicting market trends.

3.3 Generation Of Technicial Index [Ti]

By applying some technical indicators a technical index is derived.

3.4 Crawler

The best way to extract data that is spread out across many pages of a site is by building a crawler. A crawler travels to every page of a site looking for other pages that match. Crawlers play a vital role when we require huge data from relevant unknown sites.

3.5 RSS Feeds from Sites

RSS stands for "Real Simple Syndication" which is specific type of file format that can be automatically read by News Readers that subscribers have installed on their computers or

access online. We require RSS Feeds as input.

3.6 Extraction of Feeds from 'N' Number of Sites

This will extract feeds from different sites.

3.7 Classification of Reviews

Review classification is necessary to identify type of review.

3.8 Applying Classifier Algorithm

At this stage, the algorithm works to classify the feeds, reviews and opinions to generate Review Index [RI]

3.9 Creation of Database (Feeds, Reviews, Opinions)

We need to maintain database to keep record of feeds, reviews and opinions. This will be further used as input.

3.10 Aggregation

AGGREGATION is the merging and finding average of two index values generated by technical index and review index.

3.11 Displaying Output

Output will be represented in tabular format depicting the best stocks for trading.

4. CONCLUSION

1. Designing a recommendation system that will help investors to figure out which recommendation holds the best.
2. The main motivation is to develop a prediction system that decides the best stocks.
3. This system can be used by investors as well as speculators to maximize their returns.

5. REFERENCES

- [1] Technische Universität Darmstadt, Hochschulstraße 1, 64289 Darmstadt, Germany ©2013 Elsevier B.V. All rights reserved.
- [2] Dept. of Computer Science, Stony Brook University, Stony Brook, NY 11794-4400, USA
- [3] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," Foundations and Trends in Information Retrieval, vol. 2, nos. 1–2, 2008, pp. 1–135.
- [4] <http://arxiv.org/ftp/arxiv/papers/1401/1401.2618.pdf>
- [5] B. Liu, M. Hu, and J. Cheng. opinionobserver: analyzing and comparing opinions on the web. In International Conference on World Wide Web, 2005
- [6] Philips S. Yu Ding Xiaowen, Liu Bing. A holistic lexicon-based approach to opinion mining. WSDM'08, 2008.
- [7] Bruce J. Vanstone, Gavin Finnie and Clarence Tan.pdf