Volume 4, Issue 10, October 2014 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Significance of Lingpipe Using Twitter's Tweets for Stock Market Prognosis

Mr. R. V. Argiddi *
Associate Professor,
Walchand Institute Technology,
Solapur, India

Ms. S. S. Apte
Professor,
Walchand Institute Technology,
Solapur, India

Mr. V. S. Adam
PG Scholar
Walchand Institute Technology,
Solapur, India

Abstract- ***Now a day's people are mostly connected with social network for exchanging their thoughts and ideas and its popularity is growing very rapidly. Sentiment analysis is booming its popularity in domain like advertisement, online shopping and all types of E commerce. Normally human being gather information related to entity which he wants to buy from friends and relatives. But now the time has changed it is an era of social networking were lot of opinions can be collected from social network sites like twitter this article discusses an approach were large no. of tweets are collected from twitter which are processed and classified based on their emotional content as positive, negative and neutral and analyze them .This result of analysis is used for concluding stock price . The tools used are Ling Pipe and Support Vector Machine, here we compare the processed data with yesterday's closing stock data and generate final result.***

Keywords— sentiment analysis; tweets; stock prediction;

I. INTRODUCTION

The domains like social networking, blogging and micro-blogging website are gaining popularity in day today life as they have vast information available. The information generated from this social network sites can be used for scientific surveys from a social or political point of view. So various companies use this way to extract information of their product and consider it as feedback or review of their product. Now a day's customer's are also making use of this concept on predicting of particular product and intelligently purchase them. In same way movie rating are also calculated and predicted now a days. As we all know that stock values can be predicted by querying the economic experts or share market brokers. But now trend has changed, in our India alone there are 168.7 crore million users who are presently active users of twitter. Obliviously lots of raw data in format of tweets are available with us just we have to make use of them.

Usually people relay on news for stock market updates but twitter tweets are more faster than new because if we follow the important people account like important decision makers of particular company are any financial firm the direct message is received to our account and therefore large no. of tweets are collected and lot of emotions can be extracted out of it for eg: online chat activity of any social network we can determine people are talking of which book and predict the sales of book in future..

II. RELATED WORK

Prediction in stock market is always a hot topic for domain expertise in data mining. Work done by many researchers is close to accurate prediction of share price in stock market. Neural Network, Genetic Algorithm, Association, Decision Tree and Fuzzy systems are widely used to predict stock prices. As above mentioned techniques if we see from implementation point of view they are very easy and output oriented.'

In the work carried out by Johan Bollen [1], have shown that emotions can deeply influence individual behaviour and decision making. Here analysis of moods collected from large no. of twitter feeds which are linked with DJIA over time .Two mood tracking tools are used for analyzing the moods collected from twitter Opinion Finder is used to measure the moods like positive vs. negative mood were as Google-Profile of Mood is used to check moods in 6 dimensions (Calm, Alert, Sure, Vital, Kind, and Happy). At last final results are produced by granger causality and self organizing fuzzy neural network which uses historic twitter data.

Another related work done by Rowan Chakoumakos[2] here author implements predictive classifier that unit economic analysis of stocks with features based on natural language processing of twitter comments linked to each stock in a specified portfolio to enable options straddling stock trading strategies. Identical SVM model built on this combination set of features showed an average improvement of 3.5 percentages.

In another paper related work was done by BalakrishnanGokulakrishnan[3] here authors discusses about an unique classification and pre-processing approach were stream of tweets are collected from twitter which are further classified into three categories positive, negative and neutral ,based on precision value of this three categories here comparison is done with various classification algorithm.

© 2014, IJARCSSE All Rights Reserved

III. BACKGROUND

A. Twitter API

Twitter has its own API (application program interface) the REST (Representational State Transfer) architecture .This architecture can be defined as view point of network design that defines assets, way to address and access data.

REST architecture is nothing but a set of network design viewpoint that defines assets and ways to address and access data. This architecture is assumed designed logical model and not as physical model were servers, computers and other resources are arranged and there data flow is shown. For Twitter, a REST architecture works with mostly integration format, it integrates from various web sources and display data to us. And this integration format is nothing but web syndication.

Web syndication is a good and simple concept. The main purpose of application is to take data as input process it and send this processed output to various modules on web. Few Web integration processes are used on web which is named as Syndication formats. Twitter is well-suited with two of them Really Simple Syndication (RSS) and Atom Syndication Format (Atom). These 2 formats are mainly used for data recover purpose from one module and send it to another. As above mentioned both tools have very few lines of code they are well suited for twitter. Web admin use this code to develop their site and visitors are always welcomed to this syndication service , they feed and receive update every time when admin checks website .This well-known feature is used by twitter to allow the various users of its to post tweets to another users. (As a result twitter users subscribe to other users which results into third party developer partial access to its API, especially twitter has special programs that describes about twitter special services. Oblivious corporation application has desktop feed reader programmers that allow users to post and retrieve messages in twitter's network using simple, independent interface.

E.g. Outwit's windows application which permits users to access Twitter through the Outlook e-mail program

Tweet Scan: It is an application which allows users to search public Twitter posts in real time using either a customized search engine or Firefox's search box

B. Ling Pipe

Ling pipe is newly developed tool which usually does linguistic analysis based on human language with the help of java libraries Ling Pipe is a software library for natural language processing implemented in Java. This book explains the tools that are available in Ling Pipe and provides examples of how they can be used to build natural language processing (NLP) applications for multiple languages and genres, and for many kinds of applications. Ling Pipe's application programming interface (API) is tailored to abstract over low-level implementation details to enable components such as tokenizes, feature extractors, or classifiers to be swapped in a plug-and-play fashion. LingPipe contains a mixture of heuristic rule-based components and statistical components, often implementing the same interfaces, such as chunking or tokenization

Eg. Incredible India! Output: positive

C. SVM (Support Vector Machines)

In machine learning, support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. Consider example from given training set , here example are divided into categories .Firstly SVM training algorithms generate model which assigns new examples into one category which leads to non probabilistic binary linear classifier .An SVM model is used for representation of example as point in space which are mapped so that examples of separate categories are divided in crystal clear format. New examples are then mapped into same space and prediction is done based on categories which they belong

IV. METHODLOGY

Many researchers have shown variety of ways of calculating the sentiments from collected tweet feeds they used OpenNlp, Prediction IO etc and concluded the stock price. Here goal is to combine the power of various sentiment analysis tools and retain the accuracy as much as possible

A. Data Collection

As we all know that data plays very important role .Processing is not possible without help of data. Here input raw data is nothing but tweet feeds collected from twitter .Directly we cannot access it but twitter is providing some tools and API's which include various methods and functions so that user may collect the tweets .After collecting the tweets we store them into database and from data base we send it for processing.

E.g. Streaming API

B. Pre Processing

Basic principle here is to (keep track of what bloggers say about brands like TATA which are further classified into positive, negative and neutral categories. The important idea is to use Ling pipe classification framework to do classification task were separation of subjective and objective sentences are done. Second is movie reviews are also analyzed separating positive, negative and neutral one.

© 2014, IJARCSSE All Rights Reserved

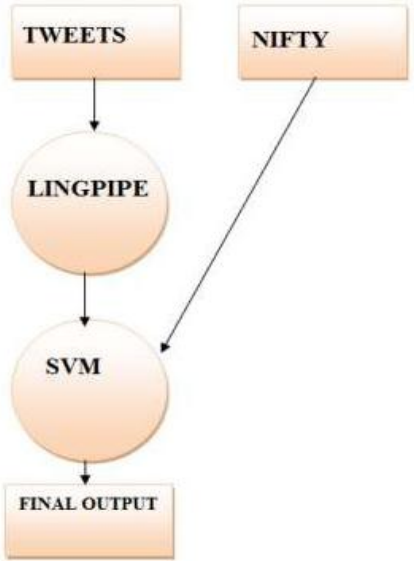
C. Data conversion

It part is very important here output from tweet classifier are combined with latest update values from stock index and further passed to SVM were exact value in percentage is obtained to us .

Sentiment	Query	Tweet
Positive	Infy	Vishal Sikka: infy will be the best IT company
Neutral	Mumbai	Ram: Week end @ Mumbai
Negative	Fsl	Jhon:fsl results are bad
Pos itive	India	Narendra Modi : India has won Achhe Din alewale hai
Neutral	nano	Ratna Tata:Nano project in Gujarat

Explanation of above tweets and their interpretation:

- Here CEO of Infosys is confidently telling that overall performance of company will be best so we can recommend buying it.
- Ram is telling that he will spend his weekend at Mumbai which has no impact on others life so consider it as neutral.
- John is explaining that First source solution company's quarterly results are bad so we can conclude to immediately sell them.
- Narendra Modi is expressing the joy by twitting that INDIA has won and good days will be coming.
- Ratan Tata is giving update news that Nano project will in installed in Gujarat which is bad news for other states but good news for Gujarat.



Above diagram states over all working of concept .It starts from tweets database processing them through lingpipe and further through SVM them compare with latest update values of stock index values and generate output.

V. CONCLUSION

Our aim is to generate closest and accurate prediction of stock values and this can be achived with the help of Ling pipe, SVM processing capabilities which helps in improving the accuracy than previous researches because here we have utilized the combined power of both mentioned tools.

REFRENCES

[1] Johan Bollen, Huina Maoa, & Xiaojun Zeng (2011)"Twitter mood predicts the stock market." Journal of Computational Science, Volume 2, Issue 1.

[2] Rowan Chakoumakos, Stephen Trusheim, Vikas Yendluri “Automated Market Sentiment Analysis of Twitter for Options Trading”

[3] Balakrishnan Gokulakrishnan , Pavalanathan Priyanthan , ThiruchittampalamRagavan ,Nadarajah Prasath, AShehan Perera “Opinion Mining and Sentiment Analysis on a Twitter Data Stream”

[4] R V Argiddi, S S Apte “A Study of Association Rule Mining in Fragmented Item-Sets for Prediction of Transactions Outcome in Stock Trading Systems”

[5] Lean Yu , Huanhuan Chen , Shouyang Wang , Kin Keung Lai “Evolving Least Squares Support Vector Machines for Stock Market Trend Mining “

© 2014, IJARCSSE All Rights Reserved

[6] Dattatray P.Gandhmal,Ranjeetsingh B. Parihar,Rajesh V. Argiddi "An Optimized Approach to Analyze Stock market using Data Mining Technique"
[7] Pham, Hung and Chien, Andrew and Lim, Youngwhan. "A Framework for Stock Prediction." 2009.
[8] Potts, Christopher. "Sentiment Symposium Tutorial." 2011.
[9] Manning, Christopher D, and Hinrich Schütze. "Foundations of Statistical Natural Language Processing".
[10] [10]Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten (2009); "The WEKA Data Mining Software"
[11] McNair, Douglas and Maurice, Lorr and Droppelman, Leo." Profile of mood states." Educational and Industrial Testing Service, San Diego, CA. 1971.
[12] Yang J., Leskovec J." Temporal Variation in Online Media". ACM International Conference on Web Search and Data Mining (WSDM '11), 2011.
[13] Zhang, Wenbin, AND Skiena, Steven. "Trading Strategies to Exploit Blog and News Sentiment" Internation
[14] Andranik Tumasjan, Timm O. Sprenger, Philipp G. Sandner, Isabell M. Welpes" Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment"

© 2014, IJARCSSE All Rights Reserved