ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 4, March 2017

A Survey on Social Media Analytics in Big data

S. Geetha
Assistant Professor
Department of Computer Science
Muthurangam Govt. Arts College(Autonomous)
Vellore- 632001
Email: geeth_s20@yahoo.com

Dr. R. Kaniezhil
Assistant professor
Department Of Computer Science
Bharathiyar College of Arts & Science for Women,
Attur-636105
Email:Kaniezhil@yahoo.co.in

Abstract

Social networks have greatly influenced the usage pattern of Internet by common man. It has paved way for the great proliferation of Internet into the life of even the ordinary people. Such an exponential rise in the use of the world wide web due to social networks started to fetch huge volume of data across diverse domains in short period of time. These characteristics by which the huge amounts of social network data are generated make them to categorize as Big Data. Stock prediction has recently grown to be a huge research area in the field of predictive analytics, big data analytics and statistical analysis. The field of stock prediction has used machine learning techniques as well as recently predictive analytics to predict stock prices. This paper gives a brief description of big data analytics and stock prediction and its methodologies.. There are varieties of social networking websites are available in www. Main objective of this survey is to how people are connected with social networks and how TV media are utilizing the social networks to rate the programs and channel.

Keywords: Social network, Big data, Stock prediction, social networks analytics, social networks with TV media

Introduction

Big data is the term for a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. The challenges include capture, curation, storage, search, sharing, transfer, analysis, and visualization [1]. It can updated every second with tera bytes of data with various information source like social media networks, scientific instruments, mobile devices, sensor technology and networks. For example considered a Market analyzers are keeping on viewing the transaction and interactions of the online users to make new trends in business.

The onset of Big Data Analytics has created a revolutionary change in the way businesses and people perceive data. [1, 2] Big data has proved too costly for rival organizations in the way the decisions have been undertaken. There exists a common misconception among the society that any huge volume of data can be classified under

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 4, March 2017

big data. But the reality is that data generated by systems which fulfills the constraint imposed in the dimensions of Volume, Velocity, Variety and Veracity can be classified as big data[3] as shown in Figure 1 When any system or organization generates data at a higher rate with diverse attributes and that data which when used to make learned decisions will fall under the hood of big data. The processing of big data is different from the way how conventional data is handled. Big data processing requires certain special techniques and tools to accomplish effective analytical decisions. The proliferation of Internet in the nook and corner of the globe has made the presence of big data felt everywhere. This made the data scientists to use these big datasets towards provisioning informed decisions. Organizations that generate big data, utilize the diverse tools and techniques to carry out sentiment analysis, predictive modeling, semantic web analysis to improve the Return on Investment (RoI). This massive buzzword "big data" is used by organizations to understand and target customers, optimize business processes and performance quantization in various domains such as health care, hospitality industry, education and learning systems, automobile, industrial automation systems, agriculture etc [4]. As a small step towards extending the footprint of the applications of big data, this paper tries to depict the machine learning techniques to perform Social network analytics that may provide a 360 degree insight into the social network data.

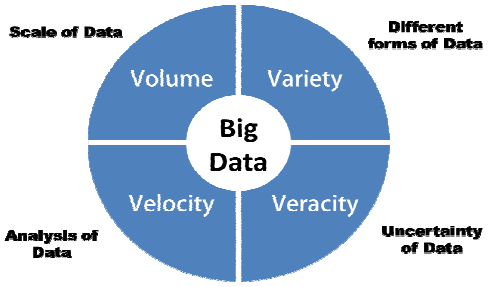

Figure 1: Attributes of Big Data

Data mining is the method to compute patterns and correlations within a large amount of data. This may involve methods such as machine learning, artificial intelligence and statistical analysis. Data mining has been used to find patterns such as shooting percentage of a basketball player or to optimize the lineup of the team [1]. Predictive analysis is the process of collection of huge amounts of data to find the underlying trends and patterns in the data to make scientific decisions in the future. There have been recent advancements in stock prediction by predictive analytics of social media [1, 2] and at the same time it has been shown that machine learning techniques can be successfully used to predict stock prices [3].Stock prediction is a key research interest in recent times and there have been many improvements in the prediction methodologies of stock. Stocks represent the claim in a company's assets.

Stock prediction is the science of determining the future value of a stock. The basic methodologies of stock prediction are classified into fundamental analysis and technical analysis. Fundamental analysis does not take into account the previous time series data and rather takes into

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 4, March 2017

account the intrinsic worth of the stock such as the earning potential, the portfolio of the company and it is in other words like a survey an analyst would do. Technical analysis on the other hand uses previous data of stock, which can be taken on a daily, weekly or monthly basis, and this time series data is analyzed to predict the future price of the stock. The random walk hypothesis on the other hand states that past data cannot be used to forecast the future data because it simply is independent of each other although it has been successfully predicted in the recent times, which our paper explores.

The time series data can be analyzed in methods such as Auto Regression Moving Average (ARMA) and Auto Regression Integrated Moving Average (ARIMA). Another architecture used is Artificial Neural Networks (ANN). ANN has an input layer and many hidden layers may be present between the input and the output layer where the final hidden layer presents the output to the output layer. The neural network is often trained repeatedly many times until the output of the network correlates with the target output required.

Social Networks with TV Media

Even though TV Medias are connected in social network by creating the separate blog for their channel interact with people. It is a means of interface to communicate with different variety of people. Social networking is connects the people together online. Ads and massive TV shows are consecutively watched on YouTube and even commented. It helps the people by way of getting comments, ideas and making their programs popular. By the recommendations of the viewers, rate the programs if it is appreciated it rates their channel higher otherwise have to rectify the problem of programs. It provides the opportunity to know the competitor channel program rating. It act as discussion boards to know the public or any person opinions easy based on likes and dislikes.

Social Networks with Marketing

By analyzing the social connections, the enterprise will probably discover that the existing organization structure is partially bypassed so that the individuals can work with others that are not part of their group or department. The enterprise may be able use the connection information to reorganize their internal organization structure to reflect the social network connections and thereby improve enterprise productivity. If you share your blog post in email or on Twitter, Facebook and other social channels to Check the Availability of Your Business Name on Social Networks. We can track the content and analyze the blog in variety of channels (Email, Facebook, GooglePlus, Linkedin, and Twitter) for viewing the information and monitor user visitor's activity. It also provides the option to find the competitors ranks.

Sentimental Analyze of Social Networks

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 4, March 2017

Opinions are central to almost all human activities because they are key influencers of our behaviors. Whenever we need to make a decision, we want to know others' opinions. In the real world, businesses and organizations always want to find consumer or public opinions about their products and services. Individual consumers also want to know the opinions of existing users of a product before purchasing it, and others' opinions about political candidates before making a voting decision in a political election. In the past, when an individual needed opinions, he/she asked friends and family. When an organization or a business needed public or consumer opinions, it conducted surveys, opinion polls, and focus groups. Acquiring public and consumer opinions has long been a huge business itself for marketing, public relations, and political campaign companies.

Machine Learning

Machine learning techniques, as implied by the term, is the process of inculcating knowledge to any machine like, PC, laptop or mobile devices to learn about a system with a set of input /dependent variables and the desired output. Any machine can perform learning under three modes. They are Supervised, Unsupervised and Reinforcement learning techniques [24]. Normally, machine learning techniques are employed in any system to carryout and produce results as part of predictive analytics and forecasting methods. Any machine learning techniques will be classified under the categories of Decision tree based, linear and logistic regression based and neural network based. Many organizations have kick started to utilize the impact of social media data in the decision making process. When social media data is utilized for such a critical decision making, it becomes necessary to process the huge datasets obtained from social networks using machine learning techniques.

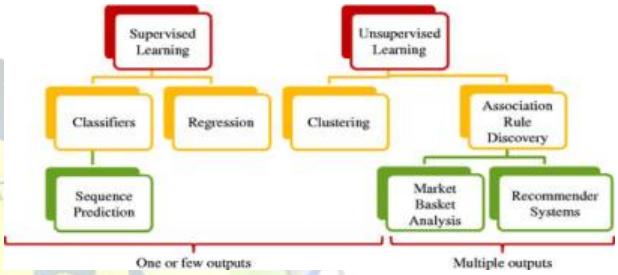

Figure 2: Algorithmic Approaches to Machine Learning

Machine learning algorithms that have played a major role in social media analysis include Decision tree learning, Naïve Bayes, Nearest Neighbor classifier, Maximum Entropy method, Support vector machine(SVM), Dynamic Language Model classifier, linear regression and logistic regression, Simple logistic classifier, Bayes Net and Multilayer Perceptron.

Stock prediction

A linear regression method to predict stock prices. Regression method is used to predict a numerical value by obtaining the past values on either daily or monthly basis, which consists of the parameters such as stock value, the opening price, closing price, lowest and highest price of the day along with the adjustments if any. To

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 4, March 2017

perform this, they calculated the correlation between the independent variables. By the comparison of their results and the actual stock values, they obtained a similarity of 61.35%

A combination of financial news as well as stock price quotes to predict the stock prices. The financial news articles, which are in text format are analyzed for the specific keywords. They specified that the bag of words approach is the easiest to implement and at the same time, it was found to be the least effective.They have obtained the proper nouns from the text in news articles to perform the analysis. With the text and the stock quotes, they built a machine-learning algorithm with support vector regression. They predict the stock price after 20 minutes. They gathered about 9,211 news articles. They only took into considerations the articles that were published during the time when the stock market was open. They obtained a result of 8.5 % return on trades. They point out that their successful prediction was mainly due to the analysis financial news article analysis. Hence, the research in the direction of financial text mining has proved to be promising.

Artificial neural networks used to forecast stock prices. Artificial intelligence method uses learning agents where the agent learns the patterns of the past events. In unsupervised learning, the agent learns these past patterns without feedback of any person where as in supervised learning; the agent is given the data with appropriate inputs and outputs to learn from. Artificial neural network is an example of such artificial intelligence based learning models. They proved that it is possible to predict stock using artificial intelligence based learning agents

Big Data & Social Network Analytics

Big data analytics can be done with the software tools commonly used as part of advanced analytics disciplines such as predictive analytics and data mining. But the unstructured data sources used for big data analytics may not fit in traditional data warehouses. The technologies associated with big data analytics include NoSQL databases, Hadoop and MapReduce. NoSQL database, also called Not Only SQL. A NoSQL database provides a mechanism for storage and retrieval of data that is modeled in means other than the tabular relations used in relational databases.

Social networks such as Face book, Twitter, LinkedIn etc pave way for generation of huge amount of diverse data in short period of time. Such social media data require the application of big data analytics to produce meaningful information to both information consumers and data generators. The impact of different big data analytics tools and techniques over processing social network data will be discussed in detail in this section of this paper.

Big data analytics techniques and tool types include all of the following such as Predictive analytics, data mining, statistical analysis, complex SQL, data visualization, artificial intelligence and

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 4, March 2017

natural language processing. The analysis of structured and unstructured data from social networks leads to social network analytics. Even blogs, micro blogs and wikis contribute to social network analytics data sets. Though there are various sources of information available in social media, we are largely concerned about the user generated contents such as sentiments, images, videos and bookmarks and interactive relationships between people, organizations and products. These two classes of information is utilized in various big data analytics tool such as Hadoop and Map Reduce Framework, Apache Pig, Apache Hive, Jaql, NoSQL etc. When user posted information is used in the analytics approach, it is called as content based analytics and when relationships between entities is used for analytics, it is known as structure based analytics. Social networks consist of millions of connected objects and analysis of data from those objects is computationally intensive and expensive. Hence there are two different approaches that shall be followed.

In parallelization approach, focus is towards dividing a huge data set into smaller sub sets and utilize the computational power through cloud computing to process the data in a parallel manner. Map Reduce and Pregel from google are pioneer in this approach. However, lots of open source initiatives in the form of Hadoop are gaining popularity in the social network analytics. Spark and Hama are also registering their market share in the research of social network data.

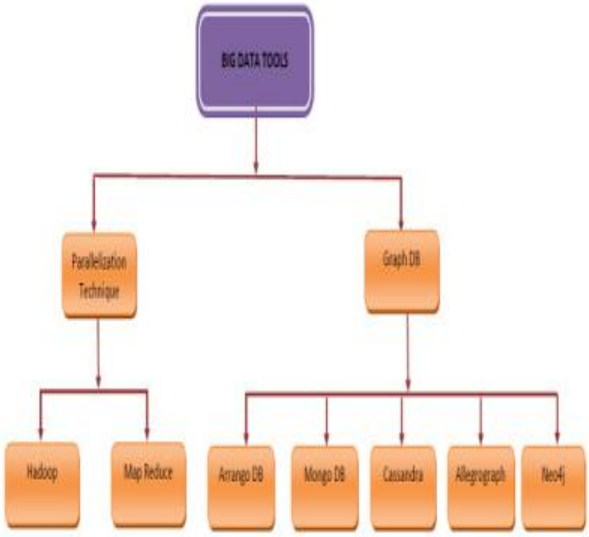

Figure 3: Categories of Big Data Tools

Everything is connected: people, information, events and places, all the more so with the advent of online social media. Social network refers to interaction among people in which exchange and / or create information Comments and likes given to those will earn more followers but in the same way, if any negatives are shared about the person it leads to hatred towards the personality. Social network otherwise we can say a user friend. It act as discuss board to find whether their thought, ideas or whatever that might is right or wrong at the same time there is a chance to convey messages, get knowledge awareness, finding new things by images, content, videos with sound. It is a not one to one communication it is a many to many communication that can be discuss or chatted with lot member of users who joined in this link There are various social networking sites are available such as Facebook, twitter, Linkedin.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 4, March 2017

The enterprise, with a much smaller number of internal social network users than the public social networks like Face book, is still generating big data, and enterprises should therefore consider what to collect and how to use the collected data to improve their internal operations. Analyzing internal social networks can provide business intelligence. Google Analytics is a great tool to monitor the performance of your blog, but you may not look at it regularly enough. One way to get around this is to set up a weekly report for delivery to you via email.

Conclusion

Thus the survey paper depicts about a survey of big data in social networking and media interaction with variety of aspects. It can update massively in the form of big data at every second. Finally coming to the conclusion that social networks equally having both merits as well as demerits. This paper addressed the importance of big data and machine learning in the first section along with the growing importance to analyze growing social media data utilizing big data and machine learning algorithms. The next section discussed about the various parallelization frameworks in the form of Hadoop and Map Reduce along with different Graph databases in the context of social network analytics. Various machine learning algorithms applied in the field of social data analysis with pros and cons is dealt with in the third section of this paper. At the same time, there have been many improvements to mine the data and analyze it accurately obtained from textual sources such as social media sites like Twitter and financial news articles. It is not just that the sentiment analysis using Twitter and news articles are performed with respect to the positive and negative word list but also with respect to litigious, superfluous words as well which provides an added measure of social media analytics in the field of stock prediction. We further note that there is a lot of research still going on to increase the accuracy of stock prediction in both sentiment analysis as well as technical analysis with different methodologies used to improve the accuracy. Hence, the research aims to make it easier to predict stock by taking into account not only the technical analysis but also the social media analytics as well as fundamental analysis and hybridized approaches, which include all the three methods to provide the maximum positive returns on trading.

References

[1] http://en.wikipedia.org

[2] Bing Liu," Sentiment Analysis and Opinion Mining" April 22, 2012

[3] V. Bharanipriya , V.Kamakshi Prasad "Web Content Mining Tools:acomparative study" January-June 2011, Volume4, No. 1, pp. 211-215

[4] Bamshad Mobasher, Robert Cooley, and Jaideep Srivastava. Creating adaptive web sites through usage-based clustering of urls. In Knowledge and Data Engineering Workshop, 1999.

[5] Bernardo Huberman, Peter Pirolli, James Pitkow, andRajan Kukose. Strong regularities in world wide web surfing. Technical report, Xerox PARC, 1998.

[6].Tania Cerquitelli, Silvia Chiusano, MirkoKampf, "Special Session on Big Data: New Trends and Applications", BigData 2013.

[7] Charu. C. Aggarwal, "Social Network Data Analytics", IBM Watson Research Center, USA [8] Ostrowski D. A, "Map Reduce Design Patterns for Social networking Analysis" Page nos. 316-319, IEEE International conference on Semantic Computing, 2014

[9] Guojin Liu, Ming zhang, Fei Yan, " Large Scale social network analysis based on MapReduce", International conference on computational aspects of social networks, Page Nos. 487-490, 2010

[10] Nan Gong, "Using Map-Reduce for large Analysis of Graph Based Data", M. S. Thesis, 2011.

[11] Pradeepa. A, Dr. Antony. S, "Significant Trends of Big Data Analytics in Social Network", In Journal of Advanced Research in Computer Science and Software Engineering, Vol-3, Issue 8, August 2013.

[12] Arinto Murdopo, "Distributed Decision Tree Learning for mining Big Data Streams", M. S. Thesis, University Politenica de Catalunya, Barcelona Tech, July 2013.

[13] Mirko Kampf, " Hadoop in Social Network Analysis-Overview on tools and some best practices", Gridka School, Karlstuhe, 2013.

[14] Boyi Xie, Rebecca J. Passonneau, Leon Wu, and Germ´ an G. Creamer. "Semantic frames to predict stock price movement" In Proc. of ACL, pages 873–883, 2013.

[15] Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. "Using structured events to predict stock price movement: An empirical investigation" In Proc. Of EMNLP, pages 1415–1425, Doha, Qatar, October 2014. Association for Computational Linguistics.