

# Role Of Various Data Mining Techniques Used For Social Media Data

Smruti Smaraki Sarangi

Department of Computer Science and Engineering, Kalinga University, Naya Raipur, Chhattisgarh.

---

## ABSTRACT

Now a day, social media plays a vital role towards communication like by interaction with friends, families and with strangers also through internet connection. The various social network sites are Facebook, Twitter, LinkedIn, Google+ etc. are very much popular and easily affordable. So social media is considered as an online communication system that connects people with each other throughout the world and participate actively in conversations. The social media not only used for communication but also used to grow the businesses spectrum through viral marketing. The social network phenomenon grows rapidly amongst the people in form of gathering information, news, opinions etc.

For detecting useful information from large datasets like patterns, rules or trends, data mining technique is applied on social media, which is considered as social network analysis. By applying data mining techniques to social media sites to make the unstructured data to structured one and to make the information better and to use the data for analysis, research and business purposes. In this paper we discuss the various data mining techniques and its impact on social media network.

**KEYWORDS:** Social Media, Social Network, Data Mining, Predictive and Descriptive Mining,

## INTRODUCTION

The analysis of social media and social network is used to find the pattern through the use of networks and graph theory. The social media data mining is the process that extracts the valuable information. The main goal is to collect various data form different sources of social media to find structured information that may be hidden. By applying different data mining techniques to social media to get better understanding on a certain concept and gain new information.

Data mining is helpful to understand large data sets. It consists of two approaches that is supervised and unsupervised that provides various algorithms to identify the patterns which are hidden in the data sets. The supervised approach depends on the inferred knowledge of previous data and unsupervised approaches characterized the data by classification method and convert them to similar elements automatically.

The drastically emerged concept of social network analysis is a significant technique in modern society which is implemented in various fields like biology, social psychology, political science, geography,

history, communication studies, information science, organization studies, development studies , business intelligence and computer science etc.

To use social media effectively various companies use the concept of social media marketing to promote a product or service and learn to allow consumers and internet users to post user generated content and use various data for analysis and improve the business strategies. The various social media websites have their own built-in data analytics tool that enables various organizations to keep track of their progress, success and advertisement campaigns. (ZafaraniR., et.al, 2014)

## **LITERATURE REVIEW**

A.L. Kavanaugh stated that the structural representation of the data of social media is unarranged and is shown in various forms like textual message, audio messages, pictures, and video messages. (Kavanaugh A.L., et.al, 2012)

Researcher Liu stated that social media is a study of viewpoint, attitude, reactions, computation, approaches, appraisal, impact, surveillance comes in textual message, survey, records, conversation, bulletins, assertion, receptions, or some other documents. (Liu, B., 2011)

Researcher Aggarwal stated that social media can be referred to as system of individual communication where communication behaves as a relationship between vertices and edges. The vertices include entities and the relationships between vertices are established through edges.(Aggarwal, C., 2009)

The researchers Chen, Chiang and Storey stated that the social media provides an excessive amount of prolonged real time data due to which traditional statistics methods are not suitable for the analysis of such huge data. So, the data mining techniques play a crucial role to overcome from the problem. (Chen, 2012).

As per the researcher Borgatti, social network is a graph that consists of vertices and edges that is used to represent social relation on social networks.(Borgatti, S. P., 2009)

As per the researchers Chen and its team, social network is responsible to refine a network service that allows users to create public and semi-public profiles.(Chen, Z. S., et.al. 2009).

As per Aggarwal, different data mining techniques are currently used to analyze viewpoint and attitudes shown on social media.(Aggarwal, C., 2009)

The researcher said that data mining technique needs enormous data sets for mining notable patterns from data; social network sites emerged as perfect for doing mining with various data mining tools (Cortizo, J.et.al., 2009).

Here the authors have studied the techniques that are currently used to analyze social media (Adedoyin-Olowe, M. et.al, 2014). In this paper the analysis of social media data has been proved to be effective, due to the capacity possessed by data mining in handling unstructured and dynamic data.

Here the author provides the analysis of available data mining techniques to mine social network data. According to author, the research will focus on the content mining where lot of human behaviour patterns

can be identified by analysing the social network profile pages and also hybrid approach by combining social network analysis which would be more useful. (FernandoS.G.S. et.al, 2014)

Here, the author has studied on the data mining techniques that are currently used to analyze social media data. Hence different kinds of data mining techniques are available in field for different applications. (ZatariT., 2015). Every data mining technique has its own strengths and limitations.

The authors have provided the summary that social network data analysis, business and management were the most active domains that require to mine the social media and the techniques used are SVM, BN, and DT. (Mohammad Noor Injadat et.al. 2016)

Here, the authors have presented a systematic data mining technique to mine intellectual knowledge from social data. (RahmanM. M. 2012).

### **VARIOUS DATA MINING TECHNIQUES USED FOR SOCIAL MEDIA**

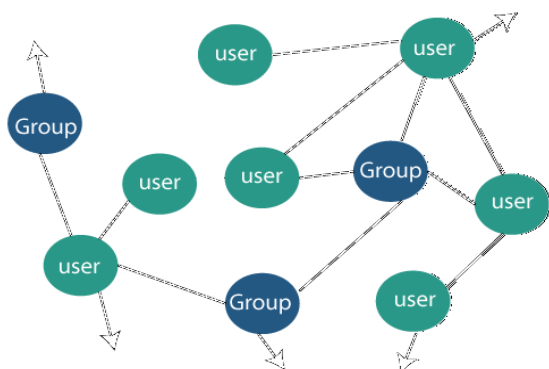
The data mining method on social media become an essential strategy which understands the recent patterns, culture and business. This is due to that social media contains the ocean of data where in various forms data are present like tweets, instagram posts and blog articles.



**Fig 1: Common Social Media**

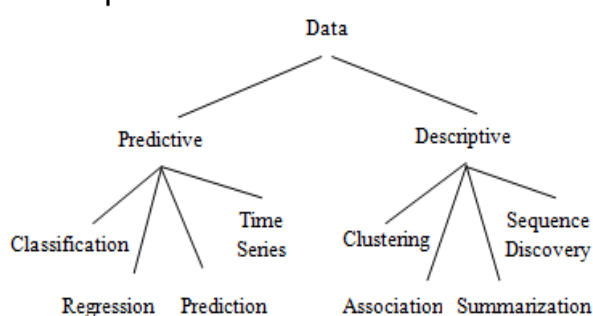
The social media data have three qualities that are: size, noise and dynamism. The data mining technique is applied on social media that make the business intelligence processes easier.

Data mining in social media analyses the raw social data and efficiently extracts patterns or trends. For example: usage of social media, determine online buying behaviours, sharing the contents, individuals connections etc. Data mining techniques are used to collect related information from unstructured data and helpful to achieve specific objectives.



**Fig 2: Hypothetical graph structure diagram for typical social media platforms**

The different data mining techniques are converted into data mining model that may be descriptive or predictive.



**Fig 3: Data Mining Model**

### DESCRIPTIVE MINING

The descriptive method provides the latest information based on past or recent events after mining the data. It produces correlation, cross tabulation, frequency etc. which are determined to find the regularities in data and reveal patterns. It focuses on summarization and conversion of the data into meaningful information.

This task describes the characteristics of data in target information. This technique is more precise and accurate. It responds to just the situations. The operations performed are reporting, query drill-down and Adhoc reporting which generate the response towards the questions what to be happened, what is the exact problem and what is the frequency of the problem.

This model includes various techniques that are: association, clustering, sequence discover and summarization.

### ASSOCIATION

In data mining, association is the most basic technique, where machine learning model is used. It is a simple correlation between same type of items to identify the patterns, analyze the data and identify the co-occurrences in databases.

### **CLUSTERING**

Cluster means diving of data objects into sub-classes. Clustering depends on the method of data segmentation on large data groups which are divided by their similarity. Clustering is the grouping of specific objects based on their characteristics and similarities. It is a unsupervised learning method.

### **SEQUENCES DISCOVER**

This technique is used to identify trends or regular occurrences of similar events. It is used to find statistical relevant patterns between data examples where values are delivered in sequential pattern. It is a special case of structured data mining. For example, with customer data, customer's behaviour can be identified.

### **SUMMARIZATION**

Summarization means abstraction of data. It is the process of shortening a set of data computationally to create a summary that represents most relevant information with in the original content.

### **PREDICTIVE MINING**

The predictive method predicts future results instead of current behaviour. It is a supervised learning function that is used for prediction of target value. The predictive method is used to allow the data miner to predict the future value of a specific variable to target one.

This task carries the induction over current and past data on which predictions can be made. It controls over the situation and respond to it.

The operations performed are predictive modelling, forecasting, simulation and alerts and generate the response towards the questions what will happen next, what is the outcome if these trends continue and what actions are required to be taken.

This model includes various techniques that are: classification, prediction, regression and time series.

### **CLASSIFICATION**

Classification technique is a complicated one which is used in various fields. It collects various attributes from a data set and combines them into visible categories. Here classification is done based on supervised learning mechanism. The various types of classification methods are: the decision tree induction, Bayesian networks, and k-nearest neighbour classifier etc.

### **PREDICTION**

This technique is used to analyze past instances and predict the future event like the component failure or to identify fraud to predict organization profits. It is the relationship between independent variables and predicts the relationship between dependent and independent variable.

In social media this technique is used to project the kind of data we want to see in future. This technique covers the predictor and predicted variables from the data and also involves the recognition of trends and patterns and generates predictions.

### **REGRESSION**

This method is used to predict a range of numeric values or continuous value given in a particular data set. It is used for mapping a data item to a real valued prediction variable. Here, target values are known.

The simple form of regression is linear regression that finds the relationship between two variables by using a straight line function  $y = mx + b$ .

The multiple regressions is used to find the relationship between multiple variables. The various types of multiple regressions are: standard multiple regressions, stepwise multiple regression, hierarchical regression and set-wise regression.

### **TIME SERIES**

This method uses the statistical technique that explains time-dependent series of data sets that generates prediction for future event that based on past event.

## **VARIOUS ANALYSIS METHOD USED IN SOCIAL NETWORK**

### **SENTIMENT ANALYSIS**

It is considered as opinion analysis, where cloud contains opinions about new products, reactions to an event, popularity of a politician or celebrity. This analysis is used to extract words or phrases from text to determine positive, negative and neutral text. This analysis is helpful for monitoring analysing the popularity of the social media.

### **MARKET TREND ANALYSIS**

It is used to analyze the consumer behaviour and current trends and passion. It is a vital one that connects with the customer and their choices and based on that companies establish their brands and make it popular.

This analysis also helpful to track the brand or product of any organization and also helpful to determine the competition present in market.

### **PREDICTIVE ANALYSIS**

This analysis used the past data and predicts the future trends. It uses the historical data to build a model to capture the necessary patterns and used the new data to predict future developments.

## **CHALLENGES IN SOCIAL MEDIA DATA MINING**

The various important challenges in social media data mining are:

### **BIG DATA**

Big Data technique is widely used technology that handles and process huge set of data. It enables to grow social media to increase the processing power and storage capacity without any downtime. It is useful to handle unlimited data and unlimited power of processing for social media.

This technique influences the growth of social media through implementing various technologies, storage, processing and analysing power to the developers.

The four ways through which Big Data put impact on social media are: personalization, decision making, effectiveness of campaign and product insights.

### **SUFFICIENCY**

This is a problem called as under fitting, which arises due to most social media networks restrict the amount of information that can be accessed in a time period. Therefore, sometimes data is not sufficient enough to generate patterns or recommendations.

### **REMOVAL OF NOISE**

In social media data, a large amount of data is always expected but to remove the noise effectively from the social data is a tricky concept. Noise is a subjective quantity which produces more error in pattern recognition.

### **EVALUATION QUANDARY**

Due to the large size of data present in social media, it is difficult to achieve a proper dataset for supervised learning mechanism and also difficult to judge the accuracy of the data set, so for evaluation unsupervised learning mechanism or clustering method is applied on data sets.

### **ISSUES IN SOCIAL MEDIA DATA MINING**

The different issues or disadvantages present in social media data mining are:

#### **PRIVACY**

It means if a user's social connection is accessed then the private information which is included with in the social connection can affect the privacy of the user. So, there is a chance to get into the confidential information which can later might be misused.

#### **SECURITY**

Security is one of the most important concerns, when individuals are usually working with large amount of information by applying social media data mining and can access the data easily but it became very dangerous if data is not used in a secure way. This alerts about data vulnerability to the user that leads to serious security issues.

#### **INITIAL COST**

This is one of the drawbacks of social media data mining. It includes hardware cost and software cost. Hardware cost like machine and its components and software cost like various software used to extract data from social media and the agreement cost etc.

#### **INCORRECT INFORMATION**

This is the main drawback of data mining system. If information is inaccurate then the result is also not exact. Using inaccurate information in social media data mining produces inaccurate result.

## CONCLUSION

Social media data mining uses a range of basic concepts from various field like computer science, data mining, machine learning and statistics. It is based on the methodology from social network analysis and network science. The various data mining technique used in social network analysis that may be supervised or unsupervised methods. By analysing social media data, especially the sentiments of users of social media became effective and useful in this paper. The capacity of data mining techniques handles noisy, large and dynamic data and useful to mine the sentiments of online users of social media.

## REFERENCES

1. Adedoyin-Olowe, M., Gaber, M. M., & Stahl F., "A Survey of Data Mining Techniques for Social Media Analysis" in *Journal of Data Mining & Digital Humanities*, PP:1-27, 2014.
2. Aggarwal, C.: *An introduction to social network data analytics*. Springer US, 2011.
3. Borgatti, S P.: "2-Mode concepts in social network analysis." *Encyclopaedia of Complexity and System Science*, 8279-8291, 2009.
4. Chen, Chiang R.H.L., Storey V.C., *Business Intelligence and Analytics: From Big Data to Big Impact*, *Mis Q.* 36 (2012) 1165– 1188.
5. Chen, Z. S., Kalashnikov, D. V. and Mehrotra, S. Exploiting context analysis for combining multiple entity resolution systems. In *Proceedings of the 2009 ACM International Conference on Management of Data (SIGMOD'09)*, 2009.
6. Cortizo J., Carrero F., Gomez J., Monsalve B., Puertas E.: *Introduction to Mining SM*. In: *Proceedings of the 1st International Workshop on Mining SM*, 1 – 3, 2009.
7. Fernando S.G.S. et.al "Empirical Analysis of Data Mining Techniques for Social Network Websites" in *COMPUSOFT, An international journal of advanced computer technology*, Volume-III, Issue-II PP:582-592, 2014.
8. Kavanaugh A.L., Fox E. A., Sheetz S.D., Yang S., Li L.T., Shoemaker D.J., et al., *Social media use by government: From the routine to the critical*, *Gov. Inf. Q.* 29 (2012) 480–491. doi:10.1016/j.giq.2012.06.002. [2] H.
9. Liu, B.: *Sentiment Analysis and Opinion Mining*. AAAI-2011, San Francisco, USA, 2011.
10. Mohammad Noor Injadat, Fadi Salo, Ali Bou Nassif "Data Mining Techniques in Social Media: A Survey" *NEUCOM17295*, Volume 214, PP:654-670, 2016.
11. Rahman, M. M, "Mining Social Data to Extract Intellectual Knowledge", in *International Journal of Intelligent Systems and Applications(IJISA)*, vol.4, no.10, PP:15-24, 2012
12. Zafarani R., Abbasi M. A., Liu H., "Social Media Mining", Cambridge University Press, 2014
13. Zatarit., "Data Mining in Social Media" in *International Journal of Scientific &*
14. *Engineering Research*, ISSN 2229-5518  
Volume 6, Issue 7, PP:152-154, 2015.