

A Survey: Data Mining Techniques for Social Media Analysis

¹Elangovan D, ²Dr.Subedha V, ³Sathishkumar R, ⁴ Ambeth kumar V D

¹Research scholar, Department of Computer Science Engineering,
Sathyabama University, India

²Professor, Department of Computer Science Engineering,
Panimalar Institute of Technology, Chennai, India

³Assistant Professor, ⁴Associate Professor, Department of Computer Science Engineering,
Panimalar Engineering College, Chennai, India

elangovan.durai@yahoo.com, subedha@gmail.com, satz.lic@gmail.com, vdambethkumar@gmail.com

Abstract—Data mining is the extraction of present information from high volume of data sets, it's a modern technology. The main intention of the mining is to extract the information from a large no of data set and convert it into a reasonable structure for further use. The social media websites like Facebook, twitter, instagram enclosed the billions of unrefined raw data. The various techniques in data mining process after analyzing the raw data, new information can be obtained. Since this data is active and unstructured, conventional data mining techniques may not be suitable. This survey paper mainly focuses on various data mining techniques used and challenges that arise while using it. The survey of various work done in the field of social network analysis mainly focuses on future trends in research.

Index Terms— Data Mining, Social media, Social network Analysis, Web Mining.

I. INTRODUCTION

Data mining is a powerful tool which will facilitate to seek out hidden patterns and various relationship between the data. Data processing discovers hidden facts from massive

databases. The overall objective of the data mining technique is to extract information from a huge data set and transform it into a comprehensible structure for more use. The different data Mining techniques are

- I. Characterization.
- II. Classification.
- III. Regression.
- IV. Association.
- V. Clustering.
- VI. Change Detection.
- VII. Deviation Detection.
- VIII. Link Analysis.
- IX. Sequential Pattern Mining.

Social network finds its application in several business activities like Co-innovation, Customer service, General promoting, increasing spoken promoting, marketing research, plan generation and new development, publicity, worker communication and reputation management.

II. LITERATURE SURVEY

| S.NO | AUTHOR | PAPERS TITLE | ALGORITHM / TOOLS | ADVANTAGES/LIMITATIONS |
|------|--|--|------------------------------|--|
| 1 | Marcelo Maia, et.al | Identifying User Behavior In Online Social Networks | Clustering Algorithm | <ul style="list-style-type: none"> ▪ For various group of users, dissimilar interaction patterns that are not similar can be experiential. ▪ Characterize and identify the user profiles in online social networks. ▪ Display more suitable advertisements based on user behavior |
| 2 | Mohamad Al-Fayoumi, Souyma Banerjee, et.al | Analysis Of Social Network Using Cleaver Ant Colony Metaphor | Cleavage Ant Colony Metaphor | <ul style="list-style-type: none"> • Using maximum group and sub grouping criteria social network structure are clustered. • Imposing performance on test run using the standard group sequence data particularly for the clustering of the instance named as Keller 6. |

| | | | | |
|----|---|---|--|--|
| 3 | David Ediger Karl Jiang, et.al | Massive Social Network Analysis: Mining Twitter For Social Good | Graph Ct Snag Algorithm | <ul style="list-style-type: none"> • To perform analysis on public data from a micro blogging network like twitter. • focus of analysts more on a much smaller data subset. |
| 4 | A.SemanBozkur et.al | Identification Of User Pattern In Social Networks By Datamining Techniques | Decision Tree Algorithm | <ul style="list-style-type: none"> • most precise results |
| 5 | StefaoBaccianella et.al | SENTIWORDNET 3.0: An Enhanced Lexical Resource For Sentiment Analyses Of Option Mining | SENTIWORD-NET 3.0 | <ul style="list-style-type: none"> • relative improvement of 19.48% for the ranking by positivity and a relative improvement of 21.96% for the ranking by negativity |
| 6 | Alexander Pak et.al | Twitter As A Corpus For Sentiment Analyses And Potation Mining | Opinion Mining And Sentiment Analysis | <ul style="list-style-type: none"> • Large data set • Improved needed increasing the amount of the training data. |
| 7 | RojalinaPriyadershini et.al | Functional Analysis Of Artificial Neural Network For Dataset Classification | Back Propagation Analysis | <ul style="list-style-type: none"> • Highly effective tool for classification. • Combination of trail, learned |
| 8 | E.Raju, Sravanthi | Analyses Of Social Networks Using The TechniqueMove Web Mining | Web miningTechnique | <ul style="list-style-type: none"> • while using the web mining for social Medias for analysis, data sampling is a huge issue. • Filtering the data is difficult. • Communities Overlapping. |
| 9 | G.Vinodhini, Rm.Chandrasekaran | Sentiment Analyses And Opinion Mining A Survey | Sentiment Classification | <ul style="list-style-type: none"> • Sentiment detection used in the different applications such as identify and classified reviews, summarize the review and other real time methods and applications . |
| 10 | Xi Ling, Wenjing Yin, et.al | Churn Analyses Of Online Social Network Uses Using Mining Technique | Decision Tree Classification And K Means Algorithm | <ul style="list-style-type: none"> • Analyze the potential mix of users. • To process huge total of data in a timely manner , decision tree-based and a k- means algorithms are used. |
| 11 | Rupam Some | A Survey On Social Network And Future Trends | Social Network Models Statically And For Analysis | <ul style="list-style-type: none"> • Link analysis and deep and dark networking Is recent trends in social networks |
| 12 | Paridhi Jain, PonnurangamKumara guru, et.al | Identifying Users Across Multiple Social Networks | Finding Nemo | <ul style="list-style-type: none"> • Identify the spammers and nemo while using this algorithm |
| 13 | Neha Mehata, Manta Kathuria et.al | Comparison Of Conventional And Fuzzy Clustering Techniques A Survey | Web documentClusteri ng | <ul style="list-style-type: none"> • More efficient in information finding process • Fuzzy clustering are suitable for handling the issues related to understandability of patterns, incomplete/noisy data, • mixed media in rank and human |

| | | | | |
|----|--|---|---|--|
| | | | | communication, and can provide approximate data |
| 14 | FarhatAroochi | Neuro Fuzzy Approach To Data Clustering A Framework For Analyses | Noro Fuzzy Clustering Analysis | <ul style="list-style-type: none"> • automatically generate the patterns |
| 15 | Mita , K.Dalal, Mukesh et.al | Atomic Classification Of Unstructured Blog Structure Text | Naïve Bayesian Model Artificial Neutral Network Model | <ul style="list-style-type: none"> • it's an domain dependent • Unstructured blog text classification |
| 16 | G.Nandi,A.Das | A Survey On Using Datamining Techniques For Online Social Network Analyses | Graph Theory | <ul style="list-style-type: none"> • with solid basis in graph theory , analysis of data is done |
| 17 | Mita K Dalal, Mukesh et.al | Semi supervised Learning Based OptimizedOpinionSummarization And Classification For Online Product Review | Semi Supervised Approach | <ul style="list-style-type: none"> • To identify opinionated sentences beginning unstructured consumer reviews with great success and categorize their orientation with acceptable accuracy |
| 18 | Asad Bukhari, et.al | Critical Review On Sentiment Analyses Technique | Sentiment Analysis Technique | <ul style="list-style-type: none"> • New robotic mechanisms are needed to help in drilling from downward to top micro posts. • Improves the sentiment analysis classifiers that are able to enhance and analyze the user sentiments |
| 19 | S.G.S Fernando | Empirical Analysis Technique For Data Mining Technique For Social Network Websites | Markov Model | <ul style="list-style-type: none"> • By combining social network analysis with content mining using Amalgam approach would be more functional . • The statistical methods like Markova models can be adopted to resolve the temporal behavior of web data. |
| 20 | Sanjeev Dhawan, Kulvinder Sing, et.al | Critical Analyses Of Social Networks With Web Data Mining | Web mining Technique | <ul style="list-style-type: none"> • Identify and Verdict communities in social networks structure, penetrating patterns in social networks and investigative overlapping communities. • online social networking websites like online photo albums, comments and blogs. |
| 21 | Sanjeev Dhawan, Kulvinder Singh, et.al | Emotion Mining Technique In Social Networking Sites | Sentiment Analysis Technique | <ul style="list-style-type: none"> • Sentimental analysis includes managing customer dealings, emotions of human ,information retrieval, natural text-to-speech system, social and literary analysis. |
| 22 | M.Vedanayaki | A Study Of Data Mining And Social Network Analysis | Knowledge Based Network Analysis | <ul style="list-style-type: none"> • Focus to find global structural patterns. • tedious to collect data |

| | | | | |
|----|---|--|---|---|
| 23 | MdAnsarulHaque, TamjidRahaman | Sentiment Analyses By Using Fuzzy Logic | Sentiment Analysis Technique | <ul style="list-style-type: none"> • Focus and analyze the extracted opinions from posted in social networks like facebook and twitter etc.. |
| 24 | Mariam Adedoyin- Olewe, et.al | A Survey Of Datamining Technique For Social Network Analyses | Tram | <ul style="list-style-type: none"> • Conducted on social network analysis using the tram technique. |
| 25 | Anais Collomb, CrinaCostea, et.al | A Survey And Comparison Of Sentiment Analyses Methods For Reputation Evaluation | Sentiment Analysis Technique | <ul style="list-style-type: none"> • using the sentimental analysis techniques , models target a simple global classification. |
| 26 | Meenu Sharma | Clustering In Defaming A Breve Review | Neuro And Fuzzy Logic Approaches | <ul style="list-style-type: none"> • Neuro and fuzzy logic approaches helps to improve the clustering review |
| 27 | Y.K. MathurAbhayaNand | Soft Computing Technique And Its Impact In Data Mining | Soft Computing Techniques | <ul style="list-style-type: none"> • Soft computing methodologies- solve data mining problems. • Discover the no pattern in the large databases. |
| 28 | Esmail FakhimiGheshlagh et.al | Datamining Technique For Web Mining : A Review | Web Mining | <ul style="list-style-type: none"> • Dynamic behavior of data. • Hybrid approach |
| 29 | Zahra ZamaniAlavijeh, et.al | The Application Of Link miningIn Social Network Analysis | Link Mining | <ul style="list-style-type: none"> • links among the data instances |
| 30 | Pooja Rohill, Ochin Sharma | Web Content Mining : A Implementation In Social Websites | Web Mining | <ul style="list-style-type: none"> • Used to extract relevant information from many websites likeonline shopping sites. • Data is extracted as per the user's criteria using web mining |
| 31 | ShilpaRadhakrishana n | A Survey In Text Filtering In Online Social Networks | Machine Learning Based Approach | <ul style="list-style-type: none"> • for text classification SVM algorithm is mostly used. • SVM algorithm has many features in social media analysis. |
| 32 | Gauri Joshi, Mr.SamadhanSonawa ne | Filter And Classification Of User And Media Data Using Metric And NativeBased Method | Naïve Bayes MultilevelClassifi er Algorithm | <ul style="list-style-type: none"> • To examining social medium statistics. |
| 33 | Hilal Ahmed Khanday, et.al | Exploring Different Aspects Of Social Network Analyses Using Web Mining Techniques | Datamining Technique | <ul style="list-style-type: none"> • An Exploratory research is carried by comparing adjacent matrices and lists with incidence matrices using graphs obtained from online social networks as data input |
| 34 | Pooja Sikka | Datamining Of Social Networks Using Clustering Based SVM | K-Means Clustering Based SVM | <ul style="list-style-type: none"> • SVM is not suitable for mining using large data sets |

| | | | | |
|----|--|--|---|---|
| 35 | FarisKateb, JugalKalita | Classifying Shot Tecta In Social Media: Twitter As Case Study | Text- Classification Technique | <ul style="list-style-type: none"> • Challenge in the classification of text in social media is that the data is streamed |
| 36 | G.Basiashavili, T.Bliadze, et.al | Classification Of Adaptive Neural Application Of Adaptive Neural Networks For The Filtration Of Spam | Multilayer Neural Technique | <ul style="list-style-type: none"> • learns to distinguish the required and redundant and unnecessary e- mails from one another |
| 37 | RetuMevali,AjithSingh, et.al | Opinion Mining Techniques On Social Media Data | OpinionMiming | <ul style="list-style-type: none"> • take out precious thing of knowledge |
| 38 | `RemyaRs, SmithaEs | Text Categorization Using Data Mining Technique On Social Media Data | Naïve Bayes Multilevel Classification | <ul style="list-style-type: none"> • At the same time it permitted one comment to fall into multiple categories. • learning analytics, instructive data mining, and education related data mining are the benefits. |
| 39 | KanikaMathur | Online Social Network Mining | Naïve Bayes Text Classification | <ul style="list-style-type: none"> • when the textual information is not structured as per the grammatical gathering, it becomes tough. |
| 40 | Rajeev Mathur | Advanced Neuro- Fuzzy Approach For Social Media Mining Methods Using Cloud | Hybrid Noro Fuzzy | <ul style="list-style-type: none"> • Hybrid Noro Fuzzy helps to solve different data mining problems |
| 41 | R. Adikka,Dr .Shaik et.al | A Survey On Data Mining Techniques For Analysis Of Social Network | Datamining Technique | <ul style="list-style-type: none"> • Discuss the various data mining techniques. |
| 42 | Thai Le Phillip Pardo, et.al | Application Of Artificial Neural Network In Social Media Data Analysis: A Case Of Lodging Business In Philadelphia | Anson | Helps to the social analysis |
| 43 | LopamudraDey, Sanjay Chakroborty | Sentiment Analysis Of Review Datasets Using Naïve Bayes' And K- NN Classifier | K-Nearest Neighbor And Native Bayes | <ul style="list-style-type: none"> • Give better results • For hotel reviews these algorithms not suitable |
| 44 | MuktaPatkar, Pooja Pawar Money Sing, et.al | A New Way For Semi Supervised Learning Based On Data Mining For Product Reviews | Semi Supervised Approach | <ul style="list-style-type: none"> • Sentiment analysis of product reviews gives optimistic and unenthusiastic reviews • Gives unbiased and constructive opinion |
| 45 | Dr.S.P.Victor, Mr.M.Xavier Rex | Analytical Implementation Of Web Structure Mining Using Data Analysis In | Web Structure Mining | <ul style="list-style-type: none"> • To identify the specified urn structure content analysis . |

| | | Educational Domain | | |
|----|--|--|--|---|
| 46 | Hemant Kumar Soni, Snjiv Sharma, et.al | Association Rule Mining: A Data Profiling And Prospective Approach | Association Rule Mining | • Displays and describes the association rule mining . |
| 47 | Heling Jiang, An Yang, Fengyun Yan , et.al | Research On Pattern Analysis And Data Classification Methodology For Data Mining And Knowledge Discovery | Pattern Analyses And Data Classification Metrology | • Using the knowledge of machine learning methodology and graph theory to analyze the organizational structure of network graphical pattern |
| 48 | Saravanan Suba And T.Chistophar | An Efficient Data Mining Method To Find Frequent Item Sets In Large Database Using TR- FCT | TR-FCTM | • For an item sets, direct frequency count and total frequency count are found |
| 49 | V.A Chakkarwar Amurta A Joshi | Semantic Web Mining Using RDF Data | Semantic Web Mining | • Extraction of number of pages is minimized by ranking technique |
| 50 | Kuldeep Sing Rathore, et.al | A Review On Web Usage Mining For Web Personalization Using Clustering Techniques | Clustering Technique | • Find hidden information. |
| 51 | Santhosh C.Pawar, et.al | Research Issues And Future Directions In Web Mining: A Survey | Web Mining | • Semi and unstructured data in challenging task |

III. CONCLUSION

This research work provides a lot of present appraisal and updates of social media network analysis .In this Literature works are reviewed supported dissimilar aspects of social networks. This work helps to studies the relevancy of the techniques and idea of web mining for social networks analysis, and reviews the connected literature concerning web mining and social networks. Data mining have several challenges during this analysis field to be resolve with improvement.

REFERENCES

- [1] Sankar K. Pal, et.al "Web Mining in Soft Computing Framework:, IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 13, NO. 5, SEPTEMBER 2002.
- [2] Tomoyuki NANN0, et.al "Automatically Collecting, Monitoring, and Mining Japanese Weblogs",.
- [3] Ralph Gross et.al "Information Revelation and Privacy in Online Social Networks ACM Workshop on Privacy in the Electronic Society (WPES), 2005.
- [4] Huan Liu et.al "Toward Integrating Feature Selection Algorithms for Classification and Clustering", IEEE Transactions 2005.
- [5] Andrea Esuli_ et.al, "SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining", Proceedings of the 5th Conference, 2006.
- [6] Marcelo Maia, et.al "Identifying User Behavior in Online Social Networks", SocialNets'08, April 1, 2008
- [7] Ai Ho, Abdou Maiga, et.al "Privacy Protection Issues in Social Networking Sites 2009
- [8] L. Dey et.al "Opinion mining from noisy text data," 2009
- [9] Mohammad Al-Fayoumi, Soumya Banerjee, Jr., et.al "Analysis of Social Network Using Clever Ant Colony Metaphor", PROCEEDINGS OF WASET 2009
- [10] David Ediger Karl Jiang et.al "Massive Social Network Analysis: Mining Twitter for Social Good", ICPP 2010
- [11] Rojalina Priyadarshini; "Functional Analysis of Artificial Neural Network for Dataset Classification IJCCT August 2010.
- [12] Selman Bozkır1 , et.al Identification of User Patterns in Social Networks by Data Mining Techniques: IMCW 2010.
- [13] Alexander Pak et.al "Twitter as a Corpus for Sentiment Analysis and Opinion mining", Proceedings of the LREC 2010.
- [14] S. Baccianella, et.al enhanced lexical resource for sentiment analysis and opinion mining," in Proceedings of LREC '10 2010
- [15] Sitaram Asur et.al " Predicting the Future With Social Media" Mar 2010.
- [16] Kuan-Yu Lin et.al, "Why people use social networking sites: An empirical study integrating network externalities and motivation theory", Computers in Human Behavior 27 (2011).
- [17] M. Taboada et.al "Lexicon-based methods for sentiment analysis .IEEE 2011
- [18] E.Raju K.Sravanthi, "Analysis of Social Networks Using the Techniques of Web Mining" IJARCCSE Volume 2, Issue 10, October 2012.
- [19] G. Vinodhini et.al, "Sentiment Analysis and Opinion Mining: A Survey", IJARCSSE, Volume2, Issue 6, June 2012.

- [20] Fatemeh Khoshnood et.al “Designing A Recommender System Based On Social Networks And Location Based Services”, IJMIT 2012.
- [21] Xi Long, Wenjing Yin, et.al “Churn Analysis of Online Social Network Users Using Data Mining Techniques”, IJEMS 2012.
- [22] Simona Vinerean et.al “The Effects of Social Media Marketing on Online Consumer Behavior”, IJBM Vol. 8, No. 14; 2013 ISSN 1833-3850.
- [23] Rupam Some, “A Survey on Social Network Analysis and its Future Trends”, (IJARCCE) Vol. 2, Issue 6, June 2013.
- [24] M. K. Dalale et.al, “Automatic classification of unstructured blog text,” IJILSA, pp. 108–114, 2013.
- [25] Paridhi Jainy et.al “Identifying Users across Multiple Online Social Networks”, IWWW 2013
- [26] Farhat Roohi, “Neuro Fuzzy Approach To Data Clustering: A Framework For Analysis”, EJS ISSN: 1857 – 7881 .
- [27] G Nandi, A Das, “A Survey on Using Data Mining Techniques for Online Social Network Analysis”, 2013.
- [28] Mita K. Dalal et.al “Semisupervised Learning Based Opinion Summarization and Classification for Online Product Reviews”, Applied Computational Intelligence and Soft Computing Volume 2013.
- [29] S.G.S Fernando, “Empirical Analysis of Data Mining Techniques for Social Network Websites”, COMPUSOFT, IACT 2013
- [30] Anaïs Collomb , et.al , “A Study and Comparison of Sentiment Analysis Methods for Reputation Evaluation ,2014.
- [31] Sanjeev Dhawan et.al “Critical Analysis of with Web Data Mining”, IJITKM Special Issue (ICFTEM2014) May 2014
- [32] M. Vedanayaki, “A Study of Data Mining and Social Network Analysis”, Indian Journal of Science and Technology, 2014.
- [33] Mariam Adedoyin-Olowe et.al “A Survey of Data Mining Techniques for Social Network Analysis”, JMDH , 2014.
- [34] Anu Sharma Dr. M.K Sharma et.al “CRITICAL REVIEW OF SENTIMENT ANALYSIS TECHNIQUES”, Proceeding of the ICAICS 2014
- [35] Meenu Sharma, “Clustering In Data Mining : A Brief Review”, (IJCEM) Volume 1, Issue 5, August 2014.
- [36] Md. Ansarul Haque1, et.al “SENTIMENT ANALYSIS BY USING FUZZY LOGIC”, (IJCEIT), 2014.
- [37] Y. K. Mathur, et.al “Soft Computing Techniques and its Impact in Data Mining – IJETAE August 2014).
- [38] Sanjeev Dhawan, et.al “Emotion Mining Techniques in Social Networking Sites”, IJICT (2014).
- [39] Haseena Rahmath P “Fuzzy based Sentiment Analysis of Online Product Reviews using Machine Learning Techniques IJCA 2015
- [40] Esmaeil Fakhimi et.al “Data Mining Techniques for Web Mining: A Review 2015
- [41] Zahra Zamani et.al “The Application of Link Mining in Social Network Analysis”, ACSIJ May 2015.
- [42] Thabit Zatari, “Data Mining in Social Media”, International Journal of Scientific & Engineering Research, Volume 6, Issue 7, July-2015.
- [43] Zohreh Madhoushi, et.al “Sentiment Analysis Techniques in Recent Works”, (SAI), 2015. IEEE, 2015.
- [44] G. Besiashvili, et.al “Application of Adaptive Neural Networks for the Filtration of Spam”, EPiC Series in Computer Science Volume 36, 2015, Pages 42–50 GCAI 2015.
- [45] Smita Bhanap, et.al “Data Mining for Business Intelligence in Social Network: A survey”, December 2015.
- [46] Literature Review and Challenges of Data Mining Techniques for Social 2015
- [47] Pooja Rohilla, Ochin Sharma, “Web Content Mining: An Implementation on Social Websites”, IJARCCCE, July 2015.
- [48] Shilpa Radhakrishnan, “A Survey on Text Filtering in Online Social Networks”, (IJCSIT) 2015
- [49] Gauri Joshiet.al “Filtering and Classification Of User Based On Social Media Data Using Memetic and Naïve Bayes Methods”, (IJCSIT) 2015
- [50] Hilal Ahmad Khanday, et.al “Exploring Different Aspects of Social Network Analysis Using Web Mining Techniques”, (IJETTCS) April 2015.
- [51] Pooja Sikka, “Data Mining Of Social Networks Using Clustering Based Svm”, Ijrrst APRIL 2015.
- [52] Faris Kateb, Jugal Kalita, “Classifying Short Text in Social Media: Twitter as Case Study”, IJCA February 2015.
- [53] Ritu, Ajit Singh et.al, “Opinion Mining Techniques on Social Media Data”, IJCA 2 015.
- [54] Remya R S, Smitha E S, “Text Categorization using Data Mining Technique on Social Media Data”, [57] (IJARET Dec. 2015)
- [55] saravanan et.al “An Efficient Data Mining Method To Find Frequent Item Sets In Large Database Using TR- FCTM ICTACT 2016.
- [56] V. A. Chakkarwar et.al “Semantic Web Mining using RDF Data”, IJCA Volume 133 – No.10, January 2016.
- [57] Kuldeep Singh Rathore, et.al, “A Review on Web Usage Mining For Web Personalization Using Clustering Techniques”, IJARCCSER Volume 6, Issue 5, May 2016.
- [58] Santosh C et.al “Research Issues and Future Directions in Web Mining: A Survey”, IJCA 2016.