An Enhanced Classification technique for Unstructured Data Analysis

D.Vivek
Assistant Professor,Dept of IT
Nandha Engineering College
Erode.
Vivekprasanth87@gmail.com

V.Saranya
UG Scholar, Dept of IT
Nandha Engineering College
Erode
saisaran448@gmail.com

S.Sangavi
UG Scholar, Dept of IT
Nandha Engineering College
Erode
sangavisaran11@gmail.com

D.Smitha
UG Scholar, Dept of IT
Nandha Engineering College
Erode
smithakrishnadhas@gmail.com

Abstract: The enormous increase of social event posting in social media, make the users and government very difficult to track about the events in the social media. To overcome this problem, the effective event attribute mining methodology is proposed for summarizing visualized and non-visualized data from the social events. Hence the feature selection and classification of social events helps to understand the evolutionary trends of the real world data set. To evaluate the effectiveness of proposed methodology, the various experiments are conducted to show improvement in qualitative and quantitative measurement in tracking of social events

I. INTRODUCTION

Social media are the latest technologies that allows the user for creating and sharing the information. Not only information but also ideas, career interest and other data through virtual communities and networks. The various stand-alone and built-in social media services recently available introduce challenges of definition. It has some common characteristics. Social media applications are interactive web 2.0 internet based applications. The content generated by the user including text posts, comments, digital photos and videos The data can be generated by means of all online interactions, which are the lifeblood of social media. The service-specific profiles can be created by the users for the website and applications which are designed and maintained by social media organization. It helps us to facilitate the development of online social networks by connecting a user's profile with those of other individuals and also groups. Social media use web-based, desktop computers and mobile technologies such as smartphones, laptops to create a highly interactive platforms through which individuals, communities, organizations can share, discuss and edit user's content which posted online. They use sequential and pervasive changes to communication between business, organizations, communities and individuals. It has changed the way of individual and large organization communicates. This helps to focus on the emerging field of technoself studies. In USA, a survey revealed that 84 percent of adolescents have an account in Facebook. About 60% of youngsters from the age of 13-17 have at least one profile on social media. They are spending more than three hours a day on social networking sites.

According to Nielsen, The users are spending more time on social media when compared to any other types of site. The total amount of time spent on social media in USA on both PCs and mobile devices increased by 99 percent to 121 billion minutes in July 2012 when compared 66 billion minutes in July 2011. For content contributors, the benefits of participating in social media have gone beyond simply social sharing to building reputation and bringing in career opportunities and monetary income. Social media totally differs from paper-based or traditional electronic media such as TV broadcasting in many ways, including quality, reach, frequency, usability, immediacy and permanence. They operate in dialogic transmission system. It is a traditional media which operates under a monologic transmission model. The monologic transmission model including paper newspaper which is delivered to many people. Some of the famous social media websites are

1. Facebook,
2. Whatsapp,
3. Instagram,
4. Twitter, Google news, YouTube, Viber.

II. LITERATURE REVIEW

The algorithm Multi-Model Event Topic Model(mmETM) for Multi-Model Event Topic Model for Social Event Analysis and its scope is to obtain the evolutionary trends of social media and generate event summary details. It also has some drawbacks such as Difficult to find the events and it is found in the summarized manner.

The algorithm TRCM to find a Survey of Data Mining Techniques for Social Network Analysis and its scope is to do Data Mining processes in terms of pre-processing, data analysis and data interpretation. The drawbacks of this algorithm includes Social media affects with its Size, Noise and Dynamism.

The algorithm Graph based models for Event Analysis in Social Media Using Clustering of Heterogeneous Information Network and its scope is to Focuses on combining multiple types of data in heterogeneous network. The drawback of this algorithm is Homogenous data are found (only one type of data is shown).

The algorithm Hofmann's hierarchical clustering/aspect model for Matching words and Pictures and its scope is Auto-annotation and Region naming. The drawback of this algorithm is Measuring the performance is difficult.

The algorithm Single Graph-Based Clustering Process for Multiscale Event Detection in Social Media and its scope is to provide different temporal and spatial scales of events in the data. The drawback of this algorithm is that it provides simultaneous data.

The algorithm Model of the normal city behaviour for Social Network Data analysis for Event Detection and its scope is minimizing the data used to maximize the total amount of usable data.

The algorithm Single Pass Clustering and thresholding model for On-Line New Event Detection and Tracking and its scope is making decisions about one-story on before subsequent stories. The drawback of this algorithm is Problems on new event detection and event tracking.

The algorithm Event based analysis for Event Analysis in Social Multimedia and its scope is to Remind the organizing data which helps in an extreme way. The drawback of this algorithm is Difficult to recall past memory.

The algorithm Event-Centric User Generated Content on social networks for Understanding Events through Analysis of Social Media and its scope is to Approach for analysing social signals by web with short messages like microblog posts and SMS.

The algorithm IEMiner for Mining Relationships Among Interval-based Events for Classification and its scope is Hierarchical representation on lossless representation. The drawback of this algorithm is No Interval-based events (Duration).

The algorithm Event Based Social Network (ESBN) for Event-based Social Networks: Linking the Online and Offline Social Networks and its scope is to provide online interactions and also includes offline interactions. The drawback of this algorithm is On Information Flow found the recommendation problem which was more cohesive than others.

III. EXISTING METHODOLOGY

In existing system, the novel multi-modal social event tracking and evolution framework, obtains the evolutionary trends of the social events and generate effective event summary details over time. The mmETM can model the multi-modal property of social event and learn the correlation between textual and visual modalities to separate the visual and non-visual representative topics. Applying this model for social event tracking, we adopt an effective incremental updating strategy. But there is a disadvantage in this methodology, here is the unstructured data is summarized as an event. It is big summary for the people who are in search of particular events. It also occupies more memory space and searching takes more time for the people. If we summarize the events after converting the unstructured data into structured data, it will be more useful. We can overcome all the disadvantages.

IV. PROPOSED METHODOLOGY

In proposed system, the drawback of the existing system is replaced using event attribute mining in social media.We will explore whether the tracking performance can be improved by considering different domains such as Facebook, Google news. The feature selection process and classification will make the proposed system more efficient when compared to existing system. The feature selection algorithm will be helpful in analysing the feature and classification is unsupervised classification algorithm. Euclidean distance and decision tree algorithm are two algorithms used. These algorithms are very effective algorithms used. Euclidean distance algorithm is an algorithm through which we can say that it is an everyday experience and perception. It says that the 1, 2 and 3 dimensional linear metric line and the distance between any two points in space corresponds to the length of straight line in between them. This algorithm helps in finding the accurate distance between any two points in space.

The decision tree algorithm is one of the most used and practical methods for inductive inference over supervised algorithm. It represents a procedure for classifying categorical data based on their attributes. It is also efficient for processing large amount of data and it is used in data mining application. The building of decision tree does require any particular knowledge or parameter

setting and it is particularly suitable for knowledge discovery. The representation of knowledge in tree form is for the easy understanding of human beings. We have to decide which attribute to test at node N by analysing the better way to separate the tuples into various classes. The dividing criteria is determined so that corresponding partitions are clear as much as possible. The divided tuples are clear, then it all belongs to the same class.

It is been classified under a rule, based on SOM algorithm (Self-Organizing Map).In unsupervised classification algorithm, the Simply Object Model (SOM) algorithm is used. The aim of this algorithm is to study a feature map from the spatially continuous input space, in which our input vectors live to the low dimensional spatially discrete output space, which is formed by arranging the computational neutrons into a grid. Depending on the aim of the application, rules may be generated from the frequent patterns discovered and the support and the confidence of each rule can be indicated.

The bar chart shows the difference between the previous algorithm and current algorithm used. There is improvement in both quantity and quality aspects of the output. In existing methodology they have used Multi-Model Event Topic Model (mmETM) algorithm Model for Social Event Analysis and its scope is to obtain the evolutionary trends of social media and generate event summary details. It also has some drawbacks such as Difficult to find the events and it is found in the summarized manner.

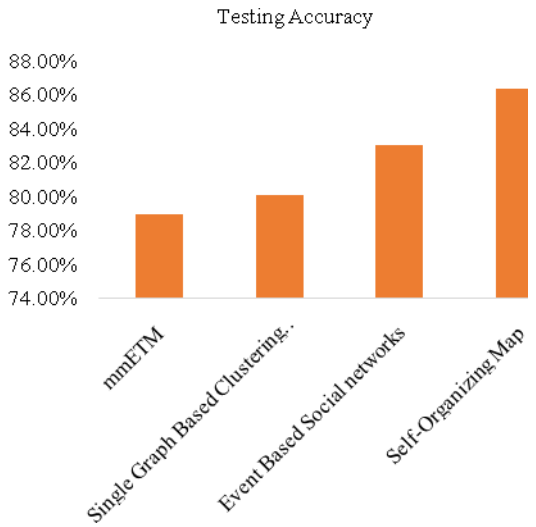

Fig 1 Analysis of Algorithm

In proposed methodology, we have used unsupervised algorithm called Simply Object Model (SOM) algorithm. This algorithm helps to achieve the best results when compared to the previous algorithm used. The SOM algorithm can be used to achieve best results instead of using Multi-Model Event Topic Model (mmETM) algorithm.

V. CONCLUSION

Thus, an enhanced unstructured data analysis in social media networking is a methodology which overcomes the disadvantage in existing methodology. In existing methodology, the unstructured data is analysed and summarized. But there is a disadvantage in this methodology, here is the unstructured data is summarized as an event. It is big summary for the people who are in search of particular events. It also occupies more memory space and searching takes more time for the people. Here, the unstructured data is converted into structured data and then it is analysed. The summarized events are mined into attributes. This helps the user to find the particular event easily.

VI. REFERENCE

] D. Patel, W. Hsu, M. L. Lee, "Mining relationships among interval-based events for classification", *Proc. SIGMOD*, pp. 393-404, 2008.

[2]. J. Allan, R. Papka, V. Lavrenko, "On-line new event detection and tracking", *Proc. SIGIR*, pp. 37-45, 1998.

[3]. M. Merler, B. Huang, L. Xie, G. Hua, A. Natsev, "Semantic model vectors for complex video event recognition", *IEEE Trans. Multimedia*, vol. 14, no. 1, pp. 88-101, Feb. 2012.

[4]. T. Zhang, C. Xu, "Cross-domain multi-event tracking via CO-PMHT", *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 10, no. 4, pp. 31:1-31:19, 2014.

[5]. X. Yang, T. Zhang, C. Xu, M. S. Hossain, "Automatic visual concept learning for social event understanding", *IEEE Trans. Multimedia*, vol. 17, no. 3, pp. 346-358, Mar. 2015.

[6]. L. Xie, "Discovering meaningful multimedia patterns with audio-visual concepts and associated text", *Proc. ICIP*, pp. 2383-2386, 2004.

[7]. D. M. Blei, J. D. Lafferty, "Dynamic topic models", *Proc. ICML*, pp. 113-120, 2006.

[8]. Y. Yang, J. Zhang, J. Carbonell, C. Jin, "Topic-conditioned novelty detection", *Proc. KDD*, pp. 688-693, 2002.

[9]. J. Makkonen, H. Ahonen-Myka, M. Salmenkivi, "Simple semantics in topic detection and tracking", *Inform. Retrieval*, vol. 7, no. 3–4, pp. 347-368, 2004.

[10]. N. Diakopoulos, M. Naaman, F. Kivran-Swaine, "Diamonds in the rough: Social media visual analytics for journalistic inquiry", *Proc. IEEE Symp. Vis. Analytics Sci. Technol.*, pp. 115-122, 2010-Oct.

[11]. X. Wu, C.-W. Ngo, A. G. Hauptmann, "Multimodal news story clustering with pairwise visual near-duplicate constraint", *IEEE Trans. Multimedia*, vol. 10, no. 2, pp. 188-199, Feb. 2008.

[12]. I. Kalamaras, A. Drosou, D. Tzovaras, "Multi-objective optimization for multimodal visualization", *IEEE Trans. Multimedia*, vol. 16, no. 5, pp. 1460-1472, Aug. 2014.

[13]. X. Chen, A. O. Hero III, S. Savarese, "Multimodal video indexing and retrieval using directed information", *IEEE Trans. Multimedia*, vol. 14, no. 1, pp. 3-16, Feb. 2012.

[14]. W. Liu, T. Mei, Y. Zhang, "Instant mobile video search with layered audio-video indexing and progressive transmission", *IEEE Trans. Multimedia*, vol. 16, no. 8, pp. 2242-2255, Dec. 2014.

[15]. C. Kang, S. Xiang, S. Liao, C. Xu, C. Pan, "Learning consistent feature representation for cross-modal multimedia retrieval", *IEEE Trans. Multimedia*, vol. 17, no. 3, pp. 370-381, Mar. 2015.