An Intelligent Agent Based Classification Model To Analyse Social Network

J Vinothini[1], P Vimala[2], J Maheswari[3]
[1,2]Assistant Professor, [3]PG Scholar
Department of Computer Science and Engineering,
Dhirajlal Gandhi College of Technology,Salem, Tamil Nadu - 636309, India

Abstract : Social network attracts the user to share the data and to communicate across the globe. There is an exponential increase in the number of users in social media like facebook, twitter etc. It facilitates user to share data, discuss events, post messages etc. In spite of above mentioned facilities it has major drawback where users struggle with some malicious behavior which is a root cause for majority of the problems in digital era. Hence there arise a necessity to categorize the users behavior based on some classifiers. This paper introduce an intelligent behaviour analysis model to overcome this drawback .The work involves the collection of dataset from users activities in the site, then label the dataset and to classify the users into positive and negative categories. Along with this the proposed hybrid classifier is used for effective classification. This proposed system works good in terms of accuracy.

Keywords: Social Network, Hybrid classifier, Classification.

I. INTRODUCTION

Recently, the utilization of social network is increasing, especially Facebook, Twitter, Google plus, Orukut etc., and it is also become one of the major way for internet users to maintain the relationship and keep communications with their friends [FAC][TWI]. According to Statistical report [STA], the number of social network users has reached 1.61 billion until late 2013, and is estimated to be around 2.33 billion user's globe, until the end of 2017. Social networking sites have gained massive eminence because of the opportunities they give people to connect to each other in an easy and timely manner. Clearly, the fastest growing ecumenical social network during the past few years is the Facebook. Serving as a medium to facilitate a large host of viewers, the 1.2 billion monthly active users make Facebook the largest social-networking site [LAF 13]. Incorporated with the attractive applications such as photo sharing, instant messaging and strong network ties of families, close friends, etc., when combined with its availability in smartphone and tablet made Facebook an requisite social network for youth. Facebook as a successful micro-blog service has become an important part of the daily life of millions of users. In addition to communicating with friends, families or acquaintances, facebook is used as recommending services, real-time news sources, marketing tools and content sharing venues. Users' posting behavior depends on many factors: interests and hobbies, time spend on line, followers' comment and so on. If a user just use twitter as a tool to communicate with friends, read latest news, post his concerned massages, his posting behavior usually reveal more random than other users. In this work we distinguish users by their posting, sharing and commenting behavior. We collect the approximate number of postings, likes, comments and shares of messages for all users, and then analyze the complexity of users' behavior in terms of real time questionnaires as features for classification. We analyze and measure the complexity of each user's behavior, using hybrid rule based classification.

The evolvement and expansion of social network is driven by users' demand or potential demand. Understanding users' behavior rule and its impacts on the social networks are an important part of social media service design and routine network optimization. Some works have been carried out on predicting the relationship among users [SMP 12] [KSA 12]. Reference [SMP 12] explored the difference among different types of relationship among users, while Reference [KS 12] predicted the different types of relationship among users by means of machine learning. Other works have presented the methods to predict the attributes of users [YNA 12][E 05][DFL 13]. In an high risk environment, data classification becomes a fundamental step in focusing behavior of user groups. The main goal of data classification is to establish a framework for classifying Facebook "share" of university students based on market segment, value and how critical it is to the marketers. Classification also comes with numerous algorithms to analyze the criteria and parameters [SMP 12]. Some of the well-known classification algorithms applied in this experiment includes K-Nearest Neighbor (KNN), Decision Tree, Naïve Bayes and Support Vector Machine (SVM).

In this paper, we propose a new behaviour analysis model for analyzing the user's behaviour on social network. The proposed model uses the existing automatic feature selection algorithm [GNK 10], SMO [HDW 94] and the proposed Rule based Classifier for categorizing the users based on their behaviour. Rest of this paper is organized as follows: Section 2 provides the survey of behaviour analysis on social network. Section 3 explains the overall architecture of the proposed system. Section 4 describes the proposed methodology in detail. Section 5 provides the result and discussion. Finally, section 6 gives the conclusion and future enhancement.

II. RELATED WORK

Many works have been done in this area by various researchers over the few past years. Among them, Duan Hu et al [DFL 13] recognized the identities of users according to the rules and characteristics of mobile phone usage behaviors of users with different identities. They have collected the record of mobile phone usage behaviors from 140 users, analyzed the behaviour differences between the users in making calls, sending SMS messages and surfing the Internet, and identified the users by machine learning algorithm according to the characteristics of user behaviours and proposed a classification method to verify the validity of user identity. Moreover they focused on analyzing the different types of characteristic values indicating social relationship (call characteristic value and SMS characteristic value) and stream characteristic values. Based on students' behaviours, Suwimon Vongsingthong [SNA 14] have analyzed the implications of Facebook "share" with respect to commercial gain. An interaction matrix of "share" interaction and profile data are composed as a dataset which are clustered into six eligible groups of commercial segment: dining, itinerary, pets, entertainment, games, and gifts/varieties. Pervasive classification algorithms: KNN, Decision Tree, Naïve Bayes and SVM are applied to explore the opportunity of target products. The data on traits of information "share" were collected, validated and explored.

Huiqi Zhang et al [HRC 11] proposed a socioscope model for social network and human-behavior analysis based on mobile-phone call-detail records and also a new index to measure the level of reciprocity between users and their communication partners by using multiple probability and statistical methods for quantifying social groups, relationships, and communication patterns and for detecting human-behavior changes has been proposed.

Su He et al [SHZ 14] introduced Multi-scale Entropy for analyzing and identifying user behavior on twitter, and separate users to different categories. Also they have identified five distinct categories of tweeting activity on twitter: individual activity, newsworthy information dissemination activity, advertising and promotion activity, automatic/robotic activity and other activities. Through the experiment they achieved good separation of different activities of these five categories based on Multi-scale Entropy of users' posting time series. The method based on Multi-scale Entropy is computationally efficient; it has many applications, including automatic spam-detection, trend identification, and trust management, user-modeling in online social media. Dae Ha Park et al [DEB 13] defined some sort of messages as "social spams", and suggested new classification method to detect them. By characterizing the problem of discovering social spams which frequently occurs in current popular SNSs, we extracted and exploited novel features that had not shown in the existing Email or web spamming prevention techniques. The dataset is the collection of various features such as behavior, celebrity, trust, common interest, etc. could incrementally been updated for SNS users.They modified the existing well-known classification techniques such as Bayesian network classifiers (BNCs) to customize for SNS features and for efficient decision making they have computed Katz or trust scores with only part of updated network topologies.

Enhua Tan et al [ELS 12] conducted a thorough analysis of a large blog trace to study the user activities in about one year. Analysis provides several new findings on the spamming behavior in blog-like UGC sites. Based on these non-textual features, they have applied several classifiers to classify UGC spammers and also proposed a spam detection scheme that design and evaluate a runtime spam. Francis T. O'Donovanet al [FCSOJJK 13] presented five clusters of users with common observed online behaviors, where these users showed correlated profile characteristics and identify some common properties of the most popular multimedia content and also presented a unique extension of prior work on clustering social network users that attempts to correlate the latent roles associated with the clusters with demographic and psychological profiles.

III. SYSTEM MODEL

The overall architecture of the proposed system is shown in Fig. 1. It consists of seven modules such as Face Book dataset, User interface module, Feature selection module, Hybrid classification module, Decision making agent, rule base and result analysis. Face Book dataset contains the Facebook user's detail. User interface module collects the necessary records from Facebook dataset and forward it to the feature selection module for further process. The feature selection module selects the necessary attributes only by the help of the existing attribute selection algorithm called Attribute Selection Algorithm and the selected featured dataset will be sent to the next module for classification.

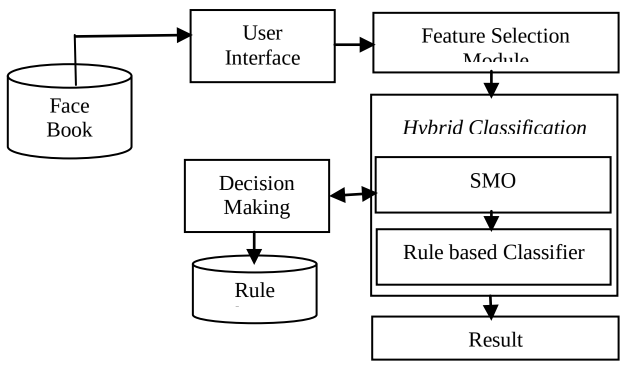

Fig. 1. System Model

The hybrid classification module classifies the records using SMO and rule based classifier. Rule based classifier contains the rules for classification which are framed by the past and present experience. The decision making agent helps to take a decision on Facebook users records by the rule base classifier. The final analysis report stored in result module.

IV. PROPOSED WORK

In this section, we propose a new behaviour analysis model for analysing the Facebook user's behaviours. The proposed model uses an attribute selection algorithm for feature selection and a hybrid classifier. The hybrid classifier is the combination of SMO and the proposed Rule based classifier.

4.1 Feature Selection

In this section, we have used the existing feature selection algorithm called Automatic Feature Selection Algorithm [GNK 10]. This algorithm selects the necessary attributes from the given input dataset. This automatic feature selection selects the necessary features based on the weights of the attributes. Some of the features were discarded which are close to zero.

4.2 Hybrid Classifier

In this section, we introduced a new hybrid classifier for effective classification that uses an existing classifier called Sequential Minimal Optimization (SMO) [HDW 94] and the proposed Rule based classifier.

4.2.1. SMO

We have used the standard SMO [HDW 94] algorithm which is built in already in WEKA for initial classification. It is a simple algorithm that can quickly solve the SVM QP problem without any extra matrix storage and without using numerical QP optimization steps at all. SMO decomposes the overall QP problem into QP sub-problems, using Osuna's theorem to ensure convergence. Unlike the previous methods, SMO chooses to solve the smallest possible optimization problem at every step. The smallest possible optimization problem involves two Lagrange multipliers, meanwhile the Lagrange multipliers must obey a linear equality constraint. There are two components to SMO: an analytic method for solving for the two Lagrange multipliers, and a heuristic for choosing which multipliers to optimize.

4.2.2. Intelligent Rule based Classifier

A new rule based classifier is proposed in this paper for effective classification. Rules were framed by the help of rule base and the recommendation of the decision making agent.

Intelligent Agent based Classification Algorithm

Input : Facebook or twitter data

Output : Categorized data (based on Behaviour)

Step 1: Read n feature selected records from data set.

Step 2: Agent selects some rules for classification for categorizing the users based on heir behaviour.

Step 3: Agent calculates the activity score for Like, Comment, Share and Post.

Step 4: Agent sets different thresholds for their each activities such as Like, Comment, Post and Share.

Step 5: Agents fire the necessary rules for decision making on each and every records of Facebook dataset.

Step 6: Agent checks for each users values (scores) of each and every activity exceeds the threshold then . Declares the particular user as a negative behavioural user

Else

Declares the particular user as a positive behavioural user

Step 7: List the users Facebook ID based on their behavioural.

The proposed classification algorithm reads the necessary classified records data from the feature selected data set by the help of feature selection algorithm. Here, categorized the users and split into several classes based on their behaviours. Activity scores are calculated based on their number of Like, Comment, Share and Post. Sets the threshold based on the average number of Like, Comment, Share and Post. Finally, categorized the users into two groups such as positive and negative behavioural users.

V. RESULTS AND DISCUSSION

This section discusses about the dataset used in this work for medical diagnosis, experimental scenario and also about obtained result and discussion of the proposed system and reason for achievements on decision making process.

5.1 Facebook Data Set

We have prepared Facebook dataset that has been collected by own from 1000 Facebook users all over India in the past one month activities. We have asked many questions regarding their frequent Like, Comment, Share and Post. The questions are including what related message/post you give like frequently?, How many like you given out of total received message/post?, How many messages you posted / shared and what type of information/post those?, etc.

5.2 Experimental Setup

We have used the Pentium IV personal computer with Intel Core i3 Processor 2.20 GHz for evaluating the proposed system. Initially, feature selection by using knowledgebase and send it for classification using WEKA tool and JAVA. WEKA is a collection of machine learning algorithms for data mining tasks. The proposed algorithms applied to the dataset from Java code and it contains tools for data analysis and predictive modelling. The input dataset of the WEKA are used in the form of CSV file.

5.3 Experimental Results

The various experiments have been conducted for evaluating the proposed behaviour analysis model. This section discusses the various results obtained by the proposed model and other models and classifiers. Table 1 shows the performance of the existing classifier SMO [HDW 94]. We have conducted five experiments for evaluating the existing algorithm and taken average. We have used Facebook dataset which is prepared by own for these experiments.

Fig. 2. shows the performance of the proposed hybrid behaviour analysis model which is the combination of SMO classifier and Rule based Classification algorithm, also it can be observed that the performance of the proposed behaviour analysis model is better than SMO with feature selection and SMO with the proposed Rule based classifier.

TABLE 1. PERFORMANCE OF SMO.

Mean absolute error	2.6679
Root mean squared error	3.1064
Relative absolute error	80.3915 %
Root relative squared error	83.0159 %
Total Number of Instances	1000

Fig. 2. Performance Analysis

Table 2. shows that the performance comparative analysis between the proposed Rule based classifier, C4.5 and DTC. From this figure, it can be observed that the performance of the proposed hybrid behaviour analysis model provides better performance than individual performance of other classifiers.

TABLE 2. COMPARATIVE ANALYSIS

Name of the Classifiers/ Model	Accuracy
C4.5	78.6
DTC	83.4
SMO	89.4
Proposed Hybrid Behavior Analysis Model	94.7

The reason for this performance difference is the uses of feature selection algorithm, Rule base, Decision making agent and the combinations of two algorithms. By the uses of feature selection gives optimal features to the classifier, so classifier can perform well with optimal information and also effectively use the less number of rules.

The overall time taken also reduced when it is compared to the existing models due to the uses of less number of features for calculating the activity scores and uses those features for making decision by the proposed classifier. The classification accuracy is automatically increases when we used the less number of features.

VI. CONCLUSION AND FUTURE WORKS

The proposed work has evolved with an efficient classification of users in to positive and negative cadres.The accuracy of classification can be enhanced in future by introducing fuzzy rules supported with time constraints to categorize users.

REFERENCES

[1] Dae-Ha Park, Eun-Ae Cho, and Byung-Won On,(2013)*:"Social Spam Discovery using Bayesian Network Classifiers based on Feature Extractions"*,12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, pp. 23-29, DEB 13.

[2] Duan Hu, Fei Sun, Lai Tu, and Benxiong Huang,(2013): *"We Know What You Are - A User Classification Based On Mobile Data"* IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing, pp.1282-1289

[3] N Eagle,(2005): *Machine Perception and Learning of Complex Social Systems*. Program in Media Arts and Sciences. Massachusetts: Massachusetts Institute of Technology, pp. 125-136.

[4] Enhua Tan, Lei Guo, Songqing Chen, Xiaodong Zhang, and Yihong (Eric) Zhao,(2012):*"Spammer Behavior Analysis and Detection in User Generated Content on Social Networks"*,32nd IEEE International Conference on Distributed Computing System, pp. 34-39.

[5] Facebook, in: http://www.facebook.com/.

[6] Francis T. O'Donovan, Connie Fournelle, Steve Gaffigan, Oliver Brdiczka, Jianqiang Shen,Juan Liu, and Kendra E. Moore,(2013): *"Characterizing User Behavior And Information Propagation On A Social Multimedia Network"*, IEEE ternational Conference on Multimedia and Expo Workshops (ICMEW), pp. 1-8.

[7] Gupta K., Nath B., Kotagiri R.,(2010*): "Layered Approach using Conditional Random Fields for Intrusion Detection"*, IEEE Transactions on Dependable and Secure Computing, vol. 7,no.1, pp. 35-49.

[8] G. Holmes; A. Donkin and I.H. Witten (1994): *"Weka: A machine learning workbench"*. Proc Second Australia and New Zealand Conference on Intelligent Information Systems, Brisbane, Australia. Retrieved 2007-06-25.

[9] Huiqi Zhang, Ram Dantu, and João W. Cangussu,(2011): *"Socioscope: Human Relationship and Behavior Analysis in Social Networks"*, IEEE Transactions on Systems, Man, and Cybernetics-part A: Systems and Humans, vol. 41, no. 6, November 2011.

[10] E. Khadangi and A. Bagheri,(2013): *Comparing MLP, SVM and KNN for predicting trust between users in Facebook*, 3rd International Conference on Computer and Knowledge Engineering (ICCKE 2013), October 31 & November 1,2013,Ferdowsi University of Mashhad pp. 466-470.

[11] Krishna Janakiraman, Sara Motahari.(2012): *How are you related? Predicting the Type of a Social Relationship using Call Graph Dat*, ASE/IEEE International Conference on Social Computing, pp.111-116.

[12] J. Lafferty. (2014, 31/3/2014). *Facebook's Q4 2013: 556M mobile DAUs*, 1.2B MAUs, $2.5B in revenue.

[13] Su He, Hui Wang, Zhi Hong Jiang,(2014*): "Identifying User Behavior on Twitter Based on Multiscale Entropy"*, International Conference on Security, Pattern Analysis, and Cybernetics (SPAC), pp. 14-18.

[14] Sara Motahari, Ole J.Mengshoel, Phyllis Reuther, Sandeep Appala, Luca Zoia, Jay Shah.(2012): *The Impact of Social Affinity on Phone Calling Patterns: Categorizing Social Ties from Call Data Records*, In Proc. of SNA-KDD-12, Beijing, China, pp. 1-9.

[15] Suwimon Vongsingthong, Nawaporn Wisitpongphan,(2014): *"Classification of University Students' Behaviors in Sharing Information on Facebook"* 11th International Joint Conference on Computer Science and Software Engineering (JCSSE), pp.134-139.

[16] Statista, in: http://www.statista.com/.

[17] Welcome to Twitter, in: http://twitter.com/.

[18] Yaniv Altshuler, Nadav Aharony, Alex ("Sandy") Pentland, Michael Fire, Yuval Elovici.(2012): *Incremental Learning with Accuracy Prediction of Social and Individual Properties from Mobile-Phone Data*, ASE/IEEE International Conference on Social Computing, pp.969-974.