Identifying User Personality in Facebook Using their Activity Data

P. Dhinesh Kumar[1] and S. Anusuya[2]
[1]PG Scholar, [2]Assistant Professor,
[1&2]Department of Computer Science, Sri Vasavi College (SFW), Tamil Nadu, India
E-Mail: dhinesh5577@gmail.com

Abstract - The facilitators of human interactions, social networks have become an interesting target of research, providing rich information for studying and modelling user's behaviour. Identification of personality-related indicators encrypted in Facebook profiles and activities are of special concern in our current research efforts. This paper explores the feasibility of modelling user personality based on a proposed set of features extracted from the Facebook data. The encouraging results of our study, exploring the suitability and performance of several classification techniques, will also be presented. Gaining insight in a web user's personality is very valuable for applications that rely on personalisation, such as recommender systems and personalised advertising. In this paper we explore the use of machine learning techniques for inferring a user's personality traits from their Facebook status updates. Even with a small set of training examples we can outperform the majority class baseline algorithm. Furthermore, the results are improved by adding training examples from another source. This is an interesting result because it indicates that personality trait recognition generalises across social media platforms.
Keywords: Social Networks, Results, Learning, Personality, Identification

I. INTRODUCTION

Social networks have become widely-used and popular mediums for information dissemination as well as facilitators of social interactions. Users' contributions and activities provide a valuable insight into individual behaviour, experiences, opinions and interests. Considering that personality, which uniquely identifies each one of us, affects a lot of aspects of human behaviour, mental processes and affective reactions, there is an enormous opportunity for adding new personality-based qualities to user interfaces. Personalized systems used in domains such as, e-learning, information filtering, collaboration and e-commerce could greatly benefit from a user interface that adapts the interaction (e.g., motivational strategies, presentation styles, interaction modalities and recommendations) according to user's personality. Having captured past user interactions is only a starting point in explaining the user behaviour from a personality point of view. This research builds upon previous interdisciplinary research works regarding personality as it pertains to the design of intelligent interactive systems. The new communication technologies have brought more information to consider, though the process of their utilization is far from straightforward. Intelligent technologies are expected to play a prominent role in

bringing these data to a new level of usability. A variety of Facebook variables were expected to play a prominent role in establishing appropriate context for our particular investigations. Facebook profiles and activities provide valuable indicators of user's personality, revealing the actual, rather than idealized or projected personality. The research has two interconnected objectives: (1) to identify the relevant personality-related indicators that are explicitly or implicitly present in Facebook user data and (2) to explore the feasibility of predictive personality modelling to support future intelligent systems. We hypothesized that increasing the relevance of what is included in the model, and considering features drawn from a variety of sources may lead to better performance of the classifiers under investigation. The choice to include a feature was based on whether the previous research had underlined the importance of such a choice and its relevance to the objectives of this research. Our research is currently focused on investigating the suitability and performance of various classification techniques for personality modelling.

II. RELATED WORK

A. Data Mining Techniques

Data mining techniques play a fundamental role in extracting correlation patterns between personality and variety of user's data captured from multiple sources. Generally, two approaches were adopted for studying personality traits of social network users. The first approach uses a variety of machine learning algorithms to build models based on social network activities only. The second one extends the personality-related features with linguistic [1].

B. Classification and Regression

Several classification and regression techniques were used to build predictive personality models along the five personality dimensions using the linguistic features of a dataset comprised of few thousand essays solicited from introductory process. The Linguistic Inquiry and Word Count – LIWC was used as a tool for linguistic analysis. The reported precisions of the classifiers were in the range of for all traits. In, SMO and Naive Bayes were used for modelling four out of five personality dimensions by extracting features from a corpus of personal web-blogs. Their results point out to the importance of the process of

feature selection in increasing the classifiers precision yielding for automatic feature selection. The point of the differences in the datasets used in these studies compared to, namely different solicitation methods and the sources from which they were collected. The correlation between users' social network activity and personality has been the focus of several studies in the last decade. Personality traits of the Chinese most popular social network users were analysed [2].

C. Personality Models

Decision Trees have shown the best results, yielding percent accuracy, for a combination of features related to users' network activity along with affective linguistic features extracted from statuses and blog posts. The work most closely related to the method. Rules and Gaussian Processes were applied to build predictive personality models. The authors consider users' Facebook data through parameters such as structural characteristics, personal info, activities and preference, in addition to the linguistic attributes extracted with LIWC from the users' statuses [3].

D. Correlations

The lack of demographic diversity in participant sampling was one of the major drawbacks for generalizing the results of the last two studies, Chinese population and authors' Facebook friends respectively. Few studies using considerably larger number of instances from the same dataset under our investigation have a rather different objective from ours, namely to examine the correlations between the personality traits and Facebook activity data and the associations between personal attributes and Facebook Likes. These studies were not meant to look at the rich linguistic patterns that occur in the language use on social networks, which is in the focus of this research [4].

E. Personality

The term personality is derived from the Latin word persona, which means the mask used by actors in a theatre. A set of attributes that characterize an individual and involves emotions, behaviour, temperament and the mind defines a personality. Due to the diversity of attributes it is crucial to gauge personality as it does not provide any definitive structure through which people can be classified and compared. The set of human emotions is vast, due to which a similar problem occurs when one tries to identify the sentiment embedded in a message (sentiment analysis), thus making it challenging to choose the basic emotions for a classification. Thus in order to automate sentiment analysis, for instance, many researchers accepts a simplified representation of sentiments by means of their polarity. Similarly for determining personality, various researchers have recognized the most essential characteristics in order to create a personality model. Personality can vary depending on different situations L [5].

III. SYSTEM WORK

A. Extraversion

People in this dimension have an inherent need to advertise their activities to others and their good mood depends on the feedback they receive from them. People in this category tend to spend more hours in social networking sites.
Particularly, in Facebook, they tend to belong to more groups and have more friends. Furthermore, they have the tendency to upload more personal photos than people belonging to other personality dimensions, share more statuses and post. Technological approaches on extracting personality traits from Facebook it is important to understand how personality traits relate to user, activities and behaviour on Facebook based on the results reported by behavioural and psychology sciences.

B. Neuroticism

The dimension describes people with the tendency to experience strongly negative emotions, such as anger, anxiety, or depression. People characterized by neuroticism, tend to be more frequent users of Facebook since they want to control the information about themselves and their environment. Thus, the most frequent activity they practice is to disseminate information or statements that they approve. In contrast they avoid publishing photos of themselves. Furthermore, neurotics tend to have fewer friends on Facebook, but at the same time, use often the like function in posts of these friends.

C. Extracting Personality Traits from Facebook

The rapid development of Facebook, compared to other social networks, and due to the enormous amount of information available for most users, many research groups have tried to acquire and exploit the log data in order to draw conclusions in relation to personality. Two main techniques are used and discussed below. Semi-automated data mining approaches utilize algorithms to extract information from public profiles on Facebook. In any case, the users involved in the study have to complete a personality questionnaire in order for the researchers to get an indication on the users' personality. Data mining algorithms and machine learning are followed in analysing and correlating the activity of users to personality traits. User's replies to the personality questionnaire are used for evaluating the models developed. These studies showed that textual elements and demographic profile information of users can provide indication of user's personality and that indeed personality is closely related to social networks usage. Although machine learning approaches are un-obstructive methods for the user, and predict user personality with high accuracy, publicly available information are getting fewer as time passes due to Facebook new privacy policies and settings. Consequently, the information one can get using this method is not rich

and similarly to the previous discussion the user is not directly getting anything back.

D. Facebook Data Extraction for User Modelling

The extraction of Facebook activity data has been done using a Facebook application, which allowed us to get users' permissions for accessing their personal data as input to the framework. The data extracted include publically available information about a user and also private activity data. Additional features have been defined by the authors that can be considered to be a list of friends of a user with whom the user 'regularly' interacts with. In order for a user to be considered as an active friend of a given user, to publish at least four posts directly on that user's wall, or appear in a Facebook activity together, during a period of a year. The reason for the four posts threshold is for excluding birthday and name-day wishes.

E. Aggregated Data

In every user model we keep a vector that consists of arithmetic normalized values of the aggregated data collected based on thresholds defined. The thresholds' values defined based on a sample of Facebook users who participated in the study presented in a following section and thus excluded from the overall evaluation sample. After considering reports of Facebook user activity, we defined the activity of 'light' to 'heavy' Facebook usage for each element. Elements can take values from, to simulate the scores of answers in a personality questionnaire. This is used in the extraction, of the user personality model, and the similarities between users in the system.

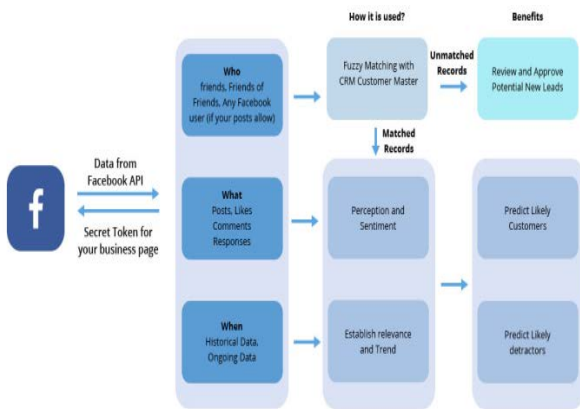## IV. ARCHITECTURE DIAGRAM

Fig. 1 Facebook Data Architecture

V. PROPOSED WORK

The purpose of the study is to propose a theoretical framework that can be used to identify the personality trait of a social media user.in this field of psychology showed that there is a correlation between personality and the linguistic behaviour of a person. This correlation can be effectively analyzed and illustrated using natural language

processing approach. Therefore, the goal of this research is to build a prediction system that can automatically predict user personality based on their activities in Facebook. There are several personality models used in predicting personality, such as Big Five Personality, MBTI (Myers-Briggs Type Indicator) or DISC (Dominance Influence Steadiness Conscientiousness). However, after some considerations and literature review process, Big Five Personality is used in this study as it is the most popular and precise in telling someone's personality traits. Traits in this model consist of Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. Personality is a way person respond to a particular situation. It is combination of characteristics that make an individual unique.

Data from Online Social Networking Sites provides a solution to this problem. The rapid growth in social media increased people perceptions towards it. It went from niche activity to vary widely and heavily processed. It has emerged as one of the most ubiquitous means of communication today. It allows individual to find like-minded ones, whether it be for romantic or social purpose. It is also being used to maintain existing social connections. The observed that online interactions generated more self-disclosures and fostered deeper personal questions than did face-to-face conversations. Now-a-days people analyze person's social profile before considering as business partner or before dating. Researchers have shown how useful social networking is among old adults, what can we learn from Facebook activity and how often it is used by famous personality. With aforementioned benefits user population using these social networking sites is increasing day-by-day. Their interaction pattern, profile data, text or multimedia content used during conversation or status update provide lot of raw data to researchers which can be used to determine personality traits.
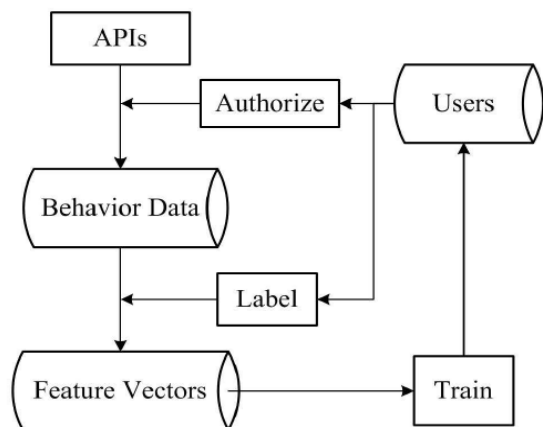

Fig. 2 Propose System

A. General Study

Assessment of personality over the past two decades in various researches has revealed that personality can be defined by five dimensions known as Big Five personality

traits. In general, study of personality considered as a psychology research based on the survey or questionnaire. But this limits the research data to less number of persons. Hence there is a need of something through which can increase the number of people involved in survey and to make the process automated.

B. Linkage Data Based Analysis

Social Network can be analyzed with mapping and measuring of relationships between various entities. Analysis is often represented using diagram as shown in figure1. It is based on network structure. Here Nodes represent actor, object, people or group. Edges represent relationship between those actors. This type of social network analysis is useful for the work related to organization development. Combining both the approaches Linkage data and content based analysis provide input to wide range of applications including in prediction of personality traits.

C. Regression Method

This is simplest and most used method. It analyzes relationship between dependable variable and predict results. It can be Linear or Non-linear. Linear model seems to describe the relation best. However, sentiment data doesn't work well in the regression model for movies. Researchers have mostly used regression algorithm like M5rule, Multivariate linear, Gaussian Process and ZeroR for the calculations.

VI. IMPLEMENTATION

A. Data Filtering

Data collected from the social media will be the text from comments posted by the user. The filtering of the text requires some phrase and pattern based techniques or term based techniques. Here, the phrase based technique is preferred because phrases carry more semantic information than terms and hence better performance can be expected . The main aim for filtering data is to remove the redundant or irrelevant data. As a result, we will get clean data which can be processed more effectively. First of all, the probable phrases and their synonyms that can occur in the comments are listed. This list helped in extracting those phrases from the text. Also, the dictionary including list of words l ike 'a', 'an', 'the', 'you', 'of', 'over' etc. is made to avoid useless text from getting processed.

B. Data Stemming

Data stemming uses the extracted phrases after data filtering. Stemming is the process for reducing the words to their stem or root form. In this, the set of words that can be treated as equivalent are identified and these multiple occurrences are replaced with their root form.

C. Personality Trait Repository

Personality trait repository is used to associate the Big five personality traits with the corresponding attributes. The attributes considered here are openness to experience, consciousness, extraversion, agreeableness, neuroticism. Each attribute included in the repository is again linked with the synonymous words. The information retrieved is the text in comments. The text is composed of phrases, certain adjectives. These phrases and adjectives will be the input to the repository where association between phrases or adjectives and synonymous words will take place.

D. Simplified Sentiments

In processing input will be provided for simplifying the sentiments. The sentiments which are associated with the text used in comment may be openness to experience, consciousness, extraversion, agreeableness, neuroticism. The input here is the stem or root form of the words or phrases used in the comments. So, it is easier to identify the corresponding sentiments. Social media is one of the most easily accessible ways to understand natural behaviour of an individual, understand user's likes and dislikes and so we can link information extracted from social media to understand personality traits of social media users. The purpose of the study is to propose a theoretical framework that can be used to identify the personality trait of a social media user. In this field of psychology showed that there is a correlation between personality and the linguistic behaviour of a person. This correlation can be effectively analyzed and illustrated using natural language processing approach. Therefore, the goal of this research is to build a prediction system that can automatically predict user personality based on their activities in Facebook.

VII. CONCLUSION

Social behaviour in online social networking sites can be used to predict User's big five personality traits. Psychologist used to follow personality questionnaire approach. This process is costly and impractical at times. With the popularity of online social networks, researches envisaged to predict the personality automatically. Researches tried to assess the personality based on internet and social network site usage. However only some of the personality traits like Extraversion and emotional stability could be assessed using this approach. Through linkage and content based analysis of these online social networking sites data, researchers were able to predict personality traits quite accurately. Based on Facebook "Likes", Network Structure like number of friends and groups, Status update, Photo upload, Tags and then using various regression and machine learning algorithms researchers were able to correlate these features with personality trait. Researchers have used multiple approaches like applying linguistic algorithm to user text and combining the results with network structure based analysis to predict with better accuracy as different traits can be best predicted with

different approaches. Some researchers have used behaviour aspect of social media like message content & type, behaviour towards friends & follower, response time etc. to correlate with personality trait.

REFERENCES

[1] L. Fast and D. Funder, "Personality as manifest in word use: correlations with self-report, acquaintance report, and behaviour", *Journal of personality and social psychology,* 2008.

[2] H. Schwartz, J. Eichstaedt, M. Kern, L. Dziurzynski, S. Ramones and M. Agrawal, "Personality, gender, and age in the language of social media: The open-vocabulary approach", 2013.

[3] Jiawei Han and Micheline Kamber, "Data Mining Concepts and Techniques, second edition", *Elsevier,* (Singapore) Pte Ltd., 2008.

[4] Han S. Kwak, H. Moon, S. Ahn, Yong-Yeol and H. Jeong, "Analysis of topological characteristics of huge online social networks services", 2007.

[5] S. Bai, T. Zhu, and L. Cheng, "Big-Five Personality Prediction Based on User Behaviours at Social Network Sites", 2012.

[6] J. Golbeck, C. Robles, and K. Turner, "Predicting Personality with Social Media", 2011.

[7] Yair Amichai-Hamburger and Gideon Vinitzky, "Social network use and personality", 2010.