

Personality Prediction with Social Behavior by Analyzing Social Media Data – A Survey

Ashish Kumar Tripathi
Department of Applied
Computer Science and Society
University of Winnipeg

tripathi-a@webmail.uwinnipeg.ca

ABSTRACT

Curiosity to predict personality, behavior and need for this is not as new as invent of social media. Personality prediction to better accuracy could be very useful for society. There are many papers and researches conducted on usefulness of the data for various purposes like in marketing, dating suggestions, organization development, personalized recommendations and health care to name a few.

With the introduction and extreme popularity of Online Social Networking Sites like Facebook, Twitter and LinkedIn numerous researches were conducted based on public data available, online social networking applications and social behavior towards friends and followers to predict the personality. Structured mining of the social media content can provide us the ability to predict some personality traits. This survey aims at providing researchers with an overview of various strategies used for studies and research concentrating on predicting user personality and behavior using online social networking site content. There positives, limitations are well summarized as reported in the literature. Finally, a brief discussion including open issues for further research in the area of social networking site based personality prediction preceding conclusion.

General Terms

Social Network, Personality, Algorithms, Human Factor.

Keywords

Personality Prediction, Social Media, Online Social Network.

1. INTRODUCTION

Personality is a way person respond to a particular situation. It is combination of characteristics that make an individual unique. 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 we can increase the number of people involved in survey and to make the process automated.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Conference '10, Month 1–2, 2010, City, State, Country.

Copyright 2010 ACM 1-58113-000-0/00/0010...\$10.00.

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 used [16]. 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 [95]. It is also being used to maintain existing social connections [96]. Tidwell and Walther [97] 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 [92]. Researchers have shown how useful social networking is among old adults [20], what can we learn from Facebook activity [27] and how often it is used by famous personality [25]. 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.

Normally by making the account on social networking site like Facebook people gives the right to collect their information [91]; based on this data Facebook research team try to monitor the user behavior. Many features and attributes of these social networking sites are useful for personality assessment. Researchers have proved this and were able to assess the personality traits by the use of social media such as Facebook [41] and Twitter [7]. Based on social behavior toward friends and based on structural information like number of friends, groups joined and likes etc. can be used to successfully predict some of the personality traits. In this paper we will try to cover the researches done to predict the personality traits using social media data, algorithm used, limitation and results.

Online social networks provide unprecedented opportunity from knowledge discovery and data mining point of view. There are many challenges that researchers face and still there is lot of scope of future research to predict it with better accuracy. We tried to cover some of those in section "A DISCUSSION" of this paper.

2. PRELIMINARIES

2.1 Study of Social Media

Social media is giving people a platform to interact with each other. Popularity of social media sites like Facebook, Twitter, and LinkedIn has increased incredibly in last few years. The core of the social networking experience centers on user's ability to perform following action:

- Online space for self-expression by means of posting self-relevant information on individualized profile page.
- Connectivity/interaction with friends and followers.

- Network expansion.

Online Social Networking sites provide web space to people and allow them to create an individualized public profile where they can express themselves and their thoughts using text, audio, pictures and video. They also provide notifications of people that we many know and those we should add as a friend using some friend recommendation system based on the principle known as homophily [65]. We can also see other people’s connections. These features help in Network expansion. Sharing of content on social sites comes with certain potential privacy risk [88, 89 and 90]. User can control it to certain degree with privacy settings provided by these sites.

Social media data can be leveraged in two type of analysis

- Content based
- Linkage data

Content based analysis: Social networking sites like Facebook, twitter and LinkedIn have tremendous amount of content in the form of text, image, audio and video. This huge database can be used for various researches.

Linkage Data based analysis: Social Network can be analyzed with mapping and measuring of relationships between various entities [11]. 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 [6].

Combining both the approaches Linkage data and content based analysis provide input to wide range of applications including in prediction of personality traits.

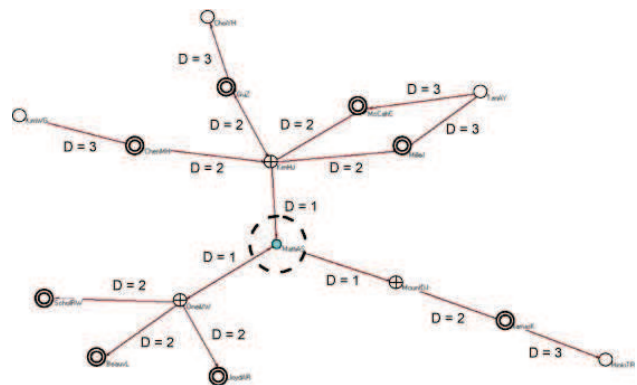


Figure1: Network Analysis Diagram [108].

2.2 Why Study Facebook?

There are three broad reasons why Facebook is of relevance to social scientist [91].

1. Activities performed on Facebook can provide concrete, observable and huge data set. Content and linkage data available here can be used for various analyses.
2. Online social networking sites bring both benefits and dangers to society, which need to be carefully examined and analyzed.
3. The universal popularity of Facebook makes it a topic worthy of study and research. Following figure2 depicts the popularity and growth of Facebook with addition of various features in it.

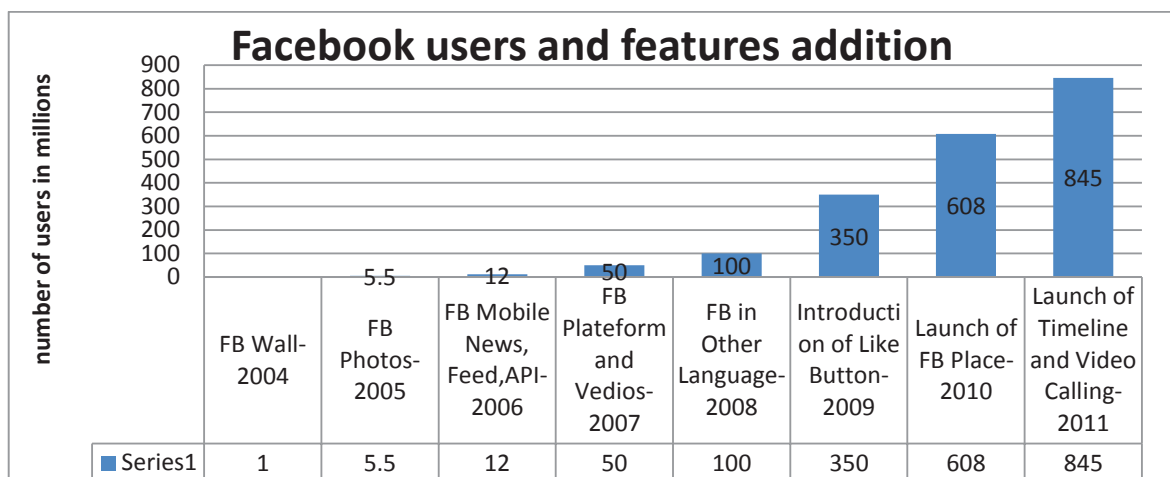


Figure 2: Facebook users: cumulative total over the year data taken from [91].

2.3 Survey of Facebook Research

There are many articles published based on Facebook. Study of Facebook is not limited to any particular discipline. From law, economics, sociology and psychology to information technology, management, and marketing are studying Facebook.

This reveals the importance of Facebook in the real world. Researchers conducted independent literature reviews and came forward with a list of the major themes from the literature study. This process resulted in the identification of five general categories that captured the major themes found throughout the

literature [91]. User Analysis and Social interaction based research covers 50% of the articles.

Table 1: Areas of Facebook Research Identified in the Literature Review [91].

Area of research	No. of articles	% of total	Associated research question
User Analysis	97	24%	Who is using Facebook and what they are doing on it.
Motivation behind use of Facebook	78	19%	Why they use Facebook?
Based on Social interaction	112	27%	How the Facebook is related to building the relations?
Related to privacy and information disclosures	75	18%	Why people are so intend to disclose their personal information despite the fact of security concern
Related to Identity	50	12%	How are people presenting themselves on Facebook?

2.4 Personality Assessment Fundamentals

There are two important fundamentals on which most of scholarly work on the field of personality assessment with social media data is based first one is personality model used and second one is prediction algorithm to calculate personality trait results.

2.4.1 Personality Models

There are two widely accepted personality models: Myers-Briggs, a four factor model, and the Big Five, a five factor model.

Big Five Personality Model: This is considered as most wide spread and accepted measures of personality structure in recent year. The five-factor model of personality is a hierarchical organization of personality traits in terms of five basic dimensions: Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness to Experience.

Extraversion: It relates with person’s tendency to get involved with external world. People high on extraversion tend to be more outgoing, friendly and socially active. Those with low score are likely to be solitary and reserved.

Agreeableness: It is a measure of maintaining positive social relationship. People high on Agreeableness tend to be cooperative, friendly, compassionate and adaptive. Low scorers are highly disagreeable people and are suspicious, distant and uncooperative the place.

Conscientiousness: highly conscientiousness people are like achiever; they are always good in discipline, responsible and prefer proper planning ahead. High Conscientiousness suggests a strong ability to regulate and control behavior.

Neuroticism: This is measure of emotional stability. Highly neurotic people are more prone to negative emotions like anxiety, anger, nervousness, stress and depression, more likely to be frustrated in day to day life. Low scorers are calm and collected, emotionally stable and balanced.

Openness to experience: Relates to person’s curiosity, interest in new experience/ideas, imagination. People with high score on this trait appreciate art, adventure and new ideas whereas ones having low score tend to be conservative, conventional and close minded.

Research using both natural language adjectives and theoretically based personality questionnaires supports the comprehensiveness of the model and its applicability across observers and cultures [38]. It is consistent across the age, gender, culture. This model presented with plethora of opportunities for the researchers in the area of personality and behavior prediction. Big five traits are based on a lexical approach to personality measurement [46], [58], [59] and [60].

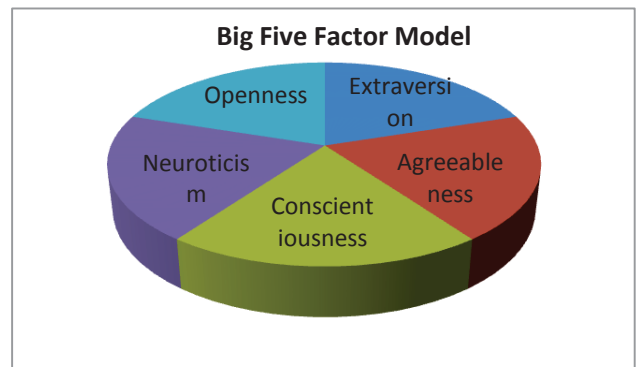


Figure 3: Five attributes of Big Five Factor Model.

Tupes and Christal [44] have considered these five as fundamental traits. Many researches [[38], [45], [46], [47] and [48]] were conducted and proved the validity of the model consistently across age, gender, and cultural lines. This model provides a conceptual framework that unifies various research findings on individual difference and personality. [[7], [1] and [24]] have attempted to relate the personality and social behavior largely based on this big five personality model.

Mostly researchers have used following approach to determine and relate with five factors model:

1. Propose hypothesis and prove the hypothesis with the result.
2. Identify the personality traits from questionnaire and prove it based on analysis of available data.
3. Compare the results and performance of proposed approach with the results obtained from other study.

2.4.2 Personality Hypothesis

Facebook behavior has been correlated with Five Factor Model [39] [40]. Rose et. al [39] started study of the relationship between personality and pattern of social network use. Based on

the questionnaire, it was first identified in 2009 that there could be a connection between Facebook data and personality [39]. The author had proposed five hypotheses.

Table 2: hypothesis used by [39] [40]

Personality Traits	Behavior on Facebook
Extraversion	Frequent user of Facebook
	More use Facebook Component
	More number of Facebook Friends
	More Facebook groups
Neuroticism	Spend more time on Facebook, more use of Facebook wall
	Less use of private message
	Share more information
Agreeableness	More number of friends
Openness	More use of Facebook for communication
	More number of components
	More knowledge of features
Conscientiousness	Limited Facebook activity

Same hypothesis were used by the author of another research[40], where they tried to prove those hypothesis

Table 3: Shows the analysis of proposed hypothesis, and analyzed results [39] [40]

Personality Traits	C. Ross, E.S. Orr, M. Arseneault, M.G. Simmering and R.R. Orr[39]	Y. Amichai-Hamburger and G. Vinitzky [40]
Extraversion	Related to number of groups. Harmonious	Related to number of friends. Harmonious
	Not related to number of friends. Contradictory	Not related to number of group joined Contradictory
Neuroticism	High neuroticism- more use of Facebook wall and prefer posting photos Harmonious	Directly proportional to Sharing of pictures Harmonious
	Prefer to post photos Harmonious	
Agreeableness	Unrelated	Directly proportional to the Facebook features used. Contradictory
Openness	Unrelated	Directly related to use of Facebook as a communication model. Harmonious
Conscientiousness	How much time individual spent on Facebook and how frequent they use their profile?	High in Conscientiousness have more number of friends and less use of photos, and spend less time. Harmonious
Limitation	Limited data (15 men, 82 women), Large proportion of females, homogenous data. It relies on self-reported answers.	It is based on Facebook data so more objective but limited data of 237 users. All college students.

2.4.3 Prediction Algorithms:

Social network sites have very rich content in the form of text, audio, video and image. Algorithm used for data mining, correlation and Prediction plays very important to get accurate results. Some of the algorithms are discussed below. Most of the researchers have used regression method only for assessment of personality traits.

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 [78]. However, sentiment data doesn't work well in the regression model for movies [79]. Researchers have mostly used regression algorithm like M5rule, Multivariate linear, Gaussian Process and ZeroR for the calculations.

according to the analysis based on the data collected from 237 student profile from Israeli University. All were student having 101 males and 136 females. Both the researches [39] and [40] have been done on different set of data from different sources. While [39] relies on self-reports by participant, [40] is based on Facebook data. So [40] is more objective approach.

Table 3 depicts the comparative analysis of both the research. This table depicts the connection between Facebook behavior of the user and personality on the scale of five traits of Five Factor Model. It's been observed that earlier work in this field [39] was able to relate only three traits of personality with the Facebook behavior of user. Some results are same as the hypothesis proposed and some were contradictory. However two of the traits of Five Factor Model, agreeableness and openness are found unrelated with the Facebook [39]. Whereas the later research [40] has successfully related with all the five traits of personality model. But again some are found to be harmonious with the hypothesis proposed and some found contradictory. In addition it has suggested some strange outcomes like introvert people provide more access of personal information than extrovert. Women with high neuroticism are good in blogging but men are unaffected from this. So this indicates the other factor affecting this models accuracy [40]. But the sample size used in both of the research is homogenous and small.

Gaussian Process Regression: Researchers like [1, 7] have used this as prediction method. Gaussian process represents underlying function obliquely and rigorously. It has three important components namely: Gaussian distribution, training data, covariance function and supervised learning. Input data is represented as Gaussian distribution. To effectively use Gaussian process we must be able to choose between different mean and covariance function in the light of data. This process is referred as training the model. Based on known functional value of training cases, predictions are made for functional value of new test inputs. Various mathematical formulas are used for these calculations.

M5rule Algorithm: M5rule is a machine learning algorithm. This algorithm uses rules to explain the data and predicts a linear model for each rule. This has been used in [41].

Multivariate Linear: We have multiple dependent variables measured with different metrics and at the same point in time.

ZeroR: ZeroR relies on target and ignores all predictors. It predicts the majority category. Algorithm is to construct a frequency table and select the most frequent value. Normally used for determining baseline performance.

Ridge regression: Used in [106]; it is a technique for analyzing data that suffer from existence of near-linear relationships among independent variable. It helps in solving mathematical problems for which no solution exists due to insufficient information and also in solving problems associated with difficulties in performing matrix inversion.

Linear support vector machine (SVM): [106] has applied this technique for data analysis. SVM is supervised learning model that analyze data and recognize patterns used for regression analysis. Based on the input data it predicts one of the two possible classes .

Logistic regression: Used in research paper [12]; it measures relationship between a dependent variable and one or more independent variable by using probability scores. Logistic function is used to calculate probability.

Clustering Algorithms: These algorithms try to cluster closely connected group of nodes. Example: K-nearest neighbor classifier- This is one of the simplest machine learning algorithms. Most of algorithm in this category use structural information. This has been shown in [99] that content and linkage both kind of information can be processed with clustering algorithm and integrated approach works better.

Hierarchical clustering and Decision Tree: In this algorithm travelling from root node to leaf, one entity will get the prediction results. Majorly structural information is used as input. Researchers have developed group recommendation system for using this model [100].

Bayesian Probabilistic Model: It is based on Bayes' theorem. Based upon the priori probability, Bays classifier uses the Bayesian formula to calculate posterior probably of the prediction event. If the prediction result is discrete, the bays classifier can be applied directly. Otherwise prediction result must be made discrete first [80].

Artificial Neural Network: Artificial Neural Network [81, 82, 83, 84, 85, 86, 87] is a computational model to simulate human brain. Neurons could belong to many interconnected group, input layer, hidden layer or output layer. Input layer is responsible for receiving raw data and transmitting them to next layer. The output layer will give us final prediction result. Choosing network structure and designing the hidden layer is the major task in such algorithms.

3. VARIOUS PERSONALITY ASSESSMENT APPROACH

As mentioned earlier that most of the researchers have used following approach for personality assessment:

- Personality questionnaires.
- Linguistic approach.
- Open Vocabulary approach.
- Based on Internet usage and propensity to use social networking site

- Linkage and content based analysis of social media data.
 - Structure/content analysis of social media data.
 - Behavior/activity based analysis of social media data.

3.1 Personality questionnaires

Most personality assessment has been based on questionnaires with scales designed for specific practical application or to measure constructs derived from personality theory [[38], [68]].

C. Ross et. al. [39] have used 28-item questionnaire related to Facebook uses and was able to relate it with personality traits of the individual.

T. Ryan et. al. [56] have used 124 questions including Facebook usage questionnaire, the big five inventory consist of 44-items that yield Big Five personality trait scores. In this research it is proved that extraverted people are more likely to use Facebook than introverted people. Facebook users have higher level of narcissism. Individuals higher on neuroticism prefer using the wall.

Both the research [39] and [56] uses Facebook usage questionnaire. Some results are harmonious like people high on neuroticism use asynchronous communication and prefer using wall. However, they contradict with each other on some aspects. For example [56] says communicative features of Facebook and it relation with Extraversion is positively correlated however [39] is contradictory on this. This contradiction might be due to different input data. [39] is having homogenous data among university student with 15 male and 82 female while [56] has conducted the research on 1635 self-selected internet users between 18 to 44 years old.

This is one of the simplest approaches towards personality prediction. However, the questionnaires used in many big five personality studies are typically lengthy. Efforts have been made to develop brief scales in psychology. In this context Gosling et al. introduced Ten Item Personality Inventory (TIPI)[75] that include ten questions to determine the Big-Five personality traits. Many studies like [74] use this TIPI to measure self-perceived personality. It is more straightforward to ask a person how extraverted he is than to ask him whether he enjoys the company of others, attends parties frequently, is talkative, outgoing, gregarious, and enthusiastic. Asking multiple questions for one traits is reduced by asking one question to avoid redundancy, boredom to give the answer and to reduce the time, so that more people can participate in survey.

Following are some limitations associated with personality questionnaire approach which may result into inaccurate or suboptimal results:

- People may fake quality for few minutes.
- Time consuming (person by person assessment).
- Expensive (Hiring a professional) as professionals charge per candidates.
- Accuracy as they are being judged by the person, so at times qualification and talent of person matters here. Besides, humans have natural tendency to prejudice so it is prone to human errors.

Table 4: List of work in Personality questionnaire category and the comparison

Publications	Input Data	Method used	Remarks/Conclusions
C. Ross et. al. [39]	Student (15 male, 82 female)	28-Item questionnaire related to Facebook uses.	Contradict with [56] on Extraversion. It is simple but input data size is small so less informative. People high on neuroticism prefer using wall.
T. Ryan et. al. [56]	1635 self-selected Internet user	124 questions including Facebook usage questionnaire and BFI questionnaire.	Extraversion is positively correlated with communicative features of Facebook. People high on neuroticism use asynchronous communication and prefer using wall.
G. Chittaranjan et al. [74]	117 participants, data collected for a continuous period of 17 months	10 questions from TIPI [Ten Item Personality Inventory] developed by Gosling et al. [75]	Less number of questions used for assessment. Might give biased results
M.D. Back et. al. [111]	Participants were 236 OSN users (ages 17–22 years) from US and 103 from Germany	Combination of TIPI , BFI-10 and NEO Five-Factor Inventory.	Users online behavior is their actual behavior not the idealized behavior

3.2 Internet and online social networking site usage

[[55], [56] and [57]] shows that some of big five personality traits are associated with total internet usage and propensity of users to use social media and social networking site.

T. Ryan et. al. [56] have made hypothesis based on Facebook usage. In this research it was proved that individual with higher score on extraversion and narcissism is more likely to be Facebook user; while people high on conscientiousness are Facebook nonusers.

Wang, Chi Yuan et. al. [102] conducted research on 132 students and concluded that those who are emotionally less stable tend to spend more time on Facebook.

T. Correa et. al. [55] concludes that extraversion and openness to experience are positively related to use of social applications on internet, emotional stability was negatively associated. He proved that age and gender also play a role in these dynamics. Extraverted men and women both are likely to be more frequent users of social media, while only men with low on emotional stability were regular users. Extraversion results were particularly true for young adult and openness emerged as important indicator of social media use for mature segment.

Amichai-Hamburger and Vinitzky[40] showed that openness is positively correlated with willingness to use Facebook as a communication tool.

Table 5: List of work based on social networking usage

Publications	Input Data	Method used	Remarks/Conclusion
T. Correa et. al. [55]	959 samples form online survey conducted on US adults. Heterogeneous data in terms of age, race, salary and education.	Instant messaging and Social media use was measured by an additive scale that calculated the frequency of usage. On 10 point response scale.	Extraversion and openness to experience are positively related to use of social applications on internet. Emotional stability and Neuroticism was negatively associated. Research also proved that age and gender play a role in these dynamics.
T. Ryan et. al. [56]	1635 self-selected Internet user	124 questions including Facebook usage questionnaire and BFI questionnaire.	Extraversion is positively correlated with communicative features of Facebook. People high on neuroticism use asynchronous communication and prefer using wall
B. Zhong et. Al. [57]	436 students	Based on time spent in a day over internet for different activity: 1. Accessing Social sites. 2. For study purpose. 3. For other activities.	Relate social media access and thinking and proves its association with personality traits. Data is homogeneous from students only so results can't be generalized.
Wang, Chi Yuan et. al. [102]	132 student with 78% female and 22% male	Survey questionnaire was used and statistical software Statistics Package for Social Scientist was used for quantitative analysis	Concluded that those who are emotionally less stable tend to spend more time on Facebook. Neuroticism is positively related with time spent on Facebook.

Limitation: It focuses upon amount of time spent instead of how individuals are using these.

3.3 Linguistic Approach

[[49], [50]] has shown that linguistic features can be used to predict personality traits. These methods can help predict the personality based on user text in social media sites. This is one of simplest analysis approach towards personality assessment.

Linguistic Inquiry and Word Count (LIWC) tool [51] is a psycholinguistic analysis tool that processes a text document and outputs the percent of words that matches pre-defined categories. It produces statistics on 81 different features of text in five categories including standard counts, psychological processes, relativity, personal concerns and other dimensions. It also counts words based on part of speech and average word length. User text in online social networking sites like Facebook status update can be analyzed with LIWC tool to analyze personality traits. Researchers have used LIWC and made some conclusion on personality traits.

Sumner et.al. [104] has used many linguistic variables, analysis of these provides some significant results about personality traits. Some conclusions drawn from this research are: Extraversion and Conscientious are positively correlated with positive emotion words and talk about family. Agreeableness and Neuroticism is positively associated with number of words per sentence. People low on agreeableness use more anger words.

Mehlet. al. [103].also proved using linguistic features that conscientious people use more positive emotional worlds.

J. Golbeck et. al. [41] have used personal written information through “About Me” and status update. This text was analyzed using LWIC tool . In this research author found that Conscientiousness has the most correlation with linguistic measures. Conscientiousness has positive correlation with words surrounding social process and negatively correlated with frequency of swear words and words . This conclusion is similar to conclusion in [104, 103] about Conscientiousness. Positive emotion words correlates with agreeableness. Surprisingly words expressing anxiety are not positively correlates with neuroticism.

GolnooshFarnadi et. al. [101] has used Facebook status update, network properties and time factor to predict personality traits. Along with other aspect, correlation between LIWC features and personality trait is also performed. As per this research Extrovert tend to use dictionary words, 2nd person and 3rd person words and not swear words. Neurotic user tend to use anger words and less likely to use social interaction words or prepositions. Agreeable users are more likely to use sexual words. Conscientious users uses five senses and prepositions and less likely to use verb. Open users update their statuses by using dictionary words, 2nd person singular and 3rd person plural pronouns.

Table 6: List of work based on Linguistic features

Publications	Input Data	Method used	Remarks
J. Golbeck et. al. [41]	Total 279 subjects, however only 167 could be used for linguistic analysis.	45-question version of the Big Five Personality Inventory and profile information of Facebook users.	Conscientiousness has positive correlation with words surrounding social process and negatively correlated with frequency of swear words. Positive emotion words correlates with agreeableness Surprisingly words expressing anxiety are not positively correlates with neuroticism.
Sumner et.al. [104]	537 Facebook users (349 female and 174 male) from 15 countries. Age range 13-111 years.	44 questions based on BFI and additional questions to capture participant’s concerns on privacy and 79 Facebook data point. Participant’s Facebook post content (wall post and photo description) was analyzed using LIWC	Extraversion and Conscientious is positively correlated with positive emotion words. Agreeableness and Neuroticism is positively associated with number of words per sentence. People low on agreeableness use more anger words
GolnooshFarnadi et. al. [101]	250 Facebook users and 9917 status update from myPersonality project.	Facebook status updates and answers to questionnaires filled by user. 81 features extracted using LIWC	Extrovert tends to use dictionary words, 2nd person and 3rd person words. Open users update their statuses by using dictionary words, 2nd person singular and 3rd person plural pronouns.

Limitation of Linguistic Approach: This research heavily relies on text analysis which is done using some tools and user text input. So the accuracy of the model might be impacted with the performance and accuracy of the tool. Also some misspelling of

the word or sub optimal performance of the tools might provide deviated output.

3.4 Open Vocabulary approach

This is an extension of Linguistic approach. In earlier section publications described use closed vocabulary technique. In

closed vocabulary technique psychologists pick a closed list of words to determine personality traits.

According to H. Andrew Schwartz et. al. [106], in open vocabulary technique; data driven collection of words, phrases and topics are extracted. Lexicon depends on words of the text under analysis. In this research author has used 700 million instances of words, phrases and topics collected from 75000 Facebook users and correlated them with gender, age and personality. Author displayed that this approach yields additional insights than tradition closed vocabulary techniques. There are two basic building blocks for this approach.

- Differential Language Analysis (DLA)
- word cloud-based technique to visualize results of LDA

Differential Language Analysis (DLA) for personality trait correlation consists of 3 steps [105]:

- Linguistic Feature extraction: Extract units like n-grams, topics etc.
- Correlation analysis: Relationship between language use and psychological variables.
- Visualization: Represent output of correlation.

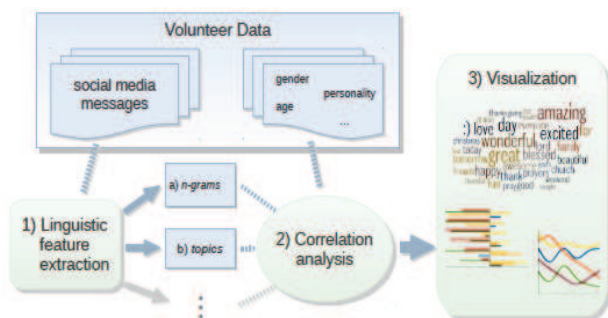


Figure 4: The differential language analysis framework used to explore connections between language and psychological variables. Source: [106]

Word cloud is generated for each of the personality trait. Below figure shows word cloud for extraversion and neuroticism as created in this research [106]. Dominant words in each cluster were consistent with prior lexical and questionnaire work. Some of the topics (automatically clustered sets of words) like “party” are not covered in LIWC but it is taken into account for personality prediction in open vocabulary approach.



Figure 5: Words, phrases, and topics most distinguishing extraversion from introversion and neuroticism from emotional stability. Source [106].

3.5 Linkage and content based analysis of social media data

Unlike questionnaires, social media data like Facebook status update allows researchers to observe people’s behavior as they freely express in their own words. Due to tremendous popularity of social media it provides unprecedented amount of data to researchers in terms of content (text, image audio and video), Profile, Network structure/linkage information and activity based information like tagging, likes. Many researchers have used this data to predict personality traits. Most of the researches use multiple approaches such as combination of linguistic features, internet usage data, questionnaires and linkage data to predict the personality traits. There can be two approaches to analyze Facebook content.

- Structure/content analysis of social media data.
- Behavior/activity based analysis of social media data.

3.5.1 Structure/content analysis of social media data

Structure/Content analysis of social media data enables researchers to predict personality automatically from freely available data. In this section we will discuss about various researches focusing on personality prediction using this data; along with their methodology, results and comparison.

Michal Kosinski et al. [12] showed that Facebook Likes can be used to automatically and accurately predict a range of highly sensitive personal attribute. Study is based on sample of 58,466 volunteers obtained through myPersonality application. Users and their likes were represented as a sparse user-Like matrix. Dimensionality of the matrix was reduced using singular-value decomposition. Numeric variable like age was predicted using linear regression model. Dichotomous variables like gender were predicted using logistic regression.

Personality trait Results based on Facebook Likes: Accuracy of predicting numeric variables as expressed by Pearson correlation coefficient between actual and predicted values. Highest correlation was obtained for age, followed by density and size of friendship network. Closely following were Openness, Extraversion. The remaining personality traits were predicted with somewhat lower accuracy. This research was able to predict some dichotomous variables with good accuracy.

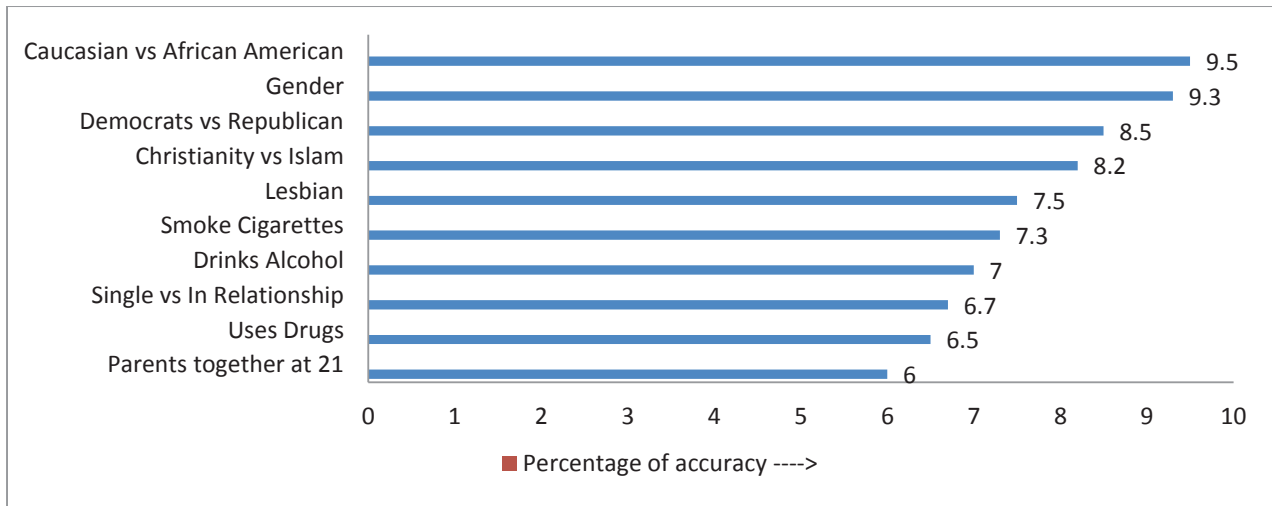


Figure 6: Prediction using Facebook like data from [12].

Golbeck et al. [41] attempted to predict personality from Facebook profile information using machine learning algorithms. Data (167 statistics including structural information, personal information, activities and preference, language features) was collected using Facebook application and 45-question version of Big Five personality inventory to users [71]. In this research 74 features per user is used to predict the score of given personality feature. Regression analysis was performed in Weka [67] using two algorithms: M5' Rules [72], a rule-based variation of the M5' algorithm [73] and Gaussian Processes. Mean absolute error for each personality factor was around 11%.

Facebook Network Structure Features: Extroverts tend to have more number of friends but their network tend to be sparse. Openness is also negatively correlated with density.

Facebook profile information: Results show significant positive correlation between neuroticism and subject's last name. It also shows that women are more conscientious, agreeable and neurotic than men. Users providing a website are positively correlated with openness.

Limitations: Number of words available per person to analyze was very small. Personality prediction using linguistic approach might not work as well as it did in other researches. Input data is limited to small sample which poses a question mark on reliability.

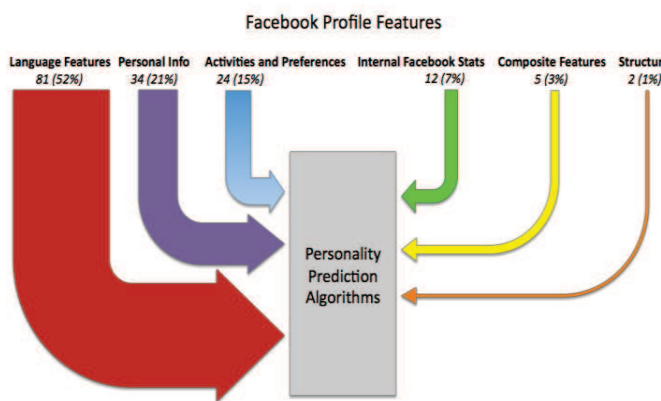


Figure 7: Facebook profile features used to predict personality traits. Source [41]

Linguistic Feature: Results discussed earlier in section 3.3.

Golnoosh Farnadiet. al. [101] has researched on 250 Facebook users and 9917 status updates frequency and time of posting collected from myPersonality project. Each user has filled a questionnaire as well. Based on the answer each user is assigned some personality traits. This research aims to predict those traits. Following features are used for prediction.

- LIWC features from Facebook status update and other text.
- Social Network features
- Time Related Features
- Others

Three algorithm used and results are compared; namely Support Vector Machine with a linear kernel (SVM), Nearest Neighbor with k=1 (kNN) and Naive Bayes (NB).

Linguistic/LIWC features: Results discussed earlier in section 3.3.

Social Network features: All social features are significantly correlated with extraversion. Result shows positive correlation with network size and negative correlation with density. These results are in lined with [41]. Conscientiousness is positively correlated with network size and negatively correlated with density. Agreeableness and Extraversion have negative correlation with transitivity while neuroticism is positively correlated. Extraversion presents opposite sign with respect to correlation of social network features with neuroticism.

Time Related Features: Results show conscientious users are less likely to update status between 00 am to 11 am.

Amichai-Hamburger and Vinitzky[40] used actual Facebook profile information. Data collected from 237 students (101 male, 136 female). Participants were requested to complete NEO-PI-R [107] to assess personality based on five factor model. Basic information, personal information, contact information and education & work information from Facebook profile is also used as input. Multivariate regression is used to predict and correlate the results.

Facebook profile information: Highly neurotic person prefer posting their photos on their Facebook profile and are less inclined to upload other's photo. People high and low on neuroticism prefer to share more basic information than moderates. Highly extroverted people demonstrated lower use of personal information. Individuals high on agreeableness have used less page features. People low and high on agreeableness tend to upload more pictures than moderates. Among females agreeableness is positively correlated with picture used. Individual high on conscientiousness demonstrated less use of picture upload feature.

Facebook Network Structure Features: Conscientiousness and Extraversion is positively correlated with number of friends.

Y. Bachrach et al. [24] uses both self-reported pattern of Facebook usage and actual Facebook profile feature with data

set of 180,000 users. Users agreed to participate on the questionnaire based on the standard five factor model questionnaire [93] [94] to obtain feedback about personality based on response. User's profile information coupled with input to the questionnaire enabled researcher to determine personality traits. The features included in this research were:

- Number of friends.
- Number of groups joined.
- Number of Facebook likes.
- Number of photos uploaded.
- Number of status updated
- Number of times others tagged user in photos

Plots are produced presenting correlations between user's personality and the property of their Facebook profile such as size and density of friendship network, number of uploaded photos, number of events attended, number of group membership and number of time user has been tagged in photos. This is called "Clustered Scatter Plots" by the author. Further a multivariate regression is used to predict the personality of the user given their Facebook profile. Following hypothesis was tested in this research:

- Openness and Neuroticism are positively correlated with number of status update, photos, groups and likes of an individual.
- Conscientiousness is negatively correlated with all aspects of Facebook use: number of friends, likes, photos etc.
- Extraversion is positively correlated with all aspects of Facebook use
- Agreeableness is positively correlated with the number of friends, groups and likes.

Table 7: Personality trait correlation as per [24]

Personality Trait	Personality Behavior	Facebook Behavior [24]
Openness	High openness means appreciation for art, adventure, more creative and has new ideas.	Positively correlated with number of likes, group association and number of status update.
Conscientiousness	Spontaneous approach in life, well organized, reliable and consistent.	Negatively correlated to number of likes and group associations but positively related to photos uploaded.
Extraversion	Prone to live in external world, positive emotions, expressive, friendly and socially active	Positively associated with status update, tendency to use "Like" and Number of Facebook friend and Facebook group.
Agreeableness	Friendly and compassionate, cooperative	Negatively correlated with the number of likes but positively co related with tagged.
Neuroticism	Related to emotions and mood swing. Negative emotions such as anger, anxiety, depression	Positively correlated with number of likes and number of groups.

Clustered Scatter Plots doesn't measure the strength and significance of relationship it display. For statistical significance two tests were conducted by author *t-distribution test* and *Mann-Whitney-Wilcox test*.

Predicting personality: Multivariate linear regression is used to determine Coefficient of determination R^2 and root mean squared error (RMSE) to prove that Extraversion, Neuroticism can be predicted with reasonable accuracy. Conscientiousness

and openness are less accurate. Agreeableness is hardest to predict using this model. This study shows the correlation between personality traits and pattern of social network using a very large sample as compared to earlier work. It further shows that by combining several features, we can make relatively accurate prediction regarding an individual's personality.

Table 8: Predicting personality traits using Facebook features through multivariate linear regression. Source [24]

Trait	R ²	RMSE
Openness	0.11	0.29
Conscientiousness	0.17	0.28
Extraversion	0.33	0.27
Agreeableness	0.01	0.29
Neuroticism	0.26	0.28

Limitations: Data used may suffer from self-selection bias. Only high level aggregate features from Facebook are used for this study.

Personality trait assessment analysis from various studies based on Facebook Features:

Table 9: List of representative work for various Facebook features “Like”, “Tag” and “Status update time”

Author	Facebook Like	Facebook Tag	Temporal Features
Michal Kosinski et al. [12]	Openness and Extraversion can be accurately predicted using Facebook like	Not used	Not used
Golbeck et al. [41]	Not used	Not used	Not used
GolnooshFarnadi et. al. [101]	Not used	Not used	Results show conscientious users are less likely to update status between 00 am to 11 am
Amichai-Hamburger et. al. [40]	Not used	Not used	Not used
Y. Bachrach et al. [24]	Openness, Neuroticism and Extraversion is positively correlated with number and tendency to use like while Conscientiousness and Agreeableness is negatively related.	Agreeableness is positively correlated with tagged	Not used

Table10: List of representative work for Facebook Profile.

Author	Prediction Algorithm	FB Profile
Michal Kosinski et al. [12]	Logistic/Linear Regression	Not used
Golbeck et al. [41]	M5’ Algorithm and Gaussian Process	Positive correlation between neuroticism and subject’s last name. It also shows that women are more conscientious, agreeable and neurotic than men. Users providing a website are positively correlated with openness.
GolnooshFarnadi et. al. [101]	Support Vector Machine with a linear kernel (SVM), Nearest Neighbor with k=1(kNN) and Naive Bayes (NB)	Not used
Amichai-Hamburger and Vinitzky[40]	Analysis of covariance (ANCOVA), t-test, two-way ANOVA and Hierarchical multiple regression	Highly neurotic person prefer posting self-photos on their Facebook profile and are less inclined to upload other’s photo. People high and low on neuroticism prefer to share more basic information than moderates. Highly extroverted people demonstrated lower use of personal information. Individuals high on agreeableness have used less page features. People low and high on agreeableness tend to upload more pictures than moderates. Among females agreeableness is positively correlated with picture used. Individual high on conscientiousness demonstrated less use of picture upload feature.
Y. Bachrach et al. [24]	Multivariate Linear Regression	Openness and Extraversion is positively correlated with number of status update. Conscientiousness is positively correlated with photos uploaded.

Table11: List of representative work for Facebook Network Structure.

Author	Input data	FB Network Features
Michal Kosinski et al. [12]	Facebook “like” of 58,466 volunteers from the United States	Not used
Golbeck et al. [41]	167 Users. Data was collected using Facebook application and 45 questions BFI.	Extroverts tend to have more number of friends but there network tend to be sparse. Openness is also negatively correlated with density.
GolnooshFarnadi et. al. [101]	250 Facebook users and 9917 status updates frequency and time of posting collected from myPersonality project	Extraversion and Conscientiousness shows positive correlation with network size and negative correlation with density. Agreeableness and Extraversion have negative correlation with transitivity while neuroticism is positively correlated. Extraversion presents opposite sign with respect to correlation of social network features with neuroticism.
Amichai-Hamburger and Vinitzky[40]	Self-reported answer for NEO-PI-R and Facebook information of 237 students.	Conscientiousness and Extraversion is positively correlated with number of friends.
Y. Bachrach et al. [24]	180,000 Users. Self-reported pattern of Facebook usage and actual Facebook profile feature.	Extraversion, Neuroticism and openness is positively correlated with number of Facebook groups. Extraversion is positively correlated with number of friends. Conscientiousness is negatively correlated

Table12: List of representative work for Facebook Status Updates.

Author	Input Data	Methodology	Purpose	Remarks
G. Farnadi1 et al. [101]	250 Facebook users, 9917 status Updates, frequency and time of posting collected from myPersonality project.	Based on questionnaire each user has assigned a trait, which verified by the research using three machine learning algorithms, SVM,KNN and NB	What part of personality: Relate personality traits of Big Five Factor with status Updates, frequency and time of posting.	Status updates do represent the personality trait of author. Combining status update with other features may improve the prediction results.
H.A. Schwartz et al. [106]	75,000 volunteers, 700 million words, phrases, and topic and correlated them with gender, age and personality.	LIWC + Open vocabulary approach. SVM + Ridge Regression Algorithms.	Predict age, gender and Big Five Personality traits.	Unlike questionnaire the benefit of Facebook is it allows user to freely present themselves. Used open Vocabulary.
E.Y.L. Ong et al. [109]	Frequency of Facebook Status Update, number of friends and number of photos user present.	Made hypothesis and proved that by regression analyses.	Examine the negative emotions and extraversion.	more narcissistic update Facebook status more frequently
L. Han et al. [110]	185 Facebook users, their status updates, friend list, network size and density.	Asked to fill questionnaire and then to share the status updates.	Analyzed positive and negative emotion in their past status updates.	Less emotional stability more use of Facebook status updates.

3.5.2 Behavior/activity based analysis of social media data

In the discussions so far, we have briefed various researches on the linkage and content based social media data. In this section we will discuss about researches focusing on behavior or activities based predictions.

Personality prediction through behavior analysis is not new. Researchers tried to assess personality with the study of

behavioral characteristics. Pianesi et al. showed that personality traits in a meeting environment can be detected using audio–visual features and supervised learning [77]. In this case personality of the participant was revealed by how participant spoke and interact. Similarly Mairesse and Walker describe an automatic procedure using NLP and audio features to detect the Big-Five traits from conversation extracts [49, 76].

These studies suggest that if we can analyze person's behavior based on data available from online social networks then this might enable us with a new perspective towards personality trait analysis. Some studies are done on personality trait prediction using social media and social network behavior like network position, behavior towards friends and followers. Asendorpf et al. [64] show impact of personality on relationships and social behavior. Klein et al. [21] study network centrality in various networks like in Friendship network, Advice network, Adversarial network. Klein showed that neuroticism and openness are negatively correlated with friendship centrality. Openness and Extraversion is positively correlated with adversarial centrality. S. Adali [17] shows that structural network features and ties can be used for personality assessment.

Burt et al. [62] show that personality varies with structural holes and correlates with network position.

Golbeck et al. [41] have displayed Activities and interest based correlation: As per this research extraversion is positively correlated with the length of reported activities. Openness is positively correlated with length of favorite book lists of user.

C. Sumner et al. [104] showed that there are significant differences in the language used between Facebook biographies, wall post and photo descriptions; which proves that linguistic analysis based on user behavior towards different Facebook features provides some extra insight for personality prediction. Author has used zero-order Spearman's correlation to measure the correlation with Big Five Personality traits. With this study they were able to show some unusual results while analyzing language in photo description and Facebook biographies. They also showed that language used in Facebook wall post and across the whole Facebook is same and obtains the similar results for personality score. 537 Facebook users from 15 countries participated in this study.

S. Adali et al. [1] developed some measures based on one's behavior towards friends and follower. This research shows that behavioral features can be used in determining personality. Authors analyzed behavior of individual in social network and considered actions in following group:

Table13: Behavior Features as per [1]

NET	Network Bandwidth	Based upon amount /distribution of activities(uniform/bursty) and size of social network
MSG	Message Content	Type of Message (Forwarded, containing URL etc.)
PAIR	Pair behavior	Behavior towards friends/followers
REC	Reciprocity of Action	Measures reciprocation by their friend
INF	Informative ness	How informative are Behavior features across all friends or group.
HOM	Homophily	Previous features computed on person's friend to understand his/her social circle

NET and MSG show the behavior of individual in public and towards friends. PAIR is indicative of decreasing social distance with respect to specific individuals. INF captures if individual is social towards some or all friends. Research was done on 71 users and all friends and followers of these users. Public tweets of given set of users were also collected. Users also completed BFI questionnaire that served as reference.

Computation: *Reciprocity* (REC) is measured using entropy. Entropy is highest when amount of action is same between two individuals. *Pairwise features* (PAIR): Captures behavior with respect to specific friend and follower. Whether they exchange direct message with each other or they forward message from a friend or follower. If propagation is not reciprocal the source is likely not a social tie. Various pair wise features like mean response time, mean balance in conversation and many others for a user towards any of the people he/she communicate with are calculated.

Distribution of behavior across friends (KL) with this author tried to find to which degree the behavior of individual differs from friend to friend (INF), especially in terms of timing and number of message. To compute the divergence from uniform, Kullback-Leibler measure from information theory is used. Higher value of KL indicates that given distribution of behavior towards friends is very informative. For all features KL divergence is computed to observe to which degree behavior of individual differs from one friend to another.

Tell me who your friends are (FF): Computes the same feature for A's friends towards their friend (HOM). Homophily hypothesis [65] is also tested with this. Normative behavior [66] pattern is computed by computing the mean of all features for each friend B of user A.

Analysis: Forward subset selection (FSS) based regression is used to identify the feature that is best predictor of the specific personality trait based on behavior for each personality trait.

Prediction: To predict the score of given personality feature regression analysis in Weka [67] is done. Algorithm used: Gaussian Process and ZeroR. Results were compared against text based prediction LIWC [51] and proved that behavioral feature perform equally well.

With this research it was shown that social behavior can be predicted based on person's behavior towards friends and follower.

Limitation: Behavior expression is complicated as same person might show different behavior on different situations. Situation leads to activation of certain personality trait which determines behavioral expression at certain moment [61]. So at different time based on situation expression of personality by behavior may display very different characteristics.

Performance is low for some of the personality features especially Neuroticism and Extraversion.

4. A DISCUSSION

Through this survey we tried to explore various aspects of personality prediction using online social networking sites.

For personality trait assessment; some researchers rely on text analysis others on static and structural information from social media sites and others on methods which focus on interpersonal behavior analysis in online social networking sites. They also tried to combine these approaches to get better results.

Researchers see lot of benefits for the society if they are able to get the results with accuracy. Some of these applications are discussed below. However, there are many limitations that researchers face before they can generalize the theory they developed using the data from online social networking sites. Some of these are discussed in section 4.2. Study about personality prediction using online social network is not yet in mature state and there is still a lot of scope for improvement and future research. Some of which is discussed in section 4.3.

4.1 Applications

As per the research [52], [53] users prefer the interfaces which most closely resembles with their own personality. This signifies the need to predict personality trait and in turn prepare a personality-oriented interface to make user most receptive. This idea can be used in many fields including marketing and advertising. Researchers demonstrated relation between marketing technique and consumer personality [54]. So if we can analyze and categorize the personality traits of the user then based on it personalized features can be provided to the individual as an interface. This will help marketing agents in building good rapport with the customers. Research paper [31] focuses on marketing using social media.

Personality traits predicted with the application of social media can be used for building better and dynamic recommender system. For example it can be used in friend recommendation system of the social sites [29]. Social networking sites are now being used for dating suggestions [18], health care [32], pre-employment screening and organization development [6]. A good recommender system in social sites if accurately implemented then it can do magical transformation in all the above areas.

Based on the performance of these works if we are able to get accurate results then it can be helpful in crime investigations, prevention and intention analysis which in turn can further empower Law and Order in the society. However there is lot of research yet to be done to reach to this level of accuracy.

Based on the user's activity on Facebook we can identify the negative emotions such as depression, humiliation and cyber bullying over a group of people and can stop it.

4.2 Limitations and Challenges

Despite having all these benefits of personality traits identifications using online social networking sites there are some significant limitations as well.

Most significant problem is about the sample data available as input. If there is a doubt about the integrity of the input data then the results can't be considered reliable. Social networking sites like Facebook doesn't directly provide needed data to researchers due to privacy reasons. Researchers need to rely on some applications or selected user inputs for this purpose. However this kind of sample data collection is suffered from self-selection bias. Also people tend to put forward idealistic behavior when they know that data is being recorded for some research purpose. Data would be most reliable if it is directly obtained from natural and prolonged behavior among users. We can hope that in future some associations or agreements between social networking sites and academic researchers will help in getting data from natural settings.

Human behavior and reaction is at times situational. This necessitate to take into account a lot many other psychological

factors before we can claim the prediction results to be accurate based on social network site behavior. In this sense behavior based approach is dependent on psychological and sociological researches as well.

There are millions of fake profiles in Facebook which might impact the accuracy of analysis.

Another limitation is about the algorithm used for getting prediction result. Although there is already lot of research done on prediction algorithm from simpler regression algorithms to complex machine learning and artificial neural network; however right use of the algorithm is very important which might affect to the result.

User behavior in social networking sites might be very dynamic and for the researches where we are taking structural information like number of likes and number of friends etc. in account these factors might already get change quite heavily in very short span of time thus making the results less reliable.

4.3 Open Research Questions and Novel Solutions

For simplicity most of the researches focuses on size and density of network, number of Likes, status updates etc. but such study doesn't focus upon detailed information like type of object "Liked" and type of group joined by a Facebook user. Most of the studies are based on quantitative data like number of likes, number of friends, number of photos and time spent on Facebook rather than qualitative features. Whether such information can be used to predict personality accurately and closely remains an open question.

Whether personality can be predicted using other potentially observable online behavior such as a user's internet browsing or web search history is another open research question [24].

Interpersonal behavior analysis based on social media data by looking at personality scores between friends can further present with plethora of opportunities. There is a scope of future research for devising parameters towards interpersonal behavior analysis. Identifying the connection between personalities, trust [70], strength of relationship [69] and other related factors and then selecting best machine learning algorithm for predictions could be the future research questions in this direction.

There is still ample scope for social behavior based analysis. Performance need to be improved for several personality features [1]. In time of uncertainty and risk, expression of personality by behavior might display very different characteristics [1]. Study of this aspect is a topic of future research.

With the evolution of smart phones, online social network site and internet usage pattern is changed quite drastically. Now users are connected all the time using the phones. Apart from this, smart phones also support many interesting features like GPS these features might add another aspect towards personality trait assessment approaches. Application of smart phone features to determine personality trait is open research topic.

With all the Facebook activities researchers try to relate the user's personality with the big five factor model but there could be other motivations like entertainment, need for social support etc. and hindrances like social norms, privacy issue and organization policy etc. which might be helpful in understanding

Facebook usage and these seems to be independent of five personality traits.

People keep on updating and changing their profile information over the period of time, this could also be one of the key to predict the personality of the user [40].

C. Sumner et.al. [104] have displayed that there is significant correlation between Facebook activity and personality type. He also noted that number of comments a person receives is correlated with number of friends a person has rather than extraversion. Further analysis is needed to support such arguments and hypothesis.

Multimedia contents are growing rapidly in online social networking sites. Multimedia content analysis might give another direction towards personality prediction. There is lot of research scope in personality prediction using multimedia content. N. Ramzan et al. [13] gives an overview of key theoretical and empirical advances in the current decade related to social media retrieval by considering not only the multimedia content analysis but also behavior analysis of the users.

Prediction algorithm used also play an important role towards the accuracy of the results by extracting meaningful data from large quantities of data. Most researchers use linear regression analysis. More prediction methods and combination of methods might lead to more accurate results [81]. More fine-grained feature selection for prediction can improve the results. This algorithm and models can help us determining the appropriate feature.

In the research paper either the Facebook data is from the limited number of users [35] or from the similar group of people like university students [40] or it suffers from self-selection bias. Users from similar group tend to influence each other and they try to show ideal behavior when they know that their inputs are recorded for some academic research purpose. This put the question on the integrity of the data. If academic institutes and researchers can directly access to social networking database without the users knowing then predictions will be more reliable but it makes it challenging in terms of security and privacy.

For some cases researcher have observed the behavior of users for period of time in these cases user may influence and change his behavior for some time. And also personality varies over time [98]. Persistency can be taken into account by looking into the data sample from longer duration. So to further improve the accuracy of the prediction from the user's data it is important to collect the user's data for the longer duration.

5. CONCLUSION

Social behavior 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 behavior aspect of social media like message content & type, behavior towards friends & follower, response time etc. to correlate with personality trait. It was also proved that network structure property like ties and structural hole can also help in personality assessment.

With the ability to predict user's personality traits several opportunities are opened for personalized services and products. Some of the opportunities are discussed on this paper however there is much work to be pursued in this area. There are some limitations in terms of accuracy when we try to correlate social network data with personality but personality prediction over a large data set through machine learning algorithm provide low cost and high efficient model. There is still a lot of scope for future research in getting the data with better accuracy; some of which we discussed in section "A DISCUSSION" under subsection "Open Research Questions and Novel Solution" of this paper. We discussed in brief about the researches which studied about predicting personality traits using online social networking sites, their benefits and limitations. These researches have provided us a low cost and efficient solution for personality trait assessment.

6. REFERENCES

- [1] S. Adali and J. Golbeck, "Predicting Personality with Social Behavior", Proc. the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012), pp. 302-309, 2012.
- [2] H.V. Zhao, W.S. Lin and K.J.R Liu, "Behavior modeling and forensics for multimedia social networks" Proc. Signal Processing Magazine, IEEE (Volume:26 , Issue: 1), pp. 118-139, 2009.
- [3] J. Sang, "Collective search and recommendation in social media", Proc. the 20th ACM international conference on Multimedia, pp. 1421-1424, 2012.
- [4] J.C. Prates, E. Fritzen, S.W.M. Sequeira, M. H. L.B. Braz and L.C.V.D. Andrade, "Contextual web searches in Facebook using learning materials and discussion messages", Computers in Human Behavior, Volume 29, Issue 2, pp. 386-394, 2013
- [5] R. Srivastava, J. Feng, S. Roy, S. Yan and T. Sim, "Don't ask me what i'm like, just watch and listen", Proc. the 20th ACM international conference on Multimedia, pp. 329-338, 2012.
- [6] N. Albaloooshi, N. Mavaridis and N. Al-Qirim, "A survey on social networks and organization development", 2012 International Conference on Collaboration Technologies and System(CTS), IEEE, pp. 539-545, 2012.
- [7] J. Golbeck, C. Robles, M. Edmondson and K. Turner, "Predicting Personality from Twitter", Privacy, security, risk and trust (passat), 2011 IEEE third international conference on and 2011 IEEE third international conference on social computing (socialcom), pp. 149-156, 2012.
- [8] G. Chittaranjan, J. Blom and D. Gatica-Perez, "Mining large-scale smartphone data for personality studies",

Journal Personal and Ubiquitous Computing Volume 17 Issue 3, pp. 433-450, 2013.

- [9] N. Caporusso, "Personality-aware interfaces for learning applications", Proc. the 37th annual ACM SIGUCCS fall conference, pp. 189-196, 2009.
- [10] S.Y. Ho, M.J. Davern and K.Y. Tam, "Personalization and choice behavior- the role of personality traits", ACM SIGMIS Database Volume 39 Issue 4, pp. 31-47, 2008.
- [11] R. Cross, S. Borgatti & A. Parker, "Making invisible using social network analysis to support strategic collaboration", CALIFORNIA MANAGEMENT REVIEW, Volume 44 Issue 2, 2002.
- [12] M. Kosinski, D. Stilwell and T. Graepel, "Private traits and attributes are predictable from digital records of human behavior", Proc. the National Academy of Science of the United State of America, volume 110, Issue 15, pp. 5802-5805, 2013.
- [13] N. Ramzan, P. Cui, F. Wang and S. Yang "Social and behavioural media access", Proc. the 2011 ACM workshop on Social and behavioural networked media access, pp. 65-70, 2011.
- [14] D. Xu, L. Zhang and J. Luo, "Understanding multimedia content using web scale social media data", Proc. the international conference on Multimedia, pp. 1777-1778, 2010.
- [15] L. Argyriou, C.Z. Patrikakos, S.C. Porter, N. Papaoulakis and C. Androulaki, "Using media related user profiles to personalize multimedia access over social networks", Proc. the 2011 ACM workshop on Social and behavioural networked media access, pp. 9-14, 2011.
- [16] A. Archambault and J. Grudin, "A Longitudinal Study of Facebook, LinkedIn, & Twitter Use", Proc. the SIGCHI Conference on Human Factors in Computing Systems, pp. 2741-2750, 2012.
- [17] S. Adali, F. Sisenda and M. Magdon-Ismail, "Actions Speak as Loud as Words: Predicting Relationships from social Behavior Data", Proc. the 21st international conference on World Wide Web, pp. 689-698, 2012.
- [18] A.T. Fiore, L.S. Taylor, G.A. Mendelsohn and M. Hearst, "Assessing Attractiveness in Online Dating Profiles", Proc. the SIGCHI Conference on Human Factors in Computing System, pp. 797-806, 2008.
- [19] K.M. Lee and C. Nass, "Designing Social Presence of Social Actors in Human Computer Interaction", Proc. the SIGCHI Conference on Human Factors in Computing Systems, pp. 289-296, 2003.
- [20] C. Bell, C. Fausset, S. Farmer, J. Nguyen, L. Harley and W.B. Fain, "Examining Social Media Use Among Older Adults", Proc. the 24th ACM Conference on Hypertext and Social Media, pp. 158-163, 2013.
- [21] J. Staiano, B. Lepri and N. Aharony, "Friends don't Lie : Inferring Personality Traits from Social Network Structure", Proc. the 2012 ACM Conference on Ubiquitous Computing, pp. 321-33-, 2012.
- [22] R. Gao, B. Hao, S. Bai, L. Li, A. Li and T. Zhu, "Improving User Profile with Personality Traits Predicted from Social Media Content", Proc. the 7th ACM conference on Recommender systems, pp. 355-358, 2013.
- [23] A. Burton-Jones and G.S. Hubona, "Individual Differences and Usage Behavior: Revisiting a Technology Acceptance Model Assumption", ACM SIGMIS Database, Volume 36, Issue 2, spring, pp. 58-77, 2005.
- [24] Y. Bachrach, M. Kosinski, T. Graepel, P. Kohli and D. Stilwell, "Personality and Patterns of Facebook Usage", Proc. the 3rd Annual ACM Web Science Conference, pp. 24-32, 2012.
- [25] D. Quercia, R. Lambiotte, D. Stilwell, M. Kosinsky and J. Crowcroft, "The Personality of Popular Facebook Users", Proc. the ACM 2012 conference on Computer Supported Cooperative Work, pp. 955-964, 2012.
- [26] F. Celli and L. Rossi, "The Role of Emotional Stability in Twitter Conversations", Proc. the Workshop on Semantic Analysis in Social Media, pp 10-17, 2012.
- [27] J. Ahn, "What Can We Learn from Facebook Activity? Using Social Learning Analytics to Observe New Media Literacy Skills", Proc. the Third International Conference on Learning Analytics and Knowledge, pp. 135-144, 2013.
- [28] N.K. Lankton and D.H. Mcknight, "What Does it Mean to Trust Facebook? Examining Technology and Interpersonal Trust Beliefs", ACM SIGMIS Database, Volume 42, Issue 2, pp. 32-54, 2011.
- [29] A. Chin, B. Xu and H. Wang, "Who Should I Add as a 'Friend'? A Study of Friend Recommendations using Proximity and Homophily", Proc. the 4th International Workshop on Modeling Social Media, Article 7, 2013.
- [30] M. Burke, R. Kraut and C. Marlow, "Social Capital on Facebook: Differentiating Uses and Users", Proc. the SIGCHI Conference on Human Factors in Computing Systems, pp. 571-580, 2011.
- [31] A. Lugmayr, "Social Media Marketing for Financial Products", Proc. the 16th International Academic MindTrek Conference, pp. 231-232, 2012.
- [32] S. Picazo-Vela, M. Fernandez-Haddad and L. F. Luna-Reyes, "IT's alive!! Social Media to Promote Public Health", Proc. the 14th Annual International Conference on Digital Government Research, pp. 111-119, 2013.
- [33] J. McCarthy, E. Pioch, J. Rowley and C. Ashworth, "Social network sites and relationship marketing communications: challenges for UK football clubs", Proc. the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, pp. 145-152, 2011.
- [34] AT and T "Corporate Social Networking in Europe", Independent Market Research Report, 2008.
- [35] L.M. Batrinca, N. Mana, B. Lepri, F. Pianesi and N. Sebe, "Please, Tell Me About Yourself: Automatic Personality Assessment Using Short Self-Presentations", Proc. the 13th international conference on multimodal interfaces, pp. 255-262, 2011.
- [36] P.T. Costa and R.R. McCrae, "Revised NEO Personality (NEO-PI-R) and NEO Five-Factor Inventory professional manual. Psychological Assessment Resources, Odessa, FL, 1992.
- [37] J. M. Digman. "Personality structure: Emergence of the five-factor model", Annual Review of Psychology, Volume 41 Issue 1, pp. 417-440, 1990.

- [38] R. McCrae and O. John, "An introduction to the five-factor model and its applications", *Journal of personality*, Volume 60 Issue 2, pp. 175–215, 1992.
- [39] C. Ross, E.S. Orr, M. Sasic, J.M. Arseneault, M.G. Simmering, and R.R. Orr, "Personality and motivations associated with facebook use", *Computers in Human Behavior*, Volume 25 Issue 2, pp. 578–586, 2009.
- [40] Y. Amichai-Hamburger and G. Vinitzky, "Social network use and personality", *Computers in Human Behavior*, Volume 26 Issue 6, pp. 1289–1295, 2010.
- [41] J. Golbeck, C. Robles, and K. Turner, "Predicting personality with social media", In *CHI*, pp. 253–262, 2011.
- [42] S.D. Gosling, A.A. Augustine, S. Vazire, N. Holtzman, and S. Gaddis, "Manifestations of personality in online social networks: Self-reported facebook-related behaviors and observable profile information. *Cyberpsychology, Behavior, and Social Networking*, 2011.
- [43] Y. R. Tausczik and J. W. Pennebaker, "The psychological meaning of words: Liwc and computerized text analysis methods", *Journal of Language and Social Psychology*, Volume 29 Issue 1, pp. 24–54, 2010.
- [44] E. Tupes and R. Christal, "Recurrent personality factors based on trait ratings", *Journal of Personality*, Volume 60 Issue 2, pp. 225–251, 1992.
- [45] O. John, "The Big Five factor taxonomy: Dimensions of personality in the natural language and in questionnaires", *Handbook of personality: Theory and research*, Volume 14, pp. 66–100, 1990.
- [46] J. Digman, "Personality structure: Emergence of the five-factor model", *Annual review of psychology*, Volume 41 Issue 1, pp. 417–440, 1990.
- [47] R. McCrae, "Why I advocate the five-factor model: Joint factor analyses of the NEO-PI with other instruments", *Personality psychology: Recent trends and emerging directions*, pp. 237–245, 1989.
- [48] R. McCrae and P. Costa, "Personality in adulthood: A five-factor theory perspective", *The Guilford Press*, 1990.
- [49] F. Mairesse, M. Walker, M. Mehl, and R. Moore, "Using linguistic cues for the automatic recognition of personality in conversation and text", *Journal of Artificial Intelligence Research*, Volume 30 Issue 1, pp. 457–500, 2007.
- [50] J. Pennebaker and L. King, "Linguistic styles: Language use as an individual difference", *Journal of personality and social psychology*, Volume 77 Issue 6, pp. 1296–1312, 1999.
- [51] J. Pennebaker, M. Francis, and R. Booth, "Linguistic inquiry and word count: LIWC 2001. Mahway: Lawrence Erlbaum Associates, 2001.
- [52] C. Nass and K. M. Lee, "Does computer-generated speech manifest personality? an experimental test of similarity-attraction", *Proc. of the SIGCHI conference on Human Factors in Computing systems*, pp. 329–336, 2000.
- [53] A. Karsvall, "Personality preferences in graphical interface design", *Proc. of the second Nordic conference on Human computer interaction*, pp. 217–218, 2002.
- [54] G. Odekerken-Schroder, K.D. Wulf, and P. Schumacher, "Strengthening outcomes of retailer-consumer relationships: The dual impact of relationship marketing tactics and consumer personality", *Journal of Business Research*, Volume 56 Issue 3, pp. 177–190, 2003.
- [55] T. Correa, A.W. Hinsley, and H.G. De Zuniga, "Who interacts on the web?: The intersection of users 'personality and social media use", *Computers in Human Behavior*, Volume 26 Issue 2, pp. 247–253, 2010.
- [56] T. Ryan and S. Xenos, "Who uses facebook? An investigation into the relationship between the big five, shyness, narcissism, loneliness, and facebook usage", *Computers in Human Behavior*, 2011.
- [57] B. Zhong, M. Hardin, and T. Sun, "Less effortful thinking leads to more social networking? The associations between the use of social network sites and personality traits", *Computers in Human Behavior*, 2011.
- [58] D. Schmitt, J. Allik, R. McCrae, and V. Benet-Martinez, "The geographic distribution of Big Five personality traits: Patterns and profiles of human self-description across 56 nations," *Journal of Cross-Cultural Psychology*, Volume 38 Issue 2, pp. 173, 2007.
- [59] B. De Raad, "The Big Five personality factors: The psycholexical approach to personality", *Hogrefe & Huber Publisher*, 2000.
- [60] L. Goldberg, "From Ace to Zombie: Some explorations in the language of personality," *Advances in personality assessment*, volume 1, pp. 203–234, 1982.
- [61] K. M. Sheldon, R. M. Ryan, L. J. Rawsthorne, and B. Ilardi, "Trait self and true self: Cross-role variation in the big-five personality traits and its relations with psychological authenticity and subjective well-being", *Journal of Personality and Social Psychology*, volume 73 Issue 6, pp. 1380–1393, 1997.
- [62] R. S. Burt, J. E. Jannotta, and J. T. Mahoney, "Personality correlates of structural holes," *Social Networks*, volume 20 Issue 1, pp. 63 – 87, 1998.
- [63] K. Klein, B. Lim, J. Saltz, and D. M. Mayer, "How do they get there? An examination of the antecedents of centrality in team networks", *Academy of Management Journal*, Volume 47 Issue 6, pp. 952–963, 2004.
- [64] J. Asendorpf and S. Wilpers, "Personality effects on social relationship", *Journal of Personality and Social Psychology*, Volume 74 Issue 6, pp. 1531–1544, 1998.
- [65] M. McPherson, L. Smith-Lovin, and J. M. Cook, "Birds of a feather: Homophily in social networks," *Annual Review of Sociology*, Volume 27, pp. 415–444, 2001.
- [66] R. A. Sherman, C. S. Nave, and D. C. Funder, "Properties of persons and situations related to overall and distinctive personality-behavior congruence", *Journal of Research in Personality*, Volume 46 Issue 1, pp. 87 – 101, 2012.
- [67] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. Witten, "The WEKA data mining software: An update", *ACM SIGKDD Explorations Newsletter*, Volume 11 Issue 1, pp. 10–18, 2009.
- [68] Goldberg, L. R. (1971). A historical survey of personality scales and inventories. In P. McReynolds (Ed.), *Advances in psychological assessment* (Vol. 2, pp. 293–336). Palo Alto, CA: Science and Behavior Books.

- [69] E. Gilbert and K. Karahalios, "Predicting tie strength with social media", Proc. of the 27th international conference on Human factors in computing systems, pp. 211–220, 2009.
- [70] J. Golbeck, "Computing and Applying Trust in Web-based Social Networks", PhD thesis, University of Maryland, College Park, MD, USA, April 2005.
- [71] O. D. John. Big five inventory, 2000.
- [72] G. Holmes, M. Hall, and E. Prank, "Generating rule sets from model trees", Advanced Topics in Artificial Intelligence, pages 1–12, 1999.
- [73] J.R. Quinlan, "Learning with continuous classes. In 5th Australian joint conference on artificial intelligence, pages 343–348. Citeseer, 1992.
- [74] G. Chittaranjan, J. Blom, D. Gatica-Perez, "Mining large-scale smartphone data for personality studies", Journal Personal and Ubiquitous Computing, Volume 17 Issue 3, 2013.
- [75] S.D. Gosling, P.J. Rentfrow, W.B. Swann Jr, "A very brief measure of the big-five personality domains", Journal of Research in Personality, Volume 37 Issue 6, pp. 504-528, 2003.
- [76] F. Mairesse, M. Walker, "Automatic recognition of personality in conversation" Proc. of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers, pp. 85-88, 2006.
- [77] F. Pianesi, N. Mana, A. Cappelletti, B. Lepri and M. Zancanaro, "Multimodal recognition of personality traits in social interactions", Proc of the 10th international conference on Multimodal interfaces, pp. 53-60, 2008.
- [78] L. Liviu, "Predicting Product Performance with Social Media," Informatics in education, Volume 15 Issue 2, pp. 46-56, 2011.
- [79] W. Zhang and S. Skiena, "Improving Movie Gross Prediction through News Analysis," in 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology, Volume 30 Issue 2, pp. 301-304, 2009.
- [80] R. Sharda and D. Delen, "Predicting box-office success of motion pictures with neural networks," Expert Systems with Applications, Volume 30 Issue 2, pp. 243-254, 2006.
- [81] S. Kak, Y. Chen and L. Wang, "Data Mining Using Surface and Deep Agents Based on Neural Networks", Proc. of the Sixteenth Annual Americas' Conference on Information Systems, 2010.
- [82] S. Kak, "On training feedforward neural networks", Pramana, Volume 40 Issue 1, pp. 35-42, 1993.
- [83] S. Kak, "New algorithms for training feedforward neural networks." Pattern Recognition Letters, Volume 15 Issue 3, pp. 295-298, 1994
- [84] S. Kak, " Three languages of the brain: quantum, reorganizational, and associative", In Learning as Self-Organization, K. Pribram and J. King, eds., Lawrence Erlbaum, Mahwah, N.J., 185—219, 1996.
- [85] S. Kak, "Faster web search and prediction using instantaneously trained neural networks", IEEE Intelligent Systems, Volume 14, pp. 79-82, 1999.
- [86] S. Kak, "A class of instantaneously trained neural networks", Information Sciences. Volume 148, pp. 97-102, 2002.
- [87] C.G. Looney, "Pattern Recognition Using Neural Networks : theory and algorithms for engineers and scientists", Oxford University Press, New York, 1997.
- [88] D. Boyd, "Facebook's privacy trainwreck: Exposure, invasion, and social convergence", International Journal of Research Into New Media Technologies, Volume 14 Issue 1, pp. 13–20, 2008.
- [89] B. Debatin, J.P. Lovejoy, A.K. Horn and B.N. Hughes, "Facebook and online privacy: Attitudes, behaviors, and unintended consequences", Journal of Computer-Mediated Communication, Volume 15 issue 1, pp. 83–108, 2009.
- [90] T. Taraszow, A. Arsoy, G. Shitta, and Y. Laoris, "How much personal and sensitive information do Cypriot teenagers reveal in Facebook?", Proc. From 7th European Conference on E-Learning, pp. 871–876, 2008.
- [91] R.E. Wilson, S.D. Gosling, and L.T. Graham, "A Review of Facebook Research in the Social Sciences", Perspectives on Psychological Science, pp. 203-220, 2012.
- [92] S. Zhao, S. Grasmuck, and J. Martin, "Identity construction on facebook: Digital empowerment in anchored relationships", Computers in Human Behavior, pp. 1816-1836, 2008.
- [93] http://www.psychometricinstitute.com.au/Psychometric-Test-Guide/Personality-Test-guide/What_do_personality_tests_measure.html
- [94] L.R. Goldberg, "A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models", Personality Psychology in Europe, Volume 7, pp. 7-28, 1999.
- [95] K.Y.A. McKenna, A.S. Green, and M.E.J. Glenson, "Relationship formation on the Internet: What's the big attraction?", Journal of Social Issues, Volume 58 Issue 1, pp. 9–31, 2002.
- [96] N.B. Ellison, C. Steinfield, and C. Lampe, "The benefits of Facebook "friends": Social capital and college students' use of online social network sites", Journal of Computer-Mediated Communication, Volume 12 Issue 4, pp. 1143-1168, 2007.
- [97] L.C. Tidwell and J.B. Walther, "Computer-mediated communication effects on disclosure, impressions, and interpersonal evaluations: Getting to know one another a bit at a time", Human Communication Research, Volume 28 Issue 3, pp. 317–348, 2002.
- [98] S. Srivastava, O. John, S. Gosling, and J. Potter, "Development of personality in early and middle adulthood: "Set like plaster or persistent change?""", Journal of Personality and Social Psychology, Volume 84 Issue 5, pp. 1041–1053, 2003.
- [99] Y. Jing and S. Baluja, "PageRank for Product Image Search". Proc. of the 17th international conference on World Wide Web, pp. 307-316, 2008.
- [100] E.-A. Baatarjav, S. Phithakkitnukoon, and R. Dantu, "Group recommendation system for facebook. On the Move to Meaningful Internet Systems: OTM 2008 Workshops

Lecture Notes in Computer Science Volume 5333, pp. 211-219, 2008.

- [101] G. Farnadi, S. Zoghbi, M.F. Moens, and M.D. Cock, "Recognising Personality Traits Using Facebook Status Updates", Workshop on computational personality recognition (WCPR13) at the 7th international AAAI conference on weblogs and social media (ICWSM13), 2013.
- [102] C.Y. Wang, G.S. Ching, "A study on the relationship of Facebook and EFL learners' Personality", International Journal of research studies in educational technology, Volume 2 Issue 2, 2013.
- [103] M.R. Mehl, S.D. Gosling and J. Pennebaker, "Personality in its natural habitat: manifestations and implicit folk theories of personality in daily life. Journal of personality and social psychology", Volume 90 Issue 5, pp. 862-877, 2006.
- [104] C. Sumner, A. Byers and M. Shearing, "Determining personality traits & privacy concerns from facebook activity". In Black Hat Briefings, pp. 1 – 29, 2011
- [105] H.A. Schwartz, J.C. Eichstaedt, L. Dziurzynski, M.L.Kern, M.E.P. Seligman, L.H. Ungar, E. Blanco, M.Kosinski and D. Stillwell, "Toward Personality Insights from Language Exploration in Social Media", AAAI spring symposium, 2013.
- [106] H.A. Schwartz, J.C. Eichstaedt, M.L.Kern, L. Dziurzynski, S.M. Ramones, Megha Agrawal, Achal Shah, M. Kosinski, D. Stillwell, M.E.P. Seligman and L.H. Ungar, "Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach", PloS one e73791, Volume 8 Issue 9, 2013
- [107] P.T. Costa, R.R. McCrae, F.L. Odessa, "NEO PI-R. Professional manual", Psychological Assessment Resources, 1992.
- [108] C. Hu and P. Racherla, "Visual representation of knowledge networks: A social network analysis of hospitality research domain", International Journal of Hospitality Management, Volume 27 Issue2, pp. 302-312, 2008.
- [109] E.Y.L. Ong, R.P. Ang, J. C.M. Ho, J.C.Y. Lim, D.H. Goh, C.S. Lee and A.Y.K. Chua, "Narcissism, extraversion and adolescents' self-presentation on Facebook", Personality and Individual Differences, Volume 50 Issue 2, pp. 180-185, 2011.
- [110] L. Han and L. Qiu, "Sharing emotion on Facebook: network size, density, and individual motivation", Proc. Extended Abstracts on Human Factors in Computing Systems, pp. 2573-2578, 2012.
- [111] M.D. Back, J.M. Stopfer, S. Vazire, S. Gaddis, S.C. Schmukle, B. Egloff and S.D. Gosling "Facebook Profiles Reflect Actual Personality, Not Self-Idealization", 2010