

Social Media Text - A Source for Personality Prediction

P.S.Dandannavar
Department. of CSE
KLS Gogte Institute of Technology
Belagavi, Karnataka, India
padmad@git.edu

S.R.Mangalwede
Department of CSE
KLS Gogte Institute of Technology
Belagavi, Karnataka, India
mangalwede@git.edu

P.M.Kulkarni
Department of MBA
KLS Gogte Institute of Technology
Belagavi, Karnataka, India
pmkulkarni@git.edu

Abstract— Social media usage has been on an ever increasing exponential rise. Usage of social media sites, such as Twitter and Facebook, for social interaction has also become a popular trend. It is estimated that on an average, around 6,000 tweets are tweeted on Twitter every second. With people spending on an average 35 minutes on Facebook each day, it is also estimated that there are about 317,000 status updates on Facebook per minute. These vast volumes of data have powerful information locked within them. This data can be analyzed for several purposes. The use of such social media data for predicting user personality is common. Prediction models have been successfully built that can predict several user attributes - age, gender, personality traits, occupation, political orientation etc. Standards in personality models such as the Big Five model, DISC and the Myers-Briggs Type Indicator have been the basis for all such personality prediction. A user's social media data can thus be used to predict his/her personality. The main objective of this work is to review the work carried out for personality prediction using social media data.

Keywords— *Personality Assessment, Personality prediction, Big Five Model, social media, Twitter.*

I. Introduction

Social media has seen an exponential growth in the last few years and has become an easy and popular method - both for social interaction and information distribution. People use social media to express themselves to the world. These expressions characterize the behavior of a user and human behavior is a reflection of his/her personality. Thus, there is a strong interconnection between a user's personality and their online behavior on social networks[1]. This implies that posts made by a user can be analyzed to obtain pieces of information which can be used to recognize his/her personality traits.

Of all the human attributes, personality has been considered the most difficult to understand. It is vital as it can be utilized to define the unique characteristics of a person [2]. A central part of human conduct/behavior, personality is made up of emotions, conduct, patterns of thoughts, feelings, which makes a person one of a kind [3]. An individual's preference of things (book, music, films etc) and interactions with other people are influenced by his/her personality.

An individual's behavior is a reflection of their inherent personality. It is thus viable to identify personality through behavior [4]. Since social media posts made by user's are reflective of their personality traits, they can be analyzed to predict the personality of a person. Thus, data from social media can be utilized to speculate the personality traits of

an individual [5]. Personality trait is not specific and generalizes across social media platforms [6].

Understanding of personality traits is very useful in the workplace and beyond. Different personality models can be used to characterize personality of an individual, for example the Big Five (BF) Model, the Myers-Briggs Type Indicator (MBTI) and the Theory of Personality types. The BF model is the one that is majorly used as it is the most well-researched and well regarded measures of personality traits [7]. The "Big Five" are broad categories of personality traits acronymed as **OCEAN** (Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism). The current trend in research is towards trying to automate personality traits prediction using content from social media. Several researcher's have contributed to this area by implementing different techniques and algorithms

I. AUTOMATIC PERSONALITY PREDICTION FROM SOCIAL MEDIA - THE NEED

Traditional methods of assessment for PP are based on "self report inventory" [4], in which a person is administered a personality test (PT). The questions for these are framed by psychologists. These methods are known to accurately return the personality score of a participant but suffer from the following deficiencies:[8]

- 1) inefficient - when the experimentation is on a large scale, manual methods tend to be costly in terms of manpower and material.
 - 2) manual computation is further needed to arrive at the results, once the answers of the inventories are returned.
 - 3) A long time is needed to complete the inventory survey.
- In short, traditional methods take time to arrive at the final results i.e., the predicted personality and are less practical. Alternatively, online questionnaires administered through websites can be used for personality tests. The user still needs to answer various questions. Recent studies undertaken have affirmed that personality traits can be gotten automatically from the text that user's wrote. Use of online environments (such as micro blogging sites) ,was thus the focus of several researcher's. Advancements in IT and unprecedented use of social media sites for social communication has paved a new way to conduct research on personality and behavior - automatic prediction of personality from social media data.

III. LITERATURE REVIEW

Human behavior is affected by personality traits [8]. There is a strong interconnection correlation between the

personality of a user and his/her online behavior [1]. The personality of a person can be recognized by analyzing their posts on social media. This analysis might help in getting information that is needed to recognize a person's personality. Several authors have looked at automatic personality prediction from social media content coming from Facebook, Twitter etc. [1-13].

Jennifer Golbeck et al attempted to answer the core research question - "Whether social media profiles can predict personality traits" [9] and presented a method as an answer to this question. Authors in [10] used the myPersonality Dataset and text classification methods (Naive Bayes, K-Nearest Neighbours and Support Vector Machine) to predict personality from tweets. Naive Bayes slightly outperformed the others. Measuring the Personality Traits of Sina microblog user's using the BF model was the work carried in [11]. The authors used multi-task regression and incremental regression algorithms to predict BF personality. Sibel Adalti et al [12] examined the degree to which behavioral measures can be used to predict personality. The authors in their previous work [9] had shown that Facebook profiles of users could be used to accurately predict personality. From text written in Bahasa Indonesia, the work of authors in [13] provided a analysis of word(s) for each personality traits. Machine learning, lexicon based, and grammatical rule approach were the three methods that were compared. While the BF model was the preferred choice of most researcher's, authors in [3] mapped the results of clustering onto the DISC (Dominance, Influence, Submission, Compliance) framework, when trying to improve user experience with personality assessment using Twitter data. The BF model, three different classification algorithms and the concept of "combined results" (majority vote of the three algorithms) were used by [5] in predicting personality of a person. In the context of PP, Basant Agarwal [2] provided a discussion on the publicly available datasets and the various methods that can be used for PP. In the work of [4], the relationship between an individual's personality and his online behavior was analyzed using Pearson's correlation coefficient and regression coefficient. Use of machine learning techniques for automatic PP was explored in [6]. The models were trained on both the essay data set and the Facebook dataset, which improved the overall results. Several classifiers, available in Weka, were compared for performance by Nor Rahayu et al [7], in an attempt to predict student's personality based on the BF model.

It is thus observed that the area of automatic PP has been well researched and that: 1) The BF personality model is the most frequently used for prediction, 2) Twitter is the most social media site for extracting data, 3) the myPersonality and essay datasets are the most commonly used datasets and 4) the most common techniques used are machine learning, classification, and linguistic features.

IV. MODELS FOR PERSONALITY ASSESSMENT

The Big Five Personality Traits, DISC, Myers Briggs Type Indicator (MBTI), Strength Finder are a few models

available for personality assessment. Characteristics of each are listed below:

1) Myers Briggs Type Indicator :

- Widely known and acknowledged test for personality assessment with a background marked by 55 years.
- Test depends on Carl Jung Typology hypothesis.
- Describes human identity in four principal measurements viz., extroversion/self preoccupation, detecting/instinct, considering/feeling, and judging/seeing
- Categorizes the human identity type, in view of basic nature and inclinations, in one of the prominent sixteen classes.

Great explanatory understanding and psycho-systematic profiling abilities are should have been ready to utilize this model

2) DISC :

- Introduced in 1928, this assessment model is simple and more intuitive.
- Four benchmarks - Dominance, Influence, Stability and Compatibility are used for assessment.
- Concentrates on behavioral preferences that's why it's more applied than MBTI.

3) The Big Five Personality Inventory:

Is one of the most well-researched and well-regarded measures of personality structure. The models identifies five personality traits (Big Five traits)- Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism. Characteristics of the Big Five traits are:

- Openness to Experience: curious, intelligent, imaginative.. High scorers will in general be creative and advanced in taste and acknowledge assorted perspectives, thoughts, and encounters.
- Conscientiousness: responsible, organized, persevering. Such individuals are greatly dependable and will in general be high achievers, diligent employees, and organizers.
- Extroversion: outgoing, amicable, assertive. Cordial and vivacious, outgoing people draw motivation from social circumstances
- Agreeableness: cooperative, helpful, nurturing. Individuals who score high in suitability are harmony attendants who are by and large hopeful and trusting of others

- Neuroticism: anxious, insecure, sensitive. Neurotics are testy, tense, and effectively tipped into encountering contrary feeling

V. THE PERSONALITY PREDICTION FRAMEWORK

Different methodologies have been adopted by different authors for personality prediction. Regardless of the methodology adopted, the general framework used for PP is the same. The various stages in the framework consist of 1) Data Extraction, 2) Pre-processing, 3) Feature extraction and 4) Classification/Prediction. Figure 1 illustrates the general pipeline of the framework.

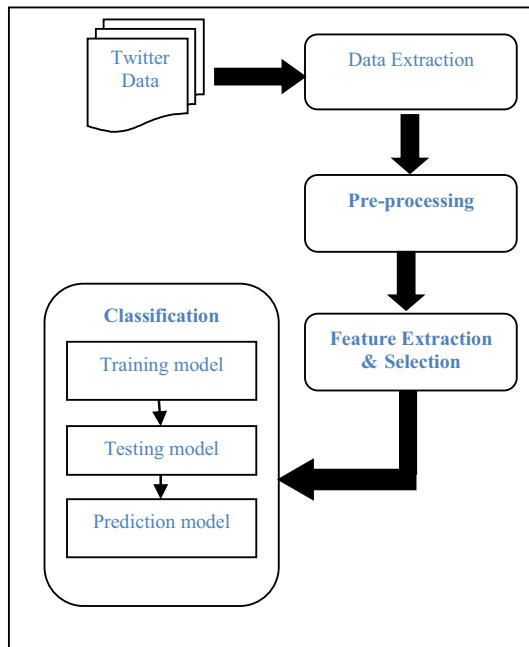


Fig. 1: The personality prediction pipeline

1. Data Extraction

This phase involves identifying and collecting data sets deemed suitable for the application being developed. Twitter has been the major source for social media data. Data sets for training and testing are collected from Twitter using suitable APIs. Other sources of data include the publicly available datasets viz, myPersonality and essays.

2. Data Pre-processing

This is a very important phase in the pipeline as it decides the efficiency of the other steps down in line. Pre-processing involves the standard steps of Case Conversion, Stop-words Removal, Punctuation Removal, Stemming, Lemmatization, POS Tagging.

3. Feature Extraction & Selection

After the relevant tweets are fetched and pre-processed, features relevant for prediction are extracted and selected. User's profile data is the most common feature set used by most researchers. Among other features used are - number of followers, number of following, twitter posts, linguistic features etc. The features extracted are dependent on the prediction to be made.

4. Classification

This phase includes modules for model training, testing and prediction. After the selected model has been trained and tested using different data sets, an unseen (new) dataset is presented to the trained model for prediction, in order to classify and predict the personality trait of new user(s). Several standard algorithms have been used by authors like Support Vector Machine, K Nearest Neighbour, Multinomial Naive Bayes, Naive Bayes, multi task regression, incremental regression algorithm etc.

VI. APPLICATIONS

Personality prediction, manual or automatic, find applications in several diverse fields, a few of which are listed below:

- 1) **Recruitment** - employers can use PP techniques to gain a deep understanding of the applicants. This helps them to find the qualified personnel they really need.
- 2) **Counseling** - personality can be used as an important assessment in career, relationship and health counseling.
- 3) **Online marketing** - a user's predicted personality can be used by online marketer's to personalize their message and presentation to suit individual preferences.
- 4) **Corporate** - for targeted advertising and marketing, employee recruitment, career and health counseling [5].
- 5) **Psychological profiling** - of user's is a useful tool for job satisfaction, career progression, selling preferences in different interfaces etc [3].
- 6) **E-commerce/ E-learning** - can benefit by a user interface that adapts the interaction according to the user's personality [2].
- 7) **Recommendation systems** - performance of such systems can be enhanced to attract more user's.
- 8) **Determining antisocial behavior** - personality traits have been found to have a close correlation with antisocial behavior. This has been revealed by the studies undertaken on personality and crime [1].

VII. CONCLUSION

Social media is being most widely used for social interaction and communication. A user's behaviour on social platforms is reflected in his/her tweets, status updates, comments, interests etc. , which in turn reveals traits of his/her personality. A user's personality is correlated with their online behaviour and hence what user's share/write on social media can be used to extract information needed to identify his/her behaviour. It has been possible to predict user attributes such as age, gender, political orientation etc using social media and models have been successfully built to achieve this. Work of several researchers has also focused on using social media data for predicting personality which has been successfully demonstrated. It is thus possible to automatically predict personality using data from social media which is more simpler, cost effective and efficient compared to the manual methods.

REFERENCES

- [1] Firoj Alam, Evgeny A. Stepanov, Giuseppe Riccardi, "Personality Traits Recognition on Social Network - Facebook", AAAI Technical Report WS-13-01 Computational Personality Recognition (Shared Task)
- [2] Basant Agarwal, "Personality Detection from Text: A Review", International Journal of Computer System (ISSN: XXXX – XXXX), Volume 01– Issue 01, September, 2014.
- [3] Nadeem Ahmad, Jawaid Siddique, "Personality Assessment using Twitter Tweets", 21st International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2017, 6-8 September 2017, Marseille, France.
- [4] Shuotian Bai a, Sha Yuana, Bibo Haoa and Tingshao Zhu , "Predicting personality traits of microblog users", Web Intelligence and Agent Systems: An International Journal 12 (2014) 249–265 249 DOI 10.3233/WIA-140295 IOS Press
- [5] Vanshika Varshney, Aman Varshney, Tameem Ahmad, Asad M. Khan, "Recognising Personality Traits using Social Media", IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI-2017) 978-1-5386-0814-2/17/\$31.00 ©2017 IEEE.
- [6] Golnoosh Farnadi1, Susana Zoghbi, Marie-Francine Moens, Martine De Cock, "Recognising Personality Traits Using Facebook Status Updates", AAAI Technical Report WS-13-01 Computational Personality Recognition (Shared Task), Copyright c 2013, Association for the Advancement of Artificial Intelligence.
- [7] Nor Rahayu Ngatirin, Zurinahni Zainol, Tan Lee Chee Yoong, "A Comparative Study of Different Classifiers for Automatic Personality Prediction", 6th IEEE International Conference on Control System, Computing and Engineering, 25–27 November 2016, Penang, Malaysia, 978-1-5090-1178-0/16 ©2016 IEEE.
- [8] Tommy Tandra, Hendro, Derwin Suhartono*, Rini Wongso, and Yen Lina Prasetyo, "Personality Prediction System from Facebook Users ", 2nd International Conference on Computer Science and Computational Intelligence 2017, ICCSCI 2017, 13-14 October 2017, Bali, Indonesia .
- [9] Jennifer Golbeck, Cristina Robles, Michon Edmondson, and Karen Turner, "Predicting Personality from Twitter", 2011 IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing, 978-0-7695-4578-3/11 © 2011 IEEE.
- [10] Bayu Yudha Pratama, Riyanarto Sarno, "Personality Classification Based on Twitter Text Using Naive Bayes, KNN and SVM ", 2015 International Conference on Data and Software Engineering, 978-1-4673-8430-8/15/ ©2015 IEEE.
- [11] Shuotian Bai, Bibo Hao, Ang Li, Sha Yuan, Rui Gao, Tingshao Zhu, "Predicting Big Five Personality Traits of Microblog Users", 2013 IEEE/WIC/ACM International Conferences on Web Intelligence (WI) and Intelligent Agent Technology (IAT), 978-1-4799-2902-3/13 © 2013 IEEE, DOI 10.1109/WI-IAT.2013.70
- [12] Sibel Adal, Jennifer Golbeck, "Predicting Personality with Social Behavior".
- [13] Louis Christy Lukito, Alva Erwin, James Purnama, and Wulan Danoeckoemo, "Social Media User Personality Classification using Computational Linguistic", 2016 8th International Conference on Information Technology and Electrical Engineering (ICITEE), Yogyakarta, Indonesia, 978-1-5090-4139-8/16 ©2016 IEEE.