Novel Algorithm for Text Based Personality Classification

Veena Parihar[1], Mr. Aishwary Kulshrestha[2]
[1]M.Tech Scholar, [2]Assistant Professor
Rajasthan College of Engineering for Women, Jaipur, Rajasthan, INDIA

Abstract: The Personality of the person is the basis of deciding about how the person is going to behave in a particular situation. In this paper we present the unique algorithm for automating the process of determining the personality of the individual using the survey and the text based approach. The results which are achieved are quite promising in the determination of personalities.

*****

I. Introduction

Personality is the product of social interaction in group life. In society every person has different traits such as skin, color, height and weight. They have different types of personalities because individuals are not alike. It refers to the habits, attitudes as well as physical traits of a person which are not same but have vary from group to group and society to society, everyone has personality, which may be good or bad, impressive or unimpressive. It develops during the process of socialization in a culture of a specific group or society. One cannot determine it of an individual exactly because it varies from culture to culture and time to time. For example, a killer is considered criminal in peace time and hero in war. The feeling and actions of an individual during interaction moulds the personality. It is the sum of total behaviors of the individual and covers both overt and covert behaviors, interests, mentality and intelligence. It is the sum of physical and mental abilities and capabilities.

Personality has been derived from the Latin word "persona" which means "mask" used by the actors to change their appearance. It is the combination of an individual's thoughts, characteristics, behaviors, attitude, idea and habits.[1]

These are divided into two types, viz., types and traits theories. Both these theories of personality focus on people's personal characteristics. However, 'type' theorists and 'trait' theorists differ in the ways they use characteristics to describe people.[1]

II. Related Work

Kanupriya Sharma, Amanpreet Kaur[1]-Twitter is a popular social media platform with millions of users. The tweets shared by these users have recently attracted the attention of researchers from diverse fields. This research focuses primarily on predicting user's personality from the analysis of tweets shared by the user. An associative study of different research works done in personality prediction shows that different techniques have been used to predict a user's personality from tweets but there are certain shortcomings which still need to be addressed. In the introduction section we have given the gist of personality prediction with Twitter research. In the next section we provide the importance and framework of predicting a user's personality with Twitter. In the consecutive section, literature survey, we have discussed briefly about the different research ideas in context to personality prediction with a side by side overview of various resources, tools and machine learning algorithms that were used alongside their potential and limitations, followed by research gap, which provides different fields of possibilities to which can be rectified to improve the efficiency of predicting personality with machine learning algorithms. Finally, a new approach is proposed to predict personality with new insights to predict personality on crucial factors such as scalability and counter measures to improve the research based on previous work by using a Logistic Regression Classifier with parameter regularization using stochastic gradient descent.

Sandeep Dang, Prof. Mahesh Kumar, Mahesh[2]-Graphology or Handwriting analysis is a scientific method of identifying, evaluating and understanding of anyone personality through the strokes and pattern revealed by handwriting. Handwriting reveals the true personality including emotional outlay, honesty, fears and defenses and etc. Handwriting stroke reflects the written trace of each individual's rhythm and Style. The image split into two areas: the signature based on three features and application form of letters digit area. In this research performance evaluation is done by calculating mean square error using Back Propagation Neural Network (BPNN).Human behavior is analyzed on the basis of signature by using neural network.

Tarika Sandhu, Shaina Kapoor[3]- Personality can be understood as "A dynamic organization, inside the person, of psychophysical systems that creates a person's characteristic patterns of behavior, thoughts, and feelings" (Carver & Scheier, 2000). The affective component thus forms an integral aspect of the structure and dynamics of personality. Emotional regulation further refers to a person's ability to understand and accept his or her emotional experience, to engage in healthy strategies to manage uncomfortable emotions when necessary, and to engage in appropriate behavior especially when distressed. Working on the assumption that person logical typifications would lend color to the psychological functioning of individual. The present study aimed at exploring how dominating personality types effect emotional self-regulation styles amongst young women. The sample of the study comprised of 200 undergraduate female students. Personality assessment was carried out by using Myers- Briggs Type Indicator by Myers & Mccaulley (1998) and emotional regulation was assessed using the Difficulties in Emotion Regulation Scale by Gratz& Roemer (2004).Results of the study reveal the significant

role of personality types in influencing typical emotional self-regulatory patterns in young women. Identifying personality types thus becomes relevant in context of social and occupational adjustment of young women, since success in their domain largely depends upon effective emotional functioning.

Vandana Korde, C Namrata Mahender[4]- As most information (over 80%) is stored as text, text mining is believed to have a high commercial potential value. knowledge may be discovered from many sources of information; yet, unstructured texts remain the largest readily available source of knowledge .Text classification which classifies the documents according to predefined categories. In this paper we are tried to give the introduction of text classification, process of text classification as well as the overview of the classifiers and tried to compare the some existing classifier on basis of few criteria like time complexity, principal and performance.

Mike Komisin and Curry Guinn[5]- Are the words that people use indicative of their personality type preferences? In this paper, it is hypothesized that word-usage is not independent of personality type, as measured by the Myers-Briggs Type Indicator (MBTI) personality assessment tool. In-class writing samples were taken from 40 graduate students along with the MBTI. The experiment utilizes naïve Bayes classifiers and Support Vector Machines (SVMs) in an attempt to guess an individual's personality type based on their word-choice. Classification is also attempted using emotional, social, cognitive, and psychological dimensions elicited by the analysis software, Linguistic Inquiry and Word Count (LIWC). The classifiers are evaluated with 40 distinct trials (leave-one-out cross validation), and parameters are chosen using leave-one-out cross validation of each trial's training set. The experiment showed that the naïve Bayes classifiers (word-based and LIWC-based) outperformed the SVMs when guessing Sensing-Intuition (S-N) and Thinking-Feeling (T-F).

Shogo Okada, Oya Aran, Daniel Gatica-Perez[6]- This paper proposes a novel feature extraction framework from multi-party multimodal conversation for inference of personality traits and emergent leadership. The proposed framework represents multi modal features as the combination of each participant's nonverbal activity and group activity. This feature representation enables to compare the nonverbal patterns extracted from the participants of different groups in a metric space. It captures how the target member outputs nonverbal behavior observed in a group (e.g. the member speaks while all members move their body), and can be available for any kind of multiparty conversation task. Frequent concurrent events are discovered using graph clustering from multimodal sequences. The proposed framework is applied for the ELEA corpus which is an audio visual dataset collected from group meetings. We evaluate the framework for binary classification task of 10 personality traits. Experimental results show that the model trained with co-occurrence features obtained higher accuracy than previously related work in 8 out of 10 traits. In addition, the co-occurrence features improve the accuracy from 2% up to 17%.

III. Problem Statement

The problem which we have consider for the analysis is the systematic study of the personalities guides with some calculations so that the predictability regarding the personality of the person can be increased and can be done more precisely. Together with that the behavior analysis of the personality by analyzing the text is the next phase, which we want to perform on the answers to specific questions of individuals.

IV. Proposed Work

Algorithm for Personality Approach

Step 1: Register the User.
Step 2: Login the user with Credentials.
Step 3: If Credentials Is OK then
Move to Step 4
Else
Exit
[End of If structure]
Step 4: Present the Questionnaire.
Step 5: Check the Answers by user.
Step 6: Get the score of each answer given by user.
Step 7: If(score>=15) then
Increase the score of "Extraversion vs. Introversion";
Else If(score>=10)
Increase the score of "Emotional stability vs. Neuroticism";
Else If (score>=8)
Increase the score of "Agreeableness vs. Disagreeable";
Else if (score>=5)
Increase the score of "Conscientiousness vs. Unconscientious";
Else
Increase the score of "Openness to experience";
[End of If structure]
Step 8: Calculate and print the percentage of all five factors.
Step 9: Stop

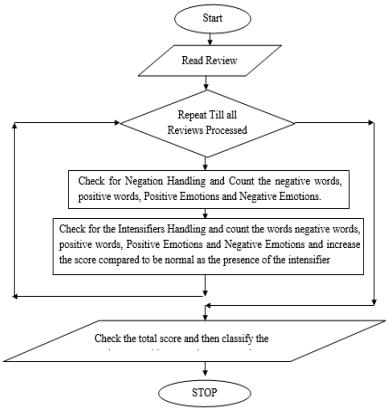

Fig 1. Text Based Analysis For Personality Classification

This implementation of the proposed work is done using the PHP MYSQL and the implementation is done in the two phases , the first implementation is the test based or opbjective questions based survery to perform the personality analaysis classification and the second phase is the text

analysis for the personality classification and using the dataset based sentiment lexicon approach.
Some of the datasets which are refereed in the implementation :

Sentiment140
This is a popular dataset, combining 160,000 tweets with emoticons pre-removed.

Twitter US Airline Sentiment
This dataset contains Twitter data on US airlines which was scraped from February 2015. Contributors classified the tweets as positive, negative, and neutral tweets.

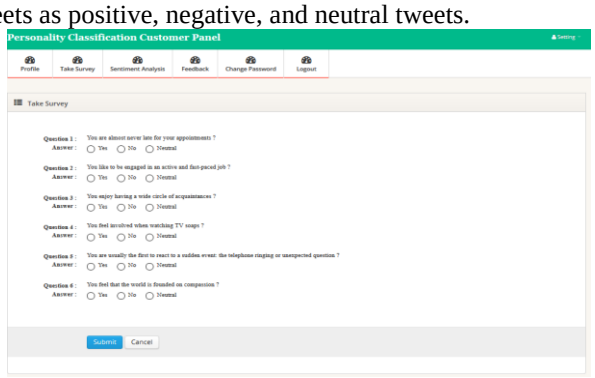

Fig 2. Implementation

V. Result Analysis

Table 1 Comparison Test Results For Positive and Negative Emotions

This Table 1 shows the comparison results for the positive and negative emotions.

Table 2 Comparison Test Results for Intensifier Handling

Review Text	Proposed Work	Sentiment Analysis of Movie Review Data Using Senti-Lexicon Algorithm by DeebhaMumtaza, et.al 2016 [16]
The life was very good.(Review 1)	2	1
The life was good.	1	1
Tom mood was too bad. (Review 2)	-2	-1
Tom mood was bad.	-1	-1
The person was behaving very badly. (Review 3)	-2	-1
The person was behaving badly.	-1	-1

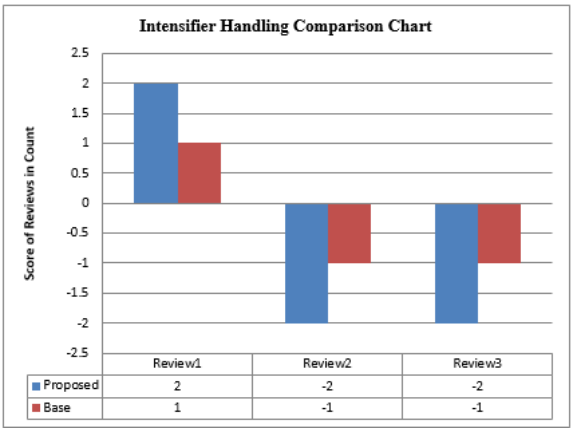

Fig 3 Comparison Test Results for Intensifier Handling

VI. Conclusion

Text analysis is a set of methods, typically implemented in computer software, that detect, measure, report, and exploit attitudes, opinions, and emotions in online, social, and enterprise information sources. Use of social networking has increased tremendously in recent times. It has become popular method for information distribution and social interaction. Personality has been considered as the most difficult human attribute to understand. It is very important as it can be used to define the uniqueness of a person. Personality detection from text means to extract the behavior characteristics of authors written the text. Personality detection models could be very useful in various domains like e-learning, information filtering, collaboration and e-commerce by a user interface that adapts the interaction according to user's personality. This dissertation using the MCQ tries to classify the personality, together using the analysis of the text input discusses the personality polarity for the input entered by the person.

Review Text	Proposed Work	Sentiment Analysis of Movie Review Data Using Senti-Lexicon Algorithm by Deebha Mumtaza, et.al 2016 [16]
Tom gesture is very satisfied .	2 (P)	1 (P)
Staff attitudes are very cheerful.	2 (P)	1 (P)
The management behavior is very irritated .	2 (N)	1 (N)
Sita get too bored in life .	2 (N)	1 (N)

Table 2 is used to show the comparison result of the intensifier handling.

VII. References

[1]. Kanupriya Sharma, Amanpreet Kaur, "Personality prediction of Twitter users with Logistic Regression Classifier learned using Stochastic Gradient Descent", IOSR Journal of Computer Engineering,2015

[2]. Sandeep Dang Prof. Mahesh Kumar, Mahesh ,"Handwriting Analysis of Human Behavior Based on Neural Network ",International Journal of Advanced Research in Computer Science and Software Engineering ,2014

[3]. Tarika Sandhu ,Shaina Kapoor, "IMPLICATIONS OF PERSONALITY TYPES FOR EMOTIONAL REGULATION IN YOUNG WOMEN",Voice of Research, Vol. 1 Issue 4, March 2013

[4]. Vandana Korde,Sardar Vallabhbhai ,"TEXT CLASSIFICATIONAND CLASSIFIERS:A SURVEY",International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.2, March 2012

[5]. Mike Komisin and Curry Guinn ," Identifying Personality Types Using Document Classification Methods", Proceedings of the Twenty-Fifth International Florida Artificial Intelligence Research Society Conference,2012

[6]. Shogo Okada ,Oya Aran ,Daniel Gatica-Perez, "Personality Trait Classification via Co-Occurrent Multiparty Multimodal Event Discovery",ACM ,2015

[7]. S. Adali, J. Golbeck, "Predicting personality with social behavior", Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining, pp. 302-309, 2012.

[8]. O. P. John and S. Srivastava, "The Big-Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives," In L. A. Pervin & O. P. John (Eds.), Handbook of personality: Theory and research, 2, 102–138 Guilford Press, New York, 2008.

[9]. J. Golbeck, C. Robles, M. Edmondson, and K. Turner, "Predicting Personality from Twitter," in Proc. IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing, 2011,pp. 149-156.

[10]. D. Quercia, M. Kosinski, D. Stillwell, and J. Crowcroft, "Our Twitter Profiles, Our Selves: Predicting Personality with Twitter," in Proc. of IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing, Boston, 2011, pp. 180-185.

[11]. C. Sumner, A. Byers, and M. Shearing, "Determining personality traits & privacy concerns from Facebook activity," in Black Hat Briefings, Abu Dhabi, United Arab Emirates, 2011,pp. 1-29.

[12]. F. Celli and L. Rossi, "The Role of Emotional Stability in Twitter Conversation," in Proc. of the 13th Conference of the European Chapter of the Association for Computational Linguistics, Avignon, France, 2012, pp. 10-17.

[13]. C. Sumner, A Byers, R. Boochever, and G. J Park, "Predicting Dark Triad Personality Traits from Twitter usage and a linguistic analysis of Tweets," in Proc. of 11th IEEE International Conference on Machine Learning and Applications, 2012, pp. 386-393.

[14]. F. Alam, E. A. Stepanov, and G. Riccardi, "Personality Traits Recognition on Social Network – Facebook," Computational Personality Recognition, pp. 6-9, 2013.

[15]. Deebha Mumtaza et al ,"Sentiment Analysis of Movie Review Data Using Senti-Lexicon Algorithm",IEEE 2016