Comparative Study on Prediction of Personality of a Person Using Text

Susmita S. Kunde[1*], A. U. Bapat[2]

[1,2] Computer Engineering Department, Goa College of Engineering, Farmagudi-Ponda, Goa, India

*Corresponding Author: susmitakunde @gmail.com

Abstract — Every person is unique in its own way; every person has a different personality. Personality detection gives an idea of behaviour of the person or gives an idea of how a person will react in a particular situation. Study of the relationship between word-use and personality traits has been successful in giving insight into human behaviour. Questionnaires is the most commonly used methods in the earlier times to detect personality traits from text but is not that effective. Due to emergence in technology different new methods are now available to detect personality of a person automatically. This paper is a summarized study of various methods used to automatically predict personality of a person from its text. Beginning with various methods used in earlier times to the methods newly emerged, this paper is a detailed study of all the different types of methods which are used for personality prediction along with different personality prediction models.

Keywords— Personality Prediction, Linguistics, LIWC, DISC, MBTI, Big Five Model.

I.   INTRODUCTION

Personality is something which plays a key role in the actions, behaviour and decision making in the life of a person. Hence personality detection can help us in getting to know the person in a more better way.

The main objective of personality detection is to differentiate a person from others effectively. A better personality detection technique will classify individuals with different traits into different categories and the individuals in the same category will be similar to each other in the maximum limit. The words spoken or written by a person are an assertion of their inner thoughts and personality [1]. We can get an unique understanding of the authors mind and thinking process by finding the patterns hidden in the style of a text-in such elements as how often certain words or categories of words is used regardless of the context , which is afar from meaningful content of the language [15].

Questionnaires is the most commonly used methods in the earlier times to detect personality traits from text of a person. But as they are managed and answered by the person itself, the results will be like a negative influence on the accuracy. With the emerging technology many methods are now available which could be used to predict personality of a person automatically by using the text written by that person. Rest of the paper includes various personality detection models in section II wherein a detailed study on personality models which are commonly used is done. Section III consists of detailed review of various methods used in

personality prediction of a person using text of the person. Further section IV gives list of datasets which could be used for personality prediction. Finally section V is the conclusion of the paper.

II.   VARIOUS PERSONALITY DETECTION MODELS

Human mind is very curious and tries to understand others behaviour by observing their behaviour on daily basis. Therefore there have been many researches on personality prediction and different inventors have invented different models of personality prediction [16].

Earlier, while observing the behavioural samples, psychologists used questionnaires as a method for gathering related data speedily and standardly. Questionnaires are found in two types, the first type describes the personality theory systems some examples are: 16PF, EPQ and NED. The other type emphasises on evaluating certain personality traits, for example the Eisenberg self-esteem scale.

But in fact the management of questionnaire method is time consuming and as it is mainly based on self-report and subjective factors of participant hence it applies negative impact on the accuracy of the testing results.

Therefore, there is need to develop innovatory methods which will help in recovering these short comings. Proliferation of information technology helps in providing solution for such problems. As social media is very much in use, the data available on social media can be a very useful

source in finding personality patterns. This could be done by applying various personality models to this data.

Some of the models widely used for personality detection are:

A. DISC

DISC is a behaviour evaluation tool which is based on the theory of Psychologist William Moulton Marston. Psychologist Walter Vernon Clarke developed this theory further into a behaviour evaluation tool.

DISC theory places an individual into four main types of personalities; these personalities are named as Dominant, Influence, Steady, and Compliance personality types.

In the DISC classification the people are first divided into two main categories called as container categories. These containers are active container and passive container. A person who is belligerent and is very determined than the other types is considered to be in active type. These people are never fully satisfied and are highly ambitious in life. Just opposite to this are the passive type people which are un-combative, who do not take a stand for themselves that much as compared to the active type. These people are mostly satisfied with whatever they have and do not have the desire of going out of the way to achieve their goals. These containers are then further divided into more two categories called as sub-containers. These sub-containers are called task oriented and people oriented. If a person if given any work are ready to accept challenges but are unfriendly with others they are associated with the task oriented sub-container. Whereas if a person is very friendly with others but not very willing to accept challenges in the work assigned they are considered under the people oriented sub-container. [14].

Using these containers and sub-containers a person is then classified into the four main categories of DISC personality assessment. Wherein the person who comes under active container and task oriented sub-container are said to be under dominant category, the person who comes under active container and people oriented sub-container are said to be of influence personality, the person who comes under passive container and people oriented sub-container are said to be under steady category, the person who comes under passive container and task oriented sub-container are said to be of compliant personality. [15].

B. MBTI

MBTI which is expanded to be Myers-Briggs Type Indicators is a psychological assessment widely used in personality prediction researches [12]. It was constructed by Katharine Cook Briggs and her daughter Isabel Briggs Myers hence the name Myers-Briggs. It is based on an ideal theory proposed by Carl Jung, who had hypothesized that the world is experienced by humans using four principal psychological functions – sensation, intuition, feeling, and thinking – and that one of these four functions is dominant for a person most of the time [10].

Myers and Briggs proposed that a person can be categorized using four key dimensions. And they are as follows:

- Extraversion/Introversion
- Sensing/Intuition
- Thinking/Feeling
- Judging/Perceiving

Each of the four dimensions is described as a dichotomy, where an either/or choice between two styles of being is given. Myers and Briggs described this as a "preference" and proposed that any individual should be able to identify a preferred style on each of the four dimensions. The sum of a person's four preferred styles becomes their personality type [12].

As per the theory of Myers and Briggs our likings on each of these four dimensions will combine to create pattern in our thoughts and behaviour which are predictable. Such that the people with similar four likings will share resemblance in the way they approach their lives like from the way they behave, to the way they think and take actions.

The four dimensions are explained as follows:

Introversion or Extraversion: This dimension is used to describe where and how a person puts its attention, time and energy.

The extraverts are very active and outgoing and are excited to be around different people. these type of people are very expressive and socializing.

The introverts are more comfortable in being with themselves and are not so outgoing. They prefer to be in small group or alone and are very quiet and reserved.

Sensing or Intuition: sensing and intuition is used to see how a person receives and processes information using its five senses or abstract thinking's.

Sensors are the ones who are more practical in nature and are mostly interested in the information they receive directly through their five senses.

Intuitivists are abstract thinkers as compared to sensors. They are interested more in abstract things like theories, and possibilities. They are the ones who live in the future and like to gather knowledge through reading.

Thinking or Feeling: Thinking and feeling dimensions are used in describing the decision making criteria of a human being.

As the name suggests the thinkers tend to take any decision by thinking about each and every aspect of that thing. They first try to understand any problem completely and then only take the decision about it.

Whereas feelers tend to take decisions using their feelings, they take decisions with their hearts. They take interest in how a person feels about a decision and these decisions affect them.

Judging or Perceiving: These dimensions describe the person using the way he approaches in his life, that is in a structured way or flexible way.
Judgers are more planned in their life; they tend to structure everything and are always well ordered.
Perceivers on the other hand are flexible in nature hence tend to make last moment changes in things. They are spontaneous and very open minded. [11, 12].
Using the combination of these four binary dimensions, a person is then classified into one of the 16 personality types [11].

C. Big five Model
The Big Five Model of personality which is also known as OCEAN model is a well-researched and commonly used measure for personality prediction in the recent years. It believes in the fact that a person's personality can be described as a rational combination of influence, behaviour, desire, and perception over time and space. It is also the most widely used personality prediction model in the literature which is popularly used amongst the language and computer science researchers. It is also used as a classic structure for personality trait identification and simulation. This model is used to characterize a particular person into five dimensions of personality [7]. In this model a human's personality is described into five different vector values which correspond to bipolar traits [20].

The Big-Five personality traits model is defined as follows:

O (Openness): People having high score in this trait tend to be Creative, Curious in nature and highly intelligent. These people are always open to new ideas and views.

C (Conscientiousness): The one who has high score in this trait is said to be very efficient and organised in doing work. These people are very hard working and are goal oriented in nature. They plan things very carefully and avoid making impulsive decisions.

E (Extraversion): High score in this trait indicates that the person is outgoing and likes to initiate conversation with new people. These kind of people are very energetic, enthusiastic and enjoy being with other people. The people scoring less in this trait are called introverts and are just opposite to this.

A (Agreeableness): The people who are cooperative and compassionate score high under this trait. These people are open minded, optimistic, peace keeping and trustworthy.

N (Neuroticism): This trait tells more of the emotional stability of a person. Person high in neuroticism tend to have more negative thoughts and emotions whereas the ones who score-less in this trait are very calm and emotionally stable [4, 9].

III. DIFFERENT PERSONALITY PREDICTION METHODS

Finding the linguistic cues from the text using data mining can be used as one of the methods for personality prediction where classification, regression and ranking models are used and results show that for some traits, any type of statistical model performs significantly better than the baseline, but ranking models perform best overall [17, 18, 19]. Data mining can also be used along with machine learning algorithms to get the correlation between text and personality [8].
In [2] a combination of fine affect and coarse affect categories of emotions is used along with SVM classifier, which shows that the use of fine emotion features leads to statistically significant improvement over a competitive baseline, whereas the use of coarse affect and specificity features does not.
Using Natural Language Processing methods to predict the personality of a person with text processing techniques and polarity based matching is also a good approach [3].

Some researchers considered features like avatars, likes, self-language usage, emoticons; that is heterogeneous information, along with text to get the personality insight of the person [4]. This heterogeneous information is derived from social networking sites like Facebook and Twitter which also provides with information like different posts of the user and hours spent online which can also be used in personality prediction [6, 10, 16].
Multi-layer perceptron neural network model along with DISC profiling method can be used in identifying and learning linear/non-linear patterns, and relationships. The model can be fit to real world data to mimic complex patterns which will enable the neural network to perform well in predicting the compatibility of the individuals [14]. DISC model along with machine learning algorithms like Sequential Minimal Optimization are used on a linguistic corpus [15].
Using Recurrent Neural Network (RNN) for building a hierarchical vector representation of words and sentences with character level and sentence level embedding for personality trait inference is done in [20].
Whereas Convolutional Neural Network (CNN) with 1, 2, 3-grams kernels for capturing the structure of the text is achieved in [4] but this method is based on generalization of stacks and text data is just a part of the whole model, the final output does not help in depicting the effectiveness of CNN. Also usage of fixed size kernel results in imposing limitation on long texts.

[21] tried a method which uses RNN to interpret reading of text and CNN to get the structure of text and then further to imitate writing. Three-dimensional convolution is used to learn the structure of an article in [5], wherein the input taken is essays dataset and these essays are encoded from word level to sentence level and then to document level. But they used max-pooling from sentence level to document level instead of CNN.

Table 1.Observation Table

Referenc e Paper	Yea r	Model	Method	Accurac y
[15]	2017	DISC	SVM	0.75
[14]	2014	DISC	Multilayer Perceptron Neural Network	0.87
[11]	2015	MBTI	Logistic Regression classifier	0.76
[3]	2016	OCEA N	NLP	0.78
[2]	2012	OCEA N	SVM	0.82
[4,5,10]	2017 , 2012	OCEA N	CNN	0.89

Table 1: Compares different reference papers with respect to the personality models used and method used for computation. The accuracy values stated are based on the various aspects such as precision and recall of all the combination of the personality traits from the model.

IV. DATASETS

The various personality detection models presented are based on different experimental settings. With the help of various resources, the researchers developed their own dataset and further tagged them manually with the help of experts from psychology. The following are some of the datasets used by many researchers commonly and which could be used to develop a personality prediction model.

A. Essay dataset
Essays dataset is a set of stream of consciousness essays. There are around 2400 unique essays from different authors. These essays were collected in the years between 1997 to 2004, and were labelled with the five dimensions of Big Five personality classes. This data is manually labelled using z-scores computed by Mairesse and then converted from scores to nominal classes. This corpus is used by different researchers as an input dataset in their work.

B. MyPersonality dataset
MyPersonality corpus (http://mypersonality.org) was collected from a very famous social networking site (Facebook). Where it contains Facebook status messages as raw text, author information, gold standard labels (both classes and scores) for classification and regression tasks. Personality traits are annotated using self-assessment questionnaire. This data is collected from around 250 different users and the number of statuses per user ranging from 1 to 223.

V. CONCLUSION

The Personality is a very complex human attribute and it also helps in describing the uniqueness of a person. Personality prediction models can be a very useful medium in domains like forensics to reduce the number of suspects, in e-commerce to make interaction with customer easier, in e-learning and information filtering and collaboration. They can also be used in fields like human resource, individual recommender system and mental illness diagnosis.

This paper is a detailed review of various methods which are used to predict personality. It is an advanced review of emerging fields in personality prediction from text. In addition to the methods various personality prediction models and datasets which could be used are also discussed here. The methods presented here include Machine Learning Algorithms, Data Mining and Neural Network methods which are Recurrent Neural Network and Convolutional Neural Network. The models discussed are DISC, MBTI and Big-Five model of personality prediction.

REFERENCES

[1] J. W. Pennebaker, L. A. King, "*Linguistic Styles: Language Use as an Individual Difference*", Journal of Personality and Social Psychology, Vol. 77, No.6, 1296-1312, 1999.
[2] S. M. Mohammad, S. Kiritchenko, "*Using Nuances of Emotion to Identify Personality*", In AAAI -2012, pp: 27-30.
[3] K. C. Pramodh, Y. Vijayalata, "*Automatic Personality Recognition of Authors UsingBig Five Factor Model*", IEEE International Conference on Advances in Computer Applications (ICACA), 2016.
[4] H. Wei, F. Zhang, N. J. Yuan, C. Cao, H. Fu, X. Xie, Y. Rui, and W. Ma, "*Beyond the words: Predicting user personality from heterogeneous information*", In Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, pages 305–314. ACM, 2017.
[5] N. Majumder, S. Poria, A. Gelbukh, and E. Cambria, "*Deep learning-based document modeling for personality detection from text*", IEEE Intelligent Systems, 32(2):74–79, 2017.
[6] L. Gou, M. X. Zhou, and H. Yang,"*Knowme and shareme: understanding automatically discovered personality traits from social media and user sharing preferences*", In Proceedings of the 32nd annual ACM conference on Human factors in computing systems, pages 955–964. ACM, 2014.
[7] J. Digman, "*Personality Structure: Emergence of the Five- Factor Model*", Ann. Rev. Psychology, vol. 41, no. 1, 1990, pp. 417–440.
[8] S. J, B. Lepri , Aharony N, Pianesi F, Sebe N, Pentland A.S., "*Friends dont Lie - Inferring Personality Traits from Social*

Network Structure", In Proceedings of International Conference on Ubiquitous Computing. 2012

[9]   S. Poria et al., "*Convolutional MKL Based Multimodal Emotion Recognition and Sentiment Analysis*", Proc. IEEE Int'l Conf. Data Mining, 2016, pp. 439–448.

[10] L. Qiu, H. Lin, J. Ramsay, F. Yang, "*You are what you tweet: Personality expression and perception on twitter*", Journal of Research in Personality, 46(6):710–718, 2012

[11] B. Plank,D. Hovy, "*Personality Traits on Twitter*", Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA 2015), pages 92–98,Lisboa, Portugal, 17 September, 2015.

[12] A. Tyagi,"*Personality Profiles Identification Using MBTI Test for Management Students: An Empirical Study*",Journal of the Indian Academy of Applied Psychology,January 2008, Vol. 34, No.1, 151-162.

[13] B. Verhoeven, W. Daelemans, "*TWISTY: a Multilingual Twitter Stylometry Corpus for Gender and Personality Profiling*", In LREC, 2014.

[14] C. A. Bhardwaj, M. Mishra, S. Hemalatha,"*An Automated Compatibility Prediction Engine using DISC Theory Based Classification and Neural Networks*", VIT, Chennai.

[15] Y. Hernández1, C. A. Peña, A. Martínez, "*Model for Personality Detection based on Text Analysis*".

[16] Kosinski, D. Stillwell, T. Graepel," *Private traitsand attributes are predictable from digital records of human behavior*", Proceedings of the National Academy of Sciences, 110(15):5802–5805,2013.

[17] Y. R. Tausczik, J. W. Pennebaker, "*The psychological meaning of words: Liwc and computerized text analysis methods*", Journal oflanguage and social psychology, 29(1):24–54, 2010.

[18] J. W. Pennebaker, M. E. Francis, R. J. Booth, "Linguistic inquiry and word count", Technical Report, Dallas, TX: Southern Methodist University, 1993.

[19] Mairesse, M. A. Walker, M. R. Mehl, R. K. Moore, "*Using linguistic cues for the automatic recognition of personalityin conversation and text*", Journal of artificial intelligence research, 30:457–500, 2007.

[20] F. Liu, J. Perez, S. Nowson, "*A language-independent and compositional model for personality trait recognition from short texts*", arXiv preprint arXiv:1610.04345, 2016.

[21] X. Sun, B. Liu, J. Cao, J. Luo, "*Who Am I? Personality Detection based on Deep Learning for Texts* ", IEEE, 978-1-5386-3180-5/18, 2018.