

Emotion Detector Using Deep Learning

Ankit Kumar

Department of Computer Science & Engg, G. L. Bajaj Institute of Technology & Management, Greater Noida, India

Ayush Verma

Department of Computer Science & Engg, G. L. Bajaj Institute of Technology & Management, Greater Noida, India

Nihal Kr. Singh

Department of Computer Science & Engg, G. L. Bajaj Institute of Technology & Management, Greater Noida, India

ABSTRACT

In 1884, Sir Francis Galton firstly came up with the idea that personality could be broken into different sub-features (PE and ST.). As time goes on, this idea got to be adapted and more completed. Then the theory of 5 Big Personality traits came out. 5 big personality traits suggest that human's personality can be described by five primary factors: extroversion, conscientiousness, agreeableness, neuroticism, and openness. Extroversion indicates how outgoing a person is. Conscientiousness describes how organized and efficient the person could be. Agreeableness represents the tendency of being cooperative versus suspicious of a person. Neuroticism describes the level of how sensitive a person is. And Openness reflects the extent of Curiosity and Creativity a person would have. Each of the five factors represents a unique aspect of human personality.

Keywords

Deep Learning, Neural Networks, Tokens, Logistic Regression

1. INTRODUCTION

Personality is a fascinating topic among all time. Personality is a combination of a person's emotion, idea's, thought and even life credo. Personality also affects how we interact with others, make decisions and deal with troubles. While this being so important, we often have little information about personality. People may hardly identify themselves personality clearly. In spoken language analysis, an utterance is the smallest unit of speech. It is well known that utterances convey a great knowledge about the speaker. Sentiment analysis is a huge category under natural language processing, since text data often carries a huge amount of information about the person emotion. However, Personality identification seems to be a less common field than other ones. With the importance of personality analysis, we think it would be interesting to do text classification based on big five personality traits using essays written by corresponding persons and analyze how different writings or words used could affect human personality.

2. OBJECTIVE AND SCOPE

The purpose of the paper is to do document-level classification based on the author's Big Five personality traits. Our approach can be summarized in the following steps:

Step 1:	Collect individual essay
Step 2:	Collect ratings on authors' Big Five personality traits
Step 3:	Extract relevant features
Step 4:	Apply classification models based on extracted features step
Step 5:	Test model and do comparison

For each document, we used three different methods for feature extraction: convolutional neural network, Mairesse features, TF-IDF. After that we applied SVM, logistic regression, neural network separately on these features and did comparison with the baseline.

3. METHODOLOGY

We reduced all letters to lowercase. All the punctuation marks according to the priorities should be dealt with. For Example: ".", ",", "?", "!" are important punctuation that must be retained while others need to be removed.

3.1 Sentence Splitting

- Each essay is separated into sentences by periods, exclamation marks, and question marks.
- Each sentence is separated into tokens by space.

3.2 Data Cleaning

- Within each sentence, punctuation tokens were removed.
- For numbers appeared in the writings, all are represented by the same token.
- Words that are not purely comprised of alphabetical characters are removed.
- Words that have length less than 2 are removed.

3.3 Unification: Lower cases

4. RESULT & DISCUSSION

In order to analyze our model, we are going to compare the performances of different classifiers based on the test data.

Table 1: Compare the performances of different classifiers

Feature Set	None	CNN	CNN+Mairesse		CNN+TF-IDF		CNN+Mairesse+TF-IDF		
			LR SVM	LR SVM	LR SVM	LR SVM	LR SVM	LR SVM	
Extroversion	55.13	53.04	52.43	53.04	53.44	55.47	54.05	56.48	57.69
Conscientiousness	55.28	52.22	52.50	55.06	54.66	58.91	58.30	58.70	55.06
Agreeableness	55.35	52.43	51.82	54.25	53.24	55.87	53.85	56.88	54.86
Neuroticism	58.09	51.82	50.63	54.86	53.64	59.51	54.66	58.10	54.66
Openness	59.57	52.43	50.81	52.23	52.02	52.83		58.10	
								54.86	

5. CONCLUSION

In conclusion, personality detection is a fascinating field with lots of unknowns. With this paper, by using more features and different models, our model accuracy surpassed baseline accuracy on Extroversion, Conscientiousness, Agreeableness and Neuroticism personality traits. For classifiers, logistic regression worked the best among all. Support Vector Machine performed well as well. Other methods such as two layers neural networks were tried but didn't work. As for feature set, CNN trained features were used as base features. Based on CNN features, adding Mairesse features increased model accuracy slightly, and adding TF-IDF features increased model accuracy more significantly. However, more features is not always better. Sometimes using feature set with CNN and TF-IDF had better performance than CNN, TF-IDF and Mairesse feature set.

REFERENCES

[1] Mayuri Pundlik Kalghatgi, Manjula Ramannavar, and Nandini S Sidnal. 2015. A neural network approach to personality prediction based on the big-five model. International Journal of Innovative Research in Advanced Engineering (IJIRAE), 2(8):56–63

[2] Wang Ling, Tiago Lu'is, Lu'is Marujo, Ramon Fernandez Astudillo, Silvio Amir, Chris Dyer, Alan W Black, and Isabel

Trancoso. 2015. Finding function in form: Compositional character models for open vocabulary word representation. arXiv preprint arXiv:1508.02096.

[3] Francois Mairesse, Marilyn A Walker, Matthias R Mehl, and Roger K Moore. 2007. Using linguistic cues for the automatic recognition of personality in conversation and text. Journal of artificial intelligence research, 30:457–500.

[4] Scott Nowson and Jon Oberlander. 2006. The identity of bloggers: Openness and gender in personal weblogs. In AAAI spring symposium: Computational approaches to analyzing weblogs, pages 163–167. Palo Alto, CA. Shrou PE and Fiske ST. Personality research, methods, and theory. Psychology Press.

[5] James W. Pennebaker and Laura A. King. 1999. Linguistic styles: Language use as an individual difference. Journal of personality and social psychology, 77(6):1296– 1312.

[6] John C. Platt. 1999. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In ADVANCES IN LARGE MARGIN CLASSIFIERS, pages 61–74. MIT Press.