

Proceedings of IYSW, (2020), vol. 9, pp 457-463.

Journal homepage: <http://journals.sdu.edu.kz/index.php/iysw>



**International
Young Scholars'
Workshop**

Prediction of Thinking-Feeling personalities of movie characters

A. Atanbekov¹, H. Shirzad²

^{1,2}Suleyman Demirel University, Kaskelen, Kazakhstan

Email: 181107036@stu.sdu.edu.kz

Abstract

Abstract. This paper is going to explore the difference between the vocabularies used by Thinking and Feeling personalities. To find out this we used the Machine Learning algorithm Naïve Bayes which showed the best accuracy in comparison with others. The concept was motivated by essays of scholars when they submitted the first time at university and to get the full psychological portrait of the student only by given text. To train the model we used a labeled dataset that was collected through a forum with real persons. This dataset contains the type of the person and their posts in social media. To test the model using another dataset which contains information about movie characters and their speech used in the movie. Psycho-type was described by Myers-Briggs Type Indicators (MBTI) which is one of the most popular typologies. To achieve better accuracy of prediction we trained the model separately for Thinking and Feeling predictors. Overall, we achieved better accuracy than previous studies and showed the difference between the vocabularies used by Thinkers and Feelers.

Keywords: MBTI, prediction

Prediction of Thinking-Feeling personalities of movie characters

MBTI (Myers-Briggs Type Indicator) - a typology of personality (or type indicator) Myers-Briggs is a holistic system that describes the individual qualities and characteristics of a person that make up 16 personality types. The main advantage of Myers-Briggs personality typology (MBTI) is that it does not require special knowledge to master it [1]. Isabelle Myers-Briggs strove to create a typology that anyone can use in his life.

Myers-Briggs personality typology (MBTI) arose from Carl Jung's ideas in the 1920s. Its authors - Isabelle Myers-Briggs and her mother Katherine Cook Briggs were able to significantly complement Jung's theory and create their own method of studying the types of people. They devoted their whole life to the development of this typology, conducting research, developing diagnostic tools, etc. The first version of the Myers-Briggs test appeared in 1942, the first version of the manual on the use of typology - in 1944. Myers-Briggs personality typology (MBTI) is currently one of the most popular methods and techniques in the United States and Europe. In Kazakhstan and the CIS, the typology is becoming increasingly well-known with the publications translated into Russian.

According to Jung theory, a person's personality can be described using four psyche preferences that add up to a certain psychological type of personality:

1. Preference for Intuition (N) or Feeling (S) – defines the process of perception, i.e. awareness of various objects, people, events and ideas
2. Preference for Mind (T) or Feeling (F) – determines the decision-making process, how we come to this or that conclusion
3. Preference for Introversion (I) or Extroversion (E) – determines interest in the inner or outer world
4. Preference for Perception (P) or Judgment (J) – determines the way of life and the way of interaction with the outside world.

Other studies, like Carol Ritberger, considered psycho-types as a classification of a humans by different colors [2]. As shown on previous works transcription can show personality traits [3].

As Golbeck [4] identified, one of the big challenges in this study is the lack of labeled dataset. Thus, we used well-known dataset from Kaggle website which originally was collected from forum.

Main contribution of the paper:

1. Getting features to differentiate Thinking-Feeling personalities
2. Using the movie characters for testing trained model
3. Getting 69% accuracy of prediction which better than previous studies

Current section described related work and subject at all. Following section describes datasets which was used in this work. Section 3 and 4 describes preprocessing steps and prediction. Section 5 and 6 describes limitations and future directions. Finally, section 7 concludes the paper.

II. Dataset

The train dataset uploaded from Kaggle which in origin was collected through the Personality Care forum and contains data about real psycho-types and their comments/posts in social media. The dataset contains over 8600 rows of data on each row is person's:

- type (these persons 4 letter MBTI type)

- section of each of the last 50 things they have posted (each entry separated by 3 pipe characters)

As test dataset we used corpus of movie lines data. This corpus comes from the paper, "Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs" [5]. This corpus contains a metadata-rich collection of fictional conversations extracted from raw movie scripts:

- 220,579 conversational exchanges between 10,292 pairs of movie characters
- involves 9,035 characters from 617 movies
- in total 304,713 utterances

III. Data preprocessing

Main concept is to differentiate vocabulary of Thinkers and Feelers. To achieve it we differentiate typology and their comments/posts. As shown on Figure 1 dataset contains near to equal count of Thinkers and Feelers. Thus, we can say that our dataset is normally distributed. It means that model will not tend to overfitting, but it depends on feature importance also. To get high accuracy in prediction necessary to realize high precision in both datasets train and test.

The cleaning data in steps:

- 1) First, keep only Thinking-Feeling function others was cleaned;
- 2) Next used regex-based pattern to (1) remove links, (2) strip punctuations, (3) remove non-words, (4) remove multiple letter repeating words and (5) transform all words to lower case

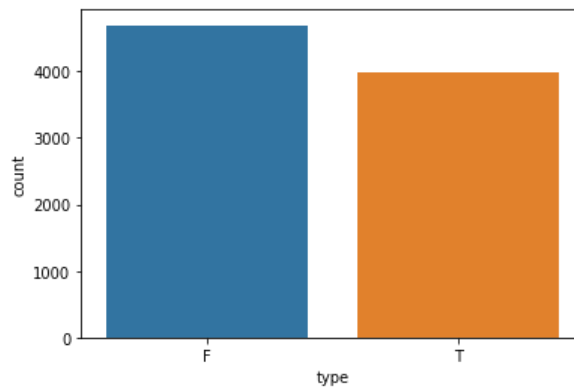


Figure 1. Amount of Thinking-Feeling personalities

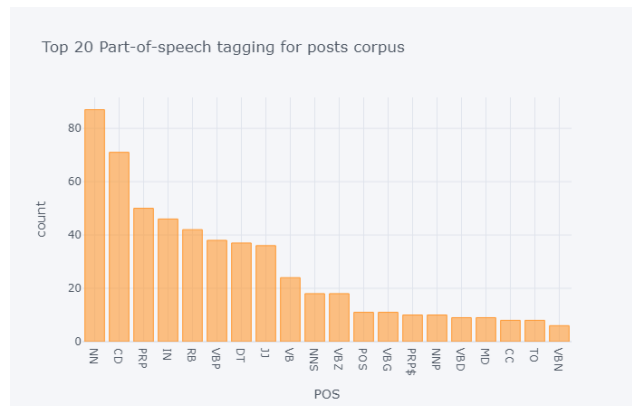


Figure 2. Top 20 Part of Speech

VI. Future Directions

In future we'd prefer to make model which be able to predict whole MBTI type not only by the given data, also by the speech, body movements and etc. The MBTI is **additionally** useful as a treatment planning tool to match individuals preferring Thinking with cognitive therapy and, potentially, individuals who preferred Feeling with humanistic therapies. Also, it'd be helpful for gathering required teams of people to realize best results.

VII. Conclusion

Movie characters was predicted with accuracy 69% and it requires to boost feature selection. Now we are able to definitely say that it's possible to predict the Thinking-Feeling character of an individual by given only text. Experiments shows that there's difference between prediction for every type. Accuracy are increase if discover more language specificity like slang or some language preferences. What we've to specialize in first is feature selection for Kazakh and Russian text. This study presents a look work using rough set analysis to grasp the differences between which words used Thinking and Feeling psychological types. this is often an alternate to the normal method of searching for a person's type. To expand on its use, we are able to also boost the prediction model the remainder typology of psycho-types. There are several important factors, using them we are able to make decisions. Furthermore, knowing the personalities of person plays a vital role in communicating and interacting with others composition.

References

- 1 Jung, 1921/1971; Myers, McCaulley, Quenk, and Hammer, 1998.
- 2 Ritberger, C. (2009). *What Color Is Your Personality?: Red, Orange, Yellow, Green...* Hay House, Inc.
- 3 F. Mairesse, M. Walker, M. Mehl, R. Moore, "Using linguistic cues for the automatic recognition of personality in conversation and text," *Journal of Artificial Intelligence Research (JAIR)*. 30(1), 457–500, 2007. Damodaran, A. *Corporate Finance: Theory and Practice*, 2nd ed. Wiley, 2001.
- 4 J. Golbeck, M. Edmondson, and K. Turner, "Predicting personality fromtwitter", pp. 149–156, Oct. 2011.[doi:10.1109/PASSAT/SocialCom.2011.33](https://doi.org/10.1109/PASSAT/SocialCom.2011.33)
Mark Yatskar and Bo Pang and Cristian Danescu-Niculescu-Mizil and Lillian Lee, "For The Sake Of Simplicity: Unsupervised Extraction