Automatic Personality Prediction with NLP

1. Introduction

Personality is a set of patterns of enduring thoughts, feelings, and behaviors that set individuals apart (Roberts & Mroczek, 2008). To understand a person is to understand their personality; thus, it has long been a goal of psychology to accomplish this task. While personality manifests itself in nearly all our actions, one of the most prominent mediums is text. With the increasingly vast amounts of user-generated text in the modern era, the prospect of deriving psychological insights through analysis of written text has become more and more tantalizing today. In this review, we discuss applications of natural language processing (NLP) techniques to the problem of automatic personality prediction.

2. Personality Models

Personality must be quantified before it can be rigorously analyzed, described, and detected. To that end, psychologists have built many models of personality. Some examples are Allport's trait theory (Allport, 1937), Cattell's 16 Factor Model (Cattell, Eber, & Tatsuoa, 1970), Eysenck Personality Questionnaire (Eysenck & Eysenck, 1975), The Big Five model (Costa & McCrae, 1992), and Myers-Briggs Type Indicator (Briggs-Myers & Myers, 1995). The latter two continue to see widespread use.

The Big Five Model, also known as the OCEAN model, pares down the list of personality traits into five broad factors, defined on a 100-point scale:

- Openness: openness to new experiences vs. conservative
- Conscientiousness: meticulous vs. disorganized
- Extraversion: sociable vs. reserved
- Agreeableness: considerate vs. competitive
- Neuroticism: vulnerable vs. stable (emotionally)

Psycholinguists have found that each dimension is exhibited through language usage. For instance, Extraverts talk more, are prone to repetition, have fewer pauses, are less formal, and so on (Mairesse & Walker, 2007). In theory, these and other patterns can be captured in NLP analysis as features for personality classification and scoring.

The Myers-Briggs Type Indicator (MBTI) model has its origins in Jungian psychology. It is similarly reductive, but along the following four dimensions:

- **E**xtraversion/**I**ntroversion
- **S**ensing/I**N**tuition
- **T**hinking/**F**eeling
- **J**udgment/**P**erception

Unlike the Big Five, which scored each dimension from 0-100, MBTI is binary for each. As a result, there are only 16 possible personality definitions under MBTI. This renders personality prediction using MBTI a simpler task, but also raises doubts about the expressivity of MBTI and how accurately it truly is for describing personality. There are also fewer studies about how different MBTI personality types differ linguistically, so it is perhaps more difficult to interpret how a personality classifier makes decisions in this context vs. the Big Five.

Modeling personality is not solved, and despite the ubiquity of the above models, there remain criticisms and concerns. This makes personality prediction an even more difficult task. Furthermore, users

may not be the best judges of their own personalities. Perhaps they will omit or misrepresent information based on their affinity towards their personalities, or what ideal personality traits they desire. These and other biases are commonplace in psychology contexts. The challenge of predicting labels we may not be depicting accurately and that may not even be accurate is one that continues to haunt this domain.

3. Psycholinguistic and NLP Features

The history of psycholinguistic analysis stretches over a century. Freud analyzed slips of the tongue as being revealing of one's psychological state and true intentions (Freud, 1901). Rorschach studied how people's descriptions of inkblots could reflect their inner thoughts and motives (Rorschach, 1921).

By the 1950s, approaches evolved to be more quantitative and methodical. Gottschalk and Gleser developed a scoring scheme meant to be applied to five-minute long stream-of-consciousness recordings from patients in order to identify psychological themes such as anxiety and hostility (Gottschalk, Gleser, Daniels, & Block, 1958). Later, building off the work of McClelland on thematic apperception tests that analyzed written descriptions of portraits (McClelland, 1979), Stone et al. built the General Inquirer program that could apply McClelland's scheme to any kind of text to presumably detect signs of personality traits and even mental disorders (Stone, Dunphy, Smith, & Ogilvie, 1966). This was the first computerized psycholinguistic analysis program. By the 1980s, we also started to see more widespread usage of statistical tools. Analysis of corpora of medical interviews identified correlations such as usage of first-person singular pronouns with depression (Weintraub, 1989), proving the applicability of NLP methods in diagnosing psychological state and personality.

To analyze text with modern machine learning and statistical techniques, it must be represented in some quantitative form. The most common examples are the bag-of-words representation and the term frequency-inverse document frequency (tf-idf) representation, but in psycholinguistics there are several more domain-specific methodologies as well.

3.1 Linguistic Inquiry and Word Count (LIWC)

In the 1980s, Pennebaker et al. performed studies on the effects of writing about emotional life events, collecting many samples of bleak stories (Pennebaker & Beall, 1986). These were judged along several axes, such as coherency, optimism, emotionality, etc. They quickly found several obstacles that are very familiar to NLP researchers. Firstly, ratings were subjective and varied among judges for the same stories. Secondly, manually rating text is expensive. Not only was it costly monetarily and temporally, but it was also draining emotionally to read such depressing texts. In search of a more impartial and efficient rater, the Linguistic Inquiry and Word Count program (LIWC) was born (Pennebaker & Francis, 2010).

LIWC consists of a processing program and several lookup dictionaries. It defines 15 categories and 91 subcategories, ranging from simple lingual classes like articles ("a", "an", "the") and pronouns to more subjective groupings such as positive and negative emotion words, social words referring to relationships, etc. (Feizi-Derakhshi, et al., 2021) The program simply analyzes the percentages of each category in given text. Clearly, the dictionaries form the meat of LIWC. Great care is taken in building and updating these dictionaries, involving reviews from panels of human judges as well as statistical NLP analysis within each category (Tausczik & Pennebaker, 2010).

While LIWC makes sense on an intuitive level – if someone is using anger-related words at a higher rate than in normal text, they are more likely to be angry – it also has its problems. Despite the effort put into the dictionaries, mistakes and ambiguity are still possible. "Mad," for instance, is usually an anger

word and negative emotion word. But in the context of "he's mad for her," it does not fall into any of those categories. Phenomena such as sarcasm and metaphor are also unrecognized by LIWC and can lead to misclassifications. Lastly, with so many categories and subcategories, interpretation of results is still highly subjective. It is nontrivial to go from LIWC outputs to a personality determination.

3.2 Machine Readable Dictionary (MRC) Psycholinguistic Database

The MRC Psycholinguistic Database is a database consisting of 27 psycholinguistic, semantic, syntactic, and phonological features labeled by humans for over 150000 words. Some examples of features include (Paetzold & Specia, 2016):

- Familiarity: Frequency with which a word is used
- Age of Acquisition: Age at which a word is typically learned
- Concreteness: Palpability of object the word describes
- Linguistic features such as part of speech, pronunciation, etc.

Unlike LIWC, there is no analytical component, so it is even less directly translatable into a personality judgment. MRC's coverage is also not as comprehensive. Not every word is scored for the more complex features. For example, only ~6% of words have a concreteness label. Furthermore, while LIWC is still being updated and sold as commercial software, MRC received its final update in 1988 (Wilson, 1988). Still, LIWC and MRC represent valuable textual features that can be used as inputs to a larger personality classifier.

3.3 Word Embeddings

Before the current decade, most NLP systems treated words as indices in a larger vocabulary, without any concept of measuring and quantifying similarity. For instance, in bag-of-words representations of "mad" and "angry," both words would be one-hot vectors that are as far from each other as any other word, despite their highly similar meaning. In LIWC, there would be separate and distinct entries for "mad" and "angry" as well. While simple and robust, these techniques can only go so far for more complex tasks, such as speech recognition or personality prediction.

In more recent years, the field has been revolutionized by the ability to generate high-quality vector representations of words that embed them in semantic space. Models such as Word2Vec (Mikolov, Chen, Corrado, & Dean, 2013) and GloVe (Pennington, Socher, & Manning, 2014) are able to represent words as real-valued vectors that not only capture word similarities and semantic meaning, but can also be manipulated using simple algebraic operations. Word analogies are a popular example. For instance, within embedding space, vector("king") - vector("man") + vector("woman") produces a vector closest to the embedding for "queen." While embeddings are a general concept that apply to all forms of text analysis, not simply automatic personality prediction, they are another source of rich textual features that can be useful for our task.

4. Automatic Personality Prediction Methods

Several attempts at personality prediction have been made using the above text representations. In 2005, using the Big Five Model, Argamon et al. performed binary classification of extraversion and neuroticism (Argamon, Dhawle, Koppel, & Pennebaker, 2005) using LIWC inputs, on a dataset of student essays and self-reported personality types (Pennebaker & King, 1999). They were able to reach 58% accuracy on both classification tasks. Mairesse et al. augmented personality self-reports with observer reports as well, and combined LIWC and MRC features (Mairesse F. , Walker, Mehl, & Moore, 2007).

Most notably, they found that observer personality scores were more predictable than self-reports. This highlights the challenge of label reliability in personality prediction.

Later approaches benefited significantly from availability of social media data as well as the development of embeddings. In fact, in 2015, models based solely on analyzing Facebook like patterns and which pages were liked were shown to outperform questionnaire-based personality judgments from close friends and family (Wu, Kosinski, & Stillwell, 2015). Park et al. combined linguistic features such as topics and word/phrase frequencies with online personality questionnaires, Facebook profile information, and other social media data to predict personality, and obtained significantly positive correlations (from 0.35 to 0.43) between model assessments and self-reports on Big Five personality traits (Park, et al., 2015), further proving the validity of the approach and the value of incorporating social media data.

Besides taking advantage of newer advances in the NLP state-of-the-art, such as transformers for language modeling, recent research in personality prediction continues to explore how to maximally utilize text data. In Personality2vec, Guan et al. look to exploit not just semantic information in user texts, but also structural (Guan, Wu, Wang, & Liu, 2020). Using linguistic inputs such as LIWC and other factors, they construct a graph representation of the possible features defining personality characteristics. In this way, users can be represented as biased walks through the network. These can then be passed to a language model that produces Big Five personality vectors for each user. Empirical results demonstrated a significant improvement over several previous baseline models across multiple datasets, showing that there is still room to go for fully leveraging text information.

5. Conclusion

Personality prediction continues to be a topic of research on both the psychological and NLP fronts. Several themes are clear from past work. Firstly, text is not only one of the primary media in which personality emerges, but it is also the most accessible, ensuring the importance of NLP in personality analysis.

Secondly, personality prediction is made challenging by continuing debate over our current theories of personality. Although we have largely settled on the Big Five currently, there are still open questions about the generalizability and expressivity of the Big Five. Recently, Kulkarni et al. used Facebook status messages and linguistic features to infer latent human traits from scratch rather than rely on existing personality models (Kulkarni, et al., 2018). They showed that these new traits generalized better to predicting aspects of personality and behavior, demonstrating how NLP can potentially even inform new directions of psychological theory.

Thirdly, there remains much unexplored research territory in utilizing text data, especially in conjunction with their contexts. Combining social media post contents with other social media metrics, for example, was shown to improve over solely relying on textual analysis in personality prediction. Personality2vec and other approaches suggest alternative representations of text can be fruitful not just for personality prediction, but other NLP tasks as well.

Despite the myriad potential directions for future work, researchers have been able to achieve useful results, sometimes even matching or exceeding human performance. Considering the many practical applications in medical and psychological treatment, improving communications through better mutual understanding, etc., hopefully progress in the field will continue.

6. References

Allport, G. W. (1937). *Personality: a psychological interpretation.* Holt.

Argamon, S., Dhawle, S., Koppel, M., & Pennebaker, J. W. (2005). Lexical predictors of personality type. *Joint Annual Meeting of the Interface and the Classification Society of North America.*

Briggs-Myers, I., & Myers, P. B. (1995). *Gifts differing: Understanding personality type.* Davies-Black Publishing.

Cattell, R. B., Eber, H. W., & Tatsuoa, M. M. (1970). *Handbook for the Sixteen Personality Factor Questionnaire.* Champaign: Institute for Personality and Ability Testing.

Coltheart, M. (1981). The MRC Psycholinguistic Database. *The Quarterly Journal of Experimental Psychology*, 85-95.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory and NEO Five-Factor Inventory: Professional Manual.* Psychological Assessment Resources.

Eysenck, H. J., & Eysenck, S. B. (1975). *Manual of the Eysenck Personality Questionnaire.* London: Hodder and Stoughton.

Feizi-Derakhshi, A.-R., Feizi-Derakhshi, M.-R., Ramezani, M., Nikzad-Khasmakhi, N., Asgari-Chenaghlu, M., Akan, T., . . . Jahanbakhsh-Naghadeh, Z. (2021). The state-of-the-art in text-based automatic personality prediction. *arXiv.org*.

Freud, S. (1901). *Psychopathology of Everyday Life.* New York: Basic Books.

Gottschalk, L. A., Gleser, G. C., Daniels, R. S., & Block, S. (1958). The Speech Patterns of Schizophrenic Patients: A Method of Assessing Relative Degree of Personal Disorganization and Social Alienation. *Journal of Nervous and Mental Disease*, 153-166.

Guan, Z., Wu, B., Wang, B., & Liu, H. (2020). Personality2vec: Network Representation Learning for Personality. *Fifth International Conference on Data Science in Cyberspace.* IEEE.

Kulkarni, V., Kern, M. L., Stillwell, D., Kosinski, M., Matz, S., Ungar, L., . . . Schwartz, H. A. (2018). Latent human traits in the language of social media: An open-vocabulary approach. *Public Library of Science One*.

Mairesse, F., & Walker, M. (2007). Personage: Personality generation for dialogue. *Proceedings of the 4th Annual Meeting of the Association for Computational Linguistics.*

Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research*, 457-500.

McClelland, D. C. (1979). Inhibited power motivation and high blood pressure in men. *Journal of Abnormal Psychology*, 182-190.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *International Conference on Learning Representations.*

Paetzold, G. H., & Specia, L. (2016). Inferring Psycholinguistic Properties of Words. *North American Chapter of the Association for Computational Linguistics - Human Language Technologies* (pp. 435-440). San Diego: Association for Computational Linguistics.

Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., . . . Seligman, M. E. (2015). Automatic Personality Assessment Through Social Media Language. *Journal of Personality and Social Psychology*, 934-952.

Pennebaker, J. W., & Beall, S. K. (1986). Confronting a traumatic event: Toward an understanding of inhibition and disease. *Journal of Abnormal Psychology*, 274-281.

Pennebaker, J. W., & Francis, M. E. (2010). Cognitive, Emotional, and Language Processes in Disclosure. *Cognition and Emotion*, 601-626.

Pennebaker, J. W., & King, L. A. (1999). Linguistic Styles: Language Use as an Individual Difference. *Journal of Personality and Social Psychology*, 1296-1312.

Pennington, J., Socher, R., & Manning, C. (2014). GloVe: Global Vectors for Word Representation. *Conference on Empirical Methods in Natural Language Processing* (pp. 1532-1543). Doha, Qatar: Association for Computational Linguistics.

Roberts, B. W., & Mroczek, D. (2008). Personality Trait Change in Adulthood. *Current Directions in Psychological Sciencce*, 31-35.

Rorschach, H. (1921). *Psychodiagnostik.* Leipzig: Ernst Bircher Verlag.

Stone, P. J., Dunphy, D. C., Smith, M. S., & Ogilvie, D. M. (1966). *The General Inquirer: A Computer Approach to Content Analysis.* Boston: MIT Press.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Methods. *Journal of Language and Social Psychology*, 24-54.

Weintraub, W. (1989). *Verbal behavior in everyday life.* New York: Springer.

Wilson, M. (1988). MRC Psycholinguistic Database: Machine-usable dictionary, version 2.00. *Behavior Research Methods, Instruments, & Computers*, 6-10.

Wu, Y., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *National Academy of Sciences*, (pp. 1036-1040).