Personality Detection from Text: A Review

Basant Agarwal

Department of Computer Engineering, Swami Keswanand Institute of Technology, India
{thebasant@gmail.com}

Abstract

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. In this paper, basic introduction to the very emerging field i.e. personality detection from text is presented. This paper discusses the state-of-art methods for personality detection. In addition, state-of-art publically available dataset is discussed.

Keywords: Personality detection, LIWC, MRC, Big-Five.

I. INTRODUCTION

Social networking on the web has grown dramatically over the last decade. Social networks have become widely-used and popular mediums for information distribution as well as for social interactions. User activities on social networking websites provide valuable insight into individual behavior, experiences, opinions and interests. It is quite similar how a person interacts socially with the human nature and behavior.

Personality is the most intricate human attribute and it also describes the uniqueness of a person. Personality is one of the fundamental aspects, by which we can understand behavioral personalities. It has been a long-term goal for psychologists to understand human personality and its impact on human behavior. Behavior involves an interaction between a person's underlying personality traits. The situation, that a person finds himself or herself in, plays a major role on his or her reaction. However, in most of the cases, people respond with respect to their underlying personality traits. It is possible to access and analyze large amounts of text samples in order to automatically identify personality types of authors and predict potential reactions and behaviors.

Humans have the tendency to understand others’ behavior on the basis of the observation of their everyday behavior. Huge number of researchers around the word has been attracted to work on this research domain from different fields especially researchers in computational linguistics, psychology, artificial intelligence, natural language processing, human-machine interaction, behavioral analytics, and machine learning.

Use of social networking websites has been increased exponentially in recent times. A survey of social networking websites estimated approximately 115 million members over all the sites on the web in January 2005, and just after five years Facebook alone has exceeded 500 million members. One of the most universal on-line environments, Facebook, is becoming an increasingly daily activity of peoples’ around the world.

Currently it facilitates daily interactions of over 800 million users spending more than 40 minutes daily on the platform on average [1]. Facebook profiles became an important source of information used to form impressions about others. In the process of creating social networking profiles, users reveal a lot about themselves both in what they share and how they say it. Through self-description, status updates, photos, and interests, much of a user’s personality comes out through their profile. For example, people examine other people’s Facebook profiles when trying to decide whether to start dating them, and they are also used when assessing job candidates.

II. APPLICATIONS

In recent years the interest of the scientific community in personality recognition has grown very fast. The current challenges are instead related to the extraction of personality from mobile social networks, from social network sites and from languages different from English. There are also many other applications that can take advantage of personality recognition, including social network analysis, recommendation systems, deception detection, authorship attribution, sentiment analysis/opinion mining, and many others. Previous research also showed that personality is correlated with many aspects of life, including job success [16], attractiveness [17], marital satisfaction [18] and happiness [19]. The further development of the discipline is beneficial for many activities that are performed by means of online facilities on a daily basis (customer support, and recommendation of services and products, etc.). Recruiters
of the HR department analyse hundreds of job applications working hard to map them to the required characteristics the future stuff should have [16]. At the same time the developers of the e-commerce resources are constantly improving the personification algorithms to help the customers obtain products and services that match the needs more precisely and present the information in a more appealing way to increase sales [20]. All these tasks will eventually involve a crucial step of implicit (mental) or explicit (through a user profile) modeling of the user personality.

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. Having captured past user interactions is only a starting point in explaining the user behavior from a personality point of view. It has been proved that personality detection models are very useful in predicting job satisfaction, professional and romantic relationship success, and even preference for different interfaces.

In today’s scenario, it is required to conduct a personality text to accurately measure users’ personality; therefore, it is impractical to use personality analysis models in many social media domains. An individual’s success depends largely on the impression made on others. Success on the job market, finding romantic partners, and gaining support and positive attention from one’s social background heavily depend on what others think of you.

The fact that people can judge each other’s personality based on Facebook profiles implies two things: an individual’s personality is manifested on their Facebook profile, and some aspects of Facebook profiles are used by people to judge others’ personalities. However, the overlap between Facebook profile features that contain the actual personality cues and features used by people to form personality judgments does not have to be perfect. It is possible that some of the actual personality cues are ignored or misinterpreted by the people, while some non-relevant features are used in the judgment. Humans are prone to biases and prejudices which may affect the accuracy of their judgments. Also, certain features of a Facebook profile are difficult for humans to grasp. For example, while the number of Facebook friends is clearly displayed on the profile, it is more difficult for a human to determine features such as the network density.

III. Big Five Model

The “Big Five” model of personality dimensions has emerged as one of the most well-researched measures of personality structure in recent years. Personality is defined as the coherent patterning of affect, behavior, cognition and desire over time and space, which are used to characterize unique individuals. The most widely used personality traits model in the literature is the “Big-Five” model, five broad personality dimensions (Matthews et al., 2003). It describes the human personality as a vector of five values corresponding to bipolar traits. This is a popular model among the language and computer science researchers and it has been used as a framework for both personality traits identification and simulations. The Big-5 personality traits model is defined as follows:

- O (Openness): Artistic, curious, imaginative, curious, intelligent, and imaginative. High scorers tend to be artistic and sophisticated in taste and appreciate diverse views, ideas, and experiences.
- C (Conscientiousness): Efficient, organized, responsible, organized, and persevering. Conscientious individuals are extremely reliable and tend to be high achievers, hard workers, and planners.
- E (Extraversion): Energetic, active, assertive, outgoing, amicable, assertive. Friendly and energetic, extroverts draw inspiration from social situations.
- A (Agreeableness): Compassionate, cooperative, cooperative, helpful, nurturing. People who score high in agreeableness are peace-keepers who are generally optimistic and trusting of others.
- N (Neuroticism): Anxious, tense, self-pitying, anxious, insecure, sensitive. Neurotics are moody, tense, and easily tipped into experiencing negative emotions.

IV. Dataset Available

Personality detection model presented so far in the literature are based on various different experimental setting. Researchers developed their own dataset by crawling the text from various resources like social media, and further they manually tagged the dataset with the help of experts from psychology. However, huge number of researchers used the following dataset and is now became state-of-art dataset to test and develop new personality detection models. The most popular labeled dataset available for the evaluation of personality detection task: Essays and myPersonality datasets.

A. Essay dataset

Essays [14] is a large dataset of stream-of-consciousness texts (about 2400, one for each author/ user), collected between 1997 and 2004 and labeled with personality classes. Texts have been produced by students who took the Big5 test. The labels, that are self-assessments, are derived by z-scores computed by Mairesse et al. [1] and converted from scores to nominal classes by authors in [25] with a median split. Since this corpus has been used by different scholars [1, 15], it has been included in the shared task as a reference to previous work.

B. myPersonality dataset

MyPersonality corpus (http://mypersonality.org) was collected from the social network (Facebook) and it contains Facebook status messages as raw text, author information, gold standard labels (both classes and scores) for classification and regression tasks. Annotation of the personality traits has been done using self-assessment questionnaire. The data was collected from 250 different users and the number of statuses per user ranges from 1 to 223.
V. PERSONALITY DETECTION MODEL

Data mining techniques play a fundamental role in extracting correlation patterns between personality and variety of user’s data captured from multiple sources. Generally, two approaches were adopted for studying personality traits of social network users. The first approach uses a variety of machine learning algorithms to build models based on social network activities only [10]. The second one extends the personality-related features with linguistic cues [1], [2].

In recent years, the interest of the scientific community has been attracted towards automatic personality detection mainly due to its applications in languages different from English [11], [12], and also learning personality of users in social networks [3], [5]. This interest is due to the fact that personality detection is also very useful in social network analysis and opinion mining that is large and developing fields of research. Online social networks are huge repositories of written data which is suitable for personality recognition; Still, there are some problems in using them for building such models. (1) Social network data is generally not publicly available, (2) provided data is unlabeled, (3) it is very difficult to annotate with personality judgements and (4) Generally, it is in a lot of different languages.

In [26], authors present an automatic personality trait recognition model based on social network (Facebook) using users’ status text. They used machine learning algorithms viz. Support Vector Machine, Bayesian Logistic Regression (BLR) and Multinomial Naïve Bayes (MNB). In [27], authors developed three machine learning algorithms i.e. support vector machine, Nearest neighbour with k=1 (kNN) and Naive Bayes for inferring the personality traits of users on the basis of their Facebook updates.

Several classification techniques were used to build predictive personality models along the five personality dimensions using the linguistic features of a dataset comprised of few thousand essays solicited from introductory psychology students [1]. Authors in [1], built personality recognition model in both conversation and text via Big5. They exploited two lexical resources as features, LIWC [14] and MRC [15], and predicted both personality scores and classes using Support Vector Machines (SVMs) and M5 trees respectively. They also reported a long list of correlations between Big5 personality traits and two lexical resources they used. The Linguistic Inquiry and Word Count – LIWC (http://www.liwc.net) was used as a tool for linguistic analysis.

In [12], authors followed the work presented in [1], and developed a supervised personality detection model for modern Greek with linguistic features (like Part-of-Speech tags) and psychological features (like in LIWC). They used SVM classifier for building the machine learning model, they demonstrated that personality and language can be successfully ported from English to other languages. In [2], authors used n-gram features from a corpus of personal web-blogs for modeling four out of five personality dimensions. They built their model with SMO and Naïve Bayes machine learning methods. Their results point out the importance of the feature selection in increasing the classifiers precision yielding 83%-93% for automatic feature selection. The correlation between users’ social network activity and personality has been the focus of several studies in the last [3, 11]. In [13], authors extracted word n-grams as features from a large corpus of blogs with different feature vector construction settings, such as the presence/ absence of stop words or inverse document frequency. They found that bigrams, treated as boolean features and keeping stop words, yield very good results using SVMs as learning algorithm.

Golbeck et al. [3] proposed a model to predict personality from Facebook profile with linguistic (such as word count) and social network features (like friends count) information using machine learning algorithms. They predicted personality scores of 279 Facebook users, exploiting both linguistic features (from LIWC) and social features (i.e. friend count, relationship status). In [4], authors also predicted the personality of 279 Twitter users with the help of LIWC, structural features (i.e. hastags, links) and sentiment features, and using a Gaussian Process (GP) as learning algorithm. Tomlinson et al. [24] studied the Conscientiousness trait to detect goal, motivation, and the way the author perceives control over the described situations. They performed the analysis of event structures of textual user status updates in a Facebook dataset.

In [5], authors used network features (like followers, following, etc.) to build M5 rules based learning model for the prediction of personality scores of 335 Twitter users. Authors in [6] presented an extensive analysis of the network traits (i.e. such as size of friendship network, uploaded photos, events attended, times user has been tagged in photos) that correlate with personality of 180000 Facebook users. They predicted personality scores using multivariate linear regression (mLR), and reported good results on extraversion.

Ross et al. [7] pioneered the study of the relation between personality and patterns of social network use. They hypothesized many relationships between personality and Facebook features, including (1) positive relationship between Extraversion and Facebook use, number of Facebook friends and associations with Facebook groups; (2) positive relation between Neuroticism and revealing private information on Facebook; (3) positive correlation between Agreeableness and number of Facebook friends; (4) positive correlation between Openness and number of different Facebook features used; (5) negative relationship between Conscientiousness and overall use of Facebook. In [16], authors developed a machine learning model for neuroticism and extraversion using linguistic features such as function words, deictic, appraisal expressions and modal verbs. In [22], authors used various emotion lexicons like NRC hash tag emotion lexicon and NRC emotion lexicon for the personality detection and found key improvement in the accuracy of the PRT system.

In [23], authors proposed a new approach for personality detection which is based on incorporating the sentiment, affective and common sense knowledge from the text using resources viz. SenticNet, ConceptNet, EnoSenticNet and EmoSenticSpace. In their approach, they combined common sense knowledge based features with psycho-linguistic features and frequency based features and later the features were employed in supervised
classifiers. Further, they developed five support vector machine models for five personality traits. Their experimental results show that the use of common sense knowledge with affective and sentiment information enhances the accuracy of the existing frameworks which use only psycho-linguistic features and frequency based analysis at lexical level.

VI. CONCLUSION

Social network analysis has increased tremendously in recent times. To extract the personality of the authors on the social networking websites is very useful for much application in various domain like including job success, attractiveness, marital satisfaction and happiness. Personality detection from text means to extract the behavior characteristics of authors written the text. This paper presents state-of-art review of the emerging field i.e. personality detection from text. This paper discusses the state-of-art methods for personality detection; In addition, state-of-art publically available dataset is discussed. Two types of techniques have been employed for detection of personality from the text i.e. machine learning based approach based on social network activities and second is based on the linguistic properties present in the text.

REFERENCES


[27] G. farnadi, S. Zoghbi, M. De Cock, “Recognising Personality Traits using Facebook Status Updates”, In The Seventh International AAAI Conference on Weblogs and Social Media, Workshop on Computational Personality Recognition (Shared Task),