Mining Personality Traits from Social Text Messages

Hsin-Chang Yang[†]
Department of Information Management
National University of Kaohsiung
Kaohsiung, Taiwan
yanghc@nuk.edu.tw

Chung-Hong Lee
Department of Electrical Engineering
National Kaohsiung University of Science and Technology
Kaohsiung, Taiwan
leechung@mail.ee.nkust.edu.tw

Chia-Yi Yeh
Department of Information Management
National University of Kaohsiung
Kaohsiung, Taiwan
m1063314@mail.nuk.edu.tw

ABSTRACT

Recently, approaches of applying personality on practical applications have attracted attention from multidisciplinary areas beyond psychology. An example is the personality-based recommender systems that have been successfully applied to music, movie, and game recommendation tasks. However, finding the personality traits of a person is not a trivial task. In this work, we will propose an automatic scheme for identifying the Big Five personality traits using text mining techniques. We analyzed the social messages posted by Facebook users to find the correlations among keywords and personality traits. The personality traits of a person can then be identified by examining his social messages according to such correlations. We performed the experiments using the myPersonality dataset and obtained promising results.

CCS CONCEPTS

• Human-centered computing → Empirical studies in collaborative and social computing; • Applied computing → Psychology;

KEYWORDS

Personality Trait, Text Mining, Social Message

1 INTRODUCTION

Personality affects many aspects of life such as people's behavior and interests. There is a high potential that incorporating users' characteristics into information systems, e.g. recommender systems, to enhance their quality and user experience [15]. Therefore, personality-based systems have been developed by several researchers in last decade. The idea behind such systems is simple. People with similar personality traits should have similar interest in various aspects, including hobbies, genres of favorite music and movies, and so on. Therefore, tasks such as recommendations can be made by identifying the similarity between users' personality traits. An example is the TWIN system [23] which recommended hotels to people with similar personalities using data collected from TripAdvisor.

There are many ways to describe the personality of human beings. Trait-based approaches used a set of traits (features) to predict a person's behavior. Many types of traits have been proposed, such as Five Factor models, Eysenck's traits, Cattel's traits, and Cloninger's temperament and character traits [6]. Among these, the five factor model (FFM), or Big Five model, has been widely accepted since it is assumed to represent the basic structure behind all personality traits [17]. The FFM identified five traits, namely openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism, to describe the personality of a person. These five traits are often represented by the acronyms OCEAN or CANOE. To measure the OCEAN traits of a person, practices such as NEO PI-R [7] were generally adopted. A recent approach relied on text analysis techniques to identify the Big Five traits automatically. Linguistic terms used by a person were categorized and analyzed to reveal his Big Five traits. Evidence showed that such an approach may provide convincible results on personality traits recognition [2, 3, 11, 16, 20, 26, 27, 30]. In this work, we will describe a scheme for identifying Big Five personality traits from social messages posted by the users. The messages as well as the personality traits of their posters were used to identify the personality of each keyword. The personality traits of a message were then derived by its constituent keywords.

This article is divided into the following sections. Section 2 will briefly summarize some related work. The proposed scheme will be addressed in Section 3. Section 4 shows the design and results of the experiments. Finally, we give some conclusions in the last section.

permission and/or a fee. Request permissions from Permissions@acm.org.
MISNC2020&IEMT2020, October 31-November 2, 2020, Kaohsiung, Taiwan

ACM ISBN 978-1-4503-8945-7/20/10…$15.00
DOI: 10.1145/3429395.3429412

2 RELATED WORK

Research on personality traits has been conducted in psychology for over a century. Traits theory concerns of the measurement of traits, which can be defined as habitual patterns of behavior, thought, and emotion [10]. Traits are relatively more stable characteristics on a person, although differ significantly over different people. Personality could be composed of a series of traits that are persistent characteristics of human behavior [5]. Although there could be thousands of traits [1], five high-order traits were generally recognized to describe the human personality, namely the Five Factor model or Big Five personality traits [12]. Each trait in FFM can be described by a set of lexical terms [13, 25]. Evidence has shown that personality traits do take effect on many aspects of behavior, social interaction, and mental and physical status. For ex- ample, a study showed that the job performance of a work team is related to the personality traits of its members [14]. Another interesting effect of personality traits lies in satisfaction in romantic relationships [9].

Many schemes for identifying personality traits automatically have been devised recently [28]. One approach is to use the cues from lexical terms. Pennebaker and King [19] developed the Linguistic Inquiry and Word Count (LIWC) system and showed that some LIWC categories correspond with Big Five personality traits [27]. Yarkoni [30] analyzed the relationships between words and personality using a framework similar to Pennebaker and King [19] and revealed robust correlations between the Big Five traits and the frequency with which bloggers used different word categories. Argamon et al. [2] also studied the relations between personality and texts. They used four different sets of lexical features, including a standard function wordlist, conjunctive phrases, modality indicators, and appraisal adjectives and modifiers for this task. Support Vector Machine (SVM) was used to learn linear separators for the high and low classes. The study showed that appraisal use is the best predictor for neuroticism, and that function words work best for extraversion. Oberlander and Nowson [16] used Naïve Bayes and SVM on different sets of n-gram features to classify the personality of blog authors. They found that n-gram features exhibit an empirical relationship with personality traits. Their experiments also showed that Naïve Bayes outperformed SVM in their classification tasks. Mairesse et al. [11] provided a long list of linguistic features that correlate with the personality traits of the Big Five model. Based on this list, Celli [3] selected 12 features from the list and developed a personality recognition system. He found that different writing styles and personality models are associated with different communities using Twitter. Golbeck et al. [8] adopted the same approach and proposed a method to predict a user's personality through the publicly available information on his Twitter profile. Poria et al. [20] incorporated common sense knowledge with ordinary phycho-linguistic and frequency-based features personality classification. They found that common sense knowledge with affective and sentiment information could enhance the accuracy of frameworks that used only psycho-linguistic features and frequency-based analysis at the lexical level. Schwartz et al. [26] proposed an open-vocabulary approach differing from traditional a priori lexicon approaches such as LIWC for personality prediction. They conducted experiments on a large scale dataset composed of 700 million words, phrases, and topic instances collected from the Facebook messages of 75,000 volunteers. Their result suggested the largest relative improvement between open-vocabulary approaches and LIWC for personality. An interesting work by Zhu and Fang [31] proposed a revised lexical scheme for recognition of game personality and user experience by analyzing online game reviews. They used factor analysis to discover 9 and 6 factors to describe the game personality and user experience in gameplay, respectively.

3 PROPOSED SCHEME

3.1 Methodology Overview

In this work, we try to identify the personality traits from social text messages. Fig. 1 depicts the architecture of the proposed method. The details of each step will be addressed in the following subsections.

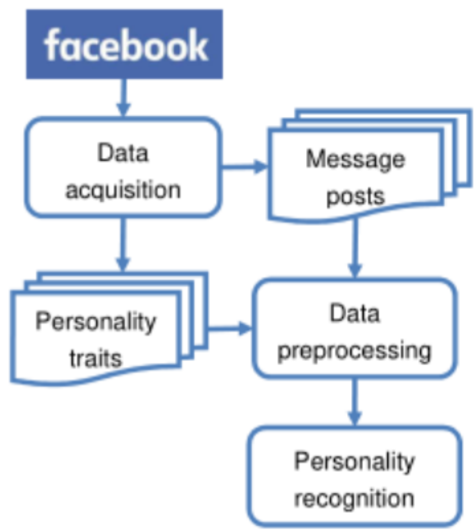


Figure 1: The system architecture of the proposed method

3.2 Data Acquisition

In this work, we used the myPersonlity dataset [4] for personality recognition tasks. The myPersonality dataset collected 9917 status updates of 250 users in Facebook. Their OCEAN traits were also identified. Both OCEAN scores and labels were included in the dataset. An excerpt of the dataset is shown in Fig. 2.

3.3 Data Preprocessing

The message posts are textual data that need to be processed and transformed into suitable representation for later processing. We adopted several text preprocessing steps on these posts prior to personality recognition. A post was first segmented into a set of words using common word segmentation tools. We removed all numbers and dates that should be irrelevant to the recognition of

#AUTHID	STATUS	sEXT	sNEU	sAGR	sCON	sOPN	cEXT	cNEU	cAGR	cCON	cOPN
b7b7764cfa1c523e4e93ab2a79a946c4	likes the sound of thunder.	2.65	3	3.15	3.25	4.4	n	y	n	n	y
b7b7764cfa1c523e4e93ab2a79a946c4	is so sleepy it's not even funny that's she can't get to sleep	2.65	3	3.15	3.25	4.4	n	y	n	n	y
b7b7764cfa1c523e4e93ab2a79a946c4	is sore and wants the knot of muscles at the base of her ne	2.65	3	3.15	3.25	4.4	n	y	n	n	y
b7b7764cfa1c523e4e93ab2a79a946c4	likes how the day sounds in this new song.	2.65	3	3.15	3.25	4.4	n	y	n	n	y
b7b7764cfa1c523e4e93ab2a79a946c4	is home. <3	2.65	3	3.15	3.25	4.4	n	y	n	n	y
b7b7764cfa1c523e4e93ab2a79a946c4	www.thejokerblogs.com	2.65	3	3.15	3.25	4.4	n	y	n	n	y
b7b7764cfa1c523e4e93ab2a79a946c4	saw a nun zombie, and liked it. Also, *PROPNAME* + T	2.65	3	3.15	3.25	4.4	n	y	n	n	y
b7b7764cfa1c523e4e93ab2a79a946c4	is in Kentucky. 421 miles into her 1100 mile journey hom	2.65	3	3.15	3.25	4.4	n	y	n	n	y
b7b7764cfa1c523e4e93ab2a79a946c4	was about to finish a digital painting before her tablet wer	2.65	3	3.15	3.25	4.4	n	y	n	n	y
b7b7764cfa1c523e4e93ab2a79a946c4	is celebrating her new haircut by listening to swinger mus	2.65	3	3.15	3.25	4.4	n	y	n	n	y
b7b7764cfa1c523e4e93ab2a79a946c4	has a crush on the Green Lantern.	2.65	3	3.15	3.25	4.4	n	y	n	n	y
b7b7764cfa1c523e4e93ab2a79a946c4	has magic on the brain.	2.65	3	3.15	3.25	4.4	n	y	n	n	y

personality traits. Punctuation marks were also removed. The remaining keywords will be used to represent these posts.

Figure 2: A part of the myPersonality dataset. We excluded social network attributes that were not used in this work. Each record contains the author’s ID(`#AUTHID`), the status update (`STATUS`), the OCEAN scores (`sEXT`, `sNEU`, `sAGR`, `sCON`, `sOPN`), and the OCEAN classes (`cEXT`, `cNEU`, `cAGR`, `cCON`, `cOPN`).

3.4 Personality Traits Identification by Text Mining

In this work, we obtained the personality traits of games by two methods. The first method used the Personality Recognizer tool developed by Mairesse et al. [11] which computed the Pearson’s Correlation Coefficients between the OCEAN personality traits and LIWC [18]. The Personality Recognizer is a Java command-line application that reads a set of text files and computes estimates of personality scores along the OCEAN traits. It provided four prediction models, namely linear regression, M5’ model tree, M5’ regression tree, and support vector machine with the linear kernel (SMOreg). In this work, we adopted the M5’ Regression Tree [29] which produced a better result in Roshchina [22]’s report. The Personality Recognizer will compute a score for each trait on a scale from 1 to 7, e.g. where 7 is strongly extravert. A running example is depicted in Fig. 3. In this work, the Personality Recognizer will be adopted as a benchmark standard for the proposed approach.

filename	Extraversion	Emotional stability	Agreeableness	Conscientiousness	Openness to experience
John-Wall\\Desktop\\PersonalityRecognizer\\examples\\1.txt'	4.858445	3.982087	4.594376	5.048187	4.712929
John-Wall\\Desktop\\PersonalityRecognizer\\examples\\2.txt'	5.071071	3.918494	4.142983	4.285166	4.805988
John-Wall\\Desktop\\PersonalityRecognizer\\examples\\3.txt'	4.70873	3.880085	4.874448	5.075438	4.955217
John-Wall\\Desktop\\PersonalityRecognizer\\examples\\4.txt'	4.260188	3.976325	4.158244	4.299845	4.947857
John-Wall\\Desktop\\PersonalityRecognizer\\examples\\5.txt'	4.834099	3.384951	4.63651	4.46637	4.743176
ohn-Wall\\Desktop\\PersonalityRecognizer\\examples\\Amy.txt'	4.772383	3.931771	5.040254	4.616022	4.434089
ohn-Wall\\Desktop\\PersonalityRecognizer\\examples\\John.txt'	4.747514	4.055487	4.861711	4.737157	4.662447
ohn-Wall\\Desktop\\PersonalityRecognizer\\examples\\Matt.txt'	4.830341	3.66499	4.654016	5.113624	4.828811
ohn-Wall\\Desktop\\PersonalityRecognizer\\examples\\Rob.txt'	4.953764	4.908476	4.829551	5.271501	4.037335

Figure 3: An example of the result of Personality Recognizer.

In spite of Personality Recognizer, we also developed a novel scheme to identify the personality traits from texts. We first calculated the OCEAN scores for each keyword using the myPersonlity dataset. We used the status updates as well as their associated author’s personality traits to calculate the OCEAN scores for each keyword. However, the status updates contain function words, redundant words, and punctuation symbols which are not helpful. We also met some minor errors in the data and corrected them manually. For ease of processing and better result, we performed several preprocessing steps on the myPersonality dataset which will be addressed in Sec. 4. The OCEAN scores of a keyword v were calculated by:

$$\mathbf{v} = \left(s_t \mid t \in V_P \right), \tag{1}$$

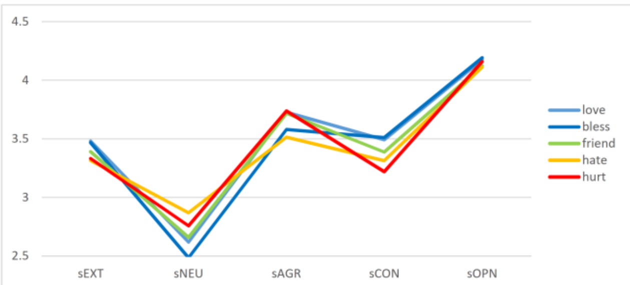
$$s_t = \frac{\mathbf{O} \cdot \mathbf{S}_t}{\|\mathbf{O}\|_1}, \tag{2}$$

where $\mathbf{O}$ is the occurrence vector whose i-th element indicates the number of times v occurs in the i-th status update. For example, $\mathbf{O} = (1, 0, 0, 1, 0, 0, 1, 0, 0, 1, \cdots)$ for v = ’like’ according to Fig. 2. $\mathbf{S}_t$ denotes the vector constructed by scores of personality trait t over the dataset. For example, $\mathbf{S}_t = (2.65, 2.65, 2.65, 2.65, 2.65, 2.65, 2.65, 2.65, 2.65, \cdots)$ for t = ’`sEXT`’ (for score of Extraversion) in Fig. 2. In this example, the elements have the same value since these updates were posted by the same author. In addition to scores, myPersonality dataset also provides trait classes for each status update. In this scope, the i-th element of $\mathbf{S}_t$ will have value 1 if the i-th status update belongs to class t. For example, $\mathbf{S}_t = (0, 0, 0, 0, 0, 0, 0, 0, 0, \cdots)$ for t = ’`cEXT`’ (for class of Extraversion) in Fig. 2. Naturally, the elements all have values of 0 for the same author. $\|\mathbf{O}\|_1$ is the Euclidean 1-norm of $\mathbf{O}$ which is the sum of all its elements. Table 1 shows the OCEAN scores of example keywords. Fig. 4 depicts the score distributions for keywords in Table 1 using trait score and trait class schemes. We can observe that trait score scheme seems to be more consistent in representing keywords’ traits such that words with similar affective polarity (love and bless, hurt and hate) tempted to have similar traits. However, it is necessary to conduct further research to give affirmative conclusions. In this work, we just adopted the OCEAN scores in Eq. 1.

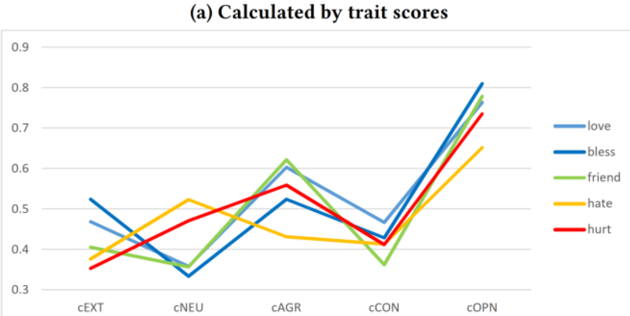
The personality traits of a message post were then obtained by aggregating the OCEAN scores of its constituent keywords. We should obtain the personality traits of message M_i by

Table 1: The OCEAN scores of keywords. # denotes the number of occurrences of keyword *v*.

v	#	Trait score scheme					Trait class scheme				
		sEXT	sNEU	sAGR	sCON	sOPN	cEXT	cNEU	cAGR	cCON	cOPN
love	461	3.48	2.62	3.73	3.49	4.18	0.47	0.36	0.60	0.47	0.76
bless	21	3.47	2.48	3.58	3.51	4.19	0.52	0.33	0.52	0.43	0.81
friend	185	3.39	2.66	3.72	3.39	4.12	0.41	0.36	0.62	0.36	0.78
hate	109	3.31	2.87	3.51	3.31	4.11	0.38	0.52	0.43	0.41	0.65
hurt	34	3.33	2.76	3.74	3.22	4.16	0.35	0.47	0.56	0.41	0.74



(a) Calculated by trait scores



(b) Calculated by trait classes

Figure 4: The distributions of keyword trait scores using different schemes.

$$\mathbf{P}_{M_i} = \frac{\sum_{v \in V_{M_i}} w_v \mathbf{v}}{\left|V_{M_i}\right|}, \tag{3}$$

where V_{M_i} denotes the set of unique keywords that appeared in M_i. $\mathbf{v}$ is the OCEAN score vectors defined in Eq. 1. w_v denotes the weight of keyword v and was defined according to the *tf-idf* weighting scheme in information retrieval [24] as follow:

$$w_v = \frac{tf(v,i)}{\max_{v \in V_{M_i}} tf(v,i)} \log \frac{N}{n_v}, \tag{4}$$

where $tf(v, i)$ denotes the number of occurrences of v in M_i. N and n_v denote the number of messages and number of messages whose reviews contain v, respectively. We should call this scheme the *myPersonality recognizer* hereinafter.

It is straightforward to obtain the personality of a person through his posted messages. The personality traits of a user U_j can be identified as follow:

$$P_{U_j} = \frac{\sum_{M_i \in \mathcal{M}_{U_j}} \mathbf{P}_{M_i}}{\left|\mathcal{M}_{U_j}\right|}, \tag{5}$$

where $\mathcal{M}_{U_j}$ denotes the set of messages posted by U_j.

4 EXPERIMENTAL RESULT

The personality traits of messages were obtained by approaches described in Sec. 3.4. We performed some preprocessing steps on the myPersonality dataset. First, we removed all punctuation marks in all status updates. We then removed all 439 stop words compiled by the University of Tennessee. The remaining words were stemmed by Porter's stemming algorithm [21]. Table 2 lists statistics of the myPersonality dataset. The myPersonality dataset contains many unmeaningful words (e.g. typos), net-centric abbreviations (e.g. 'OMG'), elongated words (e.g. 'ugggggghhhhh'), etc. that should be filtered and processed. After inspection, we observed that these unimportant words can easily be removed by their number of occurrences. For example, if we set the minimum number of occurrences to 10, the number of remaining keywords will be 1223, i.e less than 10% of the original number 12827. These frequent words contain meaningful and popular words with a negligible amount of above-mentioned unimportant words. However, we kept all words in our experiments to retain as much information as possible.

Table 2: Statistics of myPersonality dataset. All keywords are shown in their stemmed form. The keyword 'propnam' is the stemmed form of '*PROPNAME*' which stands for replaced proper names.

Number of records	9917
Number of authors	250
Maximum number of status updates per author	223
Minimum number of status updates per author	1
Number of unprocessed words	148,481
Number of keywords	12827
Number of multiple occurrence keywords	5165
Top 20 most occurred keywords	propnam thi dai ar wa im ha love want todai think need hi night feel dont happi tomorrow peopl home

We applied the myPersonality recognizer on the myPersonality dataset and tried to obtain the personality traits of the users. The messages posted by each of the 250 users in the dataset were first collected. We then applied the myPersonality recognizer on each message to obtain its personality trait scores. The personality trait scores of a user were then obtained by Eq. 5. We also applied the Personality Recognizer on the dataset using the same approach.

The results of both recognizers were then compared to the true personality provided in the dataset. Table 3 shows the mean square errors (MSE) of each trait score after applying the two recognizers. The myPersonality recognizer has an overall smaller MSE comparing to Personality Recognizer. Furthermore, the small MSE values demonstrate that the proposed approach may effectively identify the Big Five personality traits.

Table 3: The mean square errors of each trait obtained by applying the two recognizers.

Trait	myPersonality recognizer	Personality Recognizer
sEXT	0.089	0.120
sNEU	0.113	0.107
sAGR	0.122	0.131
sCON	0.095	0.112
sOPN	0.117	0.106
Average	0.107	0.115

5 CONCLUSIONS

In this work, we proposed a scheme for personality traits identification based on social text messages. The personality trait scores of each keyword were first identified using myPersonality dataset. The personality traits of messages were then identified by their constituent keywords. Finally, the personality trait scores of a person were derived by aggregating the scores of all his posts. The proposed myPersonality recognizer was applied to the myPersonality dataset to identify the Big Five personality trait scores of 250 users. We compared the result to Mairesse’s Personality Recognizer and demonstrated that the proposed approach outperformed the Personality Recognizer.

ACKNOWLEDGMENTS

The work is supported by the Ministry of Science and Technology under Grant No.: NSC 102-2221-E-390-016.

REFERENCES

[1] Gordon Willard Allport and Henry Sebastian Odbert. 1936. *Trait-names: A Psycho-lexical Study*. Psychological Review Co.

[2] Shlomo Argamon, Sushant Dhawle, Moshe Koppel, and James W. Pennebaker. 2005. Lexical Predictors Of Personality Type. In *Proceedings of the Joint Annual Meeting of the Interface and the Classification Society of North America*.

[3] Fabio Celli. 2011. *Mining User Personality in Twitter*. Technical Report. Language, Interaction and Computation Laboratory (CLIC), University of Trento, Italy.

[4] Fabio Celli, Fabio Pianesi, David Stillwell, and Michal Kosinski. 2013. Workshop on Computational Personality Recognition: shared task. In *Proceedings of the Workshop on Computational Personality Recognition*.

[5] Daniel Cervone and Lawrence A. Pervin. 2015. *Personality: Theory and Research* (13th edition ed.). John Wiley & Sons, Inc., NJ.

[6] Philip J Corr and Gerald Matthews. 2009. *The Cambridge handbook of personality psychology*. Cambridge University Press New York.

[7] Paul T Costa and Robert R MacCrae. 1992. *Revised NEO personality inventory (NEO PI-R) and NEO five-factor inventory (NEO-FFI): Professional manual*. Psychological Assessment Resources, Incorporated.

[8] Jennifer Golbeck, Cristina Robles, Michon Edmondson, and Karen Turner. 2011. Predicting Personality from Twitter. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on*. IEEE, 149–156.

[9] Ashley S Holland and Glenn I Roisman. 2008. Big Five personality traits and relationship quality: Self-reported, observational, and physiological evidence. *Journal of Social and Personal Relationships* 25, 5 (2008), 811–829.

[10] Saul Kassin. 2004. *Psychology* (fourth edition ed.). Prentice Hall, Upper Saddle River, NJ.

[11] François 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 (2007), 457–500.

[12] Gerald Matthews, Ian J. Deary, and Martha C. Whiteman. 2009. *Personality Traits*. Cambridge University Press, Cambridge, UK.

[13] R. R. McCrae and P. T. Costa. 1987. Validation of a Five-Factor Model of Personality across Instruments and Observers. *Journal of Personality and Social Psychology* 52 (1987), 81–90.

[14] George A Neuman, Stephen H Wagner, and Neil D Christiansen. 1999. The relationship between work-team personality composition and the job performance of teams. *Group & Organization Management* 24, 1 (1999), 28–45.

[15] Maria Augusta SN Nunes and Rong Hu. 2012. Personality-based recommender systems: an overview. In *Proceedings of the sixth ACM conference on Recommender systems*. ACM, 5–6.

[16] Jon Oberlander and Scott Nowson. 2006. Whose thumb is it anyway?: classifying author personality from weblog text. In *Proceedings of the COLING/ACL on Main conference poster sessions*. Association for Computational Linguistics, 627–634.

[17] Brian P. O’Connor. 2002. A Quantitative Review of the Comprehensiveness of the Five-Factor Model in Relation to Popular Personality Inventories. *Assessment* 9, 2 (2002), 188–203.

[18] James W Pennebaker, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. 2015. *The Development and Psychometric Properties of LIWC2015*. Technical Report. University of Texas at Austin, Austin, TX.

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

[20] Soujanya Poria, Alexandar Gelbukh, Basant Agarwal, Erik Cambria, and Newton Howard. 2013. Common sense knowledge based personality recognition from text. In *Mexican International Conference on Artificial Intelligence*. Springer, 484–496.

[21] M. F. Porter. 1980. An algorithm for suffix stripping. *Program* 14, 3 (1980), 130–137.

[22] Alexandra Roshchina. 2012. *TWIN: Personality-based Recommender System*. Master’s thesis. Institute of Technology Tallaght, Dublin, Ireland.

[23] Alexandra Roshchina, John Cardiff, and Paolo Rosso. 2015. TWIN: personality-based intelligent recommender system. *Journal of Intelligent & Fuzzy Systems* 28, 5 (2015), 2059–2071.

[24] G. Salton and M. J. McGill. 1983. *Introduction to Modern Information Retrieval*. McGraw-Hill, New York.

[25] Gerard Saucier. 1994. Mini-Markers: A Brief Version of Goldberg’s Unipolar Big-Five Markers. *Journal of Personality Assessment* 63, 3 (1994), 506–516.

[26] H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Lukasz Dziurzynski, Stephanie M Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin EP Seligman, and others. 2013. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one* 8, 9 (2013), e73791.

[27] Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology* 29, 1 (2010), 24–54.

[28] Alessandro Vinciarelli and Gelareh Mohammadi. 2014. A survey of personality computing. *IEEE Transactions on Affective Computing* 5, 3 (2014), 273–291.

[29] Yong Wang and Ian H. Witten. 1997. Inducing Model Trees for Continuous Classes. In *Proceedings of the 9th European Conference on Machine Learning Poster Papers*. 128–137.

[30] Tal Yarkoni. 2010. Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality* 44, 3 (2010), 363–373.

[31] Miaoqi Zhu and Xiaowen Fang. 2015. A lexical approach to study computer games and game play experience via online reviews. *International Journal of Human-Computer Interaction* 31, 6 (2015), 413–426.