Predicting Personality Traits from Social Media Using Text Semantics

Mariam Hassanein Wedad Hussein Sherine Rady Tarek F. Gharib
Information Systems Department
Ain Shams University, Faculty of Computer and Information Sciences
Cairo, Egypt
Mariam.hassanein@cis.asu.edu.eg Wedad.hussein@fcis.asu.edu.eg srady@cis.asu.edu.eg tfgharib@cis.asu.edu.eg

Abstract

Online social networks are becoming a very rich source of user generated content. This content motivates different types of applications that rely on personalization; such as recommender systems and online marketing. Detecting personalities through mining publicly available social data immerges as an important related issue that can assist web-based systems. Some approaches have been introduced to use publicly available social data to infer user's personality. This paper presents an approach for personality traits inference based on text semantic analysis. Different representations of user text combined with several semantic based measures are proposed to predict users' personality through their Facebook status updates. The proposed approach has been tested and validated on data released by the myPersonality project for the Workshop on Computational Personality Recognition. The results prove that the information content-based measure achieves the best average personality trait prediction with an accuracy of 64%.

Keywords

Automatic Personality Recognition, Big Five model, Text Analysis, Semantic similarity, Social media

I. INTRODUCTION

Since the emergence of Online Social Networks (OSNs), people have been using them dramatically as a way of expressing their preferences, feelings and opinions freely without being judged. Statistically, more than 2.4 billion users are active on social networks, making (OSN) a very rich source for User Generated Content (UGC), which is represented in various forms, such as: text, images, videos, links or through interactions. This content motivated different types of applications; among which is personality assessment. Knowledge about an individual personality can help in building personalized and adaptive systems [1][2] to enhance recommendation process. Psychology suggests that an individual's personality can be assessed in terms of five bipolar major dimensions or traits, called the Big Five personality traits. These traits are namely openness to experience, conscientiousness, extraversion, agreeableness and neuroticism (also called emotional stability), forming the abbreviation (O.C.E.A.N) [3][4]. Those dimensions determine one's propensity and the embodiment of an individual's being. Each trait, however, exhibits different features. Table 1 presents a summary of the most dominant characteristics of each of the five personality traits. Usually, Psychology researchers use traditional assessment techniques to assess a personality, mostly using Big Five Inventory (BFI) and the Revised NEO Personality Inventory (NEO PI-R) models [5]. Both techniques require the user to answer a self- report questionnaire to measure the individual's characteristics based on the given answers. Unfortunately, this approach is time consuming and impractical for a long set of questions. For example, the 300-item International Personality Item Pool (IPIP) and the 240- item NEO-PI-R need users to spend a long time to complete them[6]. To overcome the mentioned drawbacks, a process called "Automatic Personality Recognition" (APR) was suggested to provide an alternative computerized solution [7]. Many applications took advantages of APR, including opinion mining and sentiment analysis [8], deception detection [9] , social networks analysis [10] and anomalies detection [11].

The information available on users' online profiles in social media is believed to be a reflection of the users' actual personalities rather than their most desirable ones. So recently UGC on OSN has been used in the APR. Inference of users' personality traits is done by extracting and analyzing features from the publicly available content. Researches in this field have employed a common research design to explore digital footprints left by people on social media[12]. First, users should solve self-report questionnaires for personality assessment to act as a ground truth. Then, digital footprints from users' social media profiles are collected. From this content single or multiple features are extracted to be used in predictive models. Finally, the accuracy of prediction is evaluated based on the predicting model and ground truth. Some studies explored various machine learning techniques to automatically infer the user's personality traits based on available features, thus, improve the prediction accuracy. This paper presents an approach for personality assessment in OSN, as an automated assessment method rather than undertaking questionnaires. The approach is based on semantic similarity measurement between the user's text and the words that are best representative for personality traits. In this paper, three different representations for user text are suggested. The vector space model commonly employed in text indexing , is used while integrating knowledge-based metrics[13] and consulting WordNet [14]. The employed metrics provide measures that can detect to what extent the user text is close to each of the five personality traits, and hence an automated trait predictor solution. The remainder of this paper is structured as follows: Section 2 presents overview on related works. Section 3 presents the proposed semantic based approach for traits prediction. Section 4 presents the experiments conducted, and finally, section 5 presents the conclusion and future work.

Table 1:Personality Traits Characteristics

Personality Trait	Characteristics
Openness	Imaginative, Insightful
Conscientiousness	Ambitious, Self-disciplined
Extraversion	Sociable, Outgoing
Agreeableness	Trusting, Modest
Neuroticism	Nervous, Anxious

II. RELATED WORK

In the last years, several works started taking advantage from the enormous amount of data available on social networks for automatic personality recognition. Some of these works investigated visual features extracted from images on Flicker, Instagram and Facebook [15][16][17]. Some works investigated extracting statistical features from individual's online social networks profile. Features such as the size and density of friendship network and the number of group memberships, were studied to show significant relationships between personality traits and various features of profiles[18][19]. Other works merged different types of features; textual and non-textual features targeting the enhancement of personality traits prediction, such as [20] where Textual , Social network and time related features were merged. Other approaches analyzed individuals' texts by using Natural Language Processing (NLP) techniques. For example, work done in [21] used tweets with self-identified personalities obtained from twitter, to explore most related linguistic features for the purpose of distinguishing different personality types. A new linguistic tool was designed where three categories of features were extracted, namely bag of n-grams, POS tags and word vectors. By analyzing those features, the usage of language for different personalities was observed. For instance, for, introverts' some of the most commonly used words are "My God", "kind of" and "I don't". Researchers in [22] aimed to develop a model which required a reduced amount of data for personality modeling. The model showed high accuracy results when integrating word embedding features with Gaussian Processes regression.

Moreover, by using "Linguistic Inquiry and Word Count" (LIWC) tool on a standard dataset from twitter, the work in [23] was able to extract 81 different linguistic features to train a machine learning model for personality classification. Additional statistical features about users; such as number of followers, number of followings, and number of mentions were introduced aiming to enhance accuracy of prediction.

Conscientiousness trait prediction was introduced in [4] by analyzing verbs chosen to express a specific event. The work suggested that the way an individual describes a situation reveals his/her personality characteristics. For Conscientiousness detection, features like word objectivity and specificity were extracted using SentiWordNet [24] which is a lexical resource for opinion mining and sentiment classification Other work in personality detection integrated semantic features. Using Latent Semantic Analysis (LSA), the work in [25] tested whether the characteristics of a personality can be assessed through analyzing the semantic content of an individual written text. Initially, participants in this experiment were asked to solve big five personality tests and to write essays describing their alleged feelings and actions in five different scenarios. Each scenario was made to measure one of the five traits. LSA was trained using 7 training corpora generating 7 semantic spaces. Results found significant correlations relating semantic content with 3 traits; Extraversion, Neuroticism and openness.

Similar work in [26] used LSA on Facebook status updates to predict users' personality traits focusing on predicting Neuroticism and Extraversion. As a ground truth, each participant was asked to solve EPQR-S test to measure Extraversion and Neuroticism. For the used dataset, the latest Facebook posts were collected in addition to Facebook psychometric variables like number of friends, frequency of status updates, and time spent on Facebook. Results showed that extraversion trait is positively correlated with the number of Facebook friends and the frequency of an individual's status updates while neuroticism is negatively correlated with time spent on Facebook.

Another approach for personality assessment was introduced in [27] based on Vectorial Semantic Models (VSM) [28], which assumes that the meaning of a word can be recognized by analyzing words currently appearing with the target word in a given context. User text was analyzed by representing it in a vector form and measuring its similarity to set of pre-defined vectors of personality traits. Continuing the previous work, work in [29] used the same approach for automatically identifying Personality Disorders namely Paranoid personality disorder, Narcissistic personality disorder, Schizotypal personality disorder and Depression. The study was mostly used for identification of school shooters. It was found that School shooters acquired distinctively higher score on the Narcissistic personality disorder. Moreover ,for recruitment purposes, [30] used linguistic features extracted from Facebook status updates. Those Features were extracted using (LIWC) tool to train a regression model for detecting the dark side personalities namely psychopathy, narcissism and Machiavellianism. Relationships between the dark side of personality and language used in Facebook status updates were investigated.

III. SEMANTIC BASED APPROACH FOR PERSONALITY PREDICTION

For the automatic personality assessment, the Vector Space Model [28] is used. In this model, user text is represented in the form of a vector consisting of the counts of each word in the user text. In the proposed approach, semantic metrics are used to measure the similarity between user's generated vector and the vectors representing personality traits. The used vectors for representing personality traits are populated as a list of adjectives provided by Trapnell and Wiggins[27].

In this representation and for each of the five traits, two vectors are constructed; one that represents the existence of the trait and another that represents the opposite trait. This is shown in table 2. Semantic similarity is measured using knowledge-based semantic similarity measures consulting the WordNet ontology. Knowledge based metrics use a hierarchal semantic network to measure relatedness between two concepts. WordNet is the most commonly used lexical database for English words to link semantic relationships. In WordNet, words are grouped into sets of synonyms known as synsets. Most of the words are polysemious (i.e. most words have multiple meanings). WordNet provides different knowledge-based measures for measuring the semantic similarity or relatedness between two-word senses (concepts), such as those relying on path length and information content (IC).

Detecting the proper sense of a word in a given context to reduce the semantic gap is called word sense disambiguation. To handle the mentioned challenge, most frequent sense is considered while measuring relatedness between two words. Semantic similarity can be measured between the text and each vector producing ten different similarity scores. This is more illustrated in the mathematical representation in section B. The calculated scores reflect the degree to which a personality trait appears in a text. As the score gets higher, the text is expected to show a higher similarity to the personality vector.

Fig. 1 shows a block diagram for the proposed personality traits prediction approach. It consists of three main modules namely: a “Pre-processing” module, a “Similarity scores calculation” module and finally a “Traits predictor” module. In “Pre-processing” module only relevant tokens needed for the prediction process are extracted and presented in the form of vector. While in “Similarity scores calculation” module similarity between this vector and vectors representing traits are calculated. And finally, the process of detecting the presence or absence of traits is held in “Traits predictor “module. These modules are explained in more details below.

A. Data Preprocessing

Status posts for each user are concatenated in one single document. Series of pre-processing steps are applied for each document. First, the text undergoes tokenization. Then, Non-English and meaningless tokens are removed after being checked against an English dictionary. Stopping words are removed before nouns, adjectives, verbs and adverbs are extracted using a Part-of-Speech tagger for further analysis. After Pre-processing, three different vectors representations are generated to find out the most appropriate representation for the traits prediction. The first representation (Vec1) contains many of the sentence structure including nouns, adjectives, verbs and adverbs which are extracted from the user text. The second representation (Vec2) is simpler by regarding only all extracted nouns , the third representation (Vec3) will contain only emotionally charged nouns obtained by NRC emotion lexicon[31][32]. This lexicon contains 14182 words, each word is tagged with one of ten attributes joy, surprise, trust, anger, fear, disgust, anticipation and sadness, besides the polarity of a word either negative or positive. Fig. 2 demonstrates an illustrative example for the data pre-processing workflow. The input user text is U. The consecutive pre-processing steps are shown on the left-hand side, with the processed data on the right-hand side. The final pre-processed output is U*. The example shows that 10-word input post is reduced to 3 different sized vectors.

Table 2:Personality Traits Vectors according to Trapnell and Wiggins

Personality Trait	Descriptive Adjectives
Extraversion Positive	< Assertive, Dominant, Authoritarian, Forceful, Firm, Confident, Persistent, Assured >
Extraversion Negative (Introversion)	< Forceless, Nervous, Quiet, Indecisive Afraid, Modest, Shy, Calm >
Agreeableness Positive	< Tender, Gentle, Soft, Kind, Affectionate, Helpful, Sympathetic, Friendly >
Agreeableness Negative	< Cruel, Unfriendly, Negative, Mean, Brutal, Inconsiderate, Insensitive, Cold >
Neuroticism Positive	< Worried, Stressed, Anxious, Nervous, Fearful, Guilty, Insecure, Restless, Emotional >
Neuroticism Negative	< Balanced, Stable, Confident, Fearless, Calm, Relaxed, Secure, Comforted, Peaceful >
Conscientiousness Positive	< Careful, Orderly, Neat, Persistent, Systematic, Efficient, Tidy, Straight, Reliable>
Conscientiousness-Negative	< Disloyal, Distracted, Confused, Incompetent, Inefficient, Unreliable, Chaotic, Wild, Messy >
Openness-to-experience Positive	< Open, Philosophical, Curious, Imaginative, Questioning, Literary, Abstract, Individualistic, Unique >
Openness-to-experience Negative	< Thoughtless, Narrow-minded, Ordinary, Incurious, Common, Ignorant, Conventional, Restricted Concrete>

B. Personality Traits Prediction

By the end of the pre-processing stage, a vector representing the remained words for each participant’s own status updates exists. The traits prediction module uses the semantic metrics employing WordNet, to calculate semantic distances. Semantic similarity is measured between a query vector and vectors representing personality traits in Table 2.

Assuming the user vector after pre-processing to be:

- $U^* = \{w_1, w_2, w_3 \ldots. w_R\}$

Such that w_L represents a word in a user vector of length R

And the traits vectors of size M to be:

- $S = \{V_1, V_2, V_3 \ldots. V_M\}$

Where Vi represents a specific trait Vector and M is equal to 10 representing different positive and negative classes of the traits

For each personality trait we define a vector Vi:

- $V_i = \{d_1, d_2, d_3 \ldots. d_N\}$

Such that dj represents words in the personality trait vector, in addition to their synonyms, creating a vector of length N.

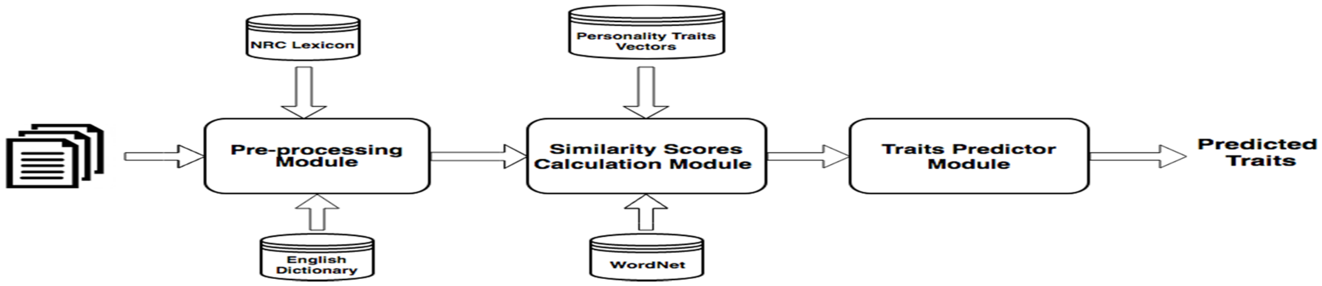

Figure 1 :Block diagram of traits prediction process

- $\forall w_L \in U^*$ *and* $\forall d_j \in V_i$, the maximum similarity between a word w_L in user text and dimension d_j in a trait vector is calculated:

$$E_{iL} = \quad \text{Max}\ (RS\ (w_L, d_j)) \qquad (1)$$

Where RS is the relatedness score function implemented using different semantic similarity measures by WordNet.
The final step calculates a total similarity score between vector U* and a given trait vector V_i

- $\forall Vi \in S$,

$$\text{Sim}\ (U^*, V_i) = \frac{\sum_{i=1}^{R} E_{iL}}{R} \qquad (2)$$

All relatedness scores are averaged to represent the semantic similarity value which states to what limit the user text is similar to a given trait. The similarity score is calculated for each trait vector (V_i) producing M similarity scores. Relatedness score measures used in this work are Path based measure and Information Content based measure. The former introduces a distance measure based on the length of the shortest path between two concepts/words like path measure defined as [33][34] :

$$sim_{path}(w_1, w_2) = \frac{1}{Path\ length(w1,w2)} \qquad (3)$$

Where "Path length" is the length of the shortest path between the concepts w1 and w2 found in a given ontology. Information content measure is based on the probability of a concept occurring in a corpus of text , such as JCN measure which is defined as [30]:

$$sim_{jcn}(w_1, w_2) = \frac{1}{IC(w1)+IC(w2)-2*IC(lcs(w1,w2))} \qquad (4)$$

Given that

$$IC(w) = \ -\log P(w) \qquad (5)$$

Where *P(w)* represents the probability of word (w) occurrence in a corpus. Fig. 3 demonstrates how the prediction process works. Processes shown on left hand side and data being processed on right hand side. The example shows that similarity measures are calculated for the ten traits vectors. Following, each trait is compared against its opposite, greater values are detected as five personality traits.

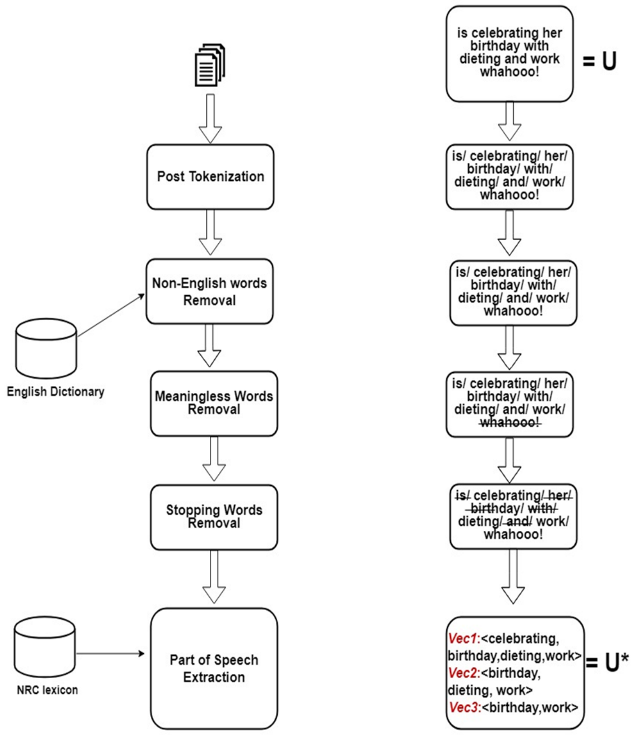

Figure 2 : Illustrative example for Pre-processing workflow

IV. Results and Discussion

A. Datset description

The used dataset contains 9917 status posts for 250 Facebook users, released by the myPersonality project for the Workshop on Computational Personality Recognition [35]. The myPersonality project is a Facebook application devoted to the academic study of personality. It contains data about Facebook users who agreed to participate by giving an access to their Facebook profiles, social network data as well as personality assessment test results. Our goal is to use the mentioned dataset to test the effectiveness of the approach for automatic personality assessment. Table 3 shows the distribution of different traits in the experimented dataset.

Table 3 : Distribution of traits in myPersonality Dataset

	Open.	Consc.	Extra.	Agree.	Neuro.
Positive	176	130	96	134	99
Negative	74	120	154	116	151

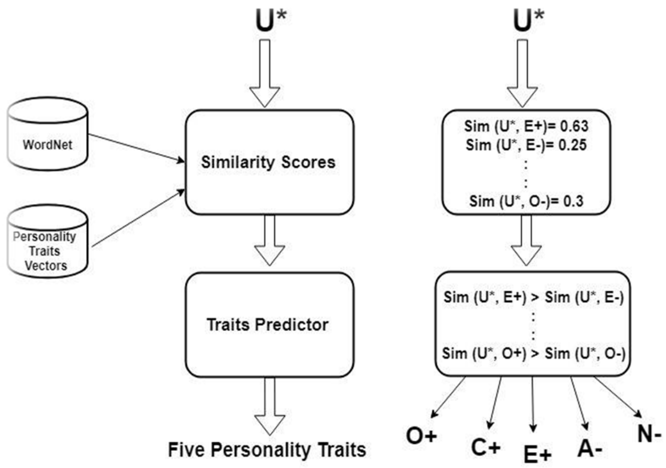

Figure 3:An illustrative example for trait prediction process

B. Evaluation Metrics

Four measures were used to evaluate the quality of prediction process; accuracy, precision, recall and F-measure. Accuracy measure is calculated by the percentage of correct predictions. Which is defined by:

$$\text{Accuracy}= \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictios} \qquad \textbf{(6)}$$

Besides the accuracy metric, precision, recall and f- measure measures were also considered which are defined in equations 7,8, and 9 respectively.

$$\text{Precision} = \frac{\mathbf{TP}}{\mathbf{TP + FP}} \qquad (7)$$

$$\text{Recall} = \frac{\mathbf{TP}}{\mathbf{TP + FN}} \qquad (8)$$

$$\text{F- measure} = \mathbf{2} * \frac{\mathbf{Recall * Precision}}{\mathbf{Recall + Precision}} \qquad (9)$$

Where TN, TP, FP and FN mentioned refer to true negative, true positive, false positive and false negative respectively.

C. Expermintal Results

The three Vectors representations (Vec1, Vec2 and Vec3) mentioned in section 3.1 were used to evaluate the effectiveness of the proposed semantic based approach for traits prediction. Applying pre-processing step as explained produced 7658 posts from the total posts, which were used for further analysis. The two semantic similarity measures "Path" and "JCN" defined by equations (3) and (4) were used for similarity measurement between user text and personality traits vectors. Prediction performance measures for each personality trait were calculated according to equation (6), (7), (8) and (9). Using Vec1 showed that IC measure (JCN) obtained higher accuracy (64%) than the path-based measure (Path) (59%). Using JCN measure, analysis yielded nearly similar results for agreeableness and neuroticism. But the most significant change was in the prediction of extraversion, conscientiousness and openness to experience as shown in fig. 4. High values of accuracy were obtained especially for extraversion (78%) and openness to experience (70%). For the second representation (Vec2) which involved working on nouns extracted from user text, JCN also obtained higher accuracy results (60%) than path measure (58%). On one hand, JCN measure showed high results for prediction of extraversion (61%) and conscientiousness (56%) traits. On the other hand, 70% was the accuracy of prediction for openness to experience trait when Path measure was used. Fig. 5 shows the results for accuracies obtained using Vec2. The third representation Vec3 contained only nouns containing personality cues, where a noun is considered emotionally charged if it is tagged with, at least, one of the ten emotions. Otherwise, the noun is considered to be neutral. Higher results were obtained using JCN measure (58%) compared to Path measure with accuracy (52%) as mentioned in fig. 6. In table 4, the average accuracy, precision, recall and F-measure values were recorded for the three mentioned representations.

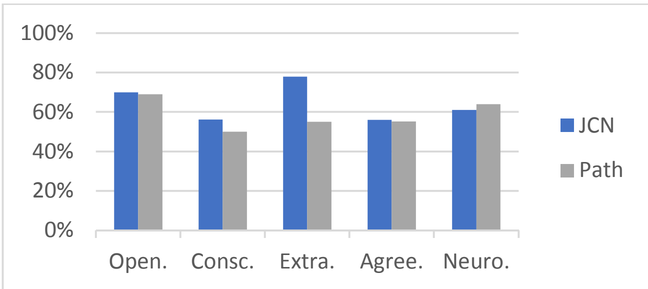

Figure 4: Accuracy Performance using Vec1

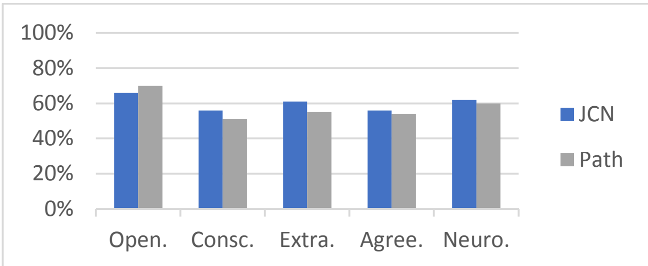

Figure 5 : Accuracy Performance using Vec2

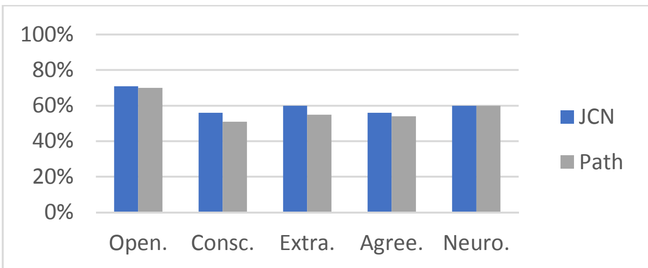

Figure 6 : Accuracy Performance using Vec3

Table 4 : Prediction Performance

	Similarity measure	Accuracy	Precision	Recall	F-measure
Vec1	***JCN***	64%	64%	75%	65%
	Path	59%	59%	45%	39%
Vec2	***JCN***	60%	57%	50%	46%
	Path	58%	52%	48%	43%
Vec3	***JCN***	61%	58%	47%	43%
	Path	55%	59%	51%	49%

As seen for JCN measure high values for both accuracy and precision were reached with all of the vectors, proving that Information Content based measures were more suitable for personality traits prediction than path-based measures. JCN measure showed high precision values for Vec1 and Vec2, while Path measure resulted higher precision value for Vec3. Recall and F-measure values showed the same behaviour as precision for the three representations. Our results proved that working on the whole user text was of higher quality than working only on nouns. And as expected, removing neutral nouns improved the results rather than working on all nouns.

V. CONCLUSION

User generated content on social networks is an effective predictor for personality traits. The paper proposes an approach using Vector Space Model and semantic similarity metrics to measure similarity between user's texts and set of vectors representing the different characteristics personality traits. Three different representations for user text were extracted and tested. The values of semantic similarity measures to what extent the user text is similar to each trait. The approach showed promising results for predicting personality traits. JCN measure which is based on measuring information content managed to achieve the best accuracy of value 64%. Using the semantic Path measure recorded an accuracy of 59%. In the future work, the study will consider additional features to the text analytics ones generated by the vector space model, in order to enhance the accuracy of traits prediction.

REFERENCES

[1] B. Ferwerda and M. Schedl, "Enhancing music recommender systems with personality information and emotional states: A proposal," *CEUR Workshop Proc.*, vol. 1181, pp. 36–44, 2014.

[2] B. Ferwerda, E. Yang, M. Schedl, and M. Tkalčič, "Personality Traits Predict Music Taxonomy Preferences," *Proc. 33rd Annu. ACM Conf. Ext. Abstr. Hum. Factors Comput. Syst. - CHI EA '15*, pp. 2241–2246, 2015.

[3] O. P. John and S. Srivastava, "The Big Five trait taxonomy: History, measurement, and theoretical perspectives," *Handb. Personal. Theory Res.*, vol. 2, no. 510, pp. 102–138, 1999.

[4] M. Tomlinson, D. Hinote, and D. Bracewell, "Predicting conscientiousness through semantic analysis of facebook posts," *Seventh Int. AAAI Conf. Weblogs Soc. Media*, pp. 1–4, 2013.

[5] P. T. Costa and R. R. McCrae, "The revised NEO personality inventory (NEO-PI-R)," in *The SAGE Handbook of Personality Theory and Assessment: Volume 2 - Personality Measurement and Testing*, no. January 2008, 2008, pp. 179–198.

[6] H. Wei *et al.*, "Beyond the Words: Predicting User Personality from Heterogeneous Information," *Proc. Tenth ACM Int. Conf. Web Search Data Min. - WSDM '17*, pp. 305–314, 2017.

[7] L. Liu, D. Preot, and L. Ungar, "Analyzing Personality through Social Media Profile Picture Choice," *AAAI Digit. Libr.*, no. Icwsm, pp. 211–220, 2016.

[8] J. Golbeck, C. Park, and D. L. Hansen, "Computing political preference among Twitter followers," *Proc. SIGCHI Conf. Hum. Factors Comput. Syst. Vancouver, BC, Canada*, pp. 1105–1108, 2011.

[9] F. Enos, S. Benus, R. L. Cautin, M. Graciarena, J. Hirschberg, and E. Shriberg, "Personality factors in human deception detection: Comparing human to machine performance," *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*, vol. 2, pp. 813–816, 2006.

[10] F. Celli and L. Rossi, "The Role of Emtional Stability in Twitter Conversations," *13th Conf. Eur. Chapter Assoc. Comput. Linguist.*, pp. 10–17, 2012.

[11] T. Tuna *et al.*, "User characterization for online social networks," *Soc. Netw. Anal. Min.*, vol. 6, no. 1, pp. 1–28, 2016.

[12] D. Azucar, D. Marengo, and M. Settanni, "Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis," *Pers. Individ. Dif.*, vol. 124, no. December 2017, pp. 150–159, 2018.

[13] D. Lin, "An Information-Theoretic Definition of Similarity," *Proc. ICML*, pp. 296–304, 1998.

[14] T. Pedersen and J. Michelizzi, "WordNet :: Similarity - Measuring the Relatedness of Concepts," *HLT-NAACL--Demonstrations '04 Demonstr. Pap. HLT-NAACL 2004*, no. July, pp. 38–41, 1998.

[15] F. Celli, E. Bruni, and B. Lepri, "Automatic Personality and Interaction Style Recognition from Facebook Profile Pictures," *Proc. ACM Int. Conf. Multimed. - MM '14*, pp. 1101–1104, 2014.

[16] B. Ferwerda and M. Schedl, "Predicting Personality Traits with Instagram Pictures," *RecSys Emp. 2015 3rd Work. Emot. Personal. Pers. Syst. 2015*, pp. 7–10, 2015.

[17] Y. Yan, J. Nie, L. Huang, Z. Li, Q. Cao, and Z. Wei, "Is Your First Impression Reliable? Trustworthy Analysis Using Facial Traits in Portraits," *Multimed. Model.*, pp. 148–158, 2015.

[18] Y. Bachrach, M. Kosinski, T. Graepel, P. Kohli, and D. Stillwell, "Personality and patterns of Facebook usage," *Proc. 3rd Annu. ACM Web Sci. Conf. - WebSci '12*, pp. 24–32, 2012.

[19] D. Quercia, R. Lambiotte, D. Stillwell, M. Kosinski, and J. Crowcroft, "The personality of popular facebook users," *Proc. ACM 2012 Conf. Comput. Support. Coop. Work - CSCW '12*, p. 955, 2012.

[20] G. Farnadi, S. Zoghbi, M. Moens, and M. De Cock, "Recognising Personality Traits Using Facebook Status Updates," *Work. Comput. Personal. Recognit. Int. AAAI Conf. weblogs Soc. media*, pp. 14–18, 2013.

[21] Y. Wang, "Understanding Personality through Social Media," 2015.

[22] P.-H. Arnoux, A. Xu, N. Boyette, J. Mahmud, R. Akkiraju, and V. Sinha, "25 Tweets to Know You: A New Model to Predict Personality with Social Media," *Proc. Elev. Int. Conf. Web Soc. Media*, pp. 25–28, 2017.

[23] J. Golbeck, C. Robles, M. Edmondson, and K. Turner, "Predicting personality from twitter," *Proc. - 2011 IEEE Int. Conf. Privacy, Secur. Risk Trust IEEE Int. Conf. Soc. Comput. PASSAT/SocialCom 2011*, pp. 149–156, 2011.

[24] S. Baccianella, A. Esuli, and F. Sebastiani, "SentiWordNet 3 . 0 : An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining," *Lr.*, vol. 10, pp. 1–12, 2010.

[25] H. H. C. M. Peter J. Kwantes a, Natalia Derbentseva a, Quan Lam a, Oshin Vartanian a, b, "Assessing the Big Five personality traits with latent semantic analysis," *Pers. Individ. Dif.*, vol. Volume 102, pp. 1–24.

[26] D. Garcia and S. Sikström, "The dark side of Facebook : Semantic representations of status updates predict the Dark Triad of personality," *Pers. Individ. Dif.*, no. April 2017, 2014.

[27] Y. Neuman and Y. Cohen, "A Vectorial Semantics Approach to Personality Assessment," *Sci. Rep.*, vol. 4, p. 4761, 2014.

[28] P. D. Turney and P. Pantel, "From Frequency to Meaning_ Vector Space Models of Semantics," *J. Artif. Intell. Res.*, vol. 37, pp. 141–188, 2010.

[29] Y. Neuman, D. Assaf, Y. Cohen, and J. L. Knoll, "Profiling school shooters: automatic text-based analysis," *Front. Psychiatry*, vol. 6, no. 86, pp. 1–5, 2015.

[30] R. Akhtar, D. Winsborough, U. Ort, A. Johnson, and T. Chamorro-Premuzic, "Detecting the dark side of personality using social media status updates," *Pers. Individ. Dif.*, vol. 132, no. October 2017, pp. 90–97, 2018.

[31] S. M. Mohammad and P. D. Turney, "Crowdsourcing a word-emotion association lexicon," *Comput. Intell.*, vol. 29, no. 3, pp. 436–465, 2013.

[32] S. M. Mohammad and P. D. Turney, "Emotions evoked by common words and phrases: using mechanical turk to create an emotion lexicon," *CAAGET '10 Proc. NAACL HLT 2010 Work. Comput. Approaches to Anal. Gener. Emot. Text*, no. June, pp. 26–34, 2010.

[33] B. T. McInnes, T. Pedersen, Y. Liu, G. B. Melton, and S. V. Pakhomov, "Knowledge-based Method for Determining the Meaning of Ambiguous Biomedical Terms Using Information Content Measures of Similarity," *AMIA Annu. Symp. Proc.*, vol. 2011, p. 895, 2011.

[34] M. Palmer and Z. Wu, "VERB SEMANTICS AND LEXICAL Zhibiao W u," *Proc. ACL*, pp. 133–138, 1994.

[35] F. Celli, F. Pianesi, D. Stillwell, and M. Kosinski, "Workshop on Computational Personality Recognition : Shared Task," vol. 2006, pp. 2–5, 2006.