AAAI Technical Report WS-13-01
Computational Personality Recognition (Shared Task)

Personality Traits Recognition on Social Network - Facebook

Firoj Alam, Evgeny A. Stepanov, Giuseppe Riccardi
Department of Information Engineering and Computer Science, University of Trento, Italy
{alam, stepanov, riccardi}@disi.unitn.it

Abstract

For the natural and social interaction it is necessary to understand human behavior. Personality is one of the fundamental aspects, by which we can understand behavioral dispositions. It is evident that there is a strong correlation between users' personality and the way they behave on online social network (e.g., Facebook). This paper presents automatic recognition of Big-5 personality traits on social network (Facebook) using users' status text. For the automatic recognition we studied different classification methods such as SMO (Sequential Minimal Optimization for Support Vector Machine), Bayesian Logistic Regression (BLR) and Multinomial Naïve Bayes (MNB) sparse modeling. Performance of the systems had been measured using macro-averaged precision, recall and F1; weighted average accuracy (WA) and un-weighted average accuracy (UA). Our comparative study shows that MNB performs better than BLR and SMO for personality traits recognition on the social network data.

Introduction

For the social communication, we interact with unknown individuals, even with machines that exhibit human-like features and behaviors such as robots, embodied virtual agents and animated characters (Nass et al., 2005). To make these automated systems more human-like, we need to understand human behavior and how it is affected by personality traits.

Personality is the most complex of all the human attributes and it also characterizes the uniqueness of a person. 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 and situational variables. 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 restpect to their underlying personality traits. With time, this area has attracted researchers from different fields. especially researchers in the human-machine interaction and behavioral analytics.

It is suggested in (Nass et al., 2005) that naturalness of an interaction with a user and its efficiency increases by matching user's personality. Studies have been done on how the style of communication like emails, blog entries (Gill et al., 2007) and the choice of particular parts of speech (Oberlander et al., 2004) depend on the author's personality.

This paper describes automatic recognition of personality traits on the social network data using different classification methods. The main contribution of this paper is studying the performance of different classification methods using bags-of-words. Another contribution is to measure the performance of the systems using two different evaluation measures: (i) macro-averaged precision and recall, F1; (ii) weighted average (WA) and un-weighted average (UA).

The paper is organized as follows: Section 2 describes the related work; Section 3 describes the corpus, which was used in this study; Section 4, defines the experimental method. Details of the classification results are given in Section 5. Finally conclusion and future study appear in Section 6.

Related Work

Personality is defined as the coherent patterning of affect, behavior, cognition and desire over time and space, which are used to characterize unique individuals. There are several theories for personality traits in the literature but the most widely used personality traits model is the Big-5, 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, etc.
C (**Conscientiousness**): Efficient, organized, etc.
E (**Extraversion**): Energetic, active, assertive, etc.
A (**Agreeableness)**: Compassionate, cooperative etc.
N (**Neuroticism**): Anxious, tense, self-pitying, etc.

There have been many works to the automatic recognition of personality traits through the use of language using acoustic and linguistic information. It has been studied in different domains, which include personality of the blogger (Mohtasseb et al., 2009), personality identification for successful relationship (Donnellan et al., 2004), understanding personality of a user in dialogue system (Ivanov et al., 2011), broadcast news (Schuller, 2012), social network (Celli, 2012; Amichai-Hamburger et al., 2010), modern Greek spontaneous text using low-level linguistic features (Kermanidis, 2012), worker's personality in crowdsourcing (Kazia et al., 2011) etc. Mairesse et al. (2007) studied personality traits using corpora annotated with self-assessed and observer ratings. From these studies, it is evident that the performance of personality traits recognition systems needs to be improved to be usable.

Corpus

In the "Workshop on Computational Personality Recognition (Shared Task)" organizer released two gold standard labeled datasets: *essays* and *myPersonality* (Celli et al., 2013). For this study we have used *myPersonality* corpus. The corpus was collected from the social network (Facebook) and 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. It is suggested in the shared task guidelines to split the data as train (66%) and test (34%). We followed the suggestion and divided the corpus into train and test set using the stratified sampling (similar proportion of classes in two sets) technique. We used only class labels for personality traits classification. A distribution of the labels in the corpus is given in Table 1. Train and test set have different distributions of positive and negative cases in different OCEAN categories. In total there are 6,545 train and 3,372 test instances after the split. From the corpus analysis, it is observed that besides words, it contains tokens such as internet-slang (e.g. WTF-what the F***), emoticons (e.g., :-D), acronyms (e.g., BRB-be right back) and various shorthand notations that people use in their status. The maximum number of tokens per user status message is 89, minimum 1 and the average is 14.

Cl	Train-set		Test-set	
	Y (%)	N (%)	Y (%)	N (%)
O	4863(74.3)	1682(25.7)	2507(74.3)	865(25.7)
C	3032(46.3)	3513 (53.7)	1524(45.2)	1848(54.8)
E	2784(42.5)	3761 (57.5)	1426(42.3)	1946(57.7)
A	3506(53.6)	3039 (46.4)	1762(52.3)	1610(47.7)
N	2449(37.4)	4096 (62.6)	1268(37.6)	2104(62.4)

Table 1: Distribution of labels in the myPersonality corpus. Y and N represent positive and negative classes. Cl represents class.

Experimental Method

We conducted several experiments to examine the performance of different classification methods for the automatic recognition of personality traits. The classification models we examine are SMO (John, 1998), BLR (Genkin et al., 2007) and MNB sparse model (Puurula, 2011).

Features

We followed bag-of-words approach and used tokens (unigrams) as features, where a classification instance is a vector of tokens appearing in the Facebook status. As discussed earlier, different kinds of tokens (internet-slangs, smiles, emoticons, etc.) are present in the corpus; our assumption is that these tokens carry distinctive information for personality traits recognition. Thus, there was no attempt to remove or normalize them. Using weka's 'string to word vector', text was converted into feature vector using TF-IDF (Manning, 2008) as feature value. The training set dictionary obtained using this scheme contains ~15,000 features; the same dictionary was used for the test set. TF-IDF feature valued representation was selected for the fact that it outperformed Boolean feature valued representation on exploratory experiments.

Classifiers

For the experiments, we used SMO with linear kernel, BLR and MNB sparse model. The choice of algorithms is driven by their different properties for classification. SMO is an optimization technique for solving quadratic optimization problems, which arise during the training of SVM and it has better generalization capability. Another reason for SMO is the high classification accuracy on different tasks reported in the literature (Schuller, 2012; Mairesse, 2007; Kermanidis, 2012) on personality traits recognition. BLR uses different priors (e.g. Laplace and Gaussian) to avoid overfitting and produces sparse predictive models for text data; moreover, it is also widely applied in text categorization. The key idea of BLR is to use prior probability distribution that favors sparseness in the fitted model. Whereas, MNB sparse model is an extension of Multinomial Naïve Bayes generative model where a sparse representation is used to reduce space and time complexity. For the feature extraction and the classification we used weka[1] (Witten, 2011).

Evaluation Methods

Performance of the system had been evaluated using myPersonality test set. In the shared task guidelines it is suggested to use precision, recall, F1 as evaluation metrics. Additionally, we computed weighted average (WA) and un-weighted average (UA), which are used in recent paralinguistic classification tasks (Schuller, 2012). UA is the average of true positive rate and true negative rate where the average of both poles is considered, whereas WA is the accuracy (Acc).

Even though, the suggestion is to use precision, recall and F1, we have computed macro-averaged precision, recall and F1 to consider both poles. Another motivation is that, macro-averaged precision, recall and F1 are inline with UA and WA metrics. Hence, we use the terms Pre-Avg, Re-Avg, F1-Avg, Acc (WA) in this paper. Since UA is the same as the average of recall, it is not reported. Pre-Avg, Re-Avg and F1-Avg are computed using the equations 1-3.

$$Pre\ (Avg) = \frac{1}{2}\left(\frac{tp}{tp+fp} + \frac{tn}{tn+fn}\right) \quad (1)$$

$$Re(Avg) = \frac{1}{2}\left(\frac{tp}{tp+fn} + \frac{tn}{tn+fp}\right) \quad (2)$$

$$F1(Avg) = 2\frac{Pre(Avg)*Re(Avg)}{Pre(Avg)+Re(Avg)} \quad (3)$$

where tp, tn, fp and fn are true positive, true negative, false positive and false negative.

Classification Results

In this section we report and discuss the performances of the classification algorithms on personality traits recognition task. Table 2 reports results for SMO, where *chance (%)* is the accuracy computed by randomly drawing class labels using prior distribution. It is computed 100 times with seed (1-100) and the mean is measured. The results of BLR and MNB sparse classifiers are reported in Tables 3 and 4, respectively. All classification algorithms perform above *chance* level baseline across all categories.

Additional to training and test set evaluation, we run 10-folds cross validation on the training set to predict variability. Table 5 reports mean±standard deviation values on the cross validation run of MNB sparse model.

[1] http://www.cs.waikato.ac.nz/ml/weka/; Version: 3-7-7

From the study of the personality traits recognition on the social network data (Facebook status messages), it is observed that MNB sparse generative model performs better than discriminative models, SMO and BLR. Comparing the cross validation results on the training set (Table 4) and the test set results (Table 5) using MNB sparse model, the conclusion is that the test set results are within the statistical variation.

Class	Pre-Avg	Re-Avg	F1-Avg	Acc	Chance (%)
O	57.46	58.28	57.68	65.84	61.78
C	58.02	58.09	57.99	58.16	50.36
E	57.47	57.57	57.49	58.21	51.05
A	58.40	58.41	58.40	58.45	50.10
N	56.89	56.99	56.92	59.25	52.94
Mean	**57.65**	**57.87**	**57.70**	**59.98**	**53.25**

Table 2: Results on the test set using the SMO (linear kernel) classifier. Chance (%) is the mean accuracy obtained by randomly drawing labels 100 times using the prior distribution.

Class	Pre-Avg	Re-Avg	F1-Avg	Acc
O	55.03	55.86	55.02	62.57
C	56.99	57.06	56.90	57.00
E	56.06	56.17	56.02	56.58
A	57.79	57.71	57.68	57.95
N	55.38	55.52	55.41	57.59
Mean	**56.25**	**56.46**	**56.21**	**58.34**

Table 3: Results on the test set using the BLR.

Class	Pre-Avg	Re-Avg	F1-Avg	Acc
O	59.83	59.71	59.77	69.48
C	59.06	59.11	59.07	59.34
E	57.99	58.13	57.98	58.57
A	59.09	58.71	58.49	59.16
N	58.84	57.90	57.95	62.40
Mean	**58.96**	**58.71**	**58.65**	**61.79**

Table 4: Results on the test set using Multinomial Naïve Bayes sparse model.

Class	Pre-Avg	Re-Avg	F1-Avg	Acc
O	58.6±1.6	58.4±1.4	58.4±1.5	68.5±1.7
C	59.2±1.4	59.2±1.3	59.2±1.3	59.4±1.4
E	58.2±1.6	58.3±1.6	58.1±1.6	58.6±1.5
A	57.2±1.6	56.9±1.5	56.7±1.5	57.6±1.5
N	59.6±2.1	58.5±1.7	58.6±1.7	63.0±1.9
Over all	**58.5±0.9**	**58.3±0.8**	**58.2±0.9**	**61.4±4.5**

Table 5: Results (mean ± standard deviation) on 10-folds cross validation run of the train set using Multinomial Naïve Bayes sparse model. Last row represents the overall mean ± standard deviation.

Since there are no published results on this particular data set, we report results on other corpora used in the personality traits recognition literature. First, Mairesse et al. (2007) report classification accuracy ranging from 52.75 to 62.52 and the overall of 57.10 with SMO classifier on the essay corpus. Second, (Kermanidis, 2012) reports precision-recall results for both poles and the average recall ranges from 49.00 to 64.50 with the overall of 58.30 on the modern Greek spontaneous text corpus with SMO. Thus, the performance of classifiers on myPersonality data reported in this paper is within the expected range.

We obtained overall macro-averaged precision - 58.96, recall - 58.71, F1 - 58.65 and accuracy 61.79 with our best model. The results of MNB are statistically significant with p<2.20E-16 when compared to SMO and BLR using Pearson's Chi-squared test.

In all of the experiments we used classifiers' default parameters; additional parameter tuning might increase the performance. Additionally, we have conducted an experiment by leave-one-user-group-out (LOUGO) cross validation method using all the data set and the obtained results are reported in Table 6. The data was randomly split into 10 user groups. LOUGO cross-validation gives an almost unbiased estimate because training and test set instances are completely independent in terms of user.

Class	**Pre-Avg**	**Re-Avg**	**F1-Avg**	**Acc-Avg**
O	59.75	59.88	59.80	69.12
C	60.44	60.42	60.40	60.71
E	59.16	59.29	59.14	59.74
A	59.10	58.73	58.57	59.33
N	59.86	58.94	59.06	63.23
Mean	**59.66**	**59.45**	**59.39**	**62.43**

Table 6: Results using LOUGO cross validation method using all data with MNB sparse model.

An extension of this study would be combining different classifiers results where an upper bound of the overall accuracy would be 76.45± 2.63, which was obtained using an oracle experiment.

Conclusion and Future Study

In this paper, we present our baseline study to automatically recognize BIG-5 personality traits on the social network data (Facebook status messages). We explored different classification methods. In this study we observed that MNB sparse model performs better than SMO and BLR. We report system performances using macro-averaged precision, recall, F1, and accuracy (WA). Future directions of this study include integrating syntactic, semantic and statistical features; studying feature selection and classifier combination methods, which may lead to provide more information to recognize personality traits.

References

Amichai-Hamburger, Y. and Vinitzky, G. 2010. Social network use and personality. In Computers in Human Behavior. 26(6): 1289–1295.

Celli, F. 2012. Unsupervised Personality Recognition for Social Network Sites. In Proc. of ICDS. Valencia.

Celli, F.; Pianesi, F.; Stillwell, D.S.; and Kosinski, M. 2013 WCPR (Shared Task). The 7th International AAAI Conference On Weblogs And Social Media. Boston.

Donnellan, B.; Conger, R. D.; and Bryant, C. M. 2004. The Big Five and enduring marriages. J. of Research in Personality 38:481–504.

Genkin, A.; Lewis, DD.; Madigan, D. 2007. Large-scale Bayesian logistic regression for text categorization. Technometrics 49 (3): 291-304.

Gill A. J. and French. R. M. 2007. Level of representation and semantic distance: Rating author personality from texts. In Proc. of the 2nd European Cognitive Science Conference.

Ivanov, A. V.; Riccardi, G.; Spork, A. J.; and Franc, J. 2011. Recognition of Personality Traits from Human Spoken Conversations. In Proc. of the Interspeech-2011, 1549-1552, Italy.

John C. Platt. 1998. Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines. Microsoft Research, Technical Report.

Kazai, G.; Kamps, J.; and Milic-Frayling, N. 2011. Worker types and personality traits in crowdsourcing relevance labels. In Proc. of the 20th ACM Conference on Information and Knowledge Management, CIKM 2011, Glasgow, United Kingdom.

Kermanidis K. L. 2012. Mining Author's personality traits from Modern Greek Spontaneous Text. In 4th International Workshop on Corpora for Research on Emotion Sentiment & Social Signals.

Mairesse F.; Walker, M. A.; Mehl M. R.; and Moore, R. K. 2007. Using linguistic cues for the automatic recognition of personality in conversation and text. J. Art. Intelligence Res. 30:457-500.

Manning, C. D.; Raghavan, P.; Schütze, H. 2008. An Introduction to Information Retrieval. Cambridge University Press Cambridge, England.

Matthews G.; Deary I.; and Whiteman, M. 2003. Personality traits. Cambridge University Press.

Mohtasseb, H.; and Ahmed, A. 2009. Mining Online Diaries for Blogger Identification. In Proc. of the ICDMKE-09, 295-302.

Nass C. and S. Brave. 2005. Wired for Speech: How Voices Activates and Advances the Human-Computer Relationship. MIT Press.

Oberlander J. and A. J. Gill. 2004. Individual differences and implicit language: Personality, parts-of-speech and pervasiveness. In Proc. of the CogSci-2004, USA.

Puurula A. 2012. Scalable Text Classification with Sparse Generative Modeling. In PRICAI2012, 458-469.

Schuller B.; Steidl S.; Batliner, A.; Nöth, E.; Vinciarelli, A.; Burkhardt, F.; Son, R. van; Weninger, F.; Eyben, F.; Bocklet, T.; Mohammadi, G.; and Weiss, B. 2012. The Interspeech 2012 Speaker Trait Challenge. In Proc. of the Interspeech 2012, USA.

Witten, I.; Frank, E.; and Hall, M. 2011. Data Mining: Practical Machine Learning Tools and Techniques. Morgan K.aufmann Publishers.