Be Conscientious, Express your Sentiment!

Fabio Celli and Cristina Zaga

University of Trento,
Corso Bettini 31, 38068 Rovereto, Italy.
fabio.celli@unitn.it
cristina.zaga@gmail.com

Abstract. This paper addresses the issue of how personality recognition can be helpful for sentiment analysis. We exploited the corpus for sentiment analysis released for the SEMEVAL 2013, we automatically annotated personality labels by means of an unsupervised system for personality recognition. We validated the automatic annotation on a small set of Twitter users, whose personality types have been collected by means of an online test. Results show that hashtag position and conscientiousness are the best predictors of sentiment in Twitter.

Keywords: Personality Recognition, Twitter, Sentiment Analysis, Data Mining

1 Introduction and Background

In psychology, personality is seen as an affect processing system [1] that characterise a unique individual [11], while sentiment analysis is a NLP task for tracking the mood of the public about products or topics [21]. Since psychologists suggest that personality is related to some aspects of mood [2], we expect that personality traits would help in a sentiment analysis task. In this paper, we exploit the correlations between language and personality provided by Golbeck et al. 2011 [6] and Quercia et al. 2011 [18] to predict personality labels in a Twitter dataset for sentiment analysis [23]. We use a system for personality recognition [4] to annotate personaliy labels in Twitter. Our goal is to test whether personality types can be good predictors of sentiment polarity.

The paper is structured as follows: in subsection 1.1 we introduce related work, in section 2 we present the dataset and describe the method used for the annotation with personality labels. In section 3 we report the results of our experiments and we draw some conclusions.

1.1 Related Work

In the last decade sentiment analysis and opinion mining strongly attracted the attention of the scientific community, and Twitter is a microblogging website that has been considered a very rich source of data for opinion mining and sentiment analysis [15]. Anyway, it is very challenging to extract linguisitc information from Twitter [12]. The 140 character limitations of tweets led to a sentence-level sentiment analysis. Kouloumpis et al. 2011 [10] has shown that in the microblogging domain, common tools for NLP may not be as useful sentiment clues as the presence of intensifiers, emoticons, abbreviations and hashtags. Given these results, rencently, more and more attention is given to the wide variety of user defined hashtags [9], [22]. The uniqueness of microblogging genre also led researchers to design NLP tools that make use of any number of domain-specific features including abbreviations, hashtags, emoticons and symbols [7], [14].

Personality recognition [11], [4] is a computational task that consists in the automatic classification of authors' personality traits from pieces of text they wrote. Most scholars use the Big5 model [5]. This model describes personality along five traits formalized as bipolar scales: extroversion (sociable or shy), neuroticism (calm or neurotic), Agreeableness (friendly or uncooperative), conscientiousness (organized or careless) and openness to experience (insightful or unimaginative).

The first applications in this field were on offline essays texts [11] and on blogs [13]. In recent years the interest of the scientific community towards the application of personality recognition in social networks, including Twitter [18], [6]. In particular, they extracted correlations between language and personality traits from Twitter, that we exploted for the annotation of the data.

2 Dataset, Annotation and Experiments

2.1 Data

We used the dataset released by Wilson et al. 2013 for the SemEval-2013 task B[1]. The purpose of this task is to classify whether a tweet is of positive, negative, or neutral. Gold standard sentiment labels are provided with data. The dataset consists of Twitter status IDs, and the task organizers provided a python script that downloads the data, if available. The final data includes the following information: tweet ID; user ID; topic; sentiment polarity; tweet text. We downloaded and cleaned the data, removing not available tweets. Data is splitted in training and test set, details are reported in Table 1. For each user in the dataset we have

set	instances	missing	total
training	5747	495	5252
test	687	123	564

Table 1. Summary of the dataset

just one text, that is not enough for the personality recognition. In order to get more tweets, we exploited user IDs and automatically collected all the tweets we found in their page. We collected an average of 12 tweets per user.

[1] http://www.cs.york.ac.uk/semeval-2013/task2/

2.2 Annotation of Personality Types

For the annotation of personality labels in the dataset, we exploited the system described in [3] and [4]. It is an unsupervised instance-based personality recognition system. Given as input a set of correlations between language cues and big5 personality traits, and a set of users and their texts, the system generates personality labels for each user, adapting the correlations to the data at hand. We

feature	ext.	agr.	con.	neu.	ope.
future	.227	-.100	-.286*	.118	.142
you	.068	.364*	.252*	-.212	-.020
article	-.039	-.139	-.071	-.154	.396*
negate	-.020	.048	-.374*	.081	.040
family	.338*	.020	-.126	.096	.215
humans	.204	-.011	.055	-.113	.251*
sad	.154	-.203	-.253*	.230	-.111
cause	.224	-.258*	-.155	-.004	.264*
certain	.112	-.117	-.069	-.074	.347*
hear	.042	-.041	.014	.335*	-.084
feel	.097	-.127	-.236*	.244*	.005
body	.031	.083	-.079	.122	-.299*
achive	-.005	-.240*	-.198	-.070	.008
religion	-.152	-.151	-.025	.383*	-.073
death	-.001	.064	-.332*	-.054	.120
filler	.099	-.186	-.272*	.080	.120
! marks	-.021	-.025	.260*	.317*	-.295*
parentheses	-.254*	-.048	-.084	.133	-.302*
? marks	.263*	-.050	.024	.153	-.114
words	.285*	-.065	-.144	.031	.200
followers	.15*	.02	.10	-.19*	.05
following	.13*	.07	.08	-.17*	.05

Table 2. Feature and Correlation set. *=p-value above .05

exploited the correlations between tweets and personality traits taken from [18] and [6]. We used only the correlations with p-value above .05, reported in Table 2. These correlations, that represent the initial model for the unsupervised system, include language-independent features, such as punctuation, Twitter-specific features, such as following and followers count, and features from LIWC [17], [20].

The outputs of the system are: one personality label for each user and the input text annotated. Labels are formalized as 5-characters strings, each one representing one trait of the Big5. Each character in the string can take 3 possible values: positive pole of the scale (y), negative pole (n) and missing/balanced (o). For example the label "ynooy" stands for an extrovert, neurotic and open mindend person. The annotation is a classificaiton task with 3 target classes.

The pipeline of the personality recognition system, depicted in Figure 1, has three phases: preprocessing, procesing and evaluation. In the preprocessing phase, the system samples 20% of the input unlabeled data, computing the average distribution of each feature of the correlation set, then assigns personality labels to the sampled data according to the correlations.

In the processing phase, the system generates one personality label for each

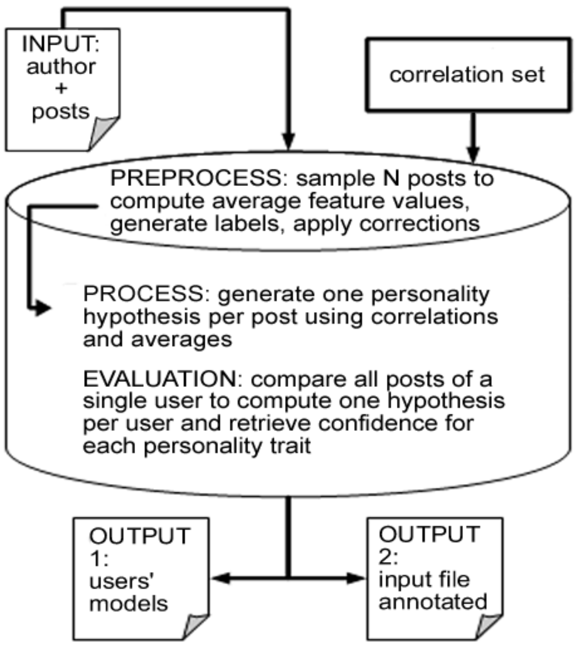

Fig. 1. System pipeline.

text in the dataset, mapping the features in the correlation set to specific personality trait poles, according to the correlations. Instances are compared to the distribution of features sampled during the preprocessing phase and filtered accordingly. Only features occurring more than the average are mapped to personality traits. For example a text containing more exclamation marks than average will fire positive correlations with conscientiousness and neuroticism and a negative correlation with openness to experience (see Table 2).

The system keeps track of the firing rate of each single feature/correlation and computes personality scores for each trait, mapping positive scores into "y", negative scores into "n" and missing or balanced values into "o" labels.

In the evaluation phase, the system compares all the personality labels generated for each single tweet of each user and retrieves one generalized label per user by computing the majority class for each trait. This is why the system can evaluate personality only for users that have at least two tweets, the other ones are discarded. In the evaluation phase the system computes average confidence and variability. Average Confidence is defined as the coverage of the majority class of the personality trait over the count of all the user's texts and gives a measure of the robustness of the personality hypothesis. Variability instead provides information about how much one author tends to write expressing the same personality traits in all the texts. It is defined as $var = \frac{avg\ conf}{T}$, where T is the the count of all the user's texts.

2.3 Validation of Personality Labels

In order to validate the annotation of the data, we developed a website[2] with a short version of the Big5 test, the BFI-10 [19]. We collected a gold-standard test set, with the personality scores of 20 Twitter users, their tweets and data. We computed random and majority baselines with 3 target classes (y, n, o), and then ran the system on the gold-standard test set. Results, reported in Table

	P	R	F1
random	0.359	0.447	0.392
majority	0.39	1	0.455
extroversion	0.595	1	0.746
neuroticism	0.595	1	0.746
agreeableness	0.371	0.5	0.426
conscientiousness	0.621	0.693	0.655
openness	0.606	0.833	0.702
avg.	0.558	0.805	0.655

Table 3. Results of the validation.

3, show that the average f-measure is in line with the results reported in [4]. Conscientiousness and openness to experience are the best predicted traits, in particular, conscientiousness has the highest precision. Agreeableness instead has a poor performance: we explain this with the fact that it is the trait for which we have fewer features.

2.4 Experiments and Discussion

We ran two different binary classification tasks, task A: subjectivity detection, and task B: sentiment polarity classification. The former is the task of distinguishing between neutral texts and texts containing sentiment, the latter is the classical opinion mining classification between positive and negative. As fea-

[2] http://personality.altervista.org/p.php

task A	task B
pronouns	verbs
proper names	hashtag final
verbs	hashtag initial
adjectives	tweets
adverbs	conscientiousness
interjections	
mentions	
urls	
emoticons	
numbers	
hashtag final	

Table 4. Feature selection

tures, we used the five personality traits, Twitter statistics (followers, following, tweets), emoticons (positive/negative), hashtag position (hashtag initial, hastag final) and Twitter Part-Of-Speech tags obtained buy means of a part-of-speech tagger designed for Twitter [7], [14].

As first experiment we ran feature selection in Weka [24], removing topics and using the correlation-based subset evaluation algorithm [8] with a greedy-stepwise feature space search. This algorithm evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them. Results are reported in Table 4: We see that hashtag position is very helpful while the only personality trait which is a good predictor of sentiment is conscientiousness. We ran a classification experiment, reported in Table 5, where we predicted the target classes using the features selected in the feature selection phase. Taking the majority baseline (zero rule), we observe that the best improvement over the baseline has been achieved in task A (distinction between neutral/subjective), while task B (positive/negative) has a very small improvement.

algorithm	task A (f1)	task B (f1)
bl (zero rule)	0.467	0.55
trees	0.619	0.571
bayes	**0.663**	0.598
svm	0.632	0.555
ripper	0.629	**0.612**

Table 5. Classification performance

3 Conclusions and Future Work

In this paper we attempted to exploit personality traits, and few other linguistic cues, including hashtags, to predict subjectivity and sentiment polarity in Twitter. The best performing team at the Semeval 2013 achieved an f1 of .889 for task A and of .69 for task B. While our results are far from the best one in task A, it is in line with the results of the shared task for task B. It is interesting the fact that conscientiousness is one of the features we exploited for task B.

The performance of the personality recognition system is far from perfect, but still we successfully exploited one specific trait of personality to classify sentiment. In the future we wish to improve the performance personality recognition system, adding more correlations, and to extend the exploitation of personality and hashtags to other domains, such as irony detection.

References

1. Adelstein J.S., Shehzad Z., Mennes M., DeYoung C.G., Zuo X-N., Kelly C., Margulies D.S., Bloomfield A., Gray J.R., Castellanos X.F. and Milham M.P. Personality Is Reflected in the Brain's Intrinsic Functional Architecture. In *PLoS ONE* 6:(11), 1–12. (2011).
2. Aitken Harris, J., and Lucia, A. The relationship between self-report mood and personality. *Personality and individual differences*, 35(8), 1903–1909. (2003).
3. Celli, F., and Rossi, L. The role of emotional stability in Twitter conversations. In *Proceedings of the Workshop on Semantic Analysis in Social Media.* (2012).
4. Celli, F. *Adaptive Personality recognition from Text.* Lambert Academic Publishing. Saarbruc̆hen. (2013).
5. Costa, P. T. and MacCrae, R. R. Normal personality assessment in clinical practice: The neo personality inventory. *Psychological assessment*, 4(1):5. (1992).
6. Golbeck J., Robles C., Edmondson M., and Turner K. Predicting Personality from Twitter. In *Proc. of International Conference on Social Computing.* (2011).
7. Gimpel K., Schneider N., O'Connor B., Das D., Mills D., Eisenstein J., Heilman M., Yogatama D., Flanigan J. and Smith N.A. Part-of-Speech Tagging for Twitter: Annotation, Features, and Experiments. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics.* (2011).
8. Hall M. A. *Correlation-based Feature Subset Selection for Machine Learning.* Hamilton, New Zealand. (1998).
9. Jiang L., Yu M., Zhou M., Liu X., and Zhao T. Target-dependent twitter sentiment classification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, (2011).
10. Kouloumpis, E., Wilson, T., and Moore, J. Twitter sentiment analysis: The Good the Bad and the OMG!. In*Proc. of ICWSM.* (2011).
11. Mairesse, F. and Walker, M. A. and Mehl, M. R., and Moore, R, K. Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text. In *Journal of Artificial intelligence Research*, 30. (2007).
12. Maynard D., Bontcheva K. and Rout D. Challenges in developing opinion mining tools for social media. In *Proceedings of NLP can u tag user generated content.* (2012).

13. Oberlander, J., and Nowson, S. Whose thumb is it anyway? classifying author personality from weblog text. In *Proceedings of the 44th Annual Meeting of the Association for Computational Linguistics ACL.* (2006).
14. Owoputi O., OConnor B., Dyer C., Gimpel K., Schneider N., Smith N.A. Improved Part-of-Speech Tagging for Online Conversational Text with Word Clusters. In *Proceedings of NAACL.* (2013).
15. Pak A. and Paroubek P. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In *proceedings of LREC.* (2010).
16. Pang B. and Lillian Lee L. Opinion Mining and Sentiment Analysis. In *Foundations and Trends in Information Retrieval.* 2(12). (2008).
17. Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., and Booth, R. J. The development and psychometric properties of LIWC2007. Austin, TX, LIWC.Net. (2007).
18. Quercia D., Kosinski M., Stillwell D. and Crowcroft J. Our Twitter Profiles, Our Selves: Predicting Personality with Twitter. In *Proceedings of SocialCom2011.* (2011).
19. Rammstedt, B., and John, O. P. Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, 41(1), 203-212. (2007).
20. Tausczik, Y. R., and Pennebaker, J. W. . The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1), 24-54. (2010).
21. Vinodhini G. and Chandrasekaran R. M. Sentiment Analysis and Opinion Mining: A Survey. In *International Journal.* 2(6). (2012).
22. Wang, X., Wei, F., Liu, X., Zhou, M., and Zhang, M. Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach. In*Proceedings of the 20th ACM international conference on Information and knowledge management.*(2011).
23. Wilson, T., Kozareva, Z., Nakov, P., Rosenthal, S., Stoyanov, V., and Ritter, A. SemEval-2013 task 2: Sentiment analysis in twitter. In *Proceedings of the International Workshop on Semantic Evaluation, SemEval.* (Vol. 13). (2013).
24. Witten I.H. and Frank E. *Data Mining. Practical Machine Learning Tools and Techniques with Java implementations.* Morgan and Kaufman, (2005).