Unsupervised Personality Recognition for Social Network Sites

Fabio Celli
CLIC-CIMeC,
University of Trento, Italy
`http://clic.cimec.unitn.it/fabio/`

Abstract

Personality is a complex of attributes that characterise a unique individual. Written sentences contain a lot of information about their authors' personality and Social Network Sites are huge, virtually infinite, corpora where authors (users) and sentences (posts) are found together in the same place. In this paper we present a system that extracts personality models without supervision from a popular Social Network Site: FriendFeed. We adopted five classes from the standard model known in psychology as the 'Big Five": extraversion, emotional stability, agreebleness, conscientiousness and openness to experience and we exploited the linguistic features associated with those classes in order to generate one personality model for each user. The system then evaluates the models by comparing all the posts of one single user (users that have one post only are discarded). The system provides accuracy (measure of the reliability of the personality model) and validity (measure of the variability of personality within a user) as evaluation measures. The analysis of a sample of 748 Italian users of FriendFeed showed that that the most frequent personality type is represented by the model of an extravert, insecure, agreeble, organized and unimaginative person. Social Network Sites, Personality Recognition, Information Extraction, Natural Language Processing

Personality is a crucial aspect of social interaction. Under the computational perspective it plays an important role in stylometry and sentiment analysis. Written sentences contain a lot of information about their authors' personality and this information is more robust and reliable than the one contained in spoken utternaces (see Mairesse *et Al.* 2007, (22)). Social Network Sites (SNSs henceforth, see Boyd and Ellison (22) for definitions and hystory) are huge, virtually infinite, corpora where authors (users) and sentences (posts) are found together in the same place. This large amount of data available makes possible the analysis of personality from text in a computational way, but there are two nontrivial problems:

1. The definition of personality, which is a very fuzzy and subjective notion;
2. The annotation of personality in the data from SNSs, that would require personality judgements by the author themselves or by other native speakers.
3. The evaluation of personality.

We are going to speak about those problems in the following paragraphs, then in section 2 we will present the system and the results of the analysis of personality. In section 3 we will conclude speaking about possible directions for future works.

Definition of Personality Psychologists describe personality along five dimensions known as the "Big Five" (see Goldberg 1992 (22)), a model introduced by Norman 1963 (22), obtained from factor analysis of personality description questionnaires that has become a standard over the years. The five dimentions are the following:

- Extraversion (E) (sociable vs shy)

- Emotional stability (S) (calm vs insecure)
- Agreebleness (A) (friendly vs uncooperative)
- Conscientiousness (C) (organized vs careless)
- Openness (O) (insightful vs unimaginative)

Those dimentions can be represented computationally as continuous numerical variables with 2 poles: one positive (1) and one negative (0), and this makes easy to use this model in a computational domain. Once we have the numerical values for each attribute (one attribute is one dimention in the "Big Five"), we can easily calculate if a user has one trait of personality (y) or not (n) or shows a balanced value (o). From this representaiton we can formalize a personality model for each user simply taking the majority class for each attribute/dimension from all posts made by one user. In the end personality models are formalized as string of five charachters: one for each attribute, which one can take three possible vaues: positive (y), negative (n) or balanced (o). We will see some examples later on.
There are a lot of factors that correlate with personality traits. In order to develop a program that is able to recognize users' personality in SNSs we need to turn those factors into features that can be extracted from text. Mairesse *et Al.* 2007 (22) provides a long list of factors and the Pearson's correlation coefficient between those factors and the personality traits, obtained from an essay corpus where subjects provided also a judgement of their personality, following the "Big Five" model. Among those factors that correlates with certain aspects of personality there are some regarding topic (for example if a person writes about job, leisure, music, other people), some regarding word usage (for example the use of negative particles, first person pronouns, fillers, swares) and some regarding psychological aspects (for example age of acquisition of the word used, length of the words used, expression of positive and negative feelings). Factors are supposed to be valid for the western culture. From the factors described in Mairesse *et Al.* 2007, we picked up 21 features formalizable in a computaional way as numerical variables. they are:

1. **all punctuation** (ap): the total of `.` `,` `;` `:` in the sentence,
2. **commas** (cm): the total of `,` in the sentence,
3. **reference to other users** (du): the frequency of the pattern `@...` in the sentence,
4. **exclamation marks** (em): the total of `!` in the sentence,
5. **external links** (el): the number of external links in the sentence,
6. **first person singular pronouns** (im): the number of first person singular pronoun forms in the sentence,
7. **negative particles** (np): the total of negative particles in the sentence,
8. **negative emotions** (ne): the total of emoticons expressing negative feelings in the sentence,
9. **numbers** (nb): the total numbers in the sentence,
10. **parenthesis** (pa): the total number of parenthetical phrases in the sentence,
11. **positive emotions** (pe): the total of emoticons expressing positive feelings in the sentence,
12. **prepositions** (pp): the total of prepositions in the sentence,
13. **pronouns** (pr): the total of pronouns in the sentence,
14. **question marks** (qm): the total of `?` in the sentence,
15. **long words** (sl): total number of words longer than 6 letters in the sentence,
16. **self reference** (sr): the total first person (singular and plural) in the sentence,
17. **swears** (sw): total number of vulgar expressions in the sentence,
18. **type/token ratio** (tt): defined as the formula below,
19. **word count** (wc): total words in the sentence,
20. **first person plural pronouns** (we): total first person plural pronoun forms in the sentence,

21. **second person singular pronouns** (yu): total second person singular pronoun forms in the sentence,

22. **mean word frequency** (mf): simple mean of the frequency of words in the sentence, defined in the formula below.

$$tt = \frac{(w1) - T}{T} \quad mf = \frac{\Sigma wf}{T}$$

where *T* is the total word count in the sentence and *wf* is the frequency count of the word in the dataset. Table 1 below shows how the linguistic features used correlate with personality traits. Correlation coefficient are reported by Mairesse *et Al.* 2007:

F.	E	S	A	C	O
ap	-.08**	-.04	-.01	-.04	-10**
cm	-.02	.01	-.02	-.01	.10**
du	-.07**	.02	.01	.01	.06**
el	-.05*	-.02	-.01	-.03	.09**
em	-.00	-.05*	.06**	.00	-.03
in	-.04*	.01	-.01	-.03	-.01
im	.05*	-.15**	.05*	.04	-.14**
np	-.08**	.12**	.11**	-.07**	.01
ne	-.03	-.18**	-.11**	-.11**	.04
nb	-.03	.05*	-.03	-.02	-.06**
pa	-.06**	.03	-.04*	-.01	.10**
pe	.07**	.07**	.05*	.02	.02
pp	.00	.06**	.04	.08**	-.04
pr	.07**	.12**	.04*	.02	-.06**
qm	-.06**	-.05*	-.04	-.06**	.08**
sr	.07**	-.14**	-.06**	-.04	-.14**
sl	-.06**	.06**	-.05*	.02	.10**
sw	-.01	.00	-.14**	-.11**	.08**
tt	-.05**	.10**	-.04*	-.05*	.09**
wc	-.01	.02	.02	-.02	.06**
we	.06**	.07**	.04*	.01	.04
yu	-.01	.03	-.06**	-.04*	.11**
mf	.05*	-.06**	.03	.06**	-.07**

Table 1: Features used in the program and their Pearson's correlation coefficients with personality traits as reported in Mairesse *et Al.* 2007. * = p smaller than .05 (weak correlation), ** = p smaller than .01 (strong correlation)

1.3 Building the Personality Model I extracted all the features that correlate with personality traits from the FriendFeed dataset, and I obtained the results summarized in table 2: In order to calculate the numeric values the system applies the following rules: if a sentence shows a feature whose frequency is higher than mean plus standard deviation and it correlates positively with one personality trait, add a +0.1 to that personality trait, if a sentence shows a feature whose frequency is higher than mean plus standard deviation and it correlates negatively with one personality trait, add a -0.1 to that personality trait except for tt, which is an inverse feature (the smaller the value, the highest the correlation with personality traits). Then the system transforms numerical values into nominal ones ("y","n" and "o") simply checking if a value is positive, negative or it is zero. In the end the majority class of each personality trait is calculated for each user and the resulting string is assumed as the user's personality model.

feature	mean	sd	min	max
ap	1	2	0	28
cm	0	1	0	19
du	0	0	0	3
el	0	0	0	3
em	0	0	0	7
im	0	0	0	3
np	0	0	0	4
ne	0	0	0	1
nb	1	4	0	64
pa	0	0	0	3
pe	0	0	0	2
pp	1	2	0	32
pr	0	0	0	8
qm	0	0	0	3
sr	0	0	0	4
sl	6	6	0	71
sw	0	0	0	1
tt	0.971	0.048	0.706	1
wc	7	7	1	79
we	0	0	0	2
yu	0	0	0	2
mf	101264	87192	68	567704

Table 2: Summary of the behavior of features associated to personality traits in the dataset.

Collection and Annotation of the Data Our dataset is a sample of 748 Italian FriendFeed users (1065 posts). It is a subset of the dataset collected by Celli *et Al.* 2010 (22). The dataset has been

collected from FriendFeed public URL[1] where new posts are displayed and can be sampled.
The dataset was already processed with a language identifier, whose performance is correct at 88%. This made easier the extraction of the italian subset from the dataset. We tried to annotate personality on the dataset by using Amazon's Mechanincal Turk[2] in order to make a supervised evaluation. Snow *et Al.* 2008 (22) showed that when many Mechanical Turk non-expert labelers agree on annotation, there is a high level of agreement between their labels and existing gold standard labels provided by expert labelers in a variety of different tasks. This task has revealed to be impracticable because Turkers could not reach the requested agreement (90%, while the default is 95%), showing that it is very hard to reach consensus on personality judgements between human annotators, because personality is a very fuzzy and subjective notion, and this proves that the task we are challenging is a nontrivial one.

Evaluation of Personality Models The evaluation method is based on the assumption that one user has one and only one complex personality, and that this personality emerges at different degrees from user's posts. Hence we evaluate the personality model comparing many posts of the same user. The drawback of this method is that we can only evaluate models for users that have more than one post in the dataset, and we have to discard all the other models. Models are expressed as strings with 5 attributes, corresponding to the 3 classes: Yes, No, Zero. This evaluation method provides two measures, accuracy (a) and validity (v), defined in the formulas below:

$$a = \frac{tp + tn}{tp + tn + fp + fn} \qquad v = 1 - \frac{a}{P}$$

Where P is the number of posts of one user; tp is the sum of each personality attribute matching within the same user (for example "y" and"y", "n" and "n", "o" and "o"); tn is the sum of opposite attributes within the same user ("y" and "n", "n" and "y"); fp is the sum of possible attributes turned to the balance value within the same user ("y" to "o" and "n" and "o") and fn is the sum of the balance attributes turned to positive or negative ("o" to "y" and "o" and "n"). Accuracy gives a measure of the reliability of the personality model and validity gives information about how much the model is valid for all the user's posts, in other words how much one user writes expressing the same personality traits. Since the personality model is built calculating the majority value for each attribute, here validity can take values from 0.5 to 1.

[1] `http://friendfeed.com/public`
[2] `https://requester.mturk.com/mturk/`

1 Analysis and Discussion

We filtered out groups and kept only single users from the dataset. Most users (592) have just one post and the models obtained from those users were not considered reliable (accuracy is set to 0). The mean accuracy (excluding the users with only one post) is 0.631, the mean validity, excluding users with one post, is 0.729, which is good result, not directly comparable with any other.
The results about the frequency of personality models in the sample, processed with the statistical tool R (see Dalgaard 2002 (22)), is reported in table 3 below: Below rank 7 models become more and more sparse, with a long tail of models appearing only once, that do not appear in table 3. The most frequent personality type in the italian subset of FriendFeed is represented by the model of an extravert, insecure, agreeble, organized and unimaginative person. It is interesting to note that the features "insecure" and "unimaginative" is present in the first four positions of the ranking and that no shy people is found in the first six positions.
Pearson's correlation test reveal that there is a strong

Rank	**Model**	**Rel.Freq.**
1	ynyyn	16.6%
2	ynyon	12.1%
3	onoyn	7.6 %
4	ooooo	7.6%
5	ynoyn	4.5%
6	yoooo	4.5%
7	ynooo	3.8%
8	ynoyo	3.8%
9	ynoon	3.2%
10	onyoo	3.2%
11+	*others*	33.1%

Table 3: Frequency of personality models.

(+0.79) and highly significant correlation (p-value = .0003) between the accuracy and personality model types, meaning that there are certain personality types that express strongly and reliably their personality in written language, and others that do not. Althought there is no correlation (p-value = .413) between personality and posting activity, once filtered out the long tail of users with sparse personality models, emerges that there is one personality type that produces more posts than others, that is the extravert, insecure, friendly, not particularly precise and unimaginative person (ynyon).

A manual look to the data reveals that there are some users (the ones with higher validity) that are focused on a topic, and sometimes this topic is clear from their username: for example "styleandthecity", or such users as "ultimora" or "cronaca24", which appear to be journalists and have a very recognizable and normalized style, but not the same personality model.

2 Conclusions and Future work

In this work I showed that it is possible to extract personality information from SNSs in an unsupervised way with acceptable accuracy, which is very good if we think about the fact that humans hardly agree on such a subjective task as personality recognition. The results reported here show that the distribution of personality models in SNSs has a high peak of people sharing the same personality traits and a long tail of people with a unique personality model. Results also show that Validity is a good measure of the recognizability of the style of a user. In the future it would be interesting to run experiments following threads of users checking for their personality in order to study how personalities interact together, and what are the features that make a post interesting to a certain personality type.

References

Boyd, D., Ellison, N. 2007. Social Network Sites: Definition, history, and scholarship. In *Journal of Computer-Mediated Communication* 13(1).

Callison-Burch, C. 2009. Fast, Cheap, and Creative: Evaluating Translation Quality Using Amazon's Mechanical Turk. In *Proceedings of EMNLP*.

Celli, F., Di Lascio, F.M.L., Magnani, M., Pacelli, B., Rossi, L. 2010. *Social Network Data and Practices: the case of Friendfeed*. Advances in Social Computing, pp. 346-353. Series: Lecture Notes in Computer Science, Springer, Berlin.

Dalgaard, P. 2002. *Introductory statistics with R*. Springer, New York, NY.

Goldberg, L., R. 1992. The Development of Markers for the Big Five factor Structure. In *Psychological Assessment*, 4(1).

Hudlicka, E. 2005. A Computational Model of Emotion and Personality: Applications to Psychotherapy Research and Practice. In *Proceedings of the 10th Annual CyberTherapy Conference: A Decade of Virtual Reality* Basel, Switzerland.

Mairesse, F., Walker, M. 2006. Words mark the nerds: computational models of personality recognition through language. In: *Proceedings of the 28th Annual Conference of the Cognitive Science Society*. Vancouver, Canada.

Mairesse, F., Walker, M. A., Mehl, M. R., Moore, R, K. 2007. Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text. In *Journal of Artificial intelligence Research*, 30.

Norman, W., T. 1963. Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality rating. In *Journal of Abnormal and Social Psychology*, 66.

Oberlander, J., Nowson, S. 2006. 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*. Sydney.

Pak, A., Paroubek P. Twitter as a corpus for sentiment analysis and opinion mining. 2010. In *Proceedings of LREC 2010* Malta.

Snow, R., O'Connor, B., Jurafsky, D., Ng, A. Y. 2008. Cheap and fast—but is it good?: evaluating non-expert annotations for natural language tasks. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Honolulu, Hawaii.