Novel Comment Spam Filtering Method on Youtube: Sentiment Analysis and Personality Recognition

Enaitz Ezpeleta[1], Iñaki Garitano[1], Ignacio Arenaza-Nuño[1], José María Gómez Hidalgo[2], and Urko Zurutuza[1]

[1] Electronics and Computing Department, Mondragon University
Goiru Kalea, 2, 20500 Arrasate-Mondragón, Spain
{eezpeleta,igaritano,iarenaza,uzurutuza}@mondragon.edu,
[2] Pragsis Technologies
Manuel Tovar, 43-53, Fuencarral - 28034 Madrid, Spain
jmgomez@pragsis.com

Abstract. The deeply entrenched use of Online Social Networks (OSNs), where millions of users share unconsciously any kind of personal data, offers a very attractive channel to attackers. They provide the possibility of sending spam messages through different channels (wall posts, comments, private messages). In this paper we propose a novel spam filtering method focused on social media spam. It aims to demonstrate that using sentiment analysis and personality recognition techniques, in order to analyze the content of the texts, the improvement of spam filtering results is possible. We add these features to each OSN spam both independently and jointly, and then we compare Bayesian spam filters with and without the new features in terms of the number of false positive and accuracy. At the end, the results of the top ten filtering classifiers have been improved, reducing also the number of false positives (26.69% on average), reaching an 82.55% of accuracy.

Keywords: spam, social spam, Youtube, polarity, security, personality

1 Introduction

The current massive publication of private information in Online Social Networks (OSNs), give the attackers the possibility of using every single information against the users. Those sites are also becoming an attractive segment to act inside them. This is a significant risk if we take into account the amount of users that the most popular OSNs have: Facebook reached 1.86 billion monthly active users as of December 31, 2016 [3]; Youtube has counted over a billion users in 2017 [4]; and Twitter has 313 million monthly active users as of June 30, 2016[5].

[3] http://newsroom.fb.com/company-info/
[4] https://www.youtube.com/yt/press/statistics.html
[5] https://about.twitter.com/company

As an example, in [1], Gao et. al. carried out a study to quantify and characterize spam campaigns launched from accounts on OSNs. Their results clearly showed that OSNs are now a major delivery platform targeted for spam.

Being selling products, creating social alarm, creating public awareness campaigns, generating traffic with viral contents, fooling users with suspicious attachments, etc. the main purpose of spam messages, those type of communications have a specific writing style that spam filtering can take advantage of. In this study we focus on the possibility of using Natural Language Processing (NLP) techniques in order to improve results obtained with current spam filtering classifiers. On the one hand, as authors demonstrate in [2], sentiment analysis of the content can help to improve email spam detection. On the other hand, in [3] results validate the possibility of using personality recognition techniques in order to obtain better results. Taking as a baseline these two methods, the main objective of this paper is to demonstrate that sentiment analysis and personality recognition techniques help to improve current spam filtering results.

First, several spam filtering classifiers and different settings are applied to a known dataset in order to identify the best ones. After that, the different sentiment analyzers and a personality recognition model are applied to create new datasets adding this features. In the next step, a combined dataset is created adding the two features together. Once, the datasets are created, the best ten classifier are applied to the different datasets to obtain all the results. Finally, a comparison and an analysis of the results is carried out.

The remainder of this paper is organized as follows. Section 2 describes the previous work conducted in the area of social media spam filtering, and sentiment and personality recognition techniques. Section 3 describes the process of the aforementioned experiments, regarding Bayesian spam filtering and spam filtering using the polarity and the personality of the texts. In Section 4, the obtained results are described, and finally, we summarize our findings and give conclusions in Section 5.

2 Related Work

2.1 Online Social Network Spam

Numerous research related with spam and OSNs has been carried out [4]. In [5] authors demonstrate that it is possible to automatically identify accounts on three large social networking sites (Facebook, Twitter and MySpace) used by spammers, and block these spam profiles. Further, a framework for spam detection which is able to run across OSNs is proposed in [6]. An equally important study is presented in [7]. The authors developed a tool that detects compromised accounts based on anomalies detected in user behaviour. Finally, in [8] authors used classification and clustering techniques to detect spam campaigns inside different OSNs such as Facebook and Twitter. Ezpeleta et al. [9] showed that personalizing spam messages using publicly available OSN profile information lead to a significantly higher success rate than conventional, non-personalized spam.

In terms of spam inside OSNs, it is important to mention that a huge amount of studies about spam in Twitter have been performed. Authors explain in [10] how criminal accounts mix into and survive in the whole Twitter space. Moreover, Song et al. [11] demonstrate how spammer detection is possible using the distance and connectivity between receiver and recipient, which are hard to manipulate by spammers.

The main problem is that although a lot of techniques has been published [12, 13], spam messages are still a significant problem in OSNs.

2.2 Sentiment Analysis

As explained in [14], the area of SA has had a huge burst of research activity during these last years, but there has been a continued interest for a while. Currently there are several research topics on opinion mining and the most important ones are explained in [15]. Among those topics we identified the document sentiment classification as a possible option for spam filtering.

The main objective of this area is classifying the positive or negative character of a document [14]. In order to classify such sentiment, some researchers use supervised learning techniques, where three classes are previously defined (positive, negative and neutral) [16]. Some other authors propose the use of unsupervised learning. In unsupervised learning techniques, opinion words or phrases are the dominating indicators for sentiment classification [17].

Moreover, authors in [18] demonstrate the possibility of using tweets sentiment analysis in order to improve spam filtering results in Twitter.

2.3 Personality Recognition

Personality is a psychological construct aimed at explaining the wide variety of human behaviors in terms of a few, stable and measurable individual characteristics [19]. As authors explain in [20], two main models to formalize personality have been defined: Myers-Briggs personality model [21], which defines the personality using four dimensions: Extroversion or Introversion, Thinking or Felling, Judging or Perceiving and Sensing or iNtuition; and the Big Five model [22] which divides the personality in 5 traits: Openness to experience, Conscientiousness, Extroversion, Agreeableness and Neuroticism.

As it is shown in [23] every text contains a lot of information about the personality of the authors, being this the reason that personality recognition became a potential tool for Natural Language Processing. During the last years, different research in personality recognition in blogs [24], offline texts [23] or online social networks [25, 26] have been published.

In [27] authors prove that personality prediction is feasible, and their email feature set can predict personality with reasonable accuracies. This work shows that it is possible to predict the personality of a writer using email messages.

Moreover, personality recognition is used in order to detect opinion spam in social media [28], and other researchers present the relationship between personality traits and deceptive communication [29].

3 Design and Implementation

As we mentioned in Section 1 first of all, the best spam filtering classifiers identified in the literature and several settings are applied to dataset composed of social media spam in order to identify the best ten. After that, original dataset is fed with sentiment, and personality features in a way that four datasets are kept for comparison: the original one, the original with a polarity feature, the original with the personality feature, and finally the aggregation of both polarity and personality features to the original dataset. Next, the ten classifiers that better discriminate the original are applied to all the datasets in order to compare the results. All the process is presented in the Figure 1.

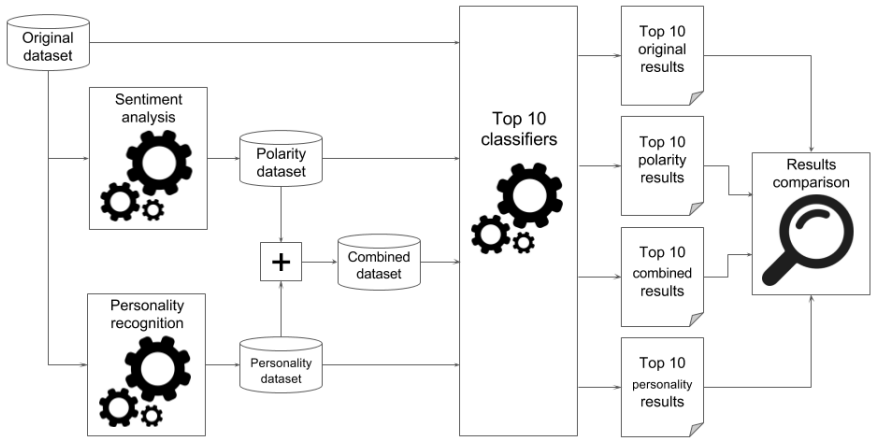

Fig. 1. Novel Social Media comments filtering method.

During those experiments 10-fold cross-validation technique is used, and the results are analyzed in terms the number of false positive and the accuracy. Accuracy is the percentage of testing set examples correctly classified by the classifier. And legitimate messages classified as spam are considered false positives.

3.1 Datasets

During this work a publicly available dataset is used:

– *Youtube Comments Dataset* [6]: Presented in [30]. This dataset contains multilingual 6,431,471 comments from a popular social media website, Youtube[7]. Among all the comments, 481,334 are marked as spam.

[6] http://mlg.ucd.ie/yt/
[7] www.youtube.com

In order to use similar number of texts to the experiments presented in [2] and [3] we created a new subset composed of 1,000 spam and 3,000 ham comments. Those texts have been selected randomly and only taking into account comments written in English.

3.2 Social Media Spam Filtering

With the objective of identifying the best spam classifiers, several spam classifiers using different settings are applied to the Youtube Comments dataset.

Following the strategy presented in [2], 7 different classifiers and 56 settings combinations per each classifiers are applied (392 combinations in total), and the best ten results are presented in Table 1.

#	Spam classifier	TP	TN	FP	FN	Accuracy (Acc)
1	NBM.c.stwv.go.ngtok	389	2911	89	611	82.50
2	NBMU.c.stwv.go.ngtok	389	2911	89	611	82.50
3	NBM.stwv.go.ngtok	370	2929	71	630	82.48
4	NBMU.stwv.go.ngtok	370	2929	71	630	82.48
5	NBM.c.stwv.go.ngtok.stemmer	379	2919	81	621	82.45
6	NBMU.c.stwv.go.ngtok.stemmer	379	2919	81	621	82.45
7	NBM.stwv.go.ngtok.stemmer	358	2936	64	642	82.35
8	NBMU.stwv.go.ngtok.stemmer	358	2936	64	642	82.35
9	CNB.stwv.go.ngtok	417	2875	125	583	82.30
10	CNB.stwv.go.ngtok.stemmer	400	2891	109	600	82.28

Table 1. Results of the best ten classifiers

During this study different nomenclatures and acronyms, which are explained in Table 2, are used. We use the same nomenclatures in this paper.

	Meaning		Meaning
CNB	Complement Naive Bayes	.stwv	String to Word Vector
NBM	Naive Bayes Multinomial	.go	General options
NBMU	Naive Bayes Multinomial Updatable	.wtok	Word Tokenizer
.c	idft F, tft F, outwc T [8]	.ngtok	NGram Tokenizer 1-3
.i.c	idft T, tft F, outwc T [8]	.stemmer	Stemmer
.i.t.c	idft T, tft T, outwc T [8]	.igain	Attribute selection using InfoGainAttributeEval

Table 2. Nomenclatures

[8] idft means Inverse Document Frequency (IDF) Transformation; tft means Term Frequency score (TF) Transformation; outwc counts the words occurrences.

Once the best classifiers and the best results are identified using the original dataset, in the following steps the objective is to improve these results. To do that, the same classifiers are applied to the new datasets, which are created adding personality and polarity features to the original dataset.

3.3 Using Sentiment Analysis To Improve Social Media Spam Filtering

The main objective of this part is to add the polarity of each message to the original dataset. To do that, we analyze the procedure shown in [2] where the best sentiment classifiers were identified to carry out the experiments.

Based on the accuracies presented in the mentioned paper, where several sentiment classifiers were applied to the Movies Review dataset [9], the best four classifiers are selected (*Adjective*, *Adjective+*, *TextBlob 0.05* and *TextBlob 0.1*). Those ones are used to annotate the text included in Youtube comments dataset which has not been annotated for sentiment. As a result, we obtain four new datasets (one per each classifier). The original one and the new four are used in the experiments.

3.4 Using Personality Recognition To Improve Social Media Spam Filtering

The next phase in our study aims to apply personality recognition techniques to each Youtube comment in order to create a new dataset, adding this feature to the original dataset.

Like in [3], in this study we use one of the most trusted personality model: Myers-Briggs personality model. This model is composed by four different dimensions (Extroversion or Introversion, Thinking or Feeling, Judging or Perceiving and Sensing or iNtuition), which are mandatory in order to determine the personality of each message. To calculate them, we use publicly available machine learning web services for text classification hosted in *uClassify*[10]. Among all the possibilities offered in this website, we focus on the Myers-Briggs functions developed by Mattias Östmar.

As the author explains, each function determines a certain dimension of the personality type according to Myers-Briggs personality model. The analysis is based on the writing style and should not be confused with the Myers-Briggs Type Indicator (MBTI) which determines personality type based on self-assessment questionnaires. Training texts are manually selected based on personality and writing style according to [31].

Those are the used functions:

- *Myers-Briggs Attitude:* Analyzes the Extroversion or Introversion dimension.
- *Myers-Briggs Judging Function:* Determines the Thinking or Feeling dimension.

[9] http://www.cs.cornell.edu/People/pabo/movie-review-data/
[10] https://www.uclassify.com

– *Myers-Briggs Lifestyle:* Determines the Judging or Perceiving dimension.
– *Myers-Briggs Perceiving Function:* Determines the Sensing or iNtuition dimension.

Each function returns a float within the range [0.0, 1.0] per each pair of characteristics of the dimension. For example, if we test a certain text and we obtain X value for Extroversion, the value for Introversion is 1-X. Thus, we only record one value per each function: Extroversion, Sensing, Thinking and Judging.

Those four values of each comment are added to the original dataset in order to create a new dataset. During the experiments, this new dataset is used in order to see the influence of the personality during the social media spam filtering. To do that, we apply the top ten classifiers mentioned previously to the original dataset and to the new one, and we compare the results.

3.5 Combining Sentiment Analysis and Personality Recognition

Finally, in order to demonstrate that the combination of different features can help in OSN spam filtering, we create a new dataset adding the polarity and the personality of each comment to the original dataset.

4 Experimental Results

4.1 Using Sentiment Analysis To Improve Social Media Spam Filtering

Descriptive Experiment To perform this experiment the sentiment analyzers identified in Section 3.3 are applied to the Youtube comments dataset in order to analyze the distribution of the comments in terms of polarity. The average of the obtained results are shown in the Figure 2.

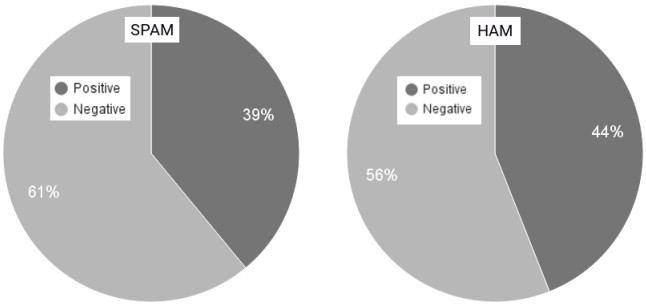

Fig. 2. Sentiment analysis of the original dataset.

Figure 2 shows that while in the previous studies such as [2] and [3], spam messages are more positive than legitimate messages, in this case, spam comments are more negative than legitimate comments.

Predictive experiments and comparison. In order to analyze the influence of the sentiment analysis in spam filtering, predictive experiments are carried out.

Then, we apply the best ten classifiers to the labeled datasets and we compare the obtained results with those obtained without polarity feature. The comparison between different results is presented in Tables 3. Tables show that sentiment analysis of the texts can help to improve the filtering results using an OSN dataset too. For instance, the best accuracy of the original dataset is improved from an 82.50% to an 82.53% using the polarity feature. Furthermore, the number of false positive are reduced in all the cases, reducing by 10% the original number in some cases (for example, from 89 to 70).

	Sentiment analyzer									
	None		*Adjective*		*Adjective+*		*TextBlob005*		*TextBlob01*	
Classifier #	**FP**	**Acc**	**FP**	**Acc**	**FP**	**Acc**	**FP**	**Acc**	**FP**	**Acc**
1	89	82.50	70	82.23	71	82.03	82	82.33	83	82.30
2	89	82.50	70	82.23	71	82.03	82	82.33	83	82.30
3	71	82.48	56	82.18	55	82.03	66	82.35	67	82.33
4	71	82.48	56	82.18	55	82.03	66	82.35	67	82.33
5	81	82.45	60	82.50	60	82.43	74	82.48	74	82.53
6	81	82.45	60	82.50	60	82.43	74	82.48	74	82.53
7	64	82.35	54	82.10	52	81.98	59	82.23	59	82.20
8	64	82.35	54	82.10	52	81.98	59	82.23	59	82.20
9	125	82.30	88	82.43	79	82.43	104	82.40	104	82.40
10	109	82.28	75	82.43	68	82.48	94	82.35	94	82.35

Table 3. Comparing original results with the results obtained using different sentiment classifiers.

4.2 Using Personality Recognition To Improve Social Media Spam Filtering

Descriptive Experiment Taking into account the personality recognition functions presented in Section 3.4, a descriptive analysis of the dataset is done. During this experiment, the different dimensions of the personality model are added to the original dataset, and a new dataset is created. The obtained results are shown in the Figure 3.

Although the differences between ham and spam comments are not significant, Figure 3 shows that the biggest difference is in terms of *thinking* feature. So in the next step, first of all a experiment using all the dimensions is carried out and after that, another test is also done adding only the *thinking* feature to the original dataset to analyze the difference.

Predictive Experiment and Comparison To analyze if personality recognition techniques help in OSNs spam filtering, on the one hand the best ten clas-

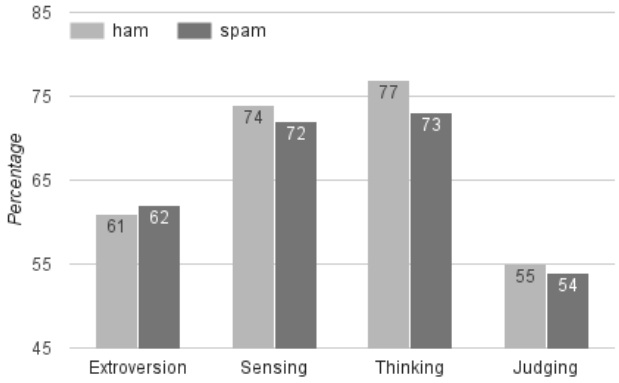

Fig. 3. Descriptive analysis in terms of personality recognition of the dataset.

sifiers identified in Section 3.2 are applied to the labeled dataset. On the other hand, taking into account the results obtained in the descriptive experiment, where we can see that the main difference between ham and spam comments is the *thinking* feature, the same experiment is carried out adding only this dimension to the original dataset. The results obtained during this experiment are presented also in Table 4.

	Original		*Personality*		*Thinking*	
Classifier #	**FP**	**Acc**	**FP**	**Acc**	**FP**	**Acc**
1	89	82.50	51	82.15	76	82.38
2	89	82.50	43	81.98	70	82.43
3	71	82.48	42	81.98	61	82.35
4	71	82.48	32	81.73	56	82.35
5	81	82.45	46	82.23	69	82.48
6	81	82.45	37	82.00	65	82.48
7	64	82.35	39	81.83	56	82.40
8	64	82.35	29	81.60	52	82.28
9	125	82.30	60	82.35	100	82.30
10	109	82.28	54	82.40	87	82.45

Table 4. Comparison of the best ten classifiers

In the first scenario (personality column), results show that while the number of false positive is reduced in every case, the accuracy is only improved in two out of ten cases.

Using only the most representative dimension (Thinking column), the accuracy is improved in more classifiers than in the previous column. The number of false positives is also reduced compared to the original dataset. Moreover, the best accuracy (82.50%) is not improved but the same percentage is obtained.

The significant reduction of the number of false positive give means to validate that personality recognition techniques help in OSNs spam filtering.

4.3 Combining Sentiment Analysis and Personality Recognition

Finally, to analyze if this new detection method could improve OSN spam filtering results, a new experiment is performed. The best ten classifiers are applied to the combined dataset, and a comparison of all the results is presented in Table 5.

	Used technique								
	None		Polarity		Personality		Comb		**FP reduction**
Classifier #	**FP**	**Acc**	**FP**	**Acc**	**FP**	**Acc**	**FP**	**Acc**	**(%)**
1	89	**82.50**	83	82.30	76	82.38	**71**	82.30	20.22
2	89	**82.50**	83	82.30	70	82.43	**66**	82.30	25.84
3	71	**82.48**	67	82.33	61	82.35	**57**	82.20	19.72
4	71	**82.48**	67	82.33	56	82.35	**51**	82.23	28.17
5	81	82.45	74	**82.53**	69	82.48	**60**	82.48	25.93
6	81	82.45	74	82.53	65	82.48	**53**	**82.55**	34.57
7	64	82.35	59	82.20	56	**82.40**	**51**	82.18	20.31
8	64	**82.35**	59	82.20	52	82.28	**46**	82.13	28.13
9	125	82.30	104	82.40	100	82.30	**84**	**82.50**	32.80
10	109	82.28	94	82.35	87	**82.45**	**75**	82.43	31.19

Table 5. Comparison of the best classifiers using the dataset of Youtube comments

Results demonstrate that the combination of different techniques improves spam filtering in both terms: accuracy and the number of false positive. The number of false positive is reduced in every case, and the best accuracy is obtained using the combined dataset (82.55%). The number of false positives is reduced by 26.69% on average.

5 Conclusions

This paper presents a new filtering method that gives the research community the opportunity of detecting non evident intent in spam. This new method consists of using polarity and personality features of each text and the combination of both. We added the features to an original dataset, and we carried out different experiments with and without the features.

As results reveal these techniques reduce the number of false positives in 26.69% (on average) and the best accuracy is improved (82.50% vs 82.55%). Despite the difference in percentage does not seem to be relevant, from 82.50% to 82.55%, if we take into account the amount of real spam traffic in OSNs, the improvement is significant. Results provided means to validate our hypothesis that it is possible to identify some insights of the intention of the texts using those techniques, and more spam texts are correctly classified.

Acknowledgments. This work has been developed by the intelligent systems for industrial systems group supported by the Department of Education, Language policy and Culture of the Basque Government. It has been partially funded by the Basque Department of Education, Language policy and Culture under the project SocialSPAM (PI 2014 1 102).

We thank Mattias Östmar for the valuable tools developed and published. And we thank Jon Kâgström (Founder of uClassify[11]) for the opportunity to use their API for research purposes.

References

1. Gao, H., Hu, J., Wilson, C., Li, Z., Chen, Y., Zhao, B.Y.: Detecting and characterizing social spam campaigns. In: Proceedings of the 17th ACM conference on Computer and communications security. CCS '10, New York, NY, USA, ACM (2010) 681–683
2. Ezpeleta, E., Zurutuza, U., Gómez Hidalgo, J.M.: Does sentiment analysis help in bayesian spam filtering? In: Hybrid Artificial Intelligent Systems: 11th International Conference, HAIS 2016, Sevilla, Spain, April 18-20, 2016, Springer (2016)
3. Ezpeleta, E., Zurutuza, U., Gómez Hidalgo, J.M.: Using personality recognition techniques to improve bayesian spam filtering. Journal Procesamiento del Lenguaje NaturalNatural (57) (2016)
4. Almaatouq, A., Shmueli, E., Nouh, M., Alabdulkareem, A., Singh, V.K., Alsaleh, M., Alarifi, A., Alfaris, A., Pentland, A.S.: If it looks like a spammer and behaves like a spammer, it must be a spammer: analysis and detection of microblogging spam accounts. International Journal of Information Security **15**(5) (2016) 475–491
5. Stringhini, G., Kruegel, C., Vigna, G.: Detecting spammers on social networks. In: Proceedings of the 26th Annual Computer Security Applications Conference. ACSAC '10, New York, NY, USA, ACM (2010) 1–9
6. Wang, D., Irani, D., Pu, C.: A social-spam detection framework. In: Proceedings of the 8th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference, ACM (2011) 46–54
7. Egele, M., Stringhini, G., Krgel, C., Vigna, G.: Compa: Detecting compromised accounts on social networks. In: NDSS, The Internet Society (2013)
8. Gao, H., Chen, Y., Lee, K., Palsetia, D., Choudhary, A.N.: Towards online spam filtering in social networks. In: NDSS, The Internet Society (2012)
9. Ezpeleta, E., Zurutuza, U., Hidalgo, J.M.G.: A study of the personalization of spam content using facebook public information. Logic Journal of the IGPL **25**(1) (2017) 30–41
10. Yang, C., Harkreader, R., Zhang, J., Shin, S., Gu, G.: Analyzing spammers' social networks for fun and profit: a case study of cyber criminal ecosystem on twitter. In: Proceedings of the 21st international conference on World Wide Web, ACM (2012) 71–80
11. Song, J., Lee, S., Kim, J.: Spam filtering in twitter using sender-receiver relationship. In: Recent Advances in Intrusion Detection, Springer (2011) 301–317
12. Wang, A.H.: Don't follow me: Spam detection in twitter. In: Security and Cryptography (SECRYPT), Proceedings of the 2010 International Conference on, IEEE (2010) 1–10

[11] https://www.uclassify.com

13. Zheng, X., Zeng, Z., Chen, Z., Yu, Y., Rong, C.: Detecting spammers on social networks. Neurocomputing **159** (2015) 27 – 34
14. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval **2**(1-2) (2008) 1–135
15. Liu, B., Zhang, L.: A survey of opinion mining and sentiment analysis. Mining Text Data (2012) 415–463
16. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up?: Sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10. EMNLP '02, Stroudsburg, PA, USA, Association for Computational Linguistics (2002) 79–86
17. Turney, P.D.: Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. ACL '02, Stroudsburg, PA, USA, Association for Computational Linguistics (2002) 417–424
18. Perveen, N., Missen, M.M.S., Rasool, Q., Akhtar, N.: Sentiment based twitter spam detection. International Journal of Advanced Computer Science and Applications(IJACSA) **7**(7) (2016) 568–573
19. Vinciarelli, A., Mohammadi, G.: A survey of personality computing. Affective Computing, IEEE Transactions on **5**(3) (2014) 273–291
20. Celli, F., Poesio, M.: PR2: A language independent unsupervised tool for personality recognition from text. CoRR **abs/1402.2796** (2014)
21. Briggs Myers, I., Myers, P.B.: Gifts differing: Understanding personality type (1980)
22. Costa, P.T., McCrae, R.R.: Normal personality assessment in clinical practice: The neo personality inventory. Psychological assessment **4**(1) (1992) 5
23. Mairesse, F., Walker, M.A., Mehl, M.R., Moore, R.K.: Using linguistic cues for the automatic recognition of personality in conversation and text. J. Artif. Int. Res. **30**(1) (November 2007) 457–500
24. Oberlander, J., Nowson, S.: Whose thumb is it anyway?: Classifying author personality from weblog text. In: Proceedings of the COLING/ACL on Main Conference Poster Sessions. COLING-ACL '06, Stroudsburg, PA, USA, Association for Computational Linguistics (2006) 627–634
25. Bai, S., Zhu, T., Cheng, L.: Big-five personality prediction based on user behaviors at social network sites. CoRR **abs/1204.4809** (2012)
26. Rangel, F., Celli, F., Rosso, P., Potthast, M., Stein, B., Daelemans, W.: Overview of the 3rd Author Profiling Task at PAN 2015. In: Working Notes Papers of the CLEF 2015 Evaluation Labs. CEUR Workshop Proceedings, CLEF and CEUR-WS.org (September 2015)
27. Shen, J., Brdiczka, O., Liu, J.: Understanding email writers: Personality prediction from email messages. In: User Modeling, Adaptation, and Personalization. Springer (2013) 318–330
28. Hernández Fusilier, D., Montes-y Gómez, M., Rosso, P., Guzmán Cabrera, R.: Detecting positive and negative deceptive opinions using pu-learning. Inf. Process. Manage. **51**(4) (July 2015) 433–443
29. Fornaciari, T., Celli, F., Poesio, M.: The effect of personality type on deceptive communication style. In: Intelligence and Security Informatics Conference (EISIC), 2013 European. (Aug 2013) 1–6
30. O'Callaghan, D., Harrigan, M., Carthy, J., Cunningham, P.: Network analysis of recurring youtube spam campaigns. CoRR **abs/1201.3783** (2012)
31. Jensen, G.H., DiTiberio, J.K.: Personality and the teaching of composition (1989)