Novel Email Spam Detection Method Using Sentiment Analysis and Personality Recognition

Enaitz Ezpeleta[1], Iñaki Velez de Mendizabal[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,ivelez,uzurutuza}@mondragon.edu,
[2] Pragsis Technologies
Manuel Tovar, 43-53, Fuencarral - 28034 Madrid, Spain
jmgomez@pragsis.com

Abstract. Unsolicited email campaigns remain as one of the biggest threats affecting millions of users per day. During the last years several techniques to detect unsolicited emails have been developed. This work provides means to validate the hypothesis that the identification of the email messages' intention can be approached by sentiment analysis and personality recognition techniques. These techniques will provide new features that improve current spam classification techniques. We combine personality recognition and sentiment analysis techniques to analyze email content. We enrich a publicly available dataset adding these features, first separately and after in combination, of each message to the dataset, creating new datasets. We apply several combinations of the best email spam classifiers and filters to each dataset in order to compare results.

Keywords: spam, polarity, personality, sentiment analysis

1 Introduction

The mass mailing of unsolicited emails has been a real threat for years. Spam campaigns have been used both for the sale of products as well as online fraud. Researchers are investigating many approaches that try to minimize this type of malicious activity that reports billions of dollars of benefits in an underground economy.

Within the spam problem, most research and products focus on improving spam classification and filtering. According to Kaspersky Lab data, the average percentage of spam in email traffic in Q2 2017 amounted to 56.97%[3]. This percentage is 2.77 percentage point higher than in Q3 2015[4], which demonstrates that spam is a current threat.

[3] https://securelist.com/spam-and-phishing-in-q2-2017/81537/

[4] https://cdn.securelist.com/files/2015/11/Q3-2015_Spam-report_final_EN.pdf

A similar study shows a dramatical increase of spam containing malicious attachments in Q1 of 2016[5]. This makes spam even more dangerous due to a gradual criminalization of it, confirmed by this growth. Several issues like social engineering, different types of attachments, diversity of languages take spam to a new level of danger.

To deal with this problem researchers started to design and develop different spam detection systems. Among others, spam filtering techniques are commonly used by both scientific and industrial communities.

This work provides means to validate the hypothesis that the identification of the messages' intention can be approached by sentiment analysis and personality recognition techniques. These techniques will provide new features that improve current spam classification techniques.

The remainder of the paper is organized as follows. Section 2 describes the previous work conducted in the area of spam filtering techniques, sentiment analysis and personality recognition. Section 3 describes the process of the experiments. In Section 4, the obtained results are described, comparing the results of the different datasets. Finally, we summarize our findings and give conclusions in Section 5.

2 Related Work

2.1 Spam filtering techniques

Different techniques to detect spam have been developed during the last years [1]. Among all proposed automatic classifying techniques, machine learning algorithms have achieved more success [2]. In [3] the authors obtained precisions up to 94.4% using those type of techniques.

In this study we focus on a specific section of machine learning algorithms; content-based filters. Those filters are based on analyzing the content of the emails in order to split messages in spam or legitimate emails as it is explained in [4]. Content-based spam filters can be separated in several types such as heuristic filtering, learning-based filtering and filtering by compression.

A comparison between various existing spam detection methods is presented in [5]: rule-based system, IP blacklist, Heuristic-based filters, Bayesian network-based filters, white list and DNS black holes. As a conclusion they define Bayesian based filters as the most effective, accurate, and reliable spam detection method.

Some of the content-based filtering techniques are also studied and analyzed in [6], and again, the Bayesian method is selected as the most effective one (classifying correctly the 96.5% of messages). Moreover, although several new approaches are obtaining good results [7–9], content-based filtering techniques still remain as one of the most efficient techniques. Furthermore, in [10] authors demonstrated that although more sophisticated methods have been implemented, Bayesian methods of text classification are still useful.

[5] https://securelist.com/analysis/quarterly-spam-reports/74682/spam-and-phishing-in-q1-2016/

2.2 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 [11]. As authors explain in [12], two main models to formalize personality have been defined: Myers-Briggs personality model [13], 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 [14] which divides the personality in 5 traits: Openness to experience, Conscientiousness, Extroversion, Agreeableness and Neuroticism.

As it is shown in [15] 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 [16], offline texts [15] or online social networks [17, 18] have been published.

In [19] 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 [20], and other researchers present the relationship between personality traits and deceptive communication [21].

2.3 Sentiment Analysis

As explained in [22], 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 [23]. 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 [22]. In order to classify such sentiment, some researchers use supervised learning techniques, where three classes are previously defined (positive, negative and neutral) [24]. Some other authors propose the use of unsupervised learning. In unsupervised learning techniques, opinion words or phrases are the dominating indicators for sentiment classification [25].

Moreover, authors in [26] demonstrate the possibility of using tweets sentiment analysis in order to improve spam filtering results in Twitter.

3 Proposed method: spam filtering using sentiment analysis and personality recognition

Taking as a baseline the previously presented studies [27, 28], the objective of this work is to validate the proposed method using two different datasets.

To do that, having an original dataset (CSDMC 2010 dataset): (1) we apply personality recognition technique to create a second dataset with this feature;

(2) we apply sentiment analysis classifiers to the original dataset and we add the obtained polarity, in order to create a third dataset; (3) we combine both techniques in the original messages and we create the fourth (combined) dataset; (4) having these four different datasets, we apply the best ten spam filtering classifiers identified in [27] to each dataset; (5) later, the top results of each dataset are analyzed; (6) finally we repeat the process using the validation dataset (TREC 2007). The full process is described in the Figure 1.

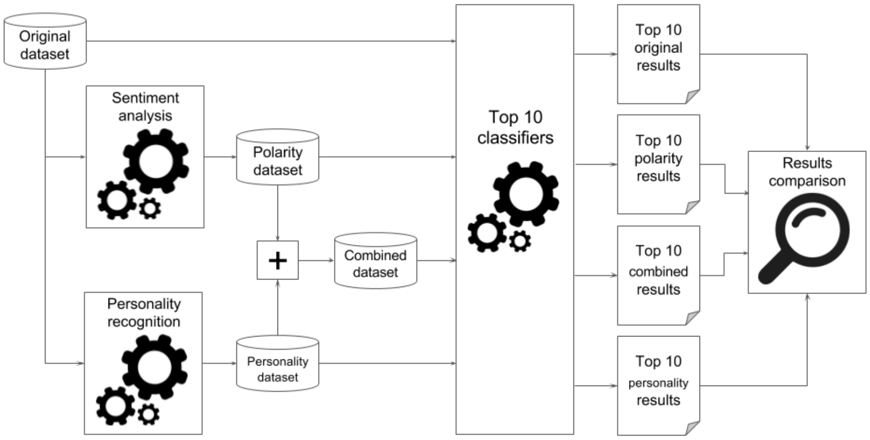

Fig. 1. Full process

During those experiments 10-fold cross-validation technique is used, and the results are analyzed in terms of number of false positive 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.

During this work publicly available dataset are used:

- *Movie Reviews*[6]: This dataset collects movie-review documents tagged in terms of polarity (positive or negative) or subjectivity rating. Also sentences are tagged with respect to their status or polarity. Among all these options the *polarity dataset v2.0* is used in this task, which is composed of 1,000 positive and 1,000 negative processed reviews introduced in [29]. This dataset is used to evaluate the effectiveness of each sentiment classifier.
- *CSDMC 2010 Spam Corpus*[7]: composed of 2,949 legitimate email messages and 1,378 spam. This dataset is used to carry out the original experiments.
- *TREC 2007 Public Corpus*[8]: This corpus contains 75,419 email messages: 25,220 ham (legitimate) and 50,199 spam emails. And we use it to repeat the

[6] http://www.cs.cornell.edu/People/pabo/movie-review-data/

[7] http://www.csmining.org/index.php/spam-email-datasets-.html

[8] http://plg.uwaterloo.ca/ gvcormac/treccorpus07/

experiment and to validate the results obtained using the previous dataset. In order to carry out the experiments using similar datasets in terms of email number, 4,000 emails are selected randomly (3,000 ham and 1,000 spam).

3.1 Spam filtering: baseline results

To analyze if our method improves Bayesian spam filtering, baseline results are needed in both cases: using the first dataset and using the validation dataset.

First dataset. In the first dataset, the best ten classifiers for spam filtering are identified taking into account the results obtained in [27]. These results are shown in the Table 1.

Table 1. Top10 Bayesian classifiers

#	**Name**	**TP**	**TN**	**FP**	**FN**	**Accuracy**
1	BLR.i.t.c.stwv.go.wtok	1,355	2,936	13	24	99.15
2	DMNBtext.c.stwv.go.wtok	1,362	2,928	21	17	99.12
3	DMNBtext.i.c.stwv.go.wtok	1,362	2,928	21	17	99.12
4	DMNBtext.i.t.c.stwv.go.wtok	1,362	2,928	21	17	99.12
5	DMNBtext.stwv.go.wtok	1,362	2,928	21	17	99.12
6	DMNBtext.c.stwv.go.stemmer	1,360	2,927	22	19	99.05
7	DMNBtext.i.c.stwv.go.stemmer	1,360	2,927	22	19	99.05
8	DMNBtext.i.t.c.stwv.go.stemmer	1,360	2,927	22	19	99.05
9	DMNBtext.stwv.go.stemmer	1,360	2,927	22	19	99.05
10	BLR.i.t.c.stwv.go.ngtok.stemmer.igain	1,351	2,935	14	28	99.03

During this paper, our main objective is to improve these results using the selected classifiers. To understand the settings of each classifier, Table 2 shows the nomenclatures used.

Table 2. Nomenclatures

	Meaning
BLR	Bayesian Logistic Regression
DMNBtext	Discriminative Multinomial Nave Bayes
.c	idft F, tft F, outwc T[9]
.i.c	idft T, tft F, outwc T[9]
.i.t.c	idft T, tft T, outwc T[9]

	Meaning
.stwv	String to Word Vector
.igain	Attribute selection using InfoGainAttributeEval
.wtok	Word Tokenizer
.ngtok	NGram Tokenizer 1-3
.stemmer	Stemmer
.go	General options

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

Second dataset. Being the main objective of this paper to validate the proposed method, we also used the previously presented *TREC2007* dataset. And the same ten classifiers that obtained the best results with the previous dataset are applied to this one in order to define the baseline results. The obtained results are shown in Table 3.

Table 3. Results of the best 10 classifiers applied to the validation dataset

#	Spam classifier	TP	TN	FP	FN	Acc
1	BLR.i.t.c.stwv.go.ngtok.stemmer.igain	976	2,983	17	24	98.98
2	DMNBtext.c.stwv.go.stemmer	979	2,979	21	21	98.95
3	DMNBtext.i.c.stwv.go.stemmer	979	2,979	21	21	98.95
4	DMNBtext.i.t.c.stwv.go.stemmer	979	2,979	21	21	98.95
5	DMNBtext.stwv.go.stemmer	979	2,979	21	21	98.95
6	DMNBtext.c.stwv.go.wtok	977	2,979	21	23	98.90
7	DMNBtext.i.c.stwv.go.wtok	977	2,979	21	23	98.90
8	DMNBtext.i.t.c.stwv.go.wtok	977	2,979	21	23	98.90
9	DMNBtext.stwv.go.wtok	977	2,979	21	23	98.90
10	BLR.i.t.c.stwv.go.wtok	972	2,978	22	28	98.75

3.2 Sentiment analysis

The objective of this phase is to carry out a sentiment classification of the dataset, in order to add the polarity of each message as a new feature for spam detection.

First a sentiment classifier is needed. So in this task two different options have been considered: (1) develop our own classifier or (2) use an existing one.

In order to design and implement our own sentiment classifier, sentiment dictionaries become useful tools. So the commonly used SentiWordNet has been chosen in this case. As shown in previous research works it is possible to obtain up to a 65% of accuracy using this dictionary [27].

SentiWordNet is a dictionary used to evaluate the polarity of a certain word depending on its grammatical properties. Using this tool, the average polarity of the email messages have been calculated.

Five sentiment classifiers have been developed with different settings. In order to evaluate *Adjectives*, *Adverbs*, *Verbs* and *Nouns*, in each classifier every word was considered to be a certain part of speech (depending on the name of the classifier), so we have obtained the polarity of those words that have that grammatical property. For instance: in the *Adjective* classifier every word was considered to be an adjective, so we have obtained the polarity of those words that can be considered as adjectives. And on the other hand, *AllPosition* classifier, which considers every part of speech per each word.

With the objective of comparing different results the existing classifier TextBlob has been used because it provides a simple API for diving into common NLP tasks. Specifically, giving a string the sentiment analyzer function returns a float value within the range [-1.0,1.0] for the polarity.

Once the classifiers have been defined, we improve the efficiency of those classifiers by changing settings and selection thresholds. For this work, a previously tagged dataset is mandatory. One commonly used dataset is called *Movie Reviews*[10]. This dataset collects movie-review documents tagged in terms of polarity (positive or negative) or subjectivity rating. Also sentences are tagged with respect to their status or polarity. Among all these options the *polarity dataset v2.0* is used in this task, which is composed by 1,000 positive and 1,000 negative processed reviews introduced in [29]. The objective is to obtain the best accuracy classifying those reviews to find the most efficient settings and thresholds. In this study the best sentiment analyzers identified in [27] are used.

3.3 Personality recognition

Following the procedure presented in [28], we use one of the most trusted personality model: Myers-Briggs personality model. This model is composed of 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. To calculate the dimensions of each text, we use publicly available machine learning web services for text classification hosted in *uClassify*[11]. 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 Jensen[30]. Those are the used functions:

- *Myers-Briggs Attitude:* Analyzes the Extroversion or Introversion dimension.
- *Myers-Briggs Judging Function:* Thinking or Feeling dimension.
- *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 SMS message 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 dimensions during the SMS 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.

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

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

3.4 Building the target dataset for validation

Being our objective to explore the possibilities to improve the spam filtering. We combine both techniques (sentiment analysis and personality recognition), we create a new dataset adding the personality dimensions and the polarity of each message to the original dataset. At the end, we apply the best ten spam filtering classifiers to compare all the results.

Finally, in the validation part, we repeat the same process but using a different dataset in order to compare the different results. Doing that, we ensure that the proposed method is valid regardless of the dataset.

4 Experimental Results

In this Section the results obtained during the previously explained experiments are shown. To carry out this experiment the *CSDMC 2010 Spam corpus* is used. Moreover, to validate the results, the other one: *TREC 2007*.

4.1 Descriptive analysis

During the data exploration part, following the processes presented in [27, 28], we apply sentiment analyzers and personality recognition techniques to the dataset, and the results presented in Figure 2 are extracted.

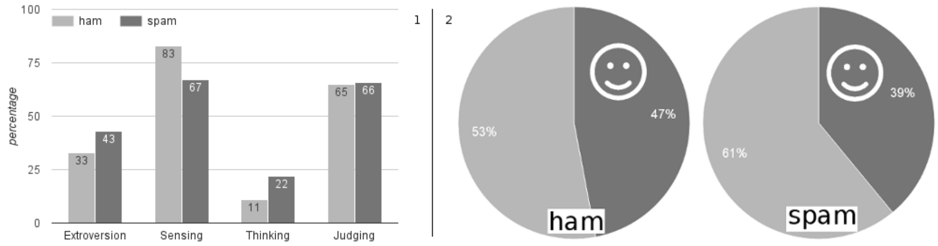

Fig. 2. Descriptive experiment of the dataset in terms of personality recognition (1) and sentiment analysis (2).

The personality dimensions of each message is extracted applying the previously explained personality recognition technique. In this point a new dataset is created by inserting the personality features extracted during the analysis. Finally the statistics about the personality dimensions in emails are calculated.

Results show that all the dimensions of the personality model have a different distribution depending on the text type. At this point we can confirm that the way emails are written varies. Furthermore, from the perspective of the effect of personality on deceptive communication the interesting thing is the difference in spam/ham messages with respect to the judging personality trait [21].

To analyze the polarity of the messages, the previously selected sentiment classifiers are used. Like in the personality part, the polarity extracted during the analysis is inserted in the dataset, creating three new datasets (one per each classifier).

Results from Figure 2 show that spam messages are mostly positive while ham messages are more negative. This means that there is a difference between spam and ham messages in terms of polarity, so it can be helpful for improving SMS spam filtering.

4.2 Evaluation and validation results

During this experiment we apply the best ten Bayesian classifiers to different datasets. These datasets are created applying the different techniques to the original *CSDMC 2010* dataset:

- Original dataset.
- Original dataset with the polarity information of each email. The best sentiment classifier identified in [27] is used to calculate the polarity score of each email.
- Original dataset with the *Sensing* feature taking into account the results published in [28].
- Original dataset with the polarity and the *Sensing* feature of each email (combining the two previous dataset) [28].

We compare the obtained results in terms of accuracy and false positive number, as it is possible to see in Table 4.

According to the obtained results, we can say that combining sentiment analysis techniques with personality recognition techniques the best result obtained in Bayesian spam filtering is improved in terms of accuracy. The combination improves (99.24% of accuracy) both the top result of the original dataset (99.15%) and the top result of the polarity analysis (99.21%). Moreover, in those cases where the best result is achieved, the combination of sentiment analysis and personality techniques reduces the false positive number.

To validate those first results, the same test is carried out using the *TREC2007* dataset and the obtained results are summarized in Table 5.

In this case, the best result of the original dataset is improved in the first step, using sentiment analysis. Later a better accuracy is obtained using personality detection techniques. Moreover, the combined dataset improves even more all the previous accuracies of each classifier, reaching 99.18% of accuracy, and validating the proposed method.

5 Conclusions

This paper presents a new filtering method that gives the research community the opportunity to detect non evident intent in spam emails. This new method consists in using a combination of the polarity feature (extracted applying sentiment analysis techniques) and the dimensions of Myers-Briggs personality model.

Table 4. Comparison of the best classifiers using the dataset CSDMC2010

	Used technique							
	None		*Sentiment analyzer*		*Personality (Sensing)*		*Combination*	
Spam classifier	**FP**	**Acc**	**FP**	**Acc**	**FP**	**Acc**	**FP**	**Acc**
BLR.i.t.c.stwv.go.wtok	13	**99.15**	14	99.12	15	99.03	15	99.03
DMNB.c.stwv.go.wtok	21	99.12	22	99.21	21	99.12	19	**99.24**
DMNB.i.c.stwv.go.wtok	21	99.12	22	99.21	21	99.12	19	**99.24**
DMNB.i.t.c.stwv.go.wtok	21	99.12	22	99.21	21	99.12	19	**99.24**
DMNB.stwv.go.wtok	21	99.12	22	99.21	21	99.12	19	**99.24**
DMNB.c.stwv. .go.stemmer	22	99.05	22	**99.15**	22	99.08	23	99.05
DMNB.i.c.stwv. .go.stemmer	22	99.05	22	**99.15**	22	99.08	23	99.05
DMNB.i.t.c.stwv. .go.stemmer	22	99.05	22	**99.15**	22	99.08	23	99.05
DMNB.stwv.go.stemmer	22	99.05	22	**99.15**	22	99.08	23	99.05
BLR.i.t.c.stwv.go. .ngtok.stemmer.igain	14	99.03	14	99.03	14	99.08	14	**99.10**

We added both features to the datasets, and we carried out the experiments with and without these features. With this combination we provided mechanisms to validate our hypothesis, that it is possible to identify some insights of the intention of the texts, and more spam texts are correctly classified.

Table 5. Comparison of the best classifiers using the dataset TREC2007

	Used technique							
	None		*Sentiment analyzer*		*Personality*		*Combination*	
Spam classifier	**FP**	**Acc**	**FP**	**Acc**	**FP**	**Acc**	**FP**	**Acc**
BLR.i.t.c.stwv.go. .ngtok.stemmer.igain	17	98.98	17	99.05	17	99.13	17	**99.18**
DMNB.c.stwv. .go.stemmer	21	98.95	20	99.05	22	98.98	21	**99.10**
DMNB.i.c.stwv. .go.stemmer	21	98.95	20	99.05	22	98.98	21	**99.10**
DMNB.i.t.c.stwv. .go.stemmer	21	98.95	20	99.05	22	98.98	21	**99.10**
DMNB.stwv.go.stemmer	21	98.95	20	99.05	22	98.98	21	**99.10**
DMNB.c.stwv.go.wtok	21	98.90	20	98.93	20	98.93	21	**98.95**
DMNB.i.c.stwv.go.wtok	21	98.90	20	98.93	20	98.93	21	**98.95**
DMNB.i.t.c.stwv.go.wtok	21	98.90	20	98.93	20	98.93	21	**98.95**
DMNB.stwv.go.wtok	21	98.90	20	98.93	20	98.93	21	**98.95**
BLR.i.t.c.stwv.go.wtok	22	98.75	22	98.68	21	98.80	22	**98.85**

Moreover, this method is validated in two different email datasets improving the best accuracy in both cases (from 99.15% to 99.24% and from 98.98% to 99.18%). Despite the difference in the percentage does not seem to be relevant, if we take into account the amount of real email traffic, the improvement is significant.

Furthermore, the same method have been applied to other types of spam. As Table 6 shows, the results are improved in all of them in terms of the best accuracy (BA), best 10 results or the number of false positives (FP). These results demonstrate that it is possible to improve spam detection applying sentiment analysis and personality recognition techniques regardless of the type of spam.

Table 6. Comparison of different spam types

	Polarity	**Personality**	**Combination**
Email	BA from 99.15% to 99.21%	9/10 results improved or equalized	BA from 99.15% to 99.24%
Email (Validation)	BA from 98.98% to 99.10%	BA from 98.98% to 99.13%	BA from 98.98% to 99.18%
SMS[31]	BA from 98.85% to 98.91%	BA from 98.85% to 98.94%	BA from 98.85% to 99.01%
SMS (Validation dataset) [32]	9/10 results improved or equalized	9/10 results improved or equalized	BA from 97.49% to 97.6%
Social Media[33]	- BA from 82.5% to 82.53% - Number of FP is reduced by 10% on average	- 5/10 results improved or equalized - Number of FP is reduced by 15% on average	- BA from 82.5% to 82.53% - Number of FP is reduced by 26% on average

Acknowledgments. This work 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 published. And we thank Jon Kâgström (uClassify[12]) for the opportunity to use their API for research purposes.

References

1. Saadat, N.: Survey on spam filtering techniques. Communications and Network (2011)
2. Cormack, G.V.: Email spam filtering: A systematic review. Foundations and Trends in Information Retrieval **1**(4) (2007) 335–455

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

3. Tretyakov, K.: Machine learning techniques in spam filtering. In: Data Mining Problem-oriented Seminar, MTAT. (2004) 60–79
4. Sanz, E.P., Hidalgo, J.M.G., Cortizo, J.C.: Email spam filtering. Advances in Computers (2008) 45–114
5. Savita Teli, S.B.: Effective spam detection method for email. In: International Conference on Advances in Engineering & Technology. (2014)
6. Malarvizhi, R.: Content-based spam filtering and detection algorithms-an efficient analysis & comparison 1. International Journal of Engineering Trends and Technology (IJETT) **4** (2013)
7. Li, L., Ren, W., Qin, B., Liu, T.: Learning document representation fordeceptive opinion spam detection. In Sun, M., Liu, Z., Zhang, M., Liu, Y., eds.: Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data, Cham, Springer International Publishing (2015) 393–404
8. He, H., Watson, T., Maple, C., Mehnen, J., Tiwari, A.: A new semantic attribute deep learning with a linguistic attribute hierarchy for spam detection. In: 2017 International Joint Conference on Neural Networks (IJCNN). (May 2017) 3862–3869
9. Bhowmick, A., Hazarika, S.M.: E-mail spam filtering: A review of techniques and trends. In Kalam, A., Das, S., Sharma, K., eds.: Advances in Electronics, Communication and Computing, Singapore, Springer Singapore (2018) 583–590
10. Eberhardt, J.J.: Bayesian spam detection. Scholarly Horizons: University of Minnesota, Morris Undergraduate Journal (2015)
11. Vinciarelli, A., Mohammadi, G.: A survey of personality computing. Affective Computing, IEEE Transactions on **5**(3) (2014) 273–291
12. Celli, F., Poesio, M.: PR2: A language independent unsupervised tool for personality recognition from text. CoRR **abs/1402.2796** (2014)
13. Briggs Myers, I., Myers, P.B.: Gifts differing: Understanding personality type (1980)
14. Costa, P.T., McCrae, R.R.: Normal personality assessment in clinical practice: The neo personality inventory. Psychological assessment **4**(1) (1992) 5
15. 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
16. 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
17. Bai, S., Zhu, T., Cheng, L.: Big-five personality prediction based on user behaviors at social network sites. CoRR **abs/1204.4809** (2012)
18. 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)
19. Shen, J., Brdiczka, O., Liu, J.: Understanding email writers: Personality prediction from email messages. In: User Modeling, Adaptation, and Personalization. Springer (2013) 318–330
20. 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

21. 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
22. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval **2**(1-2) (2008) 1–135
23. Liu, B., Zhang, L.: A survey of opinion mining and sentiment analysis. Mining Text Data (2012) 415–463
24. 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
25. 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
26. 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
27. 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)
28. Ezpeleta, E., Zurutuza, U., Gómez Hidalgo, J.M.: Using personality recognition techniques to improve bayesian spam filtering. Journal Procesamiento del Lenguaje NaturalNatural (57) (2016)
29. Pang, B., Lee, L.: A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In: Proceedings of the ACL. (2004)
30. Jensen, G.H., DiTiberio, J.K.: Personality and the teaching of composition (1989)
31. Ezpeleta, E., Zurutuza, U., Gómez Hidalgo, J.M. In: Short Messages Spam Filtering Using Sentiment Analysis. Springer International Publishing, Cham (2016) 142–153
32. Ezpeleta, E., Zurutuza, U., Hidalgo, J.M.G.: Short messages spam filtering using personality recognition. In: Proceedings of the 4th Spanish Conference on Information Retrieval. CERI '16, New York, NY, USA, ACM (2016) 7:1–7:7
33. Ezpeleta, E., Garitano, I., Arenaza-Nuo, I., Zurutuza, U., Hidalgo, J.M.G.: Novel comment spam filtering method on youtube: Sentiment analysis and personality recognition. In: Proceedings of Current Trends In Web Engineering - ICWE 2017 International Workshops, Springer International Publishing (2017)