WASSA 2013

4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis

Proceedings of the Workshop

14 June 2013
Atlanta, Georgia, U.S.A.

Endorsed by SIGNLL (ACL's Special Interest Group on Natural Language Learning)
Endorsed by SIGANN, the ACL Special Interest Group for Annotation
Sponsored by the Academic Institute for Research in Computer Science (Instituto Universitario de Investigación Informática), University of Alicante, Spain

209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-937284-47-3

Workshop Program

Friday, June 14, 2013

8:30–8:40 Opening Remarks

8:40–9:20 Invited talk: Prof. Dr. Rosalind Picard

Recent adventures with emotion-reading technology
Rosalind Picard

Session 1:Affect Recognition in Text (I)

9:20–9:45 *Bootstrapped Learning of Emotion Hashtags #hashtags4you*
Ashequl Qadir and Ellen Riloff

9:45–10:10 *Fine-Grained Emotion Recognition in Olympic Tweets Based on Human Computation*
Valentina Sintsova, Claudiu Musat and Pearl Pu

10:10–10:30 *Spanish DAL: A Spanish Dictionary of Affect in Language*
Matías Dell' Amerlina Ríos and Agustin Gravano

10:30–11:00 Break

11:00–11:40 Invited talk: Prof. Dr. Jonathan Gratch

Session 2: Affect Recognition in Text (II)

11:40–12:05 *The perfect solution for detecting sarcasm in tweets #not*
Christine Liebrecht, Florian Kunneman and Antal Van den Bosch

12:05–12:30 *Using PU-Learning to Detect Deceptive Opinion Spam*
Donato Hernández, Rafael Guzmán, Manuel Móntes y Gomez and Paolo Rosso

12:30–12:55 *Sexual predator detection in chats with chained classifiers*
Hugo Jair Escalante, Esaú Villatoro-Tello, Antonio Juárez, Manuel Montes-y-Gómez and Luis Villaseñor

12:55–14:00 Lunch Break

Using PU-Learning to Detect Deceptive Opinion Spam

Donato Hernández Fusilier[1,2]
Rafael Guzmán Cabrera
División de Ingenierías
Campus Irapuato-Salamanca.
[1]Universidad de Guanajuato
Mexico.
{donato,guzmanc}@ugto.mx

Manuel Montes-y-Gómez
Laboratorio de Tecnologías
del Lenguaje.
Instituto Nacional de
Astrofísica, Óptica y Electrónica.
Mexico.
mmontesg@inaoep.mx

Paolo Rosso
Natural Language
Engineering Lab., ELiRF.
[2]Universitat Politècnica de
València
Spain.
prosso@dsic.upv.es

Abstract

Nowadays a large number of opinion reviews are posted on the Web. Such reviews are a very important source of information for customers and companies. The former rely more than ever on online reviews to make their purchase decisions and the latter to respond promptly to their clients' expectations. Due to the economic importance of these reviews there is a growing trend to incorporate spam on such sites, and, as a consequence, to develop methods for opinion spam detection. In this paper we focus on the detection of *deceptive opinion spam*, which consists of fictitious opinions that have been deliberately written to sound authentic, in order to deceive the consumers. In particular we propose a method based on the PU-learning approach which learns only from a few positive examples and a set of unlabeled data. Evaluation results in a corpus of hotel reviews demonstrate the appropriateness of the proposed method for real applications since it reached a f-measure of 0.84 in the detection of deceptive opinions using only 100 positive examples for training.

1 Introduction

The Web is the greatest repository of digital information and communication platform ever invented. People around the world widely use it to interact with each other as well as to express opinions and feelings on different issues and topics. With the increasing availability of online review sites and blogs, costumers rely more than ever on online reviews to make their purchase decisions and businesses to respond promptly to their clients' expectations. It is not surprising that opinion mining technologies have been witnessed a great interest in recent years (Zhou et al., 2008; Mihalcea and Strapparava, 2009). Research in this field has been mainly oriented to problems such as opinion extraction (Liu B., 2012) and polarity classification (Reyes and Rosso., 2012). However, because of the current trend about the growing number of online reviews that are fake or paid by companies to promote their products or damage the reputation of competitors, the automatic detection of opinion spam has emerged as a highly relevant research topic (Jindal et al., 2010; Jindal and Liu, 2008; Lau et al., 2011; Wu et al., 2010; Ott et al., 2011; Sihong et al., 2012).

Detecting opinion spam is a very challenging problem since opinions expressed in the Web are typically short texts, written by unknown people using different styles and for different purposes. Opinion spam has many forms, e.g., fake reviews, fake comments, fake blogs, fake social network postings and deceptive texts. Opinion spam reviews may be detected by methods that seek for duplicate reviews (Jindal and Liu, 2008), however, this kind of opinion spam only represents a small percentage of the opinions from review sites. In this paper we focus on a potentially more insidious type of opinion spam, namely, *deceptive opinion spam*, which consists of fictitious opinions that have been deliberately written to sound authentic, in order to deceive the consumers.

The detection of deceptive opinion spam has been traditionally solved by means of supervised text classification techniques (Ott et al., 2011). These

Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pages 38–45,
Atlanta, Georgia, 14 June 2013.

techniques have demonstrated to be very robust if they are trained using large sets of labeled instances from both classes, deceptive opinions (positive instances) and truthful opinions (negative examples). Nevertheless, in real application scenarios it is very difficult to construct such large training sets and, moreover, it is almost impossible to determine the authenticity of the opinions (Mukherjee et al., 2011). In order to meet this restriction we propose a method that learns only from a few positive examples and a set of unlabeled data. In particular, we propose applying the PU-Learning approach (Liu et al., 2002; Liu et al., 2003) to detect deceptive opinion spam.

The evaluation of the proposed method was carried out using a corpus of hotel reviews under different training conditions. The results are encouraging; they show the appropriateness of the proposed method for being used in real opinion spam detection applications. It reached a f-measure of 0.84 in the detection of deceptive opinions using only 100 positive examples, greatly outperforming the effectiveness of the traditional supervised approach and the one-class SVM model.

The rest of the paper is organized as follows. Section 2 presents some related works in the field of opinion spam detection. Section 3 describes our adaptation of the PU-Learning approach to the task of opinion spam detection. Section 4 presents the experimental results and discusses its advantages and disadvantages. Finally, Section 5 indicates the contributions of the paper and provides some future work directions.

2 Related Work

The detection of spam in the Web has been mainly approached as a binary classification problem (spam vs. non-spam). It has been traditionally studied in the context of e-mail (Drucker et al., 2002), and web pages (Gyongyi et al., 2004; Ntoulas et al., 2006). The detection of opinion spam, i.e., the identification of fake reviews that try to deliberately mislead human readers, is just another face of the same problem (Lau et al., 2011). Nevertheless, the construction of automatic detection methods for this task is more complex than for the others since manually gathering labeled reviews –particularly truthful opinions– is very hard, if not impossible (Mukherjee et al., 2011).

One of the first works regarding the detection of opinion spam reviews was proposed by (Jindal and Liu, 2008). He proposed detecting opinion spam by identifying duplicate content. Although this method showed good precision in a review data set from Amazon[1], it has the disadvantage of under detecting original fake reviews. It is well known that spammers modify or paraphrase their own reviews to avoid being detected by automatic tools.

In (Wu et al., 2010), the authors present a method to detect hotels which are more likely to be involved in spamming. They proposed a number of criteria that might be indicative of suspicious reviews and evaluated alternative methods for integrating these criteria to produce a suspiciousness ranking. Their criteria mainly derive from characteristics of the network of reviewers and also from the impact and ratings of reviews. It is worth mentioning that they did not take advantage of reviews' content for their analysis.

Ott et al. (2011) constructed a classifier to distinguish between deceptive and truthful reviews. In order to train their classifier they considered certain types of near duplicates reviews as positive (deceptive) training data and *the rest* as the negative (truthful) training data. The review spam detection was done using different stylistic, syntactical and lexical features as well as using SVM as base classifier.

In a recent work, Sihong et al. (2012) demonstrated that a high correlation between the increase in the volume of (singleton) reviews and a sharp increase or decrease in the ratings is a clear signal that the rating is manipulated by possible spam reviews. Supported by this observation they proposed a spam detection method based on time series pattern discovery.

The method proposed in this paper is similar to Ott's et al. method in the sense that it also aims to automatically identify deceptive and truthful reviews. However, theirs shows a key problem: it depends on the availability of labeled negative instances which are difficult to obtain, and that causes traditional text classification techniques to be ineffective for real application scenarios. In contrast,

[1] http://www.Amazon.com

our method is specially suited for this application since it builds accurate two-class classifiers with only positive and unlabeled examples, but not negative examples. In particular we propose using the PU-Learning approach (Liu et al., 2002; Liu et al., 2003) for opinion spam detection. To the best of our knowledge this is the first time that this technique, or any one-class classification approach, has been applied to this task. In (Ferretti et al., 2012) PU-learning was successfully used in the task of Wikipedia flaw detection[2].

3 PU-Learning for opinion spam detection

PU-learning is a partially supervised classification technique. It is described as a two-step strategy which addresses the problem of building a two-class classifier with only positive and unlabeled examples (Liu et al., 2002; Liu et al., 2003; Zhang and Zuo, 2009). Broadly speaking this strategy consists of two main steps: i) to identify a set of reliable negative instances from the unlabeled set, and ii) to apply a learning algorithm on the refined training set to build a two-class classifier.

Figure 1 shows our adaptation of the PU-learning approach for the task of opinion spam detection. The proposed method is an iterative process with two steps. In the first step the whole unlabeled set is considered as the negative class. Then, we train a classifier using this set in conjunction with the set of positive examples. In the second step, this classifier is used to classify (automatically label) the unlabeled set. The instances from the unlabeled set classified as positive are eliminated; the rest of them are considered as the reliable negative instances for the next iteration. This iterative process is repeated until a stop criterion is reached. Finally, the latest built classifier is returned as the final classifier.

In order to clarify the construction of the opinion spam classifier, Algorithm 1 presents the formal description of the proposed method. In this algorithm P is the set of positive instances and U_i represents the unlabeled set at iteration i; U_1 is the original unlabeled set. C_i is used to represent the classifier that was built at iteration i, and W_i indicates the set of unlabeled instances classified as positive by the classifier C_i. These instances have to be removed from the training set for the next iteration. Therefore, the negative class for next iteration is defined as $U_i - W_i$. Line 4 of the algorithm shows the stop criterion that we used in our experiments, $|W_i| <= |W_{i-1}|$. The idea of this criterion is to allow a continue but gradual reduction of the negative instances.

1: $i \leftarrow 1$
2: $|W_0| \leftarrow |U_1|$
3: $|W_1| \leftarrow |U_1|$
4: **while** $|W_i| <= |W_{i-1}|$ **do**
5: $\quad C_i \leftarrow Generate\_Classifier(P, U_i)$
6: $\quad U_i^L \leftarrow C_i(U_i)$
7: $\quad W_i \leftarrow Extract\_Positives(U_i^L)$
8: $\quad U_{i+1} \leftarrow U_i - W_i$
9: $\quad i \leftarrow i + 1$
10: Return Classifier C_i

Algorithm 1: PU-Learning for opinion spam detection

4 Evaluation

4.1 Datasets

The evaluation of the proposed method was carried out using a dataset of reviews assembled by Ott et al. (2011). This corpus contains 800 opinions, 400 deceptive and 400 truthful opinions. These opinions are about the 20 most popular Chicago hotels; deceptive opinions were generated using the Amazon Mechanical Turk (AMT)[3], whereas –possible– truthful opinions were mined from a total of 6,977 reviews on TripAdvisor[4]. The following paragraphs show two opinions taken from (Ott et al., 2011). These examples are very interesting since they show the great complexity of the automatically –and even manually– detection of deceptive opinions. Both opinions are very similar and just minor details can help distinguishing one from the other. For example, in his research Ott et al. (2011) found that deceptive reviews used the words "experience", "my husband", "I", "feel", "business", and "vacation" more than genuine ones.

[2] http://www.webis.de/research/events/pan-12

[3] http://www.mturk.com

[4] http://www.tripadvisor.com

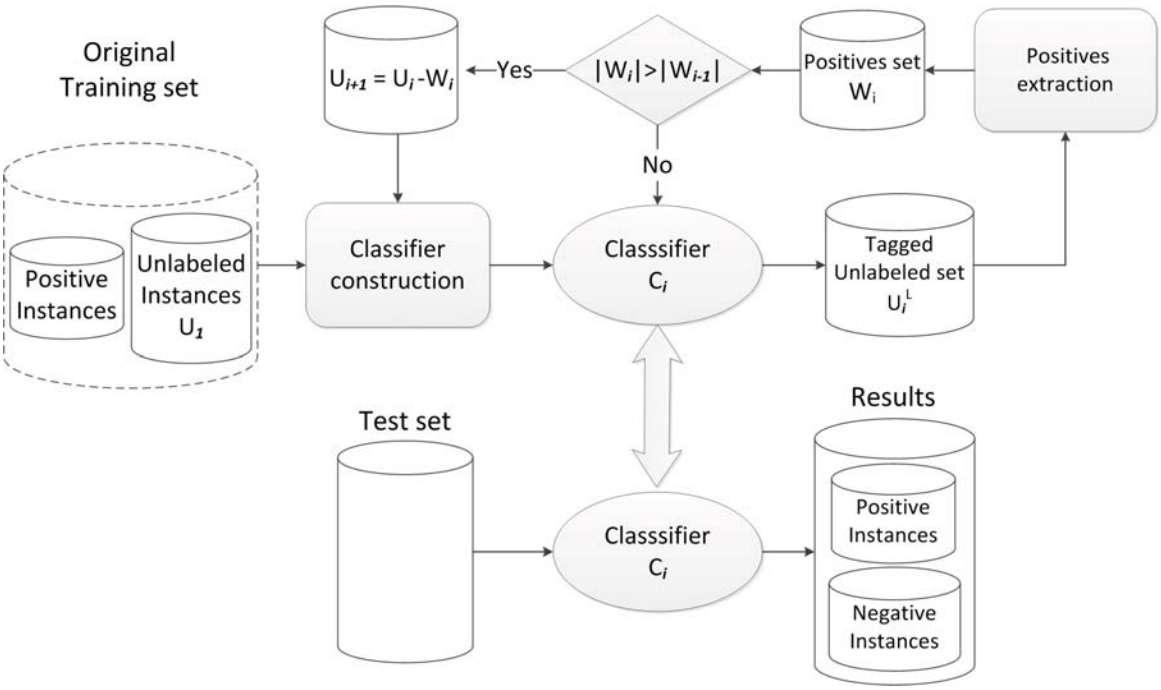

Figure 1: Classifier construction with PU-Learning approach.

Example of a truthful opinion

We stay at Hilton for 4 nights last march. It was a pleasant stay. We got a large room with 2 double beds and 2 bathrooms, The TV was Ok, a 27' CRT Flat Screen. The concierge was very friendly when we need. The room was very cleaned when we arrived, we ordered some pizzas from room service and the pizza was ok also. The main Hall is beautiful. The breakfast is charged, 20 dollars, kinda expensive. The internet access (WiFi) is charged, 13 dollars/day. Pros: Low rate price, huge rooms, close to attractions at Loop, close to metro station. Cons: Expensive breakfast, Internet access charged. Tip: When leaving the building, always use the Michigan Ave exit. It's a great view.

Example of a deceptive opinion

My husband and I stayed for two nights at the Hilton Chicago, and enjoyed every minute of it! The bedrooms are immaculate, and the linens are very soft. We also appreciated the free WiFi, as we could stay in touch with friends while staying in Chicago. The bathroom was quite spacious, and I loved the smell of the shampoo they provided-not like most hotel shampoos. Their service was amazing, and we absolutely loved the beautiful indoor pool. I would recommend staying here to anyone.

In order to simulated real scenarios to test our method we assembled several different sub-corpora from Ott's et al. (2011) dataset. First we randomly selected 80 deceptive opinions and 80 truthful opinions to build a fixed test set. The remaining 640 opinions were used to build six training sets of different sizes and distributions. They contain 20, 40, 60, 80, 100 and 120 positive instances (deceptive opinions) respectively. In all cases we used a set of 520 unlabeled instances containing a distribution of 320 truthful opinions and 200 deceptive opinions.

4.2 Evaluation Measure

The evaluation of the effectiveness of the proposed method was carried out by means of the f-measure. This measure is a linear combination of the precision and recall values. We computed this measure for both classes, deceptive and –possible– truthful opinions, nevertheless, the performance on the deceptive opinions is the only measure of real relevance. The f-measure for each opinion category O_i is defined as follows:

$$f - measure(O_i) = \frac{2 \times recall(O_i) \times precision(O_i)}{recall(O_i) + precision(O_i)} \quad (1)$$

$$recall(O_i) = \frac{number\ of\ correct\ predictions\ of\ O_i}{number\ of\ opinions\ of\ O_i} \quad (2)$$

$$precision(O_i) = \frac{number\ of\ correct\ predictions\ of\ O_i}{number\ of\ predictions\ as\ O_i} \quad (3)$$

4.3 Results

Tables 1 and 2 show the results from all the experiments we carried out. It is important to notice that we used Naïve Bayes and SVM classifiers as learning algorithms in our PU-learning method. These learning algorithms as well as the one-class implementation of SVM were also used to generated baseline results. In all the experiments we used the default implementations of these algorithms in the Weka experimental platform (Hall et al., 2009).

In order to make easy the analysis and discussion of the results we divided them in three groups: baseline results, one-class classification results, and PU-learning results. The following paragraphs describe these results.

Baseline results: The baseline results were obtained by training the NB and SVM classifiers using the unlabeled dataset as the negative class. This is a common approach to build binary classifiers in lack of negative instances. It also corresponds to the results of the first iteration of the proposed PU-learning based method. The rows named as "BASE NB" and "BASE SVM" show these results. They results clearly indicate the complexity of the task and the inadequacy of the traditional classification approach. The best f-measure in the deceptive opinion class (0.68) was obtained by the NB classifier when using 120 positive opinions for training. For the cases considering less number of training instances this approach generated very poor results. In addition we can also noticed that NB outperformed SVM in all cases.

One-class classification results: These results correspond to the application of the one-class SVM learning algorithm (Manevitz et al., 2002), which is a very robust approach for this kind of problems. This algorithm only uses the positive examples to build the classifier and does not take advantage of the available unlabeled instances. Its results are shown in the rows named as "ONE CLASS"; these results are very interesting since clearly show that this approach is very robust when there are only some examples of deceptive opinions (please refer to Table 1). On the contrary, it is also clear that this approach was outperformed by others, especially by our PU-learning based method, when more training data was available.

PU-Learning results: Rows labeled as "PU-LEA NB" and "PU-LEA SVM" show the results of the proposed method when the NB and SVM classifiers were used as base classifiers respectively. These results indicate that: i) the application of PU-learning improved baseline results in most of the cases, except when using 20 and 40 positive training instances; ii) PU-Learning results clearly outperformed the results from the one-class classifier when there were used more than 60 deceptive opinions for training; iii) results from "PU-LEA NB" were usually better than results from "PU-LEA SVM". It is also important to notice that both methods quickly converged, requiring less than seven iterations for all cases. In particular, "PU-LEA NB" took more iterations than "PU-LEA SVM", leading to greater reductions of the unlabeled sets, and, consequently, to a better identification of the subsets of reliable negative instances.

Finally, Figure 2 presents a summary of the best results obtained by each of the methods in all datasets. From this figure it is clear the advantage of the one-class SVM classifier when having only some examples of deceptive opinions for training, but also it is evident the advantage of the proposed method over the rest when having a considerable quantity of deceptive opinions for training. It is important to emphasize that the best result obtained by the proposed method (a F-meausre of 0.837 in the deceptive opinion class) is a very important results since it is comparable to the best result (0.89) reported for this collection/task, but when using 400 positive and 400 negative instances for training. Moreover, this result is also far better than the best human result obtained in this dataset, which, according to (Ott et al., 2011) it is around 60% of accuracy.

5 Conclusions and future work

In this paper we proposed a novel method for detecting deceptive opinion spam. This method adapts the PU-learning approach to this task. In contrast to traditional approaches that require large sets of labeled instances from both classes, deceptive and truthful

Original Training Set	Approach	Truthful			Deceptive			Iteration	Final Training Set
		P	R	F	P	R	F		
20-D 520-U	ONE CLASS	0.500	0.688	0.579	*0.500*	*0.313*	*0.385*		
	BASE NB	0.506	1.000	0.672	1.000	0.025	0.049		
	PU-LEA NB	0.506	1.000	0.672	1.000	0.025	0.049	5	20-D/493- U
	BASE SVM	0.500	1.000	0.667	0.000	0.000	0.000		
	PU-LEA SVM	0.500	1.000	0.667	0.000	0.000	0.000	4	20-D/518-U
40-D 520-U	ONE CLASS	0.520	0.650	0.578	*0.533*	*0.400*	*0.457*		
	BASE NB	0.517	0.975	0.675	0.778	0.088	0.157		
	PU-LEA NB	0.517	0.975	0.675	0.778	0.088	0.157	4	40-D/479-U
	BASE SVM	0.519	1.000	0.684	1.000	0.075	0.140		
	PU-LEA SVM	0.516	0.988	0.678	0.857	0.075	0.138	3	40-D/483-U
60-D 520-U	ONE CLASS	0.500	0.500	0.500	*0.500*	*0.500*	*0.500*		
	BASE NB	0.569	0.975	0.719	0.913	0.263	0.408		
	PU-LEA NB	0.574	0.975	0.722	0.917	0.275	0.423	3	60-D/449-U
	BASE SVM	0.510	0.938	0.661	0.615	0.100	0.172		
	PU-LEA SVM	0.517	0.950	0.670	0.692	0.113	0.194	3	60-D/450-U

Table 1: Comparison of the performance of different classifiers when using 20, 40 and 60 examples of deceptive opinions for training; in this table D refers to deceptive opinions and U to unlabeled opinions.

Original Training Set	Approach	Truthful			Deceptive			Iteration	Final Training Set
		P	R	F	P	R	F		
80-D 520-D	ONE CLASS	0.494	0.525	0.509	0.493	0.463	0.478		
	BASE NB	0.611	0.963	0.748	0.912	0.388	0.544		
	PU-LEA NB	0.615	0.938	0.743	*0.868*	*0.413*	*0.559*	6	80-D/267-U
	BASE SVM	0.543	0.938	0.688	0.773	0.213	0.333		
	PU-LEA SVM	0.561	0.925	0.698	0.786	0.275	0.407	3	80-D/426-U
100-D 520-U	ONE CLASS	0.482	0.513	0.497	0.480	0.450	0.465		
	BASE NB	0.623	0.950	0.752	0.895	0.425	0.576		
	PU-LEA NB	0.882	0.750	0.811	**0.783**	**0.900**	**0.837**	7	100-D/140-U
	BASE SVM	0.540	0.938	0.685	0.762	0.200	0.317		
	PU-LEA SVM	0.608	0.913	0.730	0.825	0.413	0.550	4	100-D/325-U
120-D 520-U	ONE CLASS	0.494	0.525	0.509	0.493	0.463	0.478		
	BASE NB	0.679	0.950	0.792	0.917	0.550	0.687		
	PU-LEA NB	0.708	0.850	0.773	*0.789*	*0.781*	*0.780*	5	120-D/203-U
	BASE SVM	0.581	0.938	0.718	0.839	0.325	0.468		
	PU-LEA SVM	0.615	0.738	0.670	0.672	0.538	0.597	6	120-D/169-U

Table 2: Comparison of the performance of different classifiers when using 80, 100 and 120 examples of deceptive opinions for training; in this table D refers to deceptive opinions and U to unlabeled opinions.

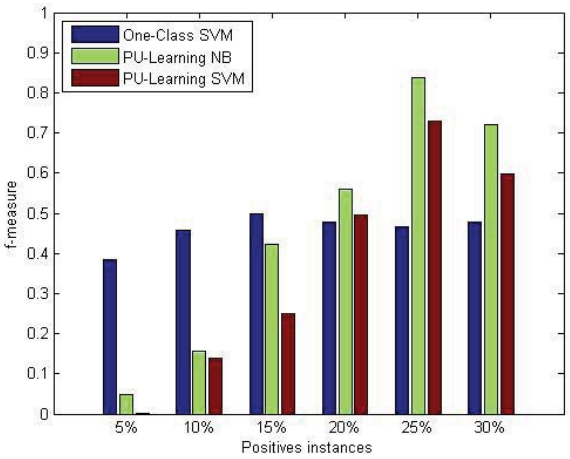

Figure 2: Summary of best F-measure results.

opinions, to build accurate classifiers, the proposed method only uses a small set of deceptive opinion examples and a set of unlabeled opinions. This characteristic represents a great advantage of our method over previous approaches since in real application scenarios it is very difficult to construct such large training sets and, moreover, it is almost impossible to determine the authenticity or truthfulness of the opinions.

The evaluation of the method in a set of hotel reviews indicated that the proposed method is very appropriate for the task of opinion spam detection. It achieved a F-meausre of 0.837 in the classification of deceptive opinions using only 100 positive examples and a bunch of unlabeled instances for training. This result is very relevant since it is comparable to previous results obtained by highly supervised methods in similar evaluation conditions.

Another important contribution of this work was the evaluation of a one-class classifier in this task. For the experimental results we can conclude that the usage of a one-class SVM classifier is very adequate for cases when there are only very few examples of deceptive opinions for training. In addition we could observe that this approach and the proposed method based on PU-learning are complementary. The one-class SVM classifier obtained the best results using less than 50 positive training examples, whereas the proposed method achieved the best results for the cases having more training examples.

As future work we plan to integrate the PU-learning and self-training approaches. Our idea is that iteratively adding some of the unlabeled instances into the original positive set may further improve the classification accuracy. We also plan to define and evaluate different stop criteria, and to apply this method in other related tasks such as email spam detection or phishing url detection.

Acknowledgments

This work is the result of the collaboration in the framework of the WIQEI IRSES project (Grant No. 269180) within the FP 7 Marie Curie. The work of the last author was in the framework the DIANA-APPLICATIONS-Finding Hidden Knowledge in Texts: Applications (TIN2012-38603-C02-01) project, and the VLC/CAMPUS Microcluster on Multimodal Interaction in Intelligent Systems.

References

H. Drucker, D. Wu and V.N. Vapnik. 2002. Support vector machines for spam categorization. Neural Networks, IEEE Transactions on, 10(5), pages 1048-1054.

Edgardo Ferretti, Donato Hernández Fusilier, Rafael Guzmán-Cabrera, Manuel Montes-y-Gómez, Marcelo Errecalde and Paolo Rosso. 2012. On the Use of PU Learning for Quality Flaw Prediction in Wikipedia. CLEF 2012 Evaluation Labs and Workshop, On line Working Notes, Rome, Italy, page 101.

Z. Gyongyi, H. Garcia-Molina and J. Pedersen. 2004. Combating web spam with trust rank. In Proceedings of the Thirtieth international conference on Very large data bases-Volume 30, pages 576-587. VLDB Endowment.

Hall Mark, Frank Eibe, Holmes Geoffrey, Pfahringer Bernhard, Reutemann Peter and Witten Ian H. 2009. The WEKA data mining software: an update. SIGKDD Explor. Newsl., pages 10-18. ACM.

N. Jindal and B. Liu. 2008. Opinion spam and analysis. In Proceedings of the international conference on Web search and web data mining, pages 219-230. ACM.

N. Jindal, B. Liu. and E. P. Lim. 2010. Finding unusual review patterns using unexpected rules. In CIKM, pages 219-230. ACM.

Raymond Y. K. Lau, S. Y. Liao, Ron Chi-Wai Kwok, Kaiquan Xu, Yunqing Xia and Yuefeng Li. 2011. Text mining and probabilistic modeling for online review spam detection. In Proceedings of the international

conference on Web search and web data mining, Volume 2 Issue 4,Article 25. pages 1-30. ACM.

E.P. Lim, V.A. Nguyen, N. Jindal, B. Liu, and H.W. Lauw. 2010. Detecting product review spammers using rating behaviors. In CIKM,pages 939-948. ACM.

B. Liu, Y. Dai, X.L. Li, W.S. Lee and Philip Y. 2002. Partially Supervised Classification of Text Documents Proceedings of the Nineteenth International Conference on Machine Learning (ICML-2002), Sydney, July 2002, pages 387-394.

B. Liu, Y. Dai, X.L. Li, W.S. Lee and Philip Y. 2003. Building Text Classifiers Using Positive and Unlabeled Examples ICDM-03, Melbourne, Florida, November 2003, pages 19-22.

B. Liu. 2012. Sentiment Analysis and Opinion Mining. Synthesis Lecture on Human Language Technologies Morgan & Claypool Publishers

Manevitz, Larry M. and Yousef, Malik 2002. One-class svms for document classification. J. Mach. Learn. Res.,January 2002, pages 139-154. JMLR.org.

R. Mihalcea and C. Strapparava. 2009. The lie detector: Explorations in the automatic recognition of deceptive language. In Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, pages 309-312. Association for Computational Linguistics.

Mukherjee Arjun, Liu Bing, Wang Junhui, Glance Natalie and Jindal Nitin. 2011. Detecting group review spam. Proceedings of the 20th international conference companion on World wide web, pages 93-94. ACM.

A. Ntoulas, M. Najork, M. Manasse and D. Fetterly. 2006. Detecting spam web pages through content analysis. Transactions on Management Information Systems (TMIS), pages 83-92. ACM.

Ott M., Choi Y., Cardie C. and Hancock J.T. 2011. Finding deceptive opinion spam by any stretch of the imagination. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, Oregon, USA, Association for Computational Linguistics (2011), pages 309-319.

Reyes A. and Rosso P. 2012. Making Objective Decisions from Subjective Data: Detecting Irony in Customers Reviews. In Journal on Decision Support Systems, vol. 53, issue 4 (Special Issue on Computational Approaches to Subjectivity and Sentiment Analysis), pages 754-760. DOI: 10.1016/j.dss.2012.05.027

Sihong Xie, Guan Wang, Shuyang Lin and Philip S. Yu. 2012. Review spam detection via time series pattern discovery. Proceedings of the 21st international conference companion on World Wide Web, pages 635-636. ACM.

G. Wu, D. Greene and P. Cunningham. 2010. Merging multiple criteria to identify suspicious reviews. RecSys10, pages 241-244. ACM.

Bangzuo Zhang and Wanli Zuo. 2009. Reliable Negative Extracting Based on KNN for Learning from Positive and Unlabeled Examples Journal of Computers, Vol. 4 No. 1., January, 2009, pages 94-101.

L. Zhou, Y. Sh and D. Zhang. 2008. A Statistical Language Modeling Approach to Online Deception Detection. IEEE Transactions on Knowledge and Data Engineering, 20(8), pages 1077-1081.