2018 4th IEEE International Conference on Information Management

Identification of Fake Reviews Using Semantic and Behavioral Features

Xinyue Wang, Xianguo Zhang, Chengzhi Jiang, Haihang Liu
College of Computer Science
Inner Mongolia University
Hohhot,China
e-mail: 2365434187@qq.com; 2595083628@qq.com; 384072007@qq.com; 1203738725@qq.com

***Abstract*—In recent years, online reviews have been playing an important role in making purchase decisions. This is because, these reviews can provide customers with large amounts of useful information about the goods or service. However, to promote factitiously or lower the quality of the products or services, spammers may forge and produce fake reviews. Due to such behavior of the spammers, customers would be misleaded and make wrong decisions. Thus detecting fake (spam) reviews is a significant problem. In this paper, we propose two types of features and apply supervised machine learning algorithms for performing classification on Yelp's real-life data. In terms of features used, there are two new semantic feature sets: readability features and topic features. Our results show that our proposed new features are more effective than n-gram features in detecting spam reviews. To improve classification on the real Yelp review data, we use a set of behavioral features about reviewers and their reviews for learning, which dramatically improves the classification result on real-life opinion spam data. For further improvement, we also ensure the number of reviewers instead of reviews is balanced.**

Keywords-fake reviews; semantic and behavioral features; supervised machine learning algorithms

I. INTRODUCTION

Nowadays with the advent of the e-commerce, an increasing number of people are taking pleasure of shopping online, and then sharing their opinions on the electronic business website. These online opinions may be used by customers and merchants when they make purchase and other decisions. For the online reviews, positive reviews play a stimulative role in reapping economic benefits and well-deserved reputation for merchants' businesses. Thus, it makes merchants have strong intentions to manoeuvre their fame and employ specialized imposters posting higher opinions on the shopping sites. Besides, there exists competition between online merchants. Consequently, the employed fraudsters may post negative opinions to defame their rivals, resulting in bad sales of products and services. Such individuals are called *opinion spammers,* and their behavior is called *opinion spamming* [1]. Because of the above activities, the online customers might be misleaded by the deceptive opinions. Therefore, opinion spam has attracted important attention from both business and research circles. And the research purposes mainly aim to identify fake reviews and let online opinions recover reliability and facticity.

Jindal and Liu [1] are the first researchers to study opinion spam by analysing the Amazon product reviews. And they provided three main types of spam reviews: false opinions, reviews on brands only, and non-reviews, along with several techniques to detect them. Based on their research, many other researchers attempt to explore different features of fake reviews and methods to efficiently solve the problem. There are three different dimensions are explored for the identification of fake reviews: detecting fake reviews [1,2,3], detecting fake reviewers [4,5], detecting fake reviewer groups [6,7]. According to the different dimensions, a variety of effective features such as textual features about content of reviews [8], behavioral features about reviewers [3,5], relational features (e.g., relationships among reviews, reviewers, and products) are proposed to detect spam reviews. In addition, there is also another category of methods to identify spam reviews, which includes text mining and Natural Language Processing (NLP) techniques. For example, n-grams methods [8,9], duplicity measure (e.g., cosine similarity), Term frequency-Inverse document frequency (Tf-Idf) measure [10], and so on are used in a number of literatures. In this paper, two angles of fake reviews and reviewers are involved in detecting fake reviews. From the angle of the reviews, we propose a new set of readability features (e.g., Automated Readability Index (ARI) and Coleman-Liau Index (CLI)), which are mainly evaluating the readability of every reviews. From another angle of reviewers, we presents a set of behavioral features, such as Restaurant Number (RN), Date Interval (DI). In addition to the above two types of features, n-gram features (e.g., unigrams and bigrams) based on the NLP techniques are also employed to classify reviews as fake or benign. After many observations and experiments on the real Yelp review data, we also present a new set of topic features (e.g., Average Topic Probabilistic (ATP) and Big Topic Probabilistic (BTP)). In general, the experimental results show that the performance of classification using topic features is a little better than that using n-gram features, however, behavior-based features is more robust.

The rest of this paper is organized as follows. Section 2 covers the relevant related work over the years. Section 3 introduces and analyses Yelp's real-life data. Section 4 presents and analysizes three types of semantic features and a set of behavioral features used in spam detection practices. In section 5, it shows the experimental results. Finally, we conclude this study in Section 6.

978-1-5386-6147-5/18/$31.00 ©2018 IEEE

II. Related Work

In recent years, a lot of research has been conducted in the field of opinion spam detection, in other words, various features and machine learning techniques are explored in most of the opinion spam researches. The relevant literatures are as follows.

Language and text analysis. Linguistic features are one of the key features for detecting opinion spam, and they help to summarize reviewers' texts in vector form. For example, bag-of-words technique is used by most of the works published, a similar approach is to use word n-grams instead of single words or unigrams [11] as used in the bag-of-words technique. Although n-grams show good results in detecting false opinions, POS tags can improve accuracies when combined with unigrams or bigrams. In [8], the researchers trained a SVM classifier on POS word features to distinguish between the writing styles of benign and malignant users. In order to gain a deeper understanding of the language used by spammers, some studies used clues from linguistics and psychology and applied LIWC (Language Query and Word Counting) as a tool to make a better understanding of deceptive text writing. The importance of the LIWC feature in language spoofing detection was proved by Ott [8], and the study showed that LIWC performed best with unigrams and bigrams on SVM classifiers. Besides the above methods, some researchers use semantic and stylistic clues to mark reviews as fake or real. Some studies, such as the research in [12], have defined semantic content overlapping to identify fake reviews with the help of WordNet lexical database. Moreover, Lai [13] used language modeling to find semantically identical terms. In conclusion, linguistic and textual features are useful to spot fake reviews.

Behavior analysis. Due to the difficulties of collecting more textual features, more and more research focuses on the behavioral activities of spammers. Studies in the literature show that the behavior of spammers is dfferent from real users. For example, spammers praise or downgrade a particular brand of product [2,14]; write a large number of reviews in a short time [4]; deviate from the majority in terms of ratings [5,15] and always provide extreme ratings (very high or very low) [6,16], duplicate ratings or content [5] and so on. Based on above activities of spammers, Mukherjee [17] proposed five behavioral characteristics. Through analysis, these features are quite discriminating. Chen [18] study the behavior of spammers on the web forum. They found that spammers write more posters than non-spammers and generally post fake reviews during working hours rather than during breaks. What's more, some investigators even detect fake reviews through spatial and temporal information [3].

III. Dataset

A. The Yelp Review Dataset

This section introduces the real-life data from Yelp. In order to ensure reviews posted on Yelp are believable, it takes a filtering algorithm to filter deceptive reviews. In recent years, Yelp's filtering algorithm has increasingly improved. Besides, Yelp is also confident enough to make its filtered reviews public. And its filter has also been claimed to be highly accurate by a study in BusinessWeek. Table 1 shows all the basic counts of reviews and reviewers.

TABLE I. Basic Counts For Reviews And Reviewers

Instance	Fake	Non-Fake	Fake Ratio	Total #
Review	8299	58588	14.1%	66887
Reviewer	7114	27988	25.4%	35102

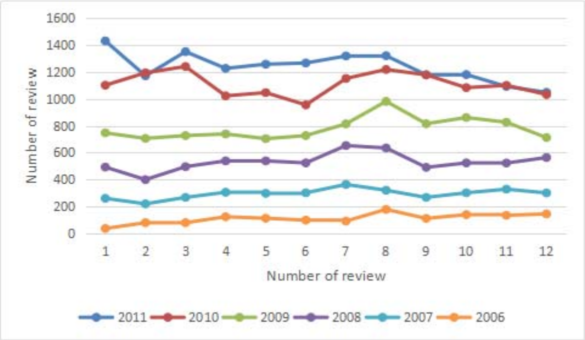

Figure 1. Number of non-fake reviews each month from 2006 to 2011.

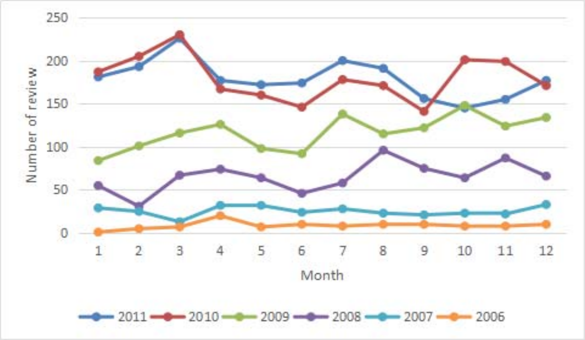

Figure 2. Number of fake reviews each month from 2006 to 2011.

By observing Table I, it can be seen that the category distribution of real Yelp data is skewed. As we all konw, highly imbalanced data usually produces bad models [19]. To build a good model for imbalanced data, under-sampling technnnique is used to randomly select a subset of instances from the majority class (Yelp's dataset labled non-fake) and combine it with the minority class (Yelp's dataset labled fake) to form a balanced dataset for model building.

B. Data Analysis

Fig. 1 shows the number of non-fake reviews in each month from January 2006 to December 2011, and Fig. 2 illustrates the number of fake reviews every month. Obviously, the number of non-fake reviews is relatively tending to be stable compared with the fake reviews from 2006 to 2011. Furthermore, we also can see that the number of fake reviews starts to increase significantly in June and November. It is well-known that shopping festival exists in the two month and more customers go shopping online during the two festivals, so producing more reivews after they received their packages. With the festival coming, spammers may be employed in producing more reviews to promote their own products or demote competitors products.

Besides, in each month, it is obvious that the number of non-fake reviews is about eight times larger than fake reviews.

IV. Feature Analysis

This section presents and describes features of fake reviews and reviewers used in our model to detect deceptive reviews in Yelp data. The feature sets of readability features, topic features and n-gram features are belong to semantic features. And the last feature set we provide consists of four behavioral features.

A. Readability Features

Many researchers use Automated Readability Index (ARI) and Coleman-Liau Index (CLI) to evaluate content quality and helpfulness of online reviews [20]. It is highly significant that consumers can understand the content of the review. The greater the readability of reviews' content, the better a customer feels. What's more, ARI and CLI indicate one's level of education, the lower the education, the higher the readability of the reviews' content. Next, ARI and CLI are defined as follows.

Automated Readability Index (ARI). Automated Readability Index is a good indicator of the readability of an English text. In order to calculate the ARI for a given review, we first calculated the total number of characters (excluding standard syntax such as hyphens and semicolons) and the total number of words. Then we need to calculate the review length. The following formula presents the Automated Readability Index:

$$ARI = 4.71 \times (\frac{characters}{words}) + 0.5 \times (\frac{words}{sentence}) - 21.43. \quad (1)$$

Coleman-Liau Index (CLI). The Coleman-Liau Index is similar to the Automated Readability Index, the only difference is that the second formula considers a more careful selection of the textual characteristics of the piece of text assessed. The following formula describes the Coleman–Liau index:

$$CLI = 5.89 \times (\frac{characters}{words}) + 0.5 \times (\frac{sentences}{word}) - 15.8. \quad (2)$$

We define AC by the following formula:

$$AC = \frac{ARI}{CLI}. \quad (3)$$

We first attempted to use ARI and CLI as two features to identify fake reviews. The performance is well in some sense. Experiments show that AC also have ability to make a contribution to the result of classification, and the performance is better than ARI and CLI. Then AC is used as a new feature. Thus we consider the above three indicators as topic features.

We use RF (Readability Features) to represent a feature set, including three features of ALI, CLI and AC.

B. Topic Features

LDA (Latent Dirichlet Allocation) model is a kind of topic model, it essentially examines the relationship between words and topics, topics and documents (text). Compared with the traditional LSI (Latent Semantic Indexing) model and PLSI (Probabilistic Latent Semantic Indexing) model, LDA model's hierarchy is more clear. And it also reduces the overfitting phenomenon, because as the number of texts increases, the bunber of topics will increase. So LDA model is more suitable to analysis large scale corpus.

The basic thought of LDA is that a text incudes a number of potential topics and each topic consists of a set number of specific words. Thus, we treat each review as a text, and we use the tool of Stanford Topic Modeling Toolbox0.4.0 to get the review-topic-distributions. Moreover, we also can obtain specific words related to each topic. Besides, we specify 10 topics for learning, which can minimize the model's perplexity. There are three new features explored by us.

Review Topic Distribution (RTD). We attempt to capture the latent differences among text contents. By applying a topic modeling tools based on LDA, the distribution of topic of each review is obtained, then we use the distributions as a new feature.

Average Topic Probabilistic (ATP). Based on the review-topic-distributions form (distribution of topic of each review), We first mark the number with a probability distribution greater than 0.1. Then we calculate the number of tags per line, and treat the number as a feature. It shows whether a review' topic distribution is balanced.

Big Topic Probabilistic (BTP). Similarly, We first mark the number with a probability distribution greater than 0.8. After that we calculate the number of tags per line. It indicates whether a review' topic is unique.

In Fig. 3, we plot the cumulative distribution function (CDF) of ATP and BTP. From Fig. 3(a), the maximum value of ATP can be 10, but the actual maximum is 7. And 45% of

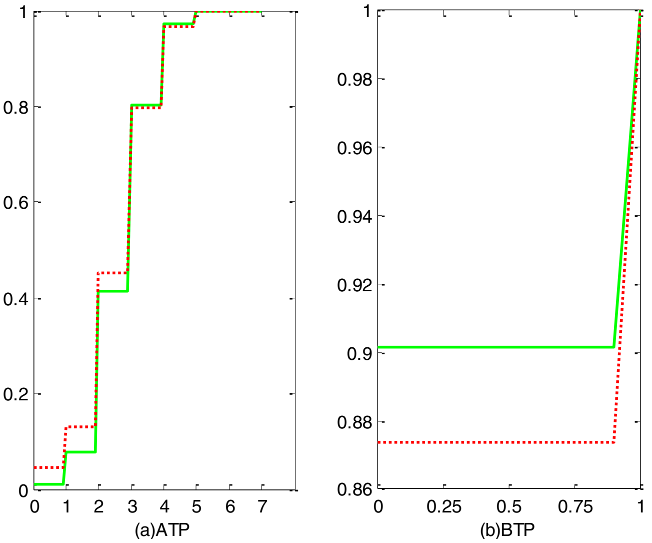

Figure 3. CDF (Cumulative Distribution Function) of topic features. Cumulative percentage of spammers (in red/dotted) and non-spammers (in green/solid) vs. topic feature value.

the fake reviews have less than 3 of feature value of ATP, however, the percentage of non-fake reviews is less than fake reviews, and the percentage is reversed when the feature value is greater than 3. This shows that the topic distribution of real reviews is more even than deceptive reviews. From Fig. 3(b), nearly 12.5% fake reviews with feature value 1, whereas true reviews with feature value 1 are only 10%. It shows that the topic of the fake reviews is more unique than the truthful reviews. Based on the above analysis, compared with the fake reviews, the topic distribution of non-fake reviews is more balanced.

We use TF (Topic Features) to represent a feature set, including three features of RTD, ATP and BTP.

C. N-gram Features

Ott [8] emphasizes the importance of bigrams (unigram and bigram), trigrams (unigram, bigram and trigram) and unigram features to classify reviews as false or benign. On that basis, we also explored the performance of bigram, trigram, and their combinations. Experimental results show that the feature of trigrams is more useful than other n-gram features.

We use NF (N-gram Features) to represent a feature set, including six features of unigram, bigram, trigram, bigrams, combination of bigram and trigram, trigrams.

D. Behavior Features

This part studies some behavioral features of fakers. For deeper analysis, we separate reviewers in our data (Table 1) into two groups, one is spammers (authors of fake reviews) group, the other is non-spammers (authors of non-fake reviews) group. The spammers post fake reviews, while the non-spammers product truthful reviews. We analyze the reviewers' profile with the following behavioral dimensions.

Percentage of Positive Reviews (PR). According to the research in [17], we use 4-5 star reviews as postive reviews. What's more, we plot the CDF of percentage of positive reviews in all reviews for spammers and non-spammers in Fig. 4(a). As can be seen, only 5% of the spammers have less than 80% of their reviews as positive, in other words, a majority (95%) of spammers rated higher than 90% of their reviews as positive. Compared with spammers, non-spammers show a rather evenly distributed trend where each kind of reviewers basically have different percentage of positive reviews. This is reasonable as in real-life, real reviewers generally have different rating levels.

Review Length (RL). Writing fake reviews requires some experience, so there is probably not much to write or a spammer may not want to spend too much time in writing. The CDF of the average number of words per review for all reviewers is presented in Fig. 4(b). A majority of spammers are bounded by 135 words in average review length which is a bit short as compared to non-spammers.

Restaurant Number (RN). In Fig. 4(c), we show the CDF of number of reviews for spammers and non-spammers. Most real reviewers (about 75%) are bounded by a restaurant number of 8. However, about 80% of spammers are bounded by 1 restaurant. The analysis shows that spammers only praise or downgrade a particular restaurant due to the cost of time and effort, or it may be because they maintain multiple accounts to avoid detection.

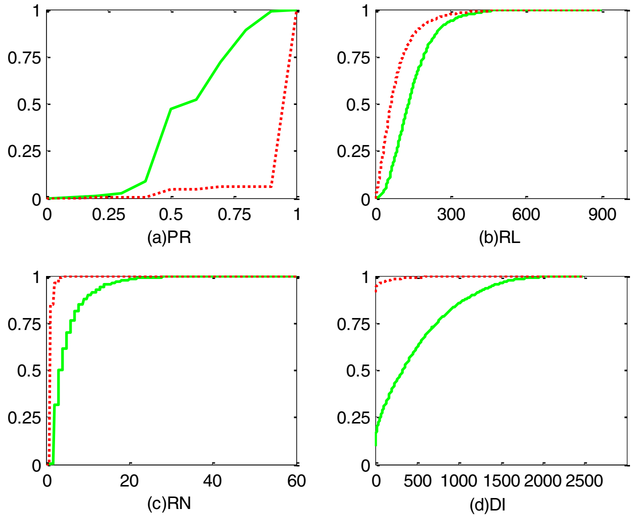

Figure 4. CDF (Cumulative Distribution Function) of behavioral features. Cumulative percentage of spammers (in red/dotted) and non-spammers (in green/solid) vs. behavioral feature value.

Date Interval (DI). We show the CDF of maximal time interval of all reviews writted by a same reviewer in Fig. 4(d). Only 8% of non-spammers have very little time gaps across the posting time of their reviews showing non-spammers always shop online each year. However, we find 90% of spammers are bounded by a time gaps of 0. This is both suspicious and abnormal.

We use BF (Behavioral Features) to represent a feature set, including four features of PR, RL, RN, DI.

V. Experiments And Results Discussion

Now we evaluate the effectiveness of the proposed model. We conduct experiments on a dataset of Yelp and report our findings.

A. Data

The initial dataset is comprised of 66887 reviews created by 35102 reviewers. To facilitate experiments, we sample this dataset to acquire a smaller and more easily evaluated dataset. In addition, to obtain a higher accurancy, we ensure the number of reviewers instead of reviews is basicly balanced by employing the technique of under-sampling. If we ensure that the number of reviews is balanced, the number of reviewers will be extremely uneven. Besides, we also manually removed some reviews considering the feature value should be meaningful. Our final dataset is consist of 33681 reviews and 10463 reviewers.

B. Classifier

Logistic Regression (LR), K-Nearest Neighbor (KNN), Naive Bayes (NB) and Support Vector Machine (SVM) classifiers are used.

C. Results

The usefulness of the proposed model is measured by the f-measure (F). F-measure is the harmonic mean of Precision

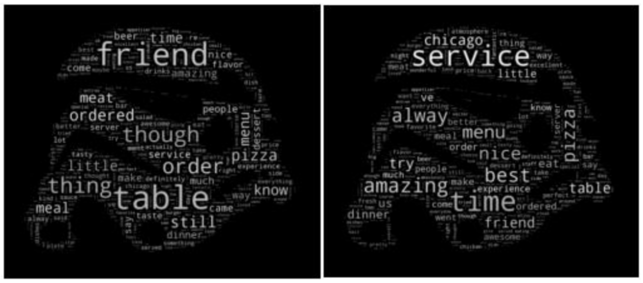

Figure 5. "non-fake" (left) and "fake" (right) words.

(P) and Recall (R) values. It is more intuitive than the arithmetic mean when calculate a mean of ratios.

Table II shows results for unigram, bigram, trigram and their combination with the classifiers of NB and SVM. On the whole, the classifier of SVM works better than NB, and SVM achieves the best recall score 0.833. Besides, for each of the n-gram features, experiments are carried out by language modeling techniques of Count and Tf-Idf measure. It can be seen that Tf-Idf measure performs better than Count. The experiment results indicate that the performance of bigrams with classifier of SVM is better.

Table III displays the performance of semantic and behavioral features. The worst feature is the readability feature set, followed by topic features, and behavioral features are the strongest features. The performance of using behavioral feature set alone is slightly higher than the performance of using all semantic and behavioral features. It demonstrates that the behavioral features are quite discriminating. Besides, although the new behavioral feature of DI is not better than PR and RN, it achieves a higher scores than RL and other sets of features. In some sense, other features also work well. Such as topic feature set, ATP and BTP are two new features and they perform equally. Besides, the results of readability features close to the n-gram features. Though feature of AC outperforms the ARI and CLI, it is slightly inferior than the combination of all readability features. In general, behavioral feature set is significantly better than semantic features. We also explored many other features, however, experimental results show that the features does not get better performance in this dataset.

Some features can provide accurate prediction. However, there may have some common words or writing habits in real reviews. Therefore, we determine to find which words play an important role in distinguishing between fake and real reviews. The left word cloud in Fig. 5 contains words with the top 1000 highest frequency and the font size of each word positively correlates to the frequency of the words. It is clear that the top frequent words are "friend" and "table", however, larger words in the right word cloud are "service" and "time". Based on the frequent words, we find that the most top words in the left wod cloud are more specific than the top words in the right word cloud. Moreover, we also find that the most genuine words represent the usual dining behavior that spammers are not likely to do. For example, we generally have lunch or dinner with our friends when we go to a restaurant or order take-outs. But spammers unlikely think of the word "friends" because of their purposeful reviews. The word of "service" is often used by all fake and non-fake reviewers, but in fact spammers didn't have dinner before they write reviews, thus they can only think of some common words, such as service. Though the real reviewers may also use the common words in their reviews, the frequent words are different from spammers' due to authentic dinner experience of non-spammers.

TABLE II. EXPERIMENTAL RESULTS USING NB AND SVM WITH N-GRAM FEATURES

Feature	NB						SVM					
	Count			Tf-Idf			Count			Tf-Idf		
	P	R	F	P	R	F	P	R	F	P	R	F
Unigram	0.795	0.821	0.799	0.846	0.811	0.726	0.767	0.761	0.764	0.796	0.820	0.801
Bigram	0.762	0.811	0.729	0.657	0.810	0.726	0.772	0.787	0.779	0.793	0.822	0.776
Trigram	0.738	0.804	0.742	0.657	0.810	0.726	0.757	0.807	0.760	0.774	0.814	0.748
Bigrams	0.752	0.810	0.726	0.657	0.810	0.726	0.791	0.797	0.794	0.808	0.832	0.801
Bigram+Trigram	0.772	0.811	0.730	0.657	0.810	0.726	0.779	0.805	0.787	0.793	0.821	0.770
Trigrams	0.733	0.810	0.726	0.657	0.810	0.726	0.783	0.805	0.798	0.811	0.833	0.800

TABLE III. EXPERIMENTAL RESULTS USING LR, KNN AND SVM WITH DIFFERENT FEATURES

Feature			LR			KNN			SVM		
			P	R	F	P	R	F	P	R	F
Semantic Analysis	**Readability Features**	**ARI**	0.657	0.810	0.726	0.702	0.759	0.725	0.657	0.810	0.726
		CLI	0.768	0.811	0.729	0.702	0.782	0.729	0.758	0.811	0.727
		AC	0.793	0.811	0.727	0.709	0.784	0.733	0.752	0.810	0.726
		RF	0.775	0.811	0.729	0.718	0.779	0.738	0.758	0.811	0.727
	Topic Features	**RTD**	0.791	0.822	0.786	0.835	0.848	0.838	0.761	0.811	0.744
		ATP	0.657	0.810	0.726	0.754	0.810	0.739	0.657	0.810	0.726
		BTP	0.657	0.810	0.726	0.657	0.810	0.726	0.657	0.810	0.726
		TF	0.807	0.829	0.788	0.831	0.846	0.834	0.761	0.811	0.744
	RF+TF		0.811	0.831	0.791	0.827	0.844	0.828	0.779	0.815	0.750
Behavior Features		**PR**	0.964	0.964	0.963	0.974	0.973	0.972	0.966	0.966	0.966
		RL	0.833	0.832	0.779	0.762	0.785	0.771	0.657	0.810	0.726
		RN	0.947	0.944	0.940	0.947	0.944	0.940	0.909	0.883	0.890
		DI	0.916	0.897	0.902	0.916	0.909	0.911	0.916	0.899	0.904
		BF	0.973	0.972	0.971	0.976	0.976	0.976	0.973	0.972	0.972
All			0.972	0.972	0.971	0.970	0.970	0.969	0.971	0.971	0.970

VI. Conclusion

In this paper, we provide two types of features, one is behavioral feature set, the other is related to semantic. We propose three feature sets by performing semantic analysis. Our study shows that although behavioral features reported very high detection accuracy with supervised classifiers, semantic features are also work well on real-life fake reviews in the commercial setting of Yelp.com. Among the semantic features, readability features, topic features and n-gram features provide approximately the same results in performance. For n-gram features, the feature vectors were constructed using Count and Tf-Idf values of the review content and classifiers of SVM and NB are used for classification process. Overall, the classifier of LR performed the best, which achieves up to 97.2% accuracy with all features.

Though the experiment result is relatively well, other aspects features are need to explore or we can develop better methods. Future research can related to improve the existed algorithm which can be used to detect fake reviews more specifically and effectively. This research will continue, for achieving even better performance using the above mentioned improvement guidelines.

Reference

[1] N. Jindal and B. Liu, "Analyzing and detecting review spam," International Conference on Web Search and Data Mining, 2007, pp. 547-552.

[2] N. Jindal and B. Liu, "Opinion spam and analysis," International Conference on Web Search and Data Mining, 2008, pp. 219-230.

[3] S. Xie, G. Wang, S. Lin, and P. S. Yu, "Review spam detection via temporal pattern discovery," ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, 2012, pp. 823-831.

[4] A. Mukherjee, A. Kumar, B. Liu, J. Wang, M. Hsu, M. Castellano, and R. Ghosh, "Spotting opinion spammers using behavioral footprints," ACM sigkdd International Conference on Knowledge Discovery and Data Mining, 2013, pp. 632-640.

[5] E. P. Lim, V. A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, "Detecting product review spammers using rating behaviors," ACM International Conference on Information and Knowledge Management, 2010, pp. 939-948.

[6] A. Mukherjee, B. Liu, and N. Glance, "Spotting fake reviewer groups in consumer reviews," International Conference on World Wide Web, 2012, pp. 191-200.

[7] A. Mukherjee, B. Liu, J. Wang, N. Glance, and N. Jindal, "Detecting group review spam," International Conference Companion on World Wide Web, 2011, pp. 93-94.

[8] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, "Finding deceptive opinion spam by any stretch of the imagination," Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 2011, pp. 309-319.

[9] M. Ott, C. Cardie, and J. T. Hancock, "Negative deceptive opinion spam," In Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2013, pp. 497-501.

[10] H. A. Najada and X. Zhu, "iSRD: Spam review detection with imbalanced data distributions," IEEE International Conference on Information Reuse and Integration , 2014, pp. 553-560.

[11] J. Li, M. Ott, C. Cardie, and E. Hovy, "Towards a general rule for identifying deceptive opinion spam," Meeting of the Association for Computational Linguistics, 2014, pp. 1566-1576.

[12] R. Y. K. Lau, S. Y. Liao, R. C. W. Kwok, K. Xu, Y. Xia, and Y. Li, "Text mining and probabilistic language modeling for online review spam detecting," ACM Transactions on Management Information Systems, vol. 2, Dec. 2011, pp. 1-30, doi:10.1145/2070710.2070716.

[13] C. L. Lai, K. Q. Xu, R. Y. K. Lau, Y. Li, and D. Song, "High-order concept associations mining and inferential language modeling for online review spam detection," IEEE International Conference on Data MiningWorkshop, 2010, pp. 1120‐1127.

[14] N. Jindal, B. Liu, and E. P. Lim, "Finding unusual review patterns using unexpected rules", ACM International Conference on Information and Knowledge Management, 2010, pp. 1549-1552.

[15] F. Li, M. Huang, Y. Yang, and X. Zhu, "Learning to identify review spam," International Joint Conference on Articial Intelligence, 2011, pp. 2488-2493.

[16] F. Song, L. Xing, A. Gogar, and Y. Choi, "Distributional footprints of deceptive product reviews," International AAAI Conference on Weblogs and Social Media , 2012, pp. 98-105.

[17] A. Mukherjee, V. Venkataraman, B. Liu, and N. Glance, "What Yelp Fake Review Filter Might Be Doing?," In Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media, 2013, pp. 409-418.

[18] Y. R. Chen and H. H. Chen, "Opinion Spam Detection in Web Forum: A Real Case Study," International Conference, 2015, pp. 173-183.

[19] N. V. Chawla, N. Japkowicz, and A. Kotcz, "Editorial: Special issue on learning from imbalanced data sets," Acm Sigkdd Explorations Newsletter, vol. 6, 2004 , 6 (1) :1-6, june. 2004, pp. 1-6, doi: 10.1145/1007730.1007733.

[20] N. Kofiatis, E. Garcia-Bariocanal, and S. S ά nchez-Alonso, "Evaluating content quality and helpfulness of online product review: The interplay of review helpfulness vs. review content," Electronic Commerce Research & Applications, vol. 11, May. 2012, pp. 205-217, doi:10.1016/j.elerap.2011.10.003.