Detecting Spamming Reviews Using Long Short-term Memory Recurrent Neural Network Framework

Chih-Chien Wang
National Taipei University
New Taipei, Taiwan
wangson@mail.ntpu.edu.tw

Min-Yuh Day
Tamkang University
New Taipei, Taiwan
myday@mail.tku.edu.tw

Chien-Chang Chen
Tamkang University
New Taipei, Taiwan
ccchen34@mail.tku.edu.tw

Jia-Wei Liou
National Taipei University
New Taipei, Taiwan
bgga1013344@gmail.com

ABSTRACT

Some unethical companies may hire workers (fake review spammers) to write reviews to influence consumers' purchasing decisions. However, it is not easy for consumers to distinguish real reviews posted by ordinary users or fake reviews post by fake review spammers. In this current study, we attempt to use Long Short-Term Memory (LSTM) Recurrent Neural Network (RNN) framework to detect spammers. In the current, we used a real case of fake review in Taiwan, and compared the analytical results of the current study with results of previous literature. We found that the LSTM method was more effective than Support Vector Machine (SVM) for detecting fake reviews. We concluded that deep learning could be use to detect fake reviews.

CCS Concepts

• **Computing methodologies → Machine learning → Machine learning approach → Neural networks**

Keywords

Fake Review; Deep Learning; Neural Network; Long Short-term Memory (LSTM); Recurrent Neural Network (RNN).

1. INTRODUCTION

Word-of-mouths play an essential role during consumers' purchase decision process. Before purchasing, consumers may consult word-of-mouths from the Internet to reduce the risk of decision making. Positive word-of-mouths will raise the level of attention to the product and build a positive image in the mind of consumers.

After purchasing or consuming product, some consumer will share their consumption experience by providing product reviews. When the consumption experience is higher than the expectation, consumers will generate positive word-of-mouths. However, if the consumption experience is lower than expected, they will generate negative word-of-mouth[1]. To earn reputation from consumers and obtain posititve word-of-mouths, companies should improve the quality of their product and service. However, some unethical companies may find that writing positive fake reviews is a short cut to "earn" reputation. They may hire workers to write fake reviews to mislead consumers.

Machine learning is common used to detect one thing from the others. Recently, the low cost GPU computing power makes deep learning a feasible analysis approach. Therefore, this study will use deep learning to detect fake reviews.

The current study using Long Short-term Memory (LSTM) that is a kind of neural network to detect fake reviews. This paper aims to compare the effectiveness of LSTM, a deep learning method, and Suppor Vector Machine (SVM), a traditional machine learning method, in detecting fake reviews.

2. LITERATURE REVIEWS

Fake reviews are posted deliberately to mislead consumers [2]. Fake reviews come in many forms and are hard to identify. Followings are literature that focused on detecting fake product review.

2.1 Supervised, Unsupervised, and Semi-supervised Learning

Spam detection approaches can be divided into supervised learning techniques, unsupervised learning techniques, and semi-supervised learning techniques methods.

Supervised learning refers to the use of labeled data. Before conducting supervised learning, researchers have to label reviews as fake or non-fake reviews. However, it is not easy to label fake reviews. Thus, literature had developed several approach to labeled fake reviews. Content repeatability can be used as a cue to detect and label spam or fake review[2]. Hiring workers to write fake reviews is anthor way to lable fake reviews. Ott et al. [3] hired workers from Amazon's Mechanical Turk to write fake reviews. Algur et al. [4] used cosine similarity on product feature and considered the repeated or close to repeated review as a fake review. When labeled fake reviews are available, researchers can use supervised learning to detect fake review. For examples, Jindal and Liu used logistic regression to detect spam. Chen and Chen [5] used SVM as a classifier to detect fake reviews. Chen and Chen adopted some features, like content characteristics, post time and sentiment on brands. Wang et al. [6] used SVM classifier and social network analysis to detect spam. The literature

ICEEG'18, June 13-15, 2018, Hong Kong, Hong Kong

ACM ISBN 978-1-4503-6490-4/18/06 …$15.00.
DOI: http://doi.org/10.1145/3234781.3234794

mentioned above are all using supervised learning methods but using different features and classifiers.

Unsupervised learning techniques refer to the use of unlabeled data. Lau et al. [7] collected Amazon review dataset and analyze content semantic by WordNet, used cluster analysis to classify reviews.

Semi-supervised learning techniques are between supervised learning and unsupervised learning. Semi-supervised learning uses a small amount of labeled data, and a large number of unlabeled data for training and classification. Ren et al. [8] used Ott's hotel review dataset [3], by some truthful reviews and lots of unlabeled reviews to build a classifier and identify deceptive reviews.

2.2 Deep Learning and Neural Network

Deep learning, also known as Deep Neural Network, was proposed by Hinton and Salakhutdinov [9]. Deep learning allows computers to make predictions more effective through in-depth learning. Deep learning is a deep (multilevel) neural network. There are some frequently used models for deep learning, such as convolution neural network (CNN), recurrent neural network (RNN) and long short-term memory (LSTM). This work adopts the LSTM method.

RNN is expert in dealing with sequential issues, such as natural language processing. As time continues to pass, however, RNN will gradually lose the ability to learn long distance memories called gradient vanishing or gradient exploding [10]. To solve this issue of RNN, Hochreiter and Schmidhuber [11] put forward LSTM architecture which a kind of RNN. LSTM has three more controllers than the RNN: Input gate, forget gate, and output gate. Thus, LSTM can remember long-term memory and is better than RNN.

There is an issue of over fitting in the training model of deep learning. The solutions to the over fitting issue include adding data sets, early stopping, normalization, and adding dropout. This study used dropout [12] which give up neurons randomly during the training process. The drop out neurons will be skipped in training, but may work in the next round. When adding dropout, results of the training model may be better than without dropout.

3. METHODOLOGY

3.1 Data Corpus

There are several ways to obtain fake reviews: Marking fake reviews manually, treating reviews with similar contents as fake, hired writers to write fake reviews, and use a ground truth of real case of fake reviews. Obtaining a real case of fake review as ground truth is the most difficult one. Nevertheless, this study got an opportunity to get the real case of fake reviews.

We used a real case of fake review in Taiwan as source of fake reviews. In this case, a Korean based international mobile phone company's Taiwan branch company hired full-time or part-time staffs to write fake reviews in an attempt to enhance the brand image and influence purchase decision of Taiwanese consumers. In April 2013, a hacker posted several internal confidential documents which mentioned details about how the company hired workers to write fake reviews and the detailed list of fake reviews and spammers. The company recognized this fact. This case became the first confirmed case of fake review spamming, and the company was fined by Fair Trade Commission, Taiwan. Based on the leaked confidential documents, we crawled the data from Mobile01.com, a large product review website in Taiwan, to get the fake reviews as well as regular reviews. We also collect list of reviews authors and distinguish them as spammers and general users. Review authors were marked as spammers when they were hired workers that mentioned in internal confidential documents. Review authors were marked as general users when they were not in the list of hired workers.

The data corpus is similar as that of Chen and Chen [6]. Table 1 shows the number of reviews we collected. The Data collected 8363 posts: 458 fake review posts and 7905 regular review posts. Also, the data consisted of 111,065 replies that included 5,245 fake review replies and 105,820 regular review replies.

Table 1. Collected fake and normal reviews

	Fake Reviews	Normal Reviews	Total
Post	458	7905	8363
Reply	5245	105820	111065

Figure 1 shows the distribution of length (number of Chinese characters) for posts and replies. Based on Figure 1, we found that the length of most replies were less than 100 characters. Nevertheless, the length of most posts was longer than 100 characters. We focus on detect fake posts rather than fake reply since replies are usually shorter than posts.

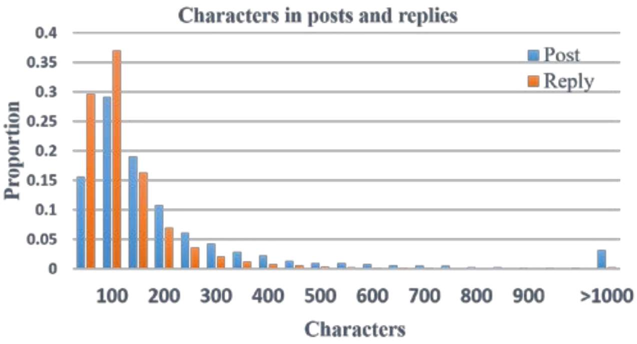

Figure 1. Characters in posts and replies

3.2 Analysis Methods

This study used LSTM methods to build deep learning model. Accuracy, precision, recall, and F-score were used to evaluate the deep learning models. This study belongs to supervised learning; we labeled spam data to 1 and labeled non-spam to 0. Our experimental process is shown in Figure 2. The first step for this study is Chinese word segmentation. To cut the Chinese sentence, we used Jieba (https://github.com/ldkrsi/jieba-zh_TW) which is open source tools for python. After segmentation, we established a dictionary based on the number of occurrences and sort from more to less.

To use the LSTM model, we have to convert review post as a vector with the same length. However, the review length is not the same to each review. Thus, we need to convert each review as the same length vector. According to Figure 1, most posts were less than 500 Chinese characters. We set the size of the vector as 500. The max length of the review should be 500. If the length of a review was longer than 500 words, only the first 500 words were included. If the length of a reviews is less than 500, we filled up 0 until the length equals 500.

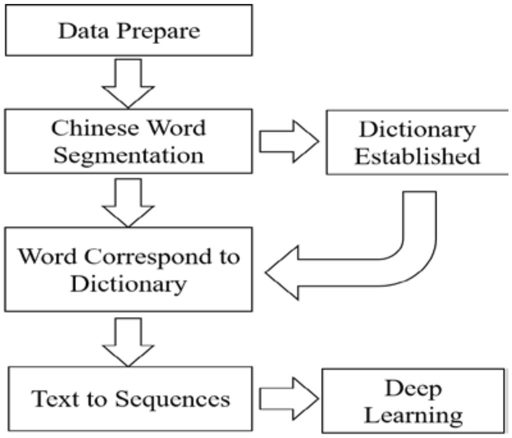

Figure 2. **Research Produce for Fake Review Detection**

We established multilayer perceptron which included input layer, hidden layer, and output layer. Input layer will receive data as a neuron, add a layer of LSTM after input layer, next is dimension reduction by the hidden layer, and output layer output one neuron. The neuron equal to 1 means that the reviewer is predicted to a spammer, equal to 0 means the reviewers is predicted to a regular user.

Table 2. Architecture of multilayer for deep learning

Layer	**Output Shape**	**Units**
Embedding	(None,500,32)	128000
LSTM	(None,128)	82432
Dropout	(None,128)	0
Dense	(None,64)	8256
Dropout	(None,64)	0
Dense	(None,32)	2080
Dropout	(None,32)	0
Dense	(None,1)	33

Table 2 shows the architecture of multilayer this research used. The input dimension is 4000, output dimension is 32, and length equal to 500. LSTM connect after embedding layer and set the shape equal to 128. Dropout was added after every dense layer to avoid over fitting. We try to set dropout from 0.2, 0.3, 0.4, 0.5 and 0.6 for each model. The total number of units trained is 220,801.

There are many parameters in the model that can be adjusted, like numbers of neuron on each layer, activation function, loss function, optimizer, and dropout. In this work, we use three kinds of activation function included Sigmoid, Tanh and Relu, two kinds of loss function included MSE and MSLE, three kinds of optimizers included Adam, RMSprop and AdaMax, and set dropout equal to 0.2, 0.3, 0.4, 0.5 and 0.6. We adjusted different parameters to get the best model.

Table 3. Imbalanced dataset

	Spam Post	Normal Post	Spam ratio
Training set	320	5533	5.46%
Testing set	138	2372	5.46%

Table 4. Balance dataset

	Spam Post	Normal Post	Spam ratio
Training set	320	960	25.00%
Testing set	138	414	25.00%

4. DATA ANALYSIS AND RESULTS

We spilt post to the training set and testing set with the ratio of 7:3, and divide into two part, the imbalanced data set presented in Table 3 and the balanced dataset presented in Table 4. We compare the results of the balanced dataset and imbalanced data set.

According to Table 3, the original data set is imbalanced; the spam ratio is 5.46%. After neural network training, we select the best five models, as shown in Table 5. We find that the accuracy of Table 5 quite high, the highest point is 0.972, the precision much higher than recall, and the highest of F-score is 0.691.

Table 5. Results of imbalanced dataset

Activation Function	Loss Function	Optimizers	Dropout	Accuracy	Precision	Recall	F-score	Time(s)
Relu	MSLE	AdaMax	0.5	0.969	0.79	0.61	0.688	2709
Relu	MSLE	AdaDelta	0.3	0.970	0.83	0.59	0.689	2833
Relu	MSLE	AdaDelta	0.6	0.969	0.81	0.58	0.675	2844
Relu	MSE	AdaDelta	0.5	0.966	0.72	0.64	0.677	2816
Sigmoid	MSE	AdaMax	0.6	0.972	0.88	0.57	0.691	2781

Table 6. Results of balanced dataset

Activation Function	Loss Function	Optimizers	Dropout	Accuracy	Precision	Recall	F-score	Time(s)
Sigmoid	MSLE	Adam	0.3	0.887	0.75	0.83	0.787	1327
Sigmoid	MSLE	Adam	0.5	0.893	0.78	0.79	0.784	1322
Sigmoid	MSLE	AdaMax	0.6	0.891	0.76	0.82	0.788	1325
TanH	MSLE	AdaMax	0.4	0.889	0.75	0.85	0.796	1331
TanH	MSLE	AdaMax	0.5	0.893	0.77	0.81	0.789	1332
Relu	MSLE	AdaMax	0.4	0.894	0.80	0.78	0.789	1322

Table 4 indicates the analysis results of the balanced dataset (spam ratio is 25.00%). We selected the best five models as shown in Table 6. We used two kinds of loss function and found the best loss function was MSLE, the highest accuracy is 0.894, precision and recall overall performance is good, and the highest of F-score is 0.796.

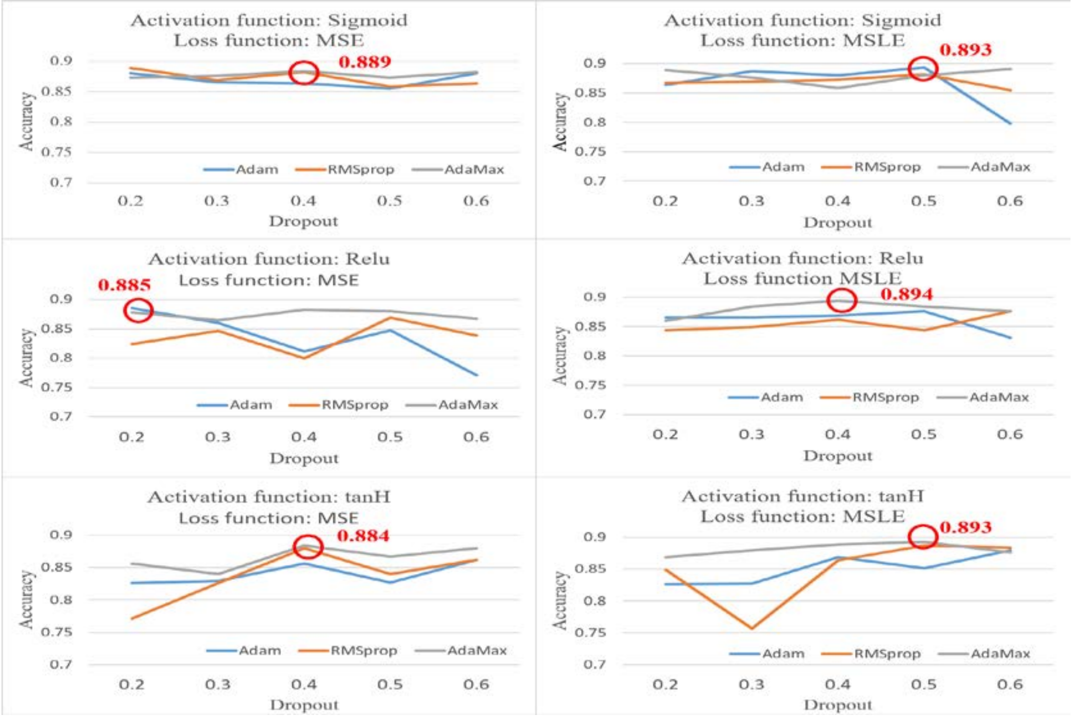

Figure 3. Accuracy of each model

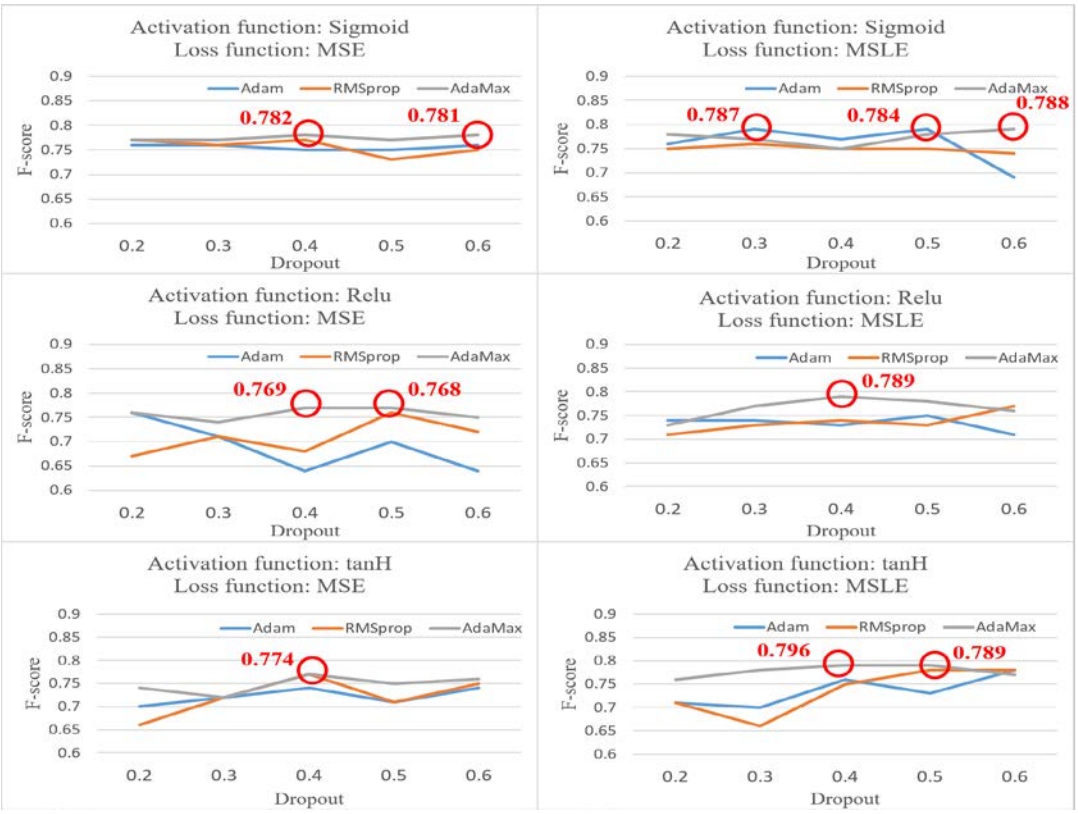

Figure 4. F-score of each model

Compared Table 5 and Table 6, we can find that the accuracy in Table 6 is smaller than that in Table 5. Thus, the accuracy for balanced data set is lower than the balanced data set.

However, when we calculated F-score, we got the opposite result: the F-score in Table 6 is better than that in Table 5. F-score for balanced data set is higher than that for balanced data set.

Figure 3 and Figure 4 provide accuracy and F-score of models of different activation function, loss function, optimizer, and dropout. We mark the models with the best accuracy. We found that MSLE is always better than MSE. The best accuracy for all models is 0.894.

Figure 4 shows that the best F-score of the models is 0.796 (MSLE model). It seems that MSLE gets better results in this study. It can also be observed AdaMax have better results in this experiment.

We also compared the results of the current study with previous literature. Table 6 shows the precision, recall, and F-score of Chen and Chen [5] and this study. We found that the precision, recall, and F-score of the current study is better than that of Chen and Chen [5]. Overall, the method used in this study was indeed effective than Chen and Chen [5].

Table 6. Comparison of previous study

Model	Spam ratio	Precision	Recall	F-score
Chen and Chen [4]	4.99%	0.6667	0.5714	0.6154
Best model on Table 5	5.46%	0. 79	0.61	0.688
Best model on Table 6	25.00%	0.75	0.85	0.796

5. CONCLUSIONS

Fake reviews mislead consumers to make the wrong purchase decision. However, it is not easy to check if a review is an ordinary review or fake review. Detecting fake review is essential.

The current study used a real case of fake review in Taiwan to test the possibility of using deep learning to detect a fake review. The current study detected spam reviews using long short-term memory. We used a variety of parameters included activation function, loss function, optimizers and dropout to test the detection performance. We found the loss function of MSLE got the best performance.

We also found that deep learning methods (LSTM) got better performance than traditional machine learning method of SVM. The detection performance of the current study is better than that of literature.

6. REFERENCE

[1] Mitchell, V. W. and McGoldrick, P.J. 1996. Consumer's risk-reduction strategies: a review and synthesis. *The International Review of Retail, Distribution and Consumer Research*, 6, 1,1-33.

[2] Jindal, N. and Liu, B. 2008. Opinion spam and analysis, in Proceedings of the *2008 International Conference on Web Search and Data Mining*. ACM: Palo Alto, California, USA. 219-230.

[3] Ott, M., et al., 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* - 1, Association for Computational Linguistics: Portland, Oregon. 309-319.

[4] Algur, S., et al., 2010. Conceptual level similarity measure based review spam detection. 416-423.

[5] Chen, Y.-R. and Chen. H.-H. 2015. Opinion spam detection in web forum: a real case study. in Proceedings of *the 24th International Conference on World Wide Web.* International World Wide Web Conferences Steering Committee.

[6] Wang, C.-C., Day, M.-Y., and Lin. Y.-R. 2016. A Real Case Analytics on Social Network of Opinion Spammers. in Information Reuse and Integration (IRI), *2016 IEEE 17th International Conference on Information Reuse and Integration (IRI)*. IEEE.

[7] Lau, R.Y., et al., 2011. Text mining and probabilistic language modeling for online review spam detecting. *ACM Transactions on Management Information Systems,* 2, 4, 1-30.

[8] Ren, Y., Ji, D., and Zhang, H. 2014. Positive Unlabeled Learning for Deceptive Reviews Detection. in EMNLP.

[9] Hinton, G.E. and Salakhutdinov, R.R. 2006. Reducing the dimensionality of data with neural networks. *Science,* 313, 5786, 504-507.

[10] Bengio, Y., Simard, P., and Frasconi, P. 1994. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, 5, 2,157-166.

[11] Hochreiter, S. and Schmidhuber, J. 1997. Long short-term memory. *Neural computation*, 9, 8,1735-1780.

[12] Hinton, G. E., et al., 2012. Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580.