Word Embedding Method of SMS Messages for Spam Message Filtering

Hyun-Young Lee
School of Software
Kookmin University
Seoul 02707, Republic of Korea
le32146@gmail.com

Seung-Shik Kang
School of Software
Kookmin University
Seoul 02707, Republic of Korea
sskang@kookmin.ac.kr

***Abstract*—SVM has been one of the most popular machine learning method for the binary classification such as sentiment analysis and spam message filtering. We explored a word embedding method for the construction of a feature vector and the deep learning method for the binary classification. CBOW is used as a word embedding technique and feedforward neural network is applied to classify SMS messages into ham or spam. The accuracy of the two classification methods of SVM and neural network are compared for the binary classification. The experimental result shows that the accuracy of deep learning method is better than the conventional machine learning method of SVM-light in the binary classification.**

Keywords—SMS Filtering; Feedforward Neural Network; Word Vector; Word Embedding; SVM light

I. INTRODUCTION

Nowadays everything is being connected over internet. As the number of smartphone users is greatly increasing, the exposure to email advertisement and SMS (Short Message Service) spam message are also tremendously increasing [1,2,3]. As a means to reducing the inconvenience from spam messages, a variety of methods for spam filtering have been explored. They are divided into rule-based, statistical model, and hybrid method [4,5,6]. Early researches on spam filtering models used the rule-based pattern matching, however there is a problem that it is difficult to adapt to the flexible changes of spam messages to get around the filtering rules. Another approaches are the traditional machine learning methods such as Naive Bayesian and SVM(Support Vector Machine) [7,8].

When deep neural network models are practically used in machine learning approaches, researchers adopted a deep learning method to natural language processing [9,10]. Lee(2017) designed a spam message filtering by using sent2vec and feedforward neural network model and implemented the model for Korean SMS message filtering system [11,12]. Besides utilizing the classifiers like SVM, text for input of the classifier need to be converted to vector representation [13,14,15]. How to represent word into a continuous vector space is a key point to get classifier achieve a good performance.

So far, the existing spam filtering system typically preprocesses an error correction to make good features, so the error correction has also been widely researched. It normally contains fixing a wrong representation of text to a correct one and automatic spacing [16]. All in all, in the conventional cases in Naive Bayesian and SVM, the preprocessing of counting the word frequency is a preparation process to extract features from the document. The feature vector construction, based on counting words such as TF-IDF, is not free on the number of feature dimensionality.

Recently using word vector, however, deep leaning technique in NLP(natural language processing) task not only resolved the curse of the dimensionality on text but also showed a good performance [17,18]. In other words, neural network not only represents a good dense feature for word into a vector space but also works well as a good classifier. As neural network-based word representation in a continuous vector space makes words, which is similar semantically and syntactically, closer to each other. The good qualitative vector representation of a word also helps classifier work well.

In this paper, we proposed a SMS spam message filtering method based on the word embedding technique of CBOW and the deep learning method of feedforward neural network has been applied to classify SMS messages into ham or spam. This work is an intermediate research result of applying the word2vec model to SVM and deep learning model and the additional experiments and final results will be published in [19].

II. NEURAL NETWORK FOR SMS MESSAGE FILTERING

The neural network is used to classify data, each node in a layer is connected to nodes in the subsequent layer. The connection, which is known as weight, is learned during training fitting to the goal of the network such as clustering or classification. To classify SMS spam message on neural network, it is needed that words consisting of SMS message is distributed to a continuous vector space. The distribution is called a word vector. After representing each word as a vector, word vectors composing a message is summed to represent a message vector. And then we classify the message vectors to ham or spam on feedforward neural network.

When representing word into a continuous vector space, input sentence is split into a sequence of tokens, where the token is a lexical token that is recognized by whitespace characters. That is, in the sentence "ㅂ r ㄷ r O1 야ㄱ l(바다 이야기)", it is divided into two tokens " ㅂ r ㄷ r " and "O1 야ㄱ l", respectively. When creating a word vector token,

978-1-5386-7789-6/19/$31.00 ©2019 IEEE

only whitespace characters are used as a delimiter, this usage of whitespace characters is intentional to also use special symbols, numbers, etc. In this paper, a token of word divided by whitespace is used to do word embedding. So word vectors of the partial changed words such as "사♥랑(lo♥ve)", "♥축하(♥ congratulations)", "경★ㅁ r(horse★racing)", "Ok 동" or "0 □동" are also used as feature vector to classify SMS message. The overall spam message filtering process is shown in Figure 1.

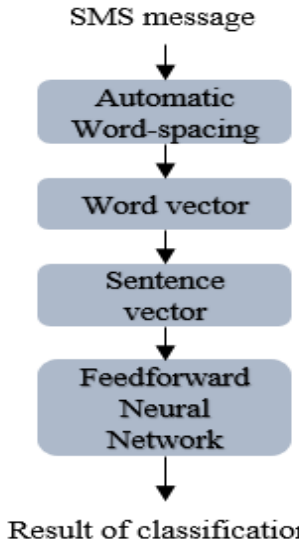

Fig. 1. Spam message filtering model

SMS messages incorporate a variety of word patterns which consist of a mixture of numbers, special characters, and multi-lingual characters such as the combination of Korean, Japanese, and English. In this paper, since the standard representation of a token for word vector is the usage of whitespace character as delimiter, automatic word-spacing is applied to make tokens from the intentional change of words like "사랑♥♥사랑(love♥♥love)" to "사랑 ♥♥ 사랑(love ♥♥ love)". There are another cases the SMS users don't put any spaces between words in a sentence. It causes a critical problem while creating a word vector. A sentence with no space in it might be regarded as a word vector which is similar to a vector representation of a rare word. In order to resolve this kind of problem, before creating word vector, we applied automatic word-spacing as a preprocessing process to get informative word vectors.

Input sentence = "좋은밤되세요내용없음"
("haveagoodnightwithnocontents")

After applying the automatic word-spacing module, the sentence becomes as follows.

Input sentence = "좋은 밤 되세요 내용 없음"
("have a good night with no contents")

The word vectors of sentences with automatic spacing are as follows. A sentence vector is the summation of word vectors consisting of the sentence.

word_vector("좋은(good)")
word_vector("밤(night)")
word_vector("되세요(have)")
word_vector("내용(contents)")
word_vector("없음(no)")

Word embedding is the most popular representation of words into a continuous vector space. We utilized the CBOW(Continuous Bag of Words) model of word embedding methodologies to represent words as vector. In order to create a word vector, we used CBOW that predicts a target word from the left and right context of the target word. And then, we generate a sentence vector summing word vectors composing a sentence. After making a sentence vector, it is passed to feedforward neural network as shown Figure 2.

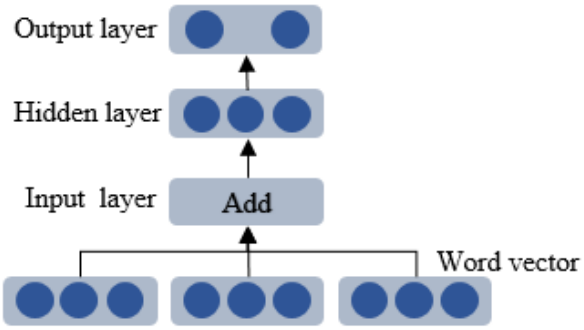

Fig. 2. Feedforward Neural network for spam filtering

In neural network system, feedforward have widely used for the classification task. The feedforward neural network is a fully connected network between a layer and the subsequent layer with weights to be learned by backpropagation. Our feedforward neural network model in Figure 2 used a sigmoid function as activation function and cross entropy as cost function for binary classification and then while making feedforward neural network deep, we measure a performance of spam filtering.

III. Experiments and Results

Training data and test data have been collected from the user's real SMS messages. As a preprocessing for machine learning, each SMS message is converted to a feature vector by word embedding technology. The result of word vectors are based on the total 109,993 sentences, which is the summation of training data and test data. We used the same number of dimensionality 300, for both word vectors and sentence vectors in the experiment of SVM light, which is a conventional machine learning model, and a feedforward neural network.

TABLE I. TRAINING AND TEST DATA SET FOR SMS SPAM FILTERING

	Train		Test		Total
	Ham	Spam	Ham	Spam	
Lines	49,993	50,000	5,000	5,000	109,993
Words	413,441	424,275	27,785	44,642	910,143

Table 2 shows the accuracy of spam message filtering about whether or not it is a spam message. It is the experimental results on SVM light according to the number of a variety of dimensionalities of CBOW word vector. The accuracy increases in higher dimensionality than the lower one.

TABLE II. SPAM MESSAGE FILTERING ACCURACY BY SVM LIGHT

	Dimensionality	Accuracy (%)
SVM light	200	93.84
	250	95.65
	300	**95.72**

Table 3 shows the accuracy of SMS filtering by deep learning method. The result of binary classification through feedforward neural network shows higher accuracy than the accuracy of binary classification using conventional SVM-light model. In addition, we confirmed that the classification accuracy on the feedforward neural network increases according to the number of layers. However, as the number of layers of the neural network increases, the accuracy may not increase proportionately.

TABLE III. SPAM MESSAGE FILTERING ACCURACY BY FNN

	Layers	Accuracy (%)
Feedforward Neural Network	1	95.40
	2	95.19
	3	**95.87**

IV. CONCLUDING REMARKS

We proposed a SMS spam message filtering method based on the word embedding technique of CBOW and the deep learning method of FNN has been applied to classify SMS messages into ham or spam. Experimental results show that the accuracy of the deep learning method is better than the conventional SVM method in the binary classification. The accuracy can be improved according to the number of hidden layers in the neural network structure. However, the accuracy does not increase proportionately as the number of hidden layers increase. Therefore, it is necessary to find out the optimal number of hidden layers that is most effective for filtering spam messages using the feedforward neural network. Also, there is a room to improve the accuracy by word vector construction through various word embedding methods of skip-gram, GloVe, FastText, and convolution neural network.

ACKNOWLEDGEMENTS

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT & Future Planning(No.NRF-2017M3C4A7068186)

REFERENCES

[1] Kang, Seung-Shik. "Junk-mail filtering by mail address validation and title-content weighting," Journal of Multimedia, Vol.9, No.2, pp.255-263, Feb, 2006.

[2] Gómez Hidalgo, J. M., Bringas, G. C., Sánz, E. P., & García, F. C. "Content based SMS spam filtering," In Proceedings of the 2006 ACM symposium on Document engineering, pp. 107-114. Oct, 2006.

[3] Lee, Seung-Jae and Deok-Jai Choi. "Personalized mobile junk message filtering system," Journal of the Korea Contents Association, Vol. 11, No. 12, pp. 122~135, Dec, 2011.

[4] Sohn, Dae-Neung, Jung-Tae Lee, Seung-Wook Lee, Joong-Hwi , Shin and Hae-Chang Rim. "Korean mobile spam filtering system considering characteristics of text messages," Journal of the Korean Academic-Industrial cooperation Society, Vol. 11, No. 7, pp.2595-2602, Jul, 2010.

[5] Kim, Sungyoon, Taesu Cha, Jaehyun Park, Jaehyun Choi, and Namyong Lee. "A technique of statistical message filtering for blocking spam message," Journal of Information Technology Services, Vol.13, No. 3, 299-308, Sep, 2014.

[6] Wu, Chih-Hung. "Behavior-based spam detection using a hybrid method of rule-based techniques and neural networks," Expert Systems with Applications, Vol. 36, No. 3, pp. 4321-4330, Apr, 2009.

[7] Androutsopoulos, Ion, Georgios Paliouras, Vangelis Karkaletsis, Georgios Sakkis, Constantine D. Spyropoulos and Panagiotis Stamatopoulos, "Learning to Filter Spam E-Mail: A Comparison of a Naive Bayesian and a Memory-Based Approach," Proceedings of the workshop on Machine Learning and Textual Information Access, pp.1-13, Sep. 2000.

[8] Amayri, Ola, and Nizar Bouguila, "A study of spam filtering using support vector machines," Artificial Intelligence Review, Volume 34, Issue 1, pp.73–108, June 2010

[9] Goldberg, Yoav. "A primer on neural network models for natural language processing," Journal of Artificial Intelligence Research(JAIR) 57, pp.345-420, 2016.

[10] Young, T., Hazarika, D., Poria, S., & Cambria, E., "Recent trends in deep learning based natural language processing," arXiv preprint arXiv:1708.02709, 2017.

[11] Lee, Hyun-Young, and Seung-Shik Kang, "Spam message filtering by using sen2vec and feedforward neural network," CSCI-ISNA, pp.1828-1829, 2017.

[12] Lee, Hyun-Young and Seung-Shik Kang, "Spam text filtering by using sen2vec and feedforward neural network," HCLT2017, pp.255-259, 2017.

[13] Mikolov, Tomas, Wen-tau Yih, and Geoffrey Zweig. "Linguistic regularities in continuous space word representations," HLT-NAACL Vol.13, 2013.

[14] Mikolov, Tomas, et al. "Efficient estimation of word representations in vector space," arXiv preprint arXiv:1301.3781, 2013.

[15] Pennington, Jeffrey, Richard Socher, and Christopher Manning, "Glove: global vectors for word representation," Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp.1532-1543, 2014.

[16] Kang, Seung-Shik and Du-Seong Chang. "Automatic error correction system for erroneous SMS strings," Journal of KISS : Software and Applications, Vol. 35, No. 6, pp. 386~391, Jun, 2008.

[17] J. Clark, I. Koprinska and J. Poon, "A neural network based approach to automated e-mail classification," In Proceedings of IEEE/WIC International Conference on Web Intelligence, pp. 702-705, Oct, 2003.

[18] Saad, Omar, Ashraf Darwish, and Ramadan Faraj. "A survey of machine learning techniques for spam filtering." International Journal of Computer Science and Network Security (IJCSNS), Vol. 12, No. 2, pp. 66-73, Feb, 2012.

[19] Lee, Hyun Young and Seung Shik Kang, "SMS text messages filtering using word embedding and deep learning techniques," The Journal of Smart Media, Vol.7, No.4, Dec. 2018.