FAMOUS: Fake News Detection Model based on Unified Key Sentence Information

Namwon Kim
Department of Electrical Engineering
Korea University
Seoul, Republic of Korea
namwon@korea.ac.kr

Deokjin Seo
Nuua
Seoul, Republic of Korea
westevil@nuua.net

Chang-Sung Jeong
Department of Electrical Engineering
Korea University
Seoul, Republic of Korea
csjeong@korea.ac.kr

Abstract— **Fake news detection causes a challenging problem due to the great influence of communication media over the public. In this paper, we shall present a new fake news detection model using unified key sentence information which can efficiently perform sentence matching between question and article by using key sentence retrieval based on bilateral multi perspective matching model. Our model makes use of one unified word vector for the key sentences of article by extracting them to the question from article and then merging the word vector for each key sentence. It can efficiently perform the sentence matching by executing matching operations between the contextual information obtained from the word vectors of question and key sentences through bidirectional long short term memory. Our model shows the competitive performance for fake news detection on the Korean article dataset over the previous result.**

Keywords-fake news detection; key sentence retrieval; sentence matching; natural language processing;

I. INTRODUCTION

A huge amount of contents shared by people make up various public opinions which sometimes greatly influence a way of thinking of the people in the society even though they are distorted information from fake news created on purpose with wrong commercial and political intent. Therefore, fake news detection becomes very important and challenging issue recently with the fast advancement of various media and communication technology. In this paper, we are concerned with a fake detection model for finding the truth of the question from a Korean article using sentence matching based on key sentence retrieval.

Sentence matching is a fundamental technique of the natural language processing(NLP) which checks whether two sentences are similar or not semantically. Recently, deep learning research has been activated by the advance of hardware such as graphics processing unit[1]. NLP techniques based on deep learning have been developed through various attempts for sentence matching. Some of them have been used as the recurrent neural network(RNN) to understand meanings between contexts of various lengths[2]. Since RNN can process a lot of data sequentially along time, it is suitable for semantic analysis of multi sentence in an article[3]. However, it has some limitations such as long-term dependency problem for capturing the relation between information too much apart vanishing and exploding gradient problem. Those problems were improved by long short term memory(LSTM) which adds forget gate to RNN[4]. Bilateral multi perspective matching model has achieved better performance for sentence matching by using two directional bidirectional long short term memory[5]. Despite these achievements, has some limitations for sentence matching in Korean due to different morphological features of Korean language[6,7]. Moreover, it has some difficulty in finding the contextual relation between two sentences too much apart in the article[8].

In this paper, we shall present a new fake news detection model using unified key sentence information(FAMOUS) which can efficiently perform sentence matching between question and article by using key sentence retrieval based on bilateral multi perspective matching(BiMPM) model in order to overcome these limitations. Our model makes use of one unified word vector for the key sentences of article by extracting them to the question from article and then merging the word vector for each key sentence. It can efficiently perform the sentence matching by executing matching operations between the contextual information obtained from the word vectors of question and key sentences through bidirectional long short term memory(BiLSTM). Our model shows the competitive performance for fake news detection on the Korean article dataset over the previous result.

The rest of our paper is composed as follows: Section II describes about related works about fake news detection and sentence matching. Section III presents our model architecture for fake news detection in detail. Section IV explains about the experiment and evaluation results for our model, and Section V gives a conclusion.

II. RELATED WORKS

A. Fake News Detection

Early approaches of fake news detection have been attempted to recognize distorted information and assess the credibility on social networking service[9,10]. Since With increasing attention on fact checking about news, and many

978-1-5386-6565-7/18 © 2018 IEEE

DOI: 10.1109/ICSESS.2018.8663864

efforts have been made to develop fact checking model[11]. Fake news detection has been used to detect distorted information which occurs with specific word aspects such as patterns of nouns, conjunctions and negative word[12,13]. Furthermore, the method of content similarity, lexical, sentiment analysis, stylistic similarity and semantic inconsistency to identify the fake reviews has been proposed[14]. These approaches are difficult to analyze complicated sentences, and not reflect contextual information. We use NLP techniques based on deep learning for fake news detection to overcome these problems.

B. Sentence Matching

Early approaches for sentence matching have been attempted to base on lexical matching techniques[15,16,17]. These approaches have difficulty in detecting similar meaning such as synonyms. This problem has been overcome through semantic similarity measures using Wordnet and corpus[18,19]. However, these approaches may not capture the rich patterns of natural language sentences which consist of complicated and sequential structures. For that reason, we exploit deep learning method in order to overcome these problems and to perform sentence matching efficiently.

The approach of sentence matching model based on deep learning has been advanced through various attempts. The matching of a pair of sentence has been attempted by Siamese neural network which measures similarity using extracted features from two different inputs[20]. Siamese architecture based model with BiLSTM has been developed for measuring the relation between a pair of sentence[21]. This model has been shown with competitive performance but loses important information which represents feature due to no interaction between a pair of sentence. A sentence matching model using interaction has been developed to solve this problem[22,23]. In addition, multi-perspective context matching(MPCM) model has been developed to perform sentence matching efficiently[24]. Then, BiMPM model has been developed to improve MPCM[5]. This model first outputs vector which encodes a pair of sentences using BiLSTM. The encoded vector is processed on multi-perspective sequentially. The processed vector is aggregated as a fixed length vector. Then, the matching result is predicted by probability using vector of the fixed length. BiMPM model has achieved the state-of-the-art performance on paraphrase identification, natural language inference and question answering. BiLSTM with attention model has been used for answering about the question to a passage and finding a similar sentence in the paragraph[25,26]. Our proposed model matches one question sentence with more than one key sentences from article rather than matching two sentences as in BiMPM.

III. Model Architecture

In this section we present a sentence matching model for fake news detection using key sentence retrieval from article based on BiMPM[5]. Given a question sentence Q and a set A of sentences in an article, our model matches Q with A in order to figure out whether Q is true or not by making use of matching information obtained from contextual information of Q and a set K of key sentences selected from A. Our model

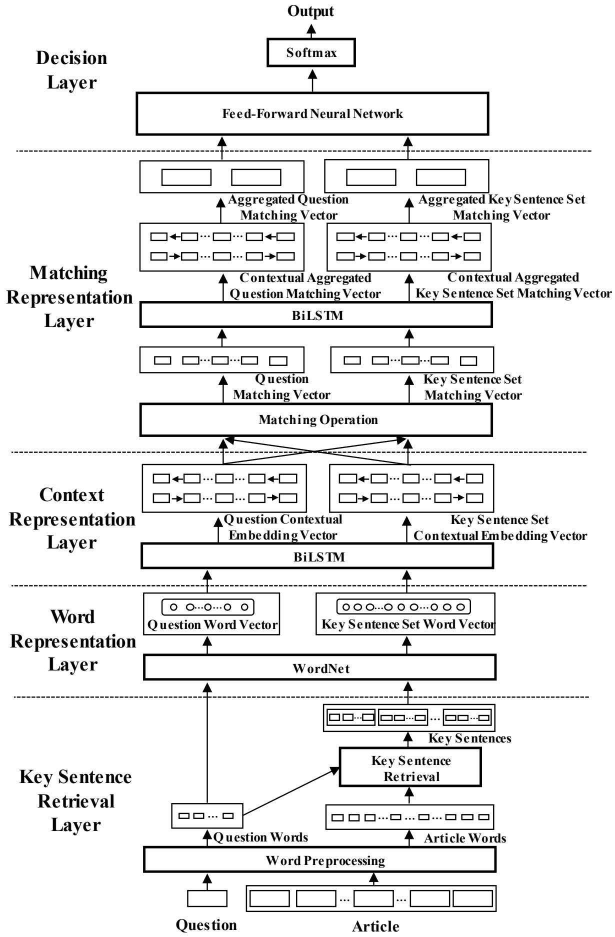

Figure 1. Fake news detection model.

consists of 5 layers: key sentence layer, word representation layer, context representation layer, matching representation layer and decision layer as shown in Fig. 1.

- Key sentence retrieval layer: A sentence in Korean is decomposed into word units each of which is formed by adding an affix to the root. A key sentence is the one in A which is most similar to Q in terms of the frequency of the same word units between them. This layer splits Q and each sentence in A into word units while eliminating special symbol, and then extracts key sentences from A by checking the appearance frequency of word units in Q in each sentence of A as in Fig. 2. Finally, it outputs a sequence QW of word units in Q, and a sequence KW of word units obtained by concatenating word units in each sentence of K.

- Word representation layer: The word representation layer generates two-word vectors, that is, question word vector QV and key sentence set word vector KV for QW and KW respectively by executing word embedding for each word unit and LSTM. We use word2vec for word embedding after training for Korean words[27].

978-1-5386-6565-7/18 © 2018 IEEE

DOI: 10.1109/ICSESS.2018.8663864

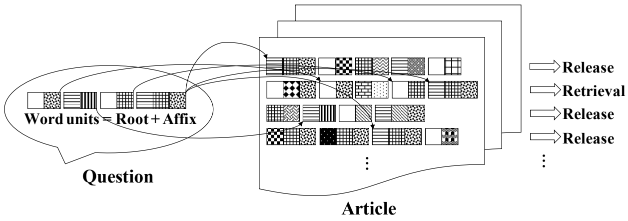

Figure 2. Key sentence retrieval.

- Context representation layer: This layer outputs a question contextual embedding vector *QCV* for the question word vector *QV* and a key sentence contextual embedding vector *KCV* for the key sentence word vector *KV* respectively through BiLSTM.
- Matching representation layer: This layer outputs two matching vectors: question matching vector *QMV* and key sentence set matching vector *KMV* for *QCV* and *KCV* respectively through matching operation. We exploit full matching and attentive matching operation[7]: In the former, each element in *QMV* is calculated by comparing each element *QCV* with the last element of *KCV*, and similarly each element in *KMV* by comparing element in *KCV* with the first element in *QCV*. Next, we generate two sequences of matching vectors through BiLSTM, and then output for aggregated matching vector by concatenating the last time matching vectors from BiLSTM models.
- Decision layer: This layer determines the probability distribution of similarity relation between the question and key sentence set by using aggregated matching vectors through two layered feed forward neural networks, and then makes a final decision by exploiting softmax function.

IV. EXPERIMENT

A. Korean Article Dataset

We use the Korean article dataset which is constructed for the evaluation of fake news detection. We evaluate fake news detection in terms of five levels as shown in Table I below. Each level includes the cases of the lower levels.

TABLE I. LEVELS OF DETECTION

Levels	Details
1	Change of subject or object on single sentence
2	Inversion of word order
3	Change words with synonyms or antonyms
4	Inference from a single sentence in an article
5	Inference from multi sentences

TABLE II. ACCURACY FOR EACH LEVEL

Levels	Accuracy of BiMPM	Accuracy of our model
1	0.70	0.74
2	0.65	0.72
3	0.63	0.66
4	0.58	0.69
5	0.62	0.64
Average	0.64	0.69

B. Evaluation Result

We evaluate our model with the Korean article dataset with training and validation sets, respectively 70% and 30% of the whole data set. We use word embeddings that are pretrained to the 300-dimensional vector from the Korean Wikipedia data, June 2018. Furthermore, we transform each word into word vector of 100 dimensions by word embedding and LSTM model. We set the dropout rate 0.2 and learning rate 0.001 within our model. Also, we use adaptive moment estimation in order to efficiently update parameters[28].

We compare the performance of our model with that of BiMPM model without key sentence retrieval. We use the measure of accuracy for evaluation as follows:

$$Accuracy = \frac{correct}{total} \times 100 \tag{1}$$

, where correct and total indicate the number of correct results and that of validation data respectively. Table II shows the comparison of accuracy compared to the previous BiMPM model for all levels. In the overall average accuracy, our model improves BiMPM from 64% to 69%. In particular, our model achieves more superior performance than BiMPM at level 1, 2 and 4 compared to level 3 and 5, since we may lose some information during key sentence retrieval in case of level 3 and 5 which use synonyms or antonyms and inference respectively.

V. CONCLUSION

In this paper, we have presented a new sentence matching model for fake news detection which can efficiently perform the sentence matching by using key sentence retrieval based on BiMPM model. Our model consists of 5 layers: In the key sentence retrieval layer, our model extracts a set of key sentences to the question from article by decomposing question and article sentences into word units and then checking the appearance frequency of word units in question in each sentence of article. Two-word vectors, question word and key sentence set word vectors, are obtained by executing word embedding for each word unit and LSTM model in word representation layer. Then, from those two vectors, question contextual embedding and a key sentence contextual embedding vectors are calculated through BiLSTM respectively in the context representation layer. Matching

978-1-5386-6565-7/18 © 2018 IEEE

DOI: 10.1109/ICSESS.2018.8663864

representation layer generates two matching vectors: question matching, and key sentence set matching vectors respectively through full matching and attentive matching operations, and then generates for aggregated matching vector by concatenating the matching vectors from BiLSTM models. Finally, decision layer makes a final decision by using aggregated matching vectors through two layered feed forward neural networks and softmax function. We have shown that our model improves the accuracy compared to the previous BiMPM model for all levels, thus upgrading the overall average accuracy from 64% to 69%. Therefore, our model shows the competitive performance for fake news detection on the Korean article dataset. As a future work, we continue to develop a more advanced model for fake news detection which applies our model for each key sentence independently, and then merges the results. Also, we plan to develop a fast-fake news detection framework which speeds up our model based on deep learning in the distributed parallel environment.

ACKNOWLEDGMENT

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2017R1D1A1B03035461), the Brain Korea 21 Plus Project in 2018, and the Institute for Information & Communications Technology Promotion(IITP) grant funded by the Korea government (MSIT) (No. 2018-0-00739, Deep learning-based natural language contents evaluation technology for detecting fake news).

REFERENCES

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature*, vol. 521, no. 7553, pp. 436-444, 2015.

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

[3] J. Elman, "Finding structure in time," *Cognitive Sci*., vol. 14, no. 2, 1990.

[4] S. Hochreiter, and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no.8, pp. 1735-1780, 1997.

[5] Z. Wang, W. Hamza, and R. Florian, "Bilateral multi-perspective matching for natural language sentences," *arXiv preprint arxiv:1702.03814*, 2017.

[6] H. M. Sohn, *The Korean Language*, Cambridge University Press, 2001.

[7] K. J. Oh, D. K. Lee, H. J. Choi, and J. Hur, "A Model of Korean Sentence Similarity Measurement Using Sense-based Morpheme Embedding and RNN Sentence Encoding," in *Big Data and Smart Computing(BigComp), 2017 IEEE International Conference on*. IEEE, 2017, pp. 430-433.

[8] K. M. Hermann, T. Kočiský, E. Grefenstette, L. Espeholt, W. Kay, M. Suleyman, and P. Blunsom, "Teaching machines to read and comprehend," in *Advances in Neural Information Prcessing Systems*, 2015, pp. 1693-1701.

[9] R. Mihalcea, and C. Strapparava, "The lie detector: Explorations in the automatic recognition of deceptive language," in *Proc. ACL-IJCNLP Conf. Assoc. Comput. Linguistics*, 2009, pp. 309-312.

[10] C. Castillo, M. Mendoza, and B. Poblete, ''Information credibility on Twitter,'' in *Proc. 20th Int. Conf. World Wide Web*, 2011, pp. 675–684.

[11] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," *ACM SIGKDD Explorations Newsletter*, vol. 19, no. 1, pp. 22-36, 2017.

[12] V. W. Feng, and G. Hirst, "Detecting deceptive opinions with profile compatibility," in *Proc. 6th Int Joint Conf. Natural Lang. Process*. (*IJCNLP*), 2013, pp. 338-346.

[13] N. J. Conroy, V. L. Rubin, and Y. Chen, "Automatic deception detection: Methods for finding fake news," *Proceedings of the Association for Information Science and Technology*, vol. 52, no.1, pp. 1-4, 2015.

[14] M. Hu, and B. Liu, "Mining and summarizing customer reviews,"in *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2004, pp. 168–177.

[15] R. Barzilay, and L. Lee, "Learning to paraphrase: an unsupervised approach using multiple-sequence alignment," in *Proceedings of NAACL 2003*, 2003, pp. 16-23.

[16] Y. Zhang, and J. Patrick, "Paraphrase identification by text canonicalization," in *proc. of the Australasian Language Technology Work shop*, 2005, pp. 160-166.

[17] Z. Kozareva, and A. Montoyo, "Paraphrase identification on the basis of supervised machine learning techniques," in *Advances in natural language processing*. Springer, Berlin, Heidelberg, 2006, pp. 524-533.

[18] A. Islam, and D. Inkpen, "Semantic similarity of short text," in *Recent Advances in Natural Language Processing V*, 2009, pp. 227-236.

[19] R. Mihalcea, C. Corley, and C. Strapparava, "Corpus-based and knowledge-based measures of text semantic similarity," in *AAAI*, 2006, pp. 775-780.

[20] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah, "Signature verification using a "Siamese" time delay neural network," in *Proc. Annu. Conf. Neural Inf. Process. Syst.*, 1994, pp. 737-744.

[21] M. Tan, B. Xiang, and B. Zhou, "Lstm-based deep learning models for non-factoid answer selection," *arXiv preprint arXiv:1511.04108*, 2015.

[22] W. Yin, H. Scütze, B. Xiang, and B. Zhou, "ABCNN: Attention based convolutional neural network for modeling sentence pairs," *Trans. Assoc. Comput. Linguisics*, vol. 4, pp. 259-272, 2016.

[23] Z. Wang, H. Mi, and A. Ittycheriah, "Sentence similarity learning by lexical decomposition and composition," *arXiv preprint arXiv:1602.07019*, 2016.

[24] Z. Wang, H. Mi, W. Hamza, and R. Florian, "Multi-perspective context matching for machine comprehension," *arXiv preprint arXiv:1612.04211*, 2016.

[25] D. Chen, J. Bolton, and C. D. Manning, "A thorough examination of the cnn/daily mail reading comprehension task," in *Association for Computational Linguistics* (ACL), 2016.

[26] D. Chen, A. Fisch, J. Weston, and A. Bordes, "Reading wikipedia to answer open-domain questions," *arXiv preprint arXiv:1704.00051*, 2017.

[27] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in neural information processing systems*, 2013, pp. 3111-3119.

[28] D. Kingma, and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1502.03167*, 2015.

978-1-5386-6565-7/18 © 2018 IEEE

DOI: 10.1109/ICSESS.2018.8663864