Journal of the Korea Industrial Information Systems Research Vol. 23 No. 3, Jun. 2018 :59-71
http://dx.doi.org./10.9723/jksiis.2018.23.3.059 ISSN:1229-3741

Korean and English Sentiment Analysis Using the Deep Learning

Adyan Marendra Ramadhani[1)], Hyung Rim Choi[2)*], and Seong Bae Lim[3)]

Abstract Social media has immense popularity among all services today. Data from social network services (SNSs) can be used for various objectives, such as text prediction or sentiment analysis. There is a great deal of Korean and English data on social media that can be used for sentiment analysis, but handling such huge amounts of unstructured data presents a difficult task. Machine learning is needed to handle such huge amounts of data. This research focuses on predicting Korean and English sentiment using deep forward neural network with a deep learning architecture and compares it with other methods, such as LDA MLP and GENSIM, using logistic regression. The research findings indicate an approximately 75% accuracy rate when predicting sentiments using DNN, with a latent Dirichelet allocation (LDA) prediction accuracy rate of approximately 81%, with the corpus being approximately 64% accurate between English and Korean.

Key Words : English Text, Korean Text, Sentiment, Deep Learning, Neural Network, Deep Neural Network

1. Introduction

In recent years, social media has become popular for sharing information worldwide. Information is shared through Facebook, Twitter, and other social media platforms. Not only is information shared, but comments are also created that reflect the commenter's opinions, whether they be positive or negative comments. This information becomes an important source for sentiment analysis using internet data. A deep learning algorithm is needed to process such statements made on the internet, to analyze the language, and to determine the opinions of the social media users, or "netizens." This process is called sentiment analysis.

Sentiment analysis can be useful for not only e-commerce but also for analyzing popular opinion on issues or problems in society. These data could be used by government sectors, such as the social and health ministry, or non-government organizations to prepare solutions to prevent these problems.

The target of sentiments is distinguishing text sentiment polarity. Sentiment analysis could be taken as a classification problem.

* Corresponding Author : hrchoi@dau.ac.kr
+ This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2015S1A3A2046781)
Manuscript received May 18, 2018 / revised June 20, 2018 / accepted June 27, 2018
1) Department of Management Information Systems, Dong-A University
2) Department of Management Information Systems, Dong-A University
3) Department of Management Information Systems, St. Mary's University

Research into sentiment analysis principally solves the two-class problem by dividing text into positive and negative sentiments. A deep neural network (DNN) and Gaussian mixture model is a robust framework and model for natural language processing. There are several papers that have already achieved accurate sentiment analysis for English and other languages, such as German and French. Several methods have been used, such as the Support Vector Machine (SVM) and naïve Bayesian classifier methods[1]. However, only a few researchers have analyzed the Korean language. Korean has become popular because of the "Korean Wave" (Hallyu or 한류), referring to the spread of Korean culture worldwide. This research uses a DNN to analyze both English and Korean texts.

The contents of the paper consist of the following: Section I introduces the research, Section II focuses on the related work and research in this field, Section III introduces the research model, and Section IV presents the research results and conclusion.

2. Related Works

Sentiment analysis is the research into how opinions and perspectives can be related to emotions and attitudes shown in natural language with respect to an occurrence or event. Current research has shown that sentiment analysis has been able to not only determine positive and negative sentiments but also behavior patterns and emotions for different topics and languages. In studies on sentiment analysis, researchers have used different techniques to predict the social opinion and emotion in text. Following are several examples:

Table 1 Techniques to Predict Social Opinion and Emotion in Text

Researchers	Examples
Joshi and Tekchandani (2016)[2]	Applied SVM, naïve Bayesian, and maximum entropy methods to compare the methods for Twitter data analysis for movie reviews from Twitter for their dataset.
Rif'at, Fatma, and Dodon (2016)[3]	Applied several methods in naïve Bayesian smoothing to analyze sentiments. An optimum result of 72.3% prediction accuracy was achieved using the Laplace smoothing technique
Duncan and Zhang (2015)[4]	Applied a neural network to classify the sentiment of tweets. The average accuracy (the number of correctly classified tweets divided by the number of incorrectly classified tweets) was 74.15%.
Lee et al. (2016)[5]	Used an unstructured text mining algorithm based on Data Dictionary.
Yun (2008)[6]	Used natural language processing based information extraction for newspapers.
Chen et al. (2018)[7]	Applied sentiment analysis for language learning. The results demonstrated that both the suggested emotion synonyms and the corresponding usage information are beneficial to learners' use of emotion vocabulary

Previous researchers have mainly used naïve Bayesian and neural networks to determine the sentiment of texts. To improve sentiment analysis and the handling of big data, a DNN using a deep learning architecture was applied. Following are several examples:

Table 2 The Handling of Big Data, a DNN using Deep Learning Architecture

Researchers	Examples
M. Y. Day and Y. D. Lin (2017)[8]	Applied deep learning to classify the Google Play consumer review sentiment analysis. They applied methods such as SVM and naïve Bayesian as prediction comparisons.
Hassan and A. Mahmood (2017)[9]	Applied deep learning to classify short sentences.
Z. Jianqiang, G. Xiaolin, and Z. Xuejun (2018)[10]	Applied deep convolution neural networks (DCNN) for Twitter sentiment analysis prediction. They achieved the highest accuracy of 87.62% using the DCNN method.

Past research has used a small dataset with naïve Bayesian and neural networks to classify text and predict sentiment. To improve prediction accuracy, a large dataset, handled via a deep learning architecture, was used in the present study with a DNN.

Based on related works and past research, this research aims at the optimum prediction for both English and Korean text using a DNN based on a deep learning architecture. The contribution of this research is to resolve and find the optimum prediction for both English and Korean text and mixed languages using a deep learning architecture for a large and complex dataset.

3. Model And Research Methods

The design of the proposed method for Korean and English text sentiments is illustrated in Fig. 1. After data have been obtained, they are cleaned and preprocessed to be better utilized to train the DNN. The final stage is sentiment testing using the test dataset.

To compare the results, Latent Dirichelet Allocation (LDA) and Multilayer Perceptron (MLP) methods are used.

A. Text Mining

Text mining is the process of discovering and extracting information from large, unstructured textual sources. Text mining can handle the unstructured data resources[11][29].

Table 3 Unstructured Data Resources

	Unstructured data resources
Information retrieval	Gather, select, and filter text from the database (Twitter, Facebook, or another database).
Information extraction	Partial, shallow, and deep language analysis.
Information extraction	Find relevant entities and facts about the entities.
Data mining	Combine and link facts.
Data mining	Discover new knowledge and facts.

In the present research, the experiment used only two of the three steps of text mining: information retrieval from the Twitter database using the Twitter Application Programming Interface (API) and information extraction using Natural Language Processing (NLP).

Fig. 1 The Flow of Korean and English Sentiment Analysis

B. Korean and English Grammar

The Korean and English languages have different sentence structures. English uses a Subject + Verb + Object (SVO) sentence structure, while Korean uses a Subject + Object + Verb (SOV) sentence structure[12].

나는 사과를 좋아해요.

Fig. 2 Sample of a Korean Sentence

I like an apple.

Fig. 3 Sample of an English Sentence

C. Text Cleaning and Preprocessing for English Text

Text cleaning is one of the text-mining processes used to reduce the words or other components of text that are hard to analyze or do not add meaning to the text. Text data contain white spaces, punctuation, stop words, etc. These characters do not add much meaning and are unnecessary to process for a sentiment analysis. For example, English stop words such as "the," "is," etc. do not provide much information about the sentiment of the text, entities mentioned in the text, or relationships between those entities[13].

Stemming or lemmatization is a way of reducing words that have the same linguistic root or stem[14].

There were several step that used for text cleaning and preprocessing for english text such as:

- Convert the text to lower case so that words such as “write” and “Write” are considered the same word for analysis (transform case)[14].
- Remove numbers[14].
- Remove English stop words, e.g., “the,” “is,” “of,”[14].
- Remove punctuation, e.g., “?”[14].
- Eliminate extra white spaces[14].
- The text cleaning and preprocessing have several steps for the normalization of a sentence, including the following:
- Cleaning the non-characters
- Cleaning the Twitter RT, @ and the links from the sentences
- Filtering the token
- Tokenizing the sentence

D. Stop Words

Stop words contain a number of terms in a variety of texts (e.g., the, is, at, which, and on). Their frequencies are quite high, and these words often affect content words that have real effects on classification[15].

There are typically two ways to filter stop words. One is to establish stop word vocabulary. The second is to filter stop words appearing in the vocabulary. Another is to set the word frequency threshold; all the words higher than the threshold value are considered stop words by direct filtration[15].

In Korean, stop words are for example, “그리고“, “하지마“, “하지마라“, and “다른“. Korean stop words are quite similar to those in English; the difference is the formal words used at the end of the sentences (입니다 /-입니까).

For stop words in this research, all the stop words will be eliminated as well as meaningless words such as "haha," "hehe," "wkwk," "emm," and "umm" for classification and prediction purposes.

E. Dataset (Training and Test)

The English dataset is downloaded from the Twitter database using its API (Stream API), and the Korean dataset is based on the Naver (Korean search engine company) movie rating dataset. Table 4 shows the dataset used in this research.

Table 4 Dataset Sample

Dataset	Total	Information
Positive Training Dataset for English	10,000	Twitter Dataset
Positive Training Dataset for Korean	150,000	Naver Dataset
Negative Training Dataset for English	10,000	Twitter Dataset
Negative Training Dataset for Korean	150,000	Naver Dataset
Test Dataset for English Text	20,000	Twitter Dataset
Test Dataset for Korean Text	10,000	Naver Dataset

Fig. 4 shows the sentence cleaning results for sample Korean and English sentenced.

No	Original Sentences	Cleaning Sentences
1	RT @ian_dian I want to eat rice, wkwkwkwk	I want eat rice
2	RT @quaannnzb 난 밥을 먹어요 ㅋㅋㅋㅋㅋ	나 밥 먹어

Fig. 4 The Tweet (Text) Cleaning for Korean and English Text

F. Text Cleaning and Preprocessing (Tokenization) for Korean Text

The text cleaning and preprocessing for the Korean text is slightly different than for the English text. The Korean text contains several levels of speech. The Korean language uses an "honorific" speech level to differentiate the speaker and the listener based on age, relationship, and formality.

These levels are as follows[16]:

- Formal (합쇼체 or 존댓말). Use when the person you are speaking to is your senior, someone important, or someone you have never met (i.e., your boss, your professor, or the lady you bumped into on the street).
- Polite (해요체, also 존댓말). Use when speaking to someone you are friendly with or when in a non-business setting (i.e., a restaurant server, a classmate, a friend, or a relative-family relationships).
- Informal (해체 or 반말). Use when speaking to someone who is your junior by a few years at least, a very close friend, a sibling, a significant other, or someone you want to insult by showing the highest form of disrespect).

There is a component at the end of Korean verbs that distinguishes the speech level. The formal one uses the -입니다 /-입니까 at the end of verbs, the polite one uses the - 아/어요 at the end of verbs, and the informal one uses the - 야 at the end of verbs.

These are examples for the formal, polite, and informal levels of speech:

- 저는 밥을 먹습니다 (Formal)
- 저는 밥을 먹어요 (Polite)
- 나는 밥을 먹어야 (Informal)

There are non-standard Korean words which are difficult to extract in data mining, such as abbreviations, slang terms, strange expressions, and emoticons[17]. Because of this difficulty, normalization will be applied in text cleaning and preprocessing for removal.

In text cleaning and preprocessing, the Korean language has several steps for the normalization of sentences[18]:

- 정규화 (Normalization)
  - o 입니닭ㅋㅋ → 입니다 ㅋㅋ
  - o 샤릉해 → 사랑해
  - o 한국어를처리하는예시입니닭ㅋㅋㅋㅋㅋ → 한국어를처리하는 예시입니다 ㅋㅋ
- 토큰화 (Tokenization)
  - o 한국어를 처리하는 예시입니다 ㅋㅋ -> 한국어 (Noun), 를 (Particle), 처리 (Noun), 하는 (Verb), 예시 (Noun), 입 (Adjective), 니다 (Eomi) ㅋㅋ (Particle)
- 어근화 (Stemming) (입니다 -> 이다)
  - o (입니다 ->이다) 한국어를 처리하는 예시 입니다 ㅋㅋ -> 한국어(Noun), 를 (Particle), 처리 (Noun), 하다 (Verb), 예시 (Noun), 이다 (Adjective), ㅋㅋ (Korean Particle).
- 어구 추출 (Phrase extraction)
  - o 한국어를 처리하는 예시입니다 ㅋㅋ -> 한국어, 처리, 예시, 처리하는 예시

G. Deep Learning Methods and DNNs

DNNs are neural networks using deep learning containing many hidden neuron layers[19][28]. In other words, Deep Learning is about learning multiple levels of representation and abstraction and a class of machine learning techniques exploiting many layers of non-linear information processing of more than supervised or unsupervised feature extraction and transformation, and for pattern analysis and classification that help to make sense of data such as images, sound, and text.

Another deep learning definition states that it is a set of algorithms in machine learning that attempt to learn on multiple levels, corresponding to different levels of abstraction. It typically uses artificial neural networks[9].

There are several types of DNNs, such as the RBM, deep feed forward neural network, deep convolutional neural network, deep belief network, and many more. Fig. 5 depicts the deep feed forward neural network, which is used in this research.

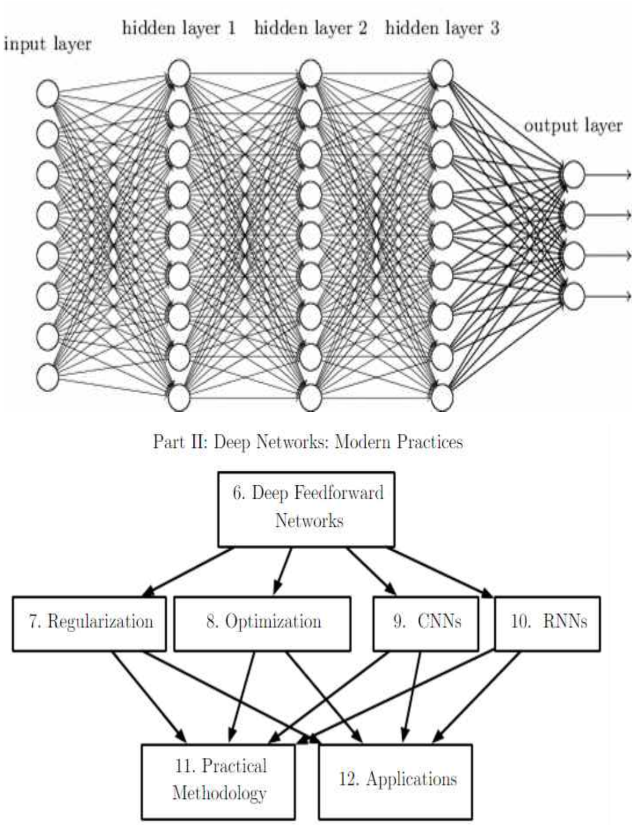

Fig. 5 Deep Learning Architecture[20]

In this paper, a DNN is applied. The specifications of the network are as follows:

- Feedforward neural network
- Three hidden layers
- Rectified linear unit (ReLU) and the sigmoid function activation

- 100-neuron input
- Mean square unit and the stochastic gradient descent

Fig. 6 explains the proposed method architecture for the neural network, including each hidden layer. Each hidden layer has its own function for classifying sentiment analysis. The first hidden layer classifies a sentiment based on its individual words, the second layer classifies it based on the complete sentence, and the third layer classifies it by the popularity of the individual words based on an online dictionary.

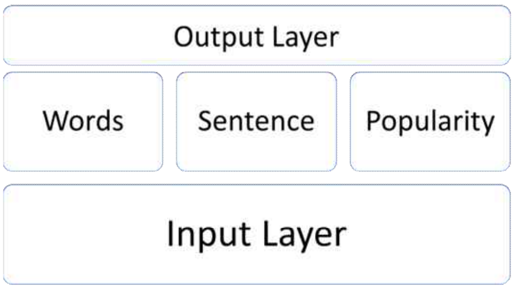

Fig. 6 The Proposed DNN Architecture

The sigmoid activation formula is :

$$s(x) = \frac{1}{1+\exp(-x)}$$

Where $x \rightarrow \pm\infty$
The ReLU activation formula is :

$$f'(x) = \begin{cases} 0 & \text{for } x < 1 \\ 1 & \text{for } x \geq 0 \end{cases}$$

The formula for the mean square unit is :

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y^i - Y_i)$$

Where the Y^i is the vector of n prediction and Y_i is the vector of observed values.
Gradient descent is a way to minimize an objective function $J(\theta)$ by a model's parameters $\theta \epsilon RD$ by updating the parameters in the opposite direction of the gradient of the objective function $\nabla_\theta J(\theta)$ w.r.t. to the parameters[21].

The Stochastic gradient descent formula [22] is :

$$\omega \leftarrow \omega - \eta(\alpha\frac{\delta R(\omega)}{\delta\omega} + \frac{\delta L(\omega^T x_i + b, y_i}{\delta\omega})$$

Where is η the learning rate that controls the step size in the parameter space. The intercept b is updated similarly but without regularization[22].

4. Comparing Methods (LDA, MLP, and Gensim)

The LDA, MLP, and Genism (corpus) methods were used to compare the classifier results

A. LDA

Blei et al.(2017) introduced and proposed the LDA method in late 2002. LDA generates a probabilistic model of unsupervised topics and uses a K-dimensional latent random variable objected Dirichlet distribution. It is based on Probabilistic Latent Semantic Analysis (PLSA)[23-24].

B. MLP

An MLP is a type of neural network that has become popular over the past several years. MLP is a class of a feedforward artificial neural network. MLPs are usually trained with an iterative gradient algorithm known as back propagation. Fig. 7 illustrates a regular MLP[19],[25].

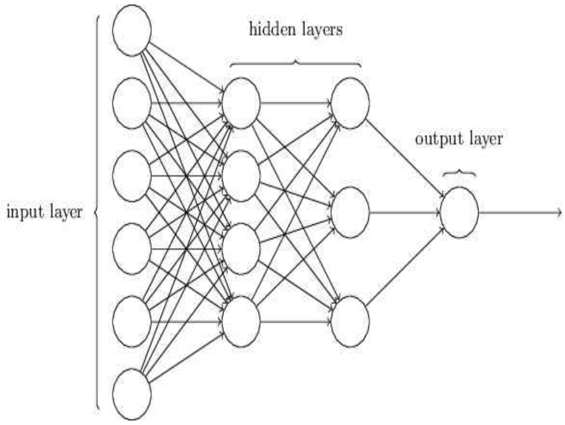

Fig. 7 MLP Architecture[26]

C. Gensim Corpus

Gensim was written by Radim Rehurek in 2008. It is an open-source topic-modeling tool that is implemented in Python. The input of Gensim is a corpus of plain text documents [27]. Gensim is designed for processing huge amounts of data. After finding semantic topics, it can be used for discovering topical similarities against other documents when text documents are queried[27].

5. Experiment Result

The standard LDA and Genism corpus settings are applied to this experiment. The Korean and English texts are preprocessed based on the explanation in Section III. For testing comparison, the precision and recall are used as the statistical comparison. These formulas are as follows:

$$Recall = \frac{TP}{TP+FN}$$

$$Precision = \frac{TP}{TP+FP}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

Where :

TP: $True\ Positive$
TN: $True\ \neg ative$
FP: $False\ Positive$
FN: $False\ \neg ative$

The precision and recall calculation in testing phase will be used after it passed traning phase. The Korean and English texts dataset will be divided into two separate parts: training and testing datasets. The experiment used a 10,000-sized dataset for positive and negative samples for English training and testing, totaling 20,000. The Korean text used a 150,000-sized dataset for the negative and positive samples. Deep learning was trained with 100 epochs at learning rates of 0.1 and 0.001. The experiment used the Tensor Flow program to create the network. The genism and NLTK python libraries were used to create the genism corpus and LDA classifier. The results are listed in Figs. 8 and 9.

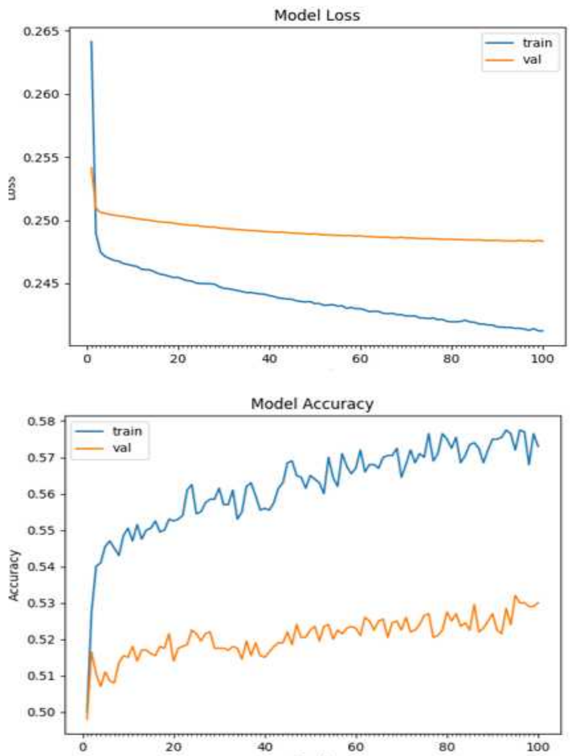

Fig. 8 Learning Rate for the 0.001 DNN

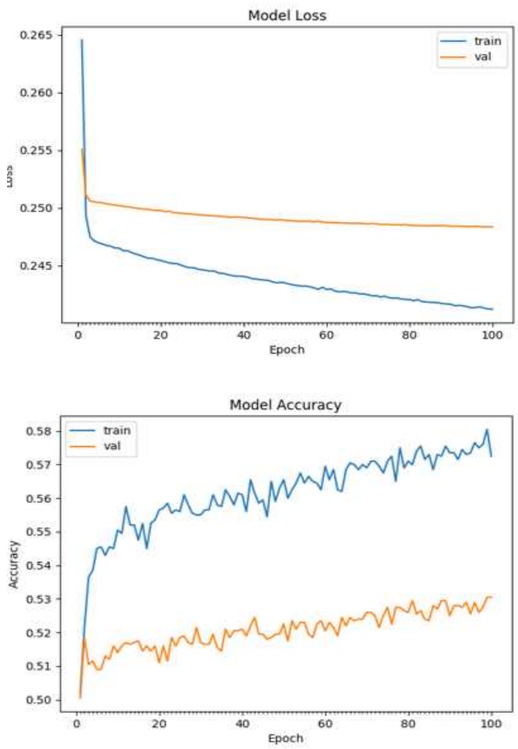

Fig. 9 Learning Rate for the 0.1 DNN

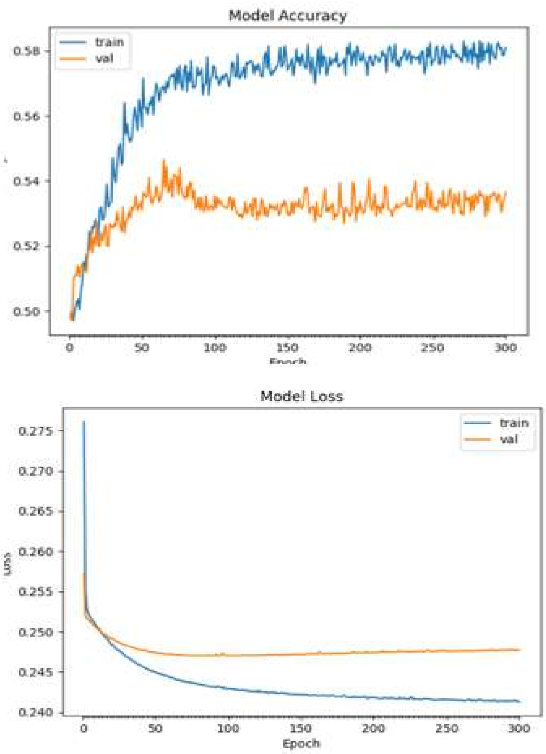

Fig. 10 Learning Rate for the MLP

Based on the graph, the training model's accuracy increases constantly with a convergence at 55%. The table shows the accuracy of the training and testing sets, with results of 77.45% for training and 75.03% testing. For MLP, the accuracy is about 67.45% for the training set and about 52.60% for the testing set. Figs. 11 and 12 show the results of the LDA classifier and the Genism (corpus) classifier. Fig. 11 explains the most informative features in the training dataset. while Fig. 12 the words represent the 여자 (Woman) words of positive sentiment based on the training dataset. Table 7 compares the LDA, Genism, MLP, and deep learning methods. In Table I4, LDA has an optimum result of 80% for the mix of English and Korean texts, and deep learning has an average of 75%. All accuracy results are based on the accuracy, precision, and recall formulas.

LDA and corpus use the linear regression as the classifier, while deep learning uses a DNN based on deep architecture as the classifier.

LDA and corpus need a feature extraction, and deep learning did not use any feature extraction. To compare with the previous research mentioned in the related works section above, for example, Duncan and Zhang (2015) used neural network and feature extraction to classify sentiment in tweets[4].

They obtained around 74 - 75% total accuracy for English sentiment, while in the current research the deep learning method obtained around 75 - 77% total accuracy using English and Korean sentences.

```
Most Informative Features
    exists(수작/Noun) = True        1 : 0    =    38.0 : 1.0
    exists(최악/Noun) = True        0 : 1    =    30.1 : 1.0
    exists(노잼/Noun) = True        0 : 1    =    22.1 : 1.0
    exists(낭비/Noun) = True        0 : 1    =    19.5 : 1.0
   exists(쓰레기/Noun) = True        0 : 1    =    19.4 : 1.0
    exists(여운/Noun) = True        1 : 0    =    18.9 : 1.0
   exists(발연기/Noun) = True        0 : 1    =    16.9 : 1.0
```

Fig. 11 Results for the Most Informative Features in the Korean Text

```
('진수/Noun', 0.4620112180709839),
('서부/Noun', 0.4616014361381531),
('풍/Adverb', 0.4575424790382385),
('한계/Noun', 0.4386754333972931),
('오락/Noun', 0.43549543619155884),
('코미디/Noun', 0.42612385749816895),
('풍기/Noun', 0.4205281138420105),
('하이/Noun', 0.41915249824523926),
('히어로/Noun', 0.4169023931026459)]
```

Fig. 12 Results for the Genism 여자 Words of Positive Sentiment in the Korean Text

Table 5 Accuracy of the Training and Testing Sets for Twitter Sentiment Analysis using Deep Learning Methods

No	Dataset	Accuracy
1	Train	77.45 %
2	Test	75.03 %

Table 6 Accuracy of the Training and Testing Sets for Twitter Sentiment Analysis using MLP

No	Dataset	Accuracy
1	Train	67.45%
2	Test	52.60%

Table 7 Comparison of LDA and Corpus Based on Recall and Precision

No	Dataset	LDA	Corpus	Neural Network	Proposed Method
1	Train	81.40	66.04	67.45	77.45
2	Test			52.60	75.03

6. Conclusions and Future Work

This paper indicates that social problems can be predicted based on the sentiment (negative or positive) of statements made on the internet. Recognizing social problems from sentiments can also be used to improve government functions for citizen services.

This paper has presented an approach for using a deep learning architecture with a neural network model to predict sentiments in short Korean and English texts. This work differs from the existing approaches in that it predicts sentiments from a mixed Korean and English language dataset.

Text mining and processing are used for the normalization and tokenization of English and Korean texts. LDA, MLP, and the genism corpus are applied to predict sentiments and compare prior research findings with those of this study. The DNN uses a deep learning architecture to predict sentiments. In LDA and the genism corpus, we applied feature extraction but not for the DNN. LDA produced a higher result, 81%, for sentiment prediction, while deep learning only produced an average of 75% and genism an even lower result of 66.04%. The reason for the higher score from LDA is the feature extraction. Although the DNN did not use feature extraction, it was still able to predict with an accuracy of 75% on average.

The limitation of this research is that short text is preprocessed in only Korean and English. Other languages included in the text data were erased. The contribution of this research is to improve the accuracy of sentiment prediction in single and mixed languages. It not only improves upon the prediction accuracy of English in previous research but also offers a prediction standard for Korean as well.

For future work, the DNN can be combined with another method, such as LDA or corpus, to increase prediction accuracy. Deep learning can be used to automatically tune up the neural network to adapt a small or large amount of data. This work will hopefully be helpful in contributing to further research.

References

[1] Lavanya, K. and Deisy, C., "Twitter Sentiment Analysis using Multi-Class SVM," 2017 International Conference on Intelligent Computing and Control (I2C2), Coimbatore, pp. 1-6, 2017.

[2] Joshi, R. and Tekchandani, R., "Comparative Analysis of Twitter Data using Supervised Classifiers," 2016 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, pp. 1 - 6, 2016.

[3] Ramadhani, R. A., Indriani, F., and Nugrahadi, D. T., "Comparison of Naive Bayes Smoothing Methods for Twitter Sentiment Analysis," 2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS), Malang, pp. 287-292, 2016.

[4] Duncan, B., and Zhang, Y., "Neural Networks For Sentiment Analysis on Twitter," 2015 IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC), Beijing, pp. 275-278, 2015.

[5] Lee, J. H. and Lee, H. K., "A Study on Unstructured Text Mining Algorithm through R Programming based on Data Dictionary," Journal of the Korea Society Industrial Information System, Vol. 20, No. 2, pp. 113-12, 2015.

[6] Yun, B. H., "Natural Language Processing-based Information Extraction for Newspapers," Journal of Korean Institute of Information Technology, Vol. 6, No. 4, pp. 188-195, 2008.

[7] Chen, M. H., Chen, W. F., and Ku, L. W., "Application of Sentiment Analysis to Language Learning," in IEEE Access, Vol. 6, pp. 24433-24442, 2018.

[8] Day, M. Y., and Lin, Y. D., "Deep Learning for Sentiment Analysis on Google Play Consumer Review," 2017 IEEE International Conference on Information Reuse and Integration (IRI), San Diego, CA, pp. 382-388, 2017.

[9] Hassan, A., and Mahmood, A., "Deep Learning Approach for Sentiment Analysis of Short Texts," 2017 3rd International Conference on Control, Automation and Robotics (ICCAR), Nagoya, pp. 705-710, 2017.

[10] Jianqiang, Z., Xiaolin, G., and Xuejun, Z., "Deep Convolution Neural Networks for Twitter Sentiment Analysis," in IEEE Access, Vol. 6, pp. 23253-23260, 2018.

[11] Deng, L. and Dong, Y., "Deep Learning: Methods and Applications," NOW Publishers, United State of America, 2014.

[12] Aaron, Basic Korean Sentence Structure, 2014. [Online]. Available at http://keytokorean.com/classes/beginner/basic-korean-sentence-structure/ [Accessed 20 May 2017].

[13] Vidhya Content Team, Quick Guide: Steps to Perform Text Data Cleaning in Python, 2015. [Online]. Available at https://www.analyticsvidhya.com/blog/2015/06/quick-guide-text-data-cleanin Goodfellow-et-al-2016 [Accessed 20 May 2017].

[14] Tomar, S .S., Text mining in R: A Tutorial, 2017 [Online]. Available at https://www.springboard.com/blog/text-mining-in-r/ [Accessed 20 May 2017].

[15] Yuhang, Z., Yue, W., and Wei, Y.,

"Research on Data Cleaning in Text Clustering," 2010 International Forum on Information Technology and Applications, Kunming, pp. 305-307, 2010.

[16] Github, Twitter-Korean-text, 2014. [Online]. Available at https://github.com/twitter/twitter-korean-text [Accessed 20 May 2017].

[17] Quora, What Are All The Speech Levels of Korean and How Are They Used?, 2012. [Online]. Available at https://www.quora.com/What-are-all-the-speech-levels-of-Korean-and-how-are-they-used [Accessed 20 May 2017].

[18] Miachel, R0., 3 Steps of Text Mining, 2012. [Online]. Available at http://www2.cs.man.ac.uk/~raym8/comp38212/main/node203.html [Accessed 20 May 2017].

[19] GoodfellowI, Y., Bengio, and Courville, B., 2016. Deep Learning. MIT Press [Online]. Available at http://www.deeplearningbook.org [Accessed 20 May 2017].

[20] Nielsen, M., Using Neural Nets to Recognize Handwritten Digits, 2017. [Online]. Available at http://neuralnetworksanddeeplearning.com/chap1.html [Accessed 20 May 2017].

[21] Ruder, S., An Overview of Gradient Descent Optimization Algorithms, 2016. [Online]. Available at http://sebastianruder.com/optimizing-gradient-descent/ [Accessed 20 May 2017].

[22] Scikit Learn Team. 2016. Stochastic Gradient Descent [Online]. Available at http://scikit-learn.org/stable/modules/sgd.html [Accessed 20 May 2017].

[23] Blei, D. M., Ng, A. Y., and Jordan, M. I., "Latent Dirichlet Allocation," Journal of Machine Learning Research, Vol. 3, No. 5, pp. 993 - 1022, 2003.

[24] Wang, D., Thint, M., and Al-Rubaie. A., "Semi-Supervised Latent Dirichlet Allocation and Its Application for Document Classification," 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, pp. 306-310, 2012.

[25] Yee, C. S. and Ahmad, A. M., "Malay Language Text-Independent Speaker Verification using Nn-Mlp Classifier with Mfcc," 2008 International Conference on Electronic Design, Penang, pp. 1 - 5, 2008.

[26] Karim, M., Deep Learning via Multilayer Perceptron Classifier - Dzone Big Data, 2018. [Online]. dzone.com. Available at https://dzone.com/articles/deep-learning-via-multilayer-perceptron-classifier [Accessed 13 June 2018].

[27] Barde, B. V. and Bainwad, A. M., "An Overview of Topic Modeling Methods and Tools," 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, pp. 745-750, 2017.

[28] Ahn, H., "A Study on Compression of Connections in Deep Artificial Neural Networks," Journal of the Korea Industrial Information Systems Research, Vol. 22, No. 5, pp. 17-24, 2017.

[29] Nalini, S. Sandhya, K. Ganesh Kumar, P., "Enhancing Gender Classification in Social Networks", 2014 The International Industrial Information Systems Conference, pp. 251-256, 2014.

마 렌 드 라 (Adyan Marendra Ramadhani)

- 정회원
- Indonesia University of Education 자연과학대학 컴퓨터 과학 교육학과 학사
- Bandung Institute of Technology 전기 공학 및 정보학대학 전기공학 석사
- 부경대학교 IT융합응용공학과 대학원정보시스템 협동과정 석사
- 동아대학교 경영대학 경영정보학과 박사 과정
- 관심분야 : 인공지능, 정보시스템, 경영정보시스템, 블록체인

최 형 림 (Hyung Rim Choi)

- 정회원
- 서울대학교 경영학과 경영학학사
- 한국과학기술원 경영과학과 경영학 석사
- 한국과학기술원 경영과학과 경영학박사
- 동아대학교 경영대학 경영정보학과 교수
- 관심분야 : 항만물류, 유비쿼터스, RFID

임 성 배 (Seong Bae Lim)

- 정회원
- 경기대학교 경영학과 학사
- 서강대학교 경영학과 석사
- 네브라스카대학교 경영학과 박사
- 미국 St. Mary’s University 경영대학 경영학과 학과장
- 관심분야 : 혁신전략, 융합경영