Journal of Intelligent & Fuzzy Systems 40 (2021) 7745–7754
DOI:10.3233/JIFS-189594
IOS Press

Age group classification to identify the progress of language development based on convolutional neural networks

Byoung-Doo Oh[a,b], Yoon-Kyoung Lee[c], Hye-Jeong Song[a,b], Jong-Dae Kim[a,b], Chan-Young Park[a,b] and Yu-Seop Kim[a,b,*]
[a]*Department of Convergence Software, Hallym University, 1, Hallymdaehak-gil, Chuncheon-si, Gangwon-do, Republic of Korea*
[b]*Bio-IT Center, Hallym University, 1, Hallymdaehak-gil, Chuncheon-si, Gangwon-do, Republic of Korea*
[c]*Division of Speech Pathology and Audiology, Hallym University, 1, Hallymdaehak-gil, Chuncheon-si, Gangwon-do, Republic of Korea*

Abstract. Speech pathology is a scientific study of speech disorders. In this field, the study also analyzes and evaluates language abilities for the purpose of improving speech and hearing. Speech therapy first performs evaluation of speech ability, which is expensive. In order to solve this problem, software methodologies have been applied to language analysis, but most of them have been applied to only part of the whole process. In this study, the degree of language development is judged by determining the age group of the speaker (Pre-school children, Elementary school, Middle and high school, Adults, and Senior citizen) using deep learning and simple statistics. We use transcription data from the counseling contents and multi-kernel CNN model. At this time, in order to understand the characteristics of Korean language belonging agglutinative languages, experiments are carried out in words, morphemes, characters, Jamo, and Jamo with POS tag-level. And we analyze the distribution of the results for each sentence of the speakers to predict their age groups and to check the degree of language development. The proposed model shows an average accuracy of about 74.6 %.

Keywords: Language analysis, age group analysis, convolutional neural networks, deep learning, statistical analysis

1. Introduction

Language is a medium of communication, a way of expressing knowledge and ideas about society [1]. It develops through experience as humans grow and is an innate ability of all humans [2]. Speech pathology is conducting various studies on disorders, including sensory and motor disturbances shown in communication. These communication disorders are difficult to define symptoms and forms. This can range from birth defects to language developmental disorder related to children's native language learning. Communication disorders can also be applied not only to children but also to adults in similar situation [1].

In this situation, the most active area of this study in recent years is the diagnosis of children's language ability and treatment of disorders, because language skills develop explosively in pre-school age. Currently, one of the most commonly used methods for determining developmental disorders is language sample analysis. The language sample analysis is a method of collecting and analyzing speeches (utterances) generated from test tools such as counseling and play. This method predicts the language age of

*Corresponding author. Yu-Seop Kim, E-mail: yskim01@hallym.ac.kr.

ISSN 1064-1246/$35.00

speakers and determines the developmental disability by performing statistical analysis based on various indicators [3–6]. However, this method is expensive and time-consuming to record, transcribe and analyze the speaker's utterance. There is also a lack of clinical experts and therapies that can use this method [7].

In the United States, about 52,870 speech therapists with a master's degree or higher were reported in 1989 and about 85,000 speech clinical researchers in 1994. However, according to research on the demand of speech therapists in Korea, the demand for speech therapists and hearing therapists in 2020 is estimated to be about 60,000, but the supply is very low [7, 8]. In Korea, many studies have been conducted on the acquisition of specialists in speech therapy [9] and on securing expertise [10, 11].

In order to solve these problems, studies have recently been conducted to introduce a methodology using the software for reducing analysis costs, such as database construction, semi-automating the transcribing process, and analyzing data [12–14]. However, these studies provided only partial solutions to the processes required for linguistic analysis, and their effectiveness was limited because of their small portion in the whole process.

This study proposes a method of identifying the progress of language development by determining the speaker's age group (Pre-school children, Elementary school, Middle and high school, Adults, and Senior citizen) using deep learning methods and simple statistics. This aims to reduce the time and cost of analyzing language skills for speaker of all ages. First, the transcription data (documents) from the counseling process for speech therapy are divide into individual utterance (sentences). At this time, each consultation content includes all the conversations between the interviewer and the interviewee (speaker). Therefore, the dataset was constructed by extracting only the utterance of the interviewee.

After that, the individual utterances are learned through a convolutional neural networks (CNN) model. CNN is a technique that has made great progress in processing images and video as one of deep learning algorithms [15]. In fact, CNN has been widely used in image processing study such as image classification and face recognition [16, 17]. Recently, CNN has shown good performance in sentence classification along with recurrent neural network (RNN) based algorithms [18] which is commonly used for time-series data such as voice and text [19]. And also, various studies have shown that the performance of sentence classification using CNN is good when compared with another machine learning and deep learning algorithms [19, 20]. Representative examples are MV-RNN [21] and SVM [22]. MV-RNN (Matrix-Vector Recursive Neural Networks) is a proposed method to solve natural language processing problem. It is proposed on the assumption that if the combination of words to combine is different, the process of combining will be different. SVM (Support Vector Machine) is a linear algorithm that creates a distinguishable line when two groups exist. This method has a very solid and sophisticated theoretical background. And it attracted a lot of attention for its outstanding performance before deep learning. In addition, the various studies prove the effectiveness of the time-series data tasks such as sentence classification through the RCNN [23] that applied the CNN together with the RNN.

Since the data used for language analysis is composed of colloquial language by free utterance rather than refined written language, there are limitations in accurately classifying various age groups. Therefore, the predicted results for each utterance of the interviewee were statistically analyzed to determine the interviewee's progress of language development. In this way, the prediction result (age group) is compared the actual age of the interviewee to diagnose and evaluate the progress of language development.

Section 2 describes how to collect data, how to extract necessary data from collected data, and how to pre-process the extracted data. Section 3 describes the structure of the CNN model used as the sentence classifier and the model's learning process. And it also describes the simple statistical analysis based on the result of individual sentence classification. Section 4 describes the experimental results about evaluating the interviewee's progress of language development based on the learned sentence classifier using the utterance data. Finally, Section 5 briefly describes this study and describes the limitations and ways to resolve them.

2. Transcription data

2.1. Data collection

The data used in this study is transcription data collected by the Department of Speech-Language Pathology and Audiology, Hallym University[1]. The data is generated by the interviewer (masters and

[1]https://slp.hallym.ac.kr/slp/index.do

Table 1
The example of transcription data. All data is written in Excel, and the format is as follows. ⓐ: turn of the conversation where the sentence was generated during the consultation. ⓑ: total number of utterances. ⓒ: all utterances in the counseling process. ⓓ: the interviewee's utterance and the data extracted for use in this study. ⓔ: the interviewer's utterance

ⓐ Turn	ⓑ Num of Utterance	ⓒ Utterance
		(KR) A: 유치원이었지?
		ⓔ Interviewer : It was kindergarten?.
1	1	(KR) B: 네.
		ⓓ Interviewee : Yes.
	2	(KR) B: 친구들이 요리하고 있었어요.
		ⓓ Interviewee : My friends were cooking.
		(KR) A: 아 친구들이 요리하고 있었구나?
		ⓔ Interviewer : Oh, your friends were cooking?.
2	3	(KR) B: 놀고 있었어요.
		ⓓ Interviewee : They were playing..
		...

undergraduates course students in the Department of Speech-Language Pathology and Audiology) recording the counseling process for interviewee and transcribing the voice records. The interviewer removes the unnecessary parts of the analysis in the recording when transcribing consultation. In addition, actions that are not general utterances or meaning less utterances are marked with special symbols. Transcription is basically conducted on individual utterances exchanged between the interviewer and the interviewee. Through the transcription process, it is possible to measure the number and the length of speaker's utterance. Table 1 is an example of transcription data. The data used in this study is transcription data for a total of 148 people.

The significant information extracted from the transcription data is the interviewee's utterance shown in ⓓ, because this study wants to analyze the language age of the interviewee. In order to facilitate the extraction of an utterance, each utterance is marked with an interviewer ("A") and an interviewee ("B"). The structure of the transcription dataset used when collecting data is the same as in Table 1, and each interviewee contains 60 utterances on average.

2.2. Preprocessing

Since free utterance includes all utterances generated by the interviewees during the counseling process, each utterance must be separated in order to use it in the analysis [24, 25]. This study defines noise utterance by analyzing collected utterances and defines noise as "sentence structure not suitable for learning" and extracts noise based on this definition. Noise is marked with special characters when it is transcribed and is removed after transcription.

For example, behavioral words, duplicated utterances, and interjections are considered noises and are removed from the transcription file. We also remove the interviewer's intervening questions. Table 3 shows examples of these noises.

In addition, this study analyzes language age using simple utterances. The reason is that simple utterance occurs in various age groups, but the frequency of occurrence varies according to age group. As shown in Fig. 1, simple utterances tend to appear in younger age groups. In other words, the person uses more simple utterances, the lower the development of language ability is. Simple utterances mainly refer to simple positive and negative responses and short utterances of 3 words or less, and examples are seen in Table 2.

3. Methodology

In this study, the procedure is as shown in Fig. 2. First, the data is extracted from the transcription data and noise removal and labeling are performed. The CNN model trains each utterance to predict its age. After that, the prediction results of each interviewee obtained through the trained model is integrated and calculated by simple statistics, and then the age group for the interviewee is determined. Table 4 shows the number of interviewees and the total number of utterances extracted by age groups.

3.1. Characteristics of Korean

Korean is an agglutinative language unlike English and Chinese, where word order is important. Therefore, by attaching affixes to roots, semantic and

Table 2
The example of simple utterance

	Single utterance	Description
KR	"네", "맞아요", ···	Simple positive
EN	"Yes", "Right", ···	
KR	"아니요", "싫어요", ···	Simple negative
EN	"No", "I don't want to", ···	
KR	"그런 것 같아요" "그것밖에 기억이 안나요" "형이랑 같이 놀아요" "그 다음엔 몰라요" ...	Sentence less than n words ($n \leq 3$)
EN	"I think so" "I can only remember it" "I play with my brother" "I don't know after that" ...	

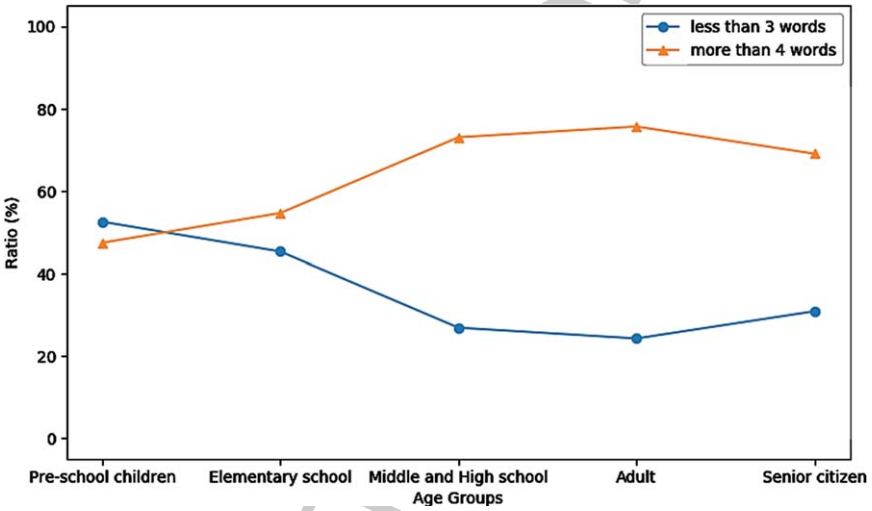

Fig. 1. The average frequency of single utterance in each age group.

Table 3
The examples of noise data

	Word	Description
	(3")	No speech for 3 seconds
KR	"(어)", "(음)", ···	Interjection
EN	"(uh)", "(um)", ···	
KR	"(네)", "(아니오)", ···	Overlapping word
EN	"(yes)", "(No)", ···	

grammatical functions are determined. For this reason, in the study of natural language processing in the Korean, morphemes are mainly the smallest unit of analysis [26, 27]. Recently, however, many studies have been attempted using characters or Jamos (letters) as well as word or morpheme for embedding [28–31]. Jamo is a segmental symbol of a phonemic writing system[2]. In Hangul (Korean alphabet), two or three Jamos are gathered together to form a character. Therefore, this study compares the performance of words, morphemes, characters, and Jamos. In addition, this study tried to check the possibility of more detailed analysis when using the part-of-speech (POS) tag together with the Jamo. Table 5 shows an example of analyzing sentence according to each unit in Korean.

3.1.1. Korean morpheme analyzers

To date, various morpheme analyzers for Korean have been developed. Among them, the most widely

[2]https://en.wikipedia.org/wiki/Letter_(alphabet)

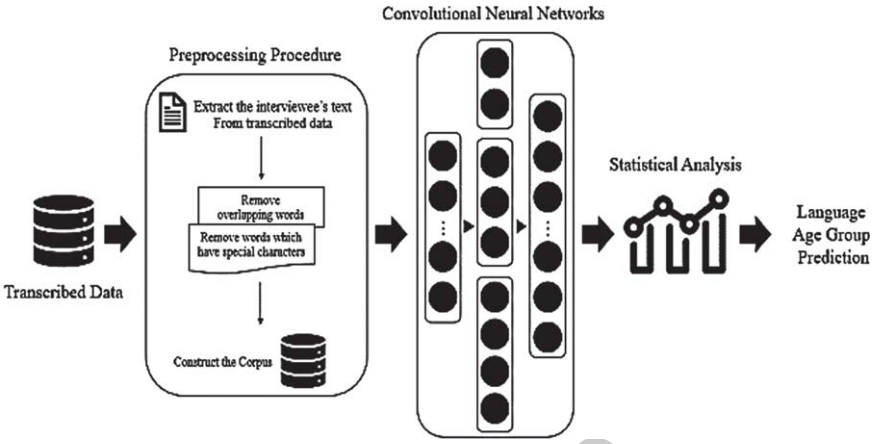

Fig. 2. The whole process of the proposed method for classifying language age groups.

Table 4
The number of utterance and interviewees by age group

Age group	Number of utterances	Number of interviewees
Pre-school children	2609	36
Elementary school	2235	25
Middle and High school	4007	28
Adult	2487	39
Senior citizen	1294	20
Total	12632	148

used analyzers are Kkma[3], Komoran[4], Hannanum[5], etc. These are all open source Korean morphological analyzer. This paper uses Komoran. Komoran is an open-source Korean morphological analyzer written in java, and has been developed by Shineware[6], since 2013.

3.2. Convolutional neural networks

This study predicts the language age of individual utterances using a convolutional neural network. In addition, the predictive results are statistically analyzed to predict the language age group of the interviewees. CNN have been used mainly in the field of image recognition and processing, but recently CNN shows good performance in the field of natural language processing. In this study, we used the model proposed in [19], which showed good performance on sentence classification.

3.2.1. Sentence matrix

The input data commonly used in CNN is an image. Therefore, there are a few things to be aware of when using natural language on CNN. First, the words (w_i) constituting the sentence are represented by a k-dimensional vector $\left(w_i \in R^k\right)$. As shown in Fig. 3, if the number of words in the sentence is i, the sentence matrix can be defined as a matrix of i x k. Second, the images used as the input of the CNN are processed with the same size. Likewise, sentences in natural language do not occur in the same length. Therefore, it is generally processed by the same sentence length using zero padding. For example, the maximum length of a sentence (S_{max}) that exists in a document is N, and the length of any one sentence (S_j) is n. And then, if $n < N$, the max length of S_j (n) is changed to N and then w_{n+1} to w_N constituting the S_j are represented as 0 (zero padding). Therefore, the length of the sentence is expressed equally.

When generating the sentence matrix, researchers can use pre-trained word vector representations (word embeddings) [32, 33]. Therefore, this study uses the pre-trained word embedding model trained by Kookmin University's Korean corpus[7] and Word2Vec [32]. Kookmin University's Korean corpus was created by scraping domestic media articles and contains about 180 million words. And Word2Vec was developed based on neural networks. This method represents words as vectors by calculating the relationship between words that appear in the context.

[3] http://kkma.snu.ac.kr/

[4] https://www.shineware.co.kr/products/komoran/

http://semanticweb.kaist.ac.kr/home/index.php/Home

[6] https://shineware.tistory.com/

[7] http://nlp.kookmin.ac.kr/kcc/

Table 5
The example of the analysis unit in Korean

Origin Sentence	KR	"나는 학교에 간다"
	EN	"I go to school"
Unit	Example	
Word	"나는", "학교에", "간다"	
Morpheme	"나"+"는", "학교"+"에", "갈"+"ㄴ다"	
Character	"나", "는", "학", "교", "에", "간", "다"	
Jamo	"ㄴ", "ㅏ", "ㄴ", "ㅡ", "ㄴ", ···	
Jamo with POS tag	"ㄴ/NP", "ㅏ/NP", "ㄴ/JX", "ㅡ/JX", "ㄴ/JX", ···	

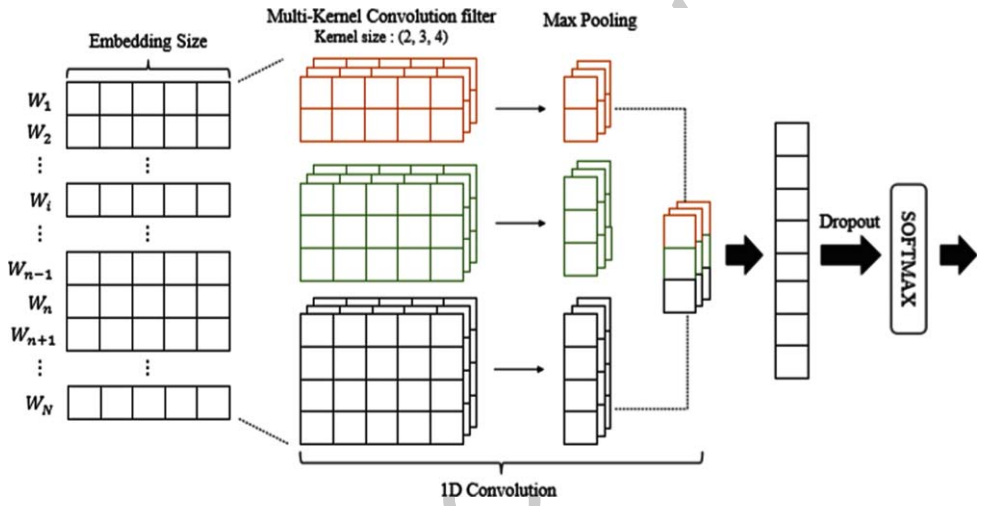

Fig. 3. The structure of multi-kernel CNN.

3.2.2. *Convolution layer*

This study utilizes the structure proposed in [19]. This structure consists of the convolution layer and the pooling layer. The convolution layer has kernels having three different sizes to generate a feature map by performing a convolution operation. And the pooling layer, the feature vector is extracted from the feature map. In [20], Zhang et al. found that a multi-kernel composed of approximations to an optimal single magnitude value shows good performance. So, this study constructed a multi-kernel CNN that approximates a single magnitude value that shows optimal performance.

Therefore, in this study, as shown in Fig. 3, the multi-kernel of the convolution layer calculates the convolution by dividing the word embedding (sentence) according to each kernel size (number of words). After that, the max-pooling layer is used to extract important features from the feature map of multi-kernel obtained through the result of the convolution layer. At this time, the important feature is the maximum value extracted according to the kernel sizes from the feature map generated by the convolution.

In addition, the convolution layer and the max pooling layer are based on 1-dimension. 1-dimension means each row in the sentence matrix. The extracted feature vector is normalized and used as input of a fully-connected layer. Fully-connected layer uses softmax function to calculate the result of the classification.

3.2.3. *Regularization*

The most common regularization method in CNN is Dropout. Dropout is a representative regularization method that solves overfitting problems in the deep neural networks (DNN) [34]. Studies have shown that the dropout ratio is optimal at 0.4 to 0.6 [34].

Therefore, in this study, regularization is performed by applying dropout method between the convolution layer and the fully-connected layer.

Table 6
The example of simple statistic results

$Interviewee_A$	
Total utterance	50
Class	P_{group} (%)
Pre-school children	6
Elementary school	4
Middle and high school	10
Adult	**70**
Senior citizen	10
True	Adult
Pred	Adult

3.3. Statistical analysis

The purpose of this study is to identify the language development by analyzing the age group through the all utterance of the interviewees. However, multi-kernel CNN models learn by sentence-level. So, the age group is predicted by sentence-level. Therefore, the final age group is determined by integrating the prediction results of the sentences of individual interviewees. For example, sentences of the $Interviewee_A$ are used as input to obtain prediction results for each sentence. The frequency is then checked by integrating the prediction results. Therefore, the age group with the highest frequency (P_{Group}) is determined as the age group of $Interviewee_A$.

$$P_{group} = \frac{num\ of\ prediction\ to\ each\ group}{Total\ utterance\ in\ Interviewee_A} \times 100$$

The denominator of P_{Group} is the total number of utterances in $Interviewee_A$. And the numerator of P_{Group} is the number of age groups among the prediction results of the individual utterances in $Interviewee_A$.

Table 6 shows the prediction result about $Interviewee_A$ through statistics of the results from CNN model. $Interviewee_A$ shows a total of fifty utterances. Fifty utterances are individually predicted of language age through a trained CNN model. As a result, 70 % of all utterances are predicted to be Adults. Therefore, $Interviewee_A$ is predicted as Adult, which turns out to be correct.

4. Experimental results

This study trained multi-kernel CNN models using the individual utterance of interviewees. First, this stu dy identified the optimal hyper-parameters for learning efficiency. In addition, the dataset was divided into train dataset, cross-validation dataset, and test dataset. When dividing the dataset, it was randomly divided. These ratios are 6:2:2 respectively. The partitioning was based on the interviewee, not on individual utterances, and the selected hyper-parameters are shown in Table 7.

In this study, the experiment was repeated 10 times for each unit of Korean. In Table 5 shows an example for each unit. So, the performance of each unit is shown in Table 8.

As a result, the performance of the model using the morphemes shows the best performance of 74.7% on average. The results predicted by the models indicate that utterances predicted as Pre-school children's utterances are consistent across all ages. The reason for this is that, as shown in Table 3, the ratio of simple utterances is relatively high in Pre-school children and Elementary school students, so it is judged that a simple utterance was trained as Pre-school children or Elementary school students. In Senior citizens, the distribution of prediction results is wide. This may be considered that the deterioration of language ability is progressing with age [35].

As shown in Table 9, in Pre-school children, everyone is classified correctly. Similarly, Middle and high school students and Adults are predicted correctly with high probability. On the other hand, Elementary school students and Senior citizens have only 40% and 31% probability of prediction, respectively. The reason for this difference is inferred as follows. First, in Elementary school students, language development is very active in terms of syntax, semantics and pragmatics. It is easy to see that some Elementary school students with fast language development already use language at the Middle and high school level. In the experiment, half of Elementary school students are identified as Elementary school students and the other half as Middle school and high school students. Second, Senior citizens are over 60 years old, so the age range is very wide. Senior citizens of this age begin to have degenerated language ability. And the degree of degeneration varies greatly depending on the state of health. In recent years, the deterioration rate of language ability in the elderly is much slower than before due to the improvement of medical and living standards. For this reason, many seniors maintain similar language abilities as adults. Therefore, it appears that many healthy seniors participated in the interview, causing this phenomenon.

Table 7
Hyper-parameters

Hyper-parameter	Best value	Experiment range
Training epochs	20	5 – 100
Batch size	32	10 – 100
Embedding dimension	100	100 – 300
Number of filters	128	50 – 300
Kernel size (1D-Convolution)	(2, 3, 4)	2 – 6
Pooling method (1D-Convolution)	Max	Max, Average, GlobalMax, GlobalAverage
Dropout rate	0.5	0.4 – 0.6

Table 8
The performance results of the proposed model. The best performances in each column is shown in bolded

Level	Acc (Average: 10 times)	Acc (Minimum)	Acc (Maximum)
Word	0.691	0.556	0.756
Morpheme	**0.747**	**0.645**	0.867
Character	0.736	0.534	**0.889**
Jamo	0.624	0.4	0.756
Jamo with POS tag	0.184	0.067	0.267

Table 9
The average distribution for the best performance of experimental results. The bolded one is the largest predictions for each age group

True \ Pred	Pre-school children	Elementary school	Middle and high school	Adult	Senior citizen
Pre-school children	**1.0**	0.0	0.0	0.0	0.0
Elementary school	0.1	0.4	**0.49**	0.01	0.0
Middle and high school	0.01	0.05	**0.94**	0.0	0.0
Adult	0.0	0.0	0.03	**0.91**	0.06
Senior citizen	0.0	0.01	0.03	**0.65**	0.31

5. Conclusion

This study proposes a method using the multi-kernel CNN model and simple statistical analysis to predict language age and developmental disorder, which is an important problem in the field of speech therapy. First, a multi-kernel CNN model is used to learn the age group for individual utterances of interviewees included in train dataset. The trained multi-kernel CNN model predicts the age group for individual utterances of interviewees included in test dataset. After that, the prediction results of each interviewee are integrated and calculated by simple statistics, and then the age group for the interviewee is predicted. In this way, instead of using various indicators used in the language sample analysis, it is possible to make identify quickly through only the conversation contents (utterances) of a person.

The methodology of this study has three advantages. First, it does not require a separate language analysis process. The existing methodology included language analysis tools in the process of using the software. In this case, the performance of the language analysis tool greatly influences the overall result. This study completely eliminated this risk. Second, significant features can be found in each age group. By analyzing the prediction result about the individual by using the model trained in this study, it can find out the significant features of the corresponding age group. Finally, not only people with late language development but also people with early language development can be identified.

However, the following new problems were found. First, there is no clear criterion for judging speech disorders. The methodology proposed in this study can predict the degree of language development, but there are limitations in determining the disorder completely. Second, the interviewee's answer is possible to be influenced by the interviewer's question.

Therefore, in the future, it is necessary to collect more data and to build a deep learning model suitable for Korean colloquial language. And the research is needed to establish a criterion for judging speech disorder based on the prediction results. It is also expected that learning the interviewer's questions together in the deep learning model will allow us to more accurately identify the progress of language development. Through this, it is expected that lan-

guage disorders can be correctly identified without complicated analysis.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2019R1A2C2006010).

References

[1] D. Crystal and V. Rosemary, Introduction to language pathology, *John Wiley & Sons*, 2013.
[2] M.Y. Kang, Language: a platonic view, *Studies in Generative Grammar* (2000), 3–36.
[3] M.J. Kim, S.Y. Pae and D.H. Ko, Syllable and phoneme frequencies in the Spontaneous speech of 2-5 year-old Korean children, *Speech Sciences* (2001), 8 [in Korean].
[4] J.Y. Shin, Phoneme frequency of 3 to 8-year-old Korean children, In *Proceedings of the KSPS conference* (2005), 15–19 [in Korean].
[5] H.J. Lee and Y.T. Kim, Measures of utterance length of normal and language-delayed children, *Communication Sciences & Disorders* **4**(1) (1999), 1–14 [in Korean].
[6] Y.H. Jung and M.S. Yoon, Mean length of utterance for typically developing children of 2 to 4 years, *Korean Journal of Early Childhood Special Education* **13**(13) (2013), 55–73 [in Korean].
[7] H.R. Yoon, S.B. Kwon, S.J. Kim, S. Pae, M.S. Shin and M. Hwang, Survey on the status of speech-language pathology curriculum and development of certification standards, *Korean Journal of Communication Sciences & Disorders* **15**(15) (2010), 271–284 [in Korean].
[8] S.I. Jeon, The importance of medical rehabilitation of problems in Korea, *Korean Journal of Disability and Employment* **5**(2) (1995), 4–8 [in Korean].
[9] Y.T. Kim, Training and certificate model of the speech-language pathologist in Korea, *Korean Journal of Rehabilitation Welfare* **6** (2002), 1–17 [in Korean].
[10] Y.S. Jeon, D.Y. Kim, Y.W. Kim and H. Kim, Core capacities of speech-language pathologists in Korea, *Communication Sciences & Disorders* **18**(1) (2013), 1–11 [in Korean].
[11] Y.T. Kim, A proposal for the national certificate system of language therapist in Korea, *Communication Sciences & Disorders* **12** (2007), 379–393 [in Korean].
[12] S.J. Kim, J.M. Kim, M.S. Yoon, M.S. Chang and J.E. Cha, The development of Korean corpus and lexical database for language assessment and intervention of preschoolers, Nazarene University, 2014 [in Korean]. Retrieved from https://www.krm.or.kr/krmts/link.html?dbGubun=SD&m201_id=10026760&res=y
[13] D.J. Won, M.S. Chang and S.M. Kang, Implement of semi-automatic labeling using transcript text, *Journal of Korean Institute of Intelligent Systems* **25**(6) (2015), 585–591 [in Korean].
[14] S. Ha, A. So and S. Pae, Speech and language development patterns of Korean two-year-old children from analysis of spontaneous utterances, *Communication Sciences & Disorders* **21**(1) (2016), 47–59 [in Korean].
[15] Y. Lecun, Y. Bengio and G.E. Hinton, Deep learning, *Nature* **521**(7553) (2015), 436.
[16] A. Krizhevsky, I. Sutskever and G.E. Hinton, ImageNet classification with deep convolutional neural networks, In *Advances in neural information processing systems*, (2012), 1097–1105.
[17] S. Lawrence, C.L. Giles, A.C. Tsoi and A.D. Back, Face recognition: A convolutional neural network approach, *IEEE Transactions on Neural Networks* **8**(1) (1997), 98–113.
[18] A. Graves, A.R. Mohamed and G.E. Hinton, Speech recognition with deep recurrent neural networks, In *2013 IEEE international conference on acustics, speech and signal processing, IEEE* (2013), 6645–6649.
[19] Y. Kim, Convolutional neural networks for sentence classification, arXiv preprint arXiv:1408.5882, 2014.
[20] Y. Zhang and B. Wallace, A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification, arXiv preprint arXiv:1510.03820, 2015.
[21] R. Socher, B. Huval, C.D. Manning and A.Y. Ng, Semantic compositionality through recursive matrix-vector spaces, In *Proceeding of the 2012 joint conference on emprirical methods in natural language processing and computational natural language learning*, (2012), 1201–1211.
[22] C. Cortes and V. Vapnik, Support-vector networks, *Machine Learning* **20**(3) (1995), 273–297.
[23] S. Lai, L. Xu, K. Liu and J. Zhao, Recurrent convolutional neural networks for text classification, In *Twenty-ninth AAAI conference on artificial intelligence*, 2015.
[24] B.S. Ahn, A study on utterance as prosodic unit for utterance phonology, *The Journal of Korean Studies* **26** (2007), 233–259 [in Korean].
[25] J.M. Kim, Utterance boundary classification in spontaneous speech of pre-school age children, *Language Therapy Research* **22** (2013), 41–54.
[26] Y.B. Kim, H. Chae, B. Snyder and Y.S. Kim, Training a Korean SRL system with rich morphological features, In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics* **2** (2014), 637–642.
[27] C.H. Han, N.R. Han, E.S. Ko and M. Palmer, Development and evaluation of a Korean treebank and its application to NLP, SIMON FRASER UNIV BURNABY (BRITISH COLUMBIA) DEPT OF LINGUISTICS, 2002.
[28] S. Choi, T. Kim, J. Seol and S.G. Lee, A syllable-based technique for word embeddings of Korean words, arXiv preprint arXiv:1708.01766, 2017.
[29] S. Kwon, Y. Ko and J. Seo, A robust named-entity recognition system using syllable bigram embedding with eojeol prefix information, In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, ACM*, (2017), 2139–2142.
[30] H.B. Shin, M.G. Seo and H.J. Byeon, Korean alphabet level convolution neural network for text classification, In *Proceedings of Korea Computer Congress 2017*, (2017), 587–589 [in Korean].
[31] W.I. Cho, S.M. Kim and N.S. Kim, Investigating an effective character-level embedding in Korean sentence classification, arXiv preprint arXiv:1905.13656, 2019.
[32] T. Mikolov, K. Chen, G. Corrado and J. Dean, Efficient estimation of word representations in vector space, arXiv preprint arXiv:1301.3781, 2013.
[33] J. Pennington, R. Socher and C. Manning, GloVe: global vectors for word representation, In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (2014), 1532–1543.

[34] N. Srivastava, G.E. Hinton, A. krizhevsky, I. Sutskever and R. Salakhutdnov, Dropout: a simple way to prevent neural networks from overfitting, *The Journal of Machine Learning Research* **15**(1) (2014), 1929–1958.

[35] J. Kim and H. Kim, Communicative ability in normal aging: a review, *Communication Science & Disorders* **14** (2019), 495–513 [in Korean].