A Novel Framework for User Intent Detect

Hyunjung Lee*, Jeongpil Lee**, Myoung-Wan Koo**
Department of {*English, **Computer Science and Engineering}
Sogang University
Seoul, Korea
hyunjlee@sogang.ac.kr, koreanfeel@gmail.com, mwkoo@sogang.ac.kr

***Abstract*—One of the main modules of Spoken Language Understanding is the user intent detection, which allows pragmatic meaning of users' utterances to be classified into a Dialogue Act class. Our study proposes a novel framework for detecting users' intent more articulately without any sophisticated techniques. For purpose of our research, we collected a set of Korean dialogues on the topics of schedule management and manually annotated the utterances with economical and practical dialogue acts. We preprocessed the basic units of Korean version of Global vector (GloVe) corpus by utilizing the technique of WordPiece Model (WPM) to segment the spaced-word units into sub-word units. It is observed that the method of segmenting words by WPM is effective to deal with Korean data, an agglutinative language with various suffixes. Our Convolutional Neural Network classifier has the accuracy of 85.2% in detecting the user's intention, which is higher than its counterpart without the WPM.**

***Index Terms*—Global Vector, WordPiece Model, Spoken Language Understanding, Dialogue Act, User Intent Detect**

I. Introduction

In spoken dialogue system (SDS), detecting a user's intention from natural language has been one of the most fundamental issues in Spoken Language Understanding (SLU) [1]. In order to achieve successful SLU, there is need to capture pragmatic intention and to extract semantic meanings from an almost infinite variety of dialogue participants' utterances. Thus, the SLU component must create a mapping between the natural language inputs and semantic representations that correspond to users' intention to correctly instruct system to know what to do in the next turn. The semantic representation must include a robust structure [2], which enables to interpret the genuine meaning of transcribed texts in the domain of specific topics. However, most prior works have focused on developing a classifier to train and learn the mapping between utterances and semantic representations, where the annotation framework still remains predefined and disorganized. In an attempt to establish a novel annotation framework, we adopted and revised the dialogue act set of the second dialogue state tracking challenge (DSTC2) due to its theoretical economy [3].

In addition to improving an annotation approach, measuring semantic similarity of words and representing them in the vector space have taken precedence in order to input refined transcribed texts to SLU. Vector space models such as Global Vector (GloVe) and Word2vec have been widely used to represent each word as a real-valued vector [7]. The vector keeps track of the context (e.g. co-occurring words) in which target terms appear in a large corpus as proxies for semantic representations, and apply geometric techniques to these vectors to measure the similarity in meaning of the corresponding words. In this study, we tackle an issue to enhance the representation ability of Korean version of GloVe with simple preprocessing of a corpus by Wordpiece Model (WPM). It is demonstrated that the result of WPM is promising to improve the speech recognition accuracy with acoustic data for languages which have no spaces between words [9]. However, no attempts have been made to investigate that the ability of WPM is applicable in tokenizing units in the corpus of vector space models. This paper will examine whether a sub-word unit segmented by WPM facilitates to extract features as an input of our convolutional neural network (CNN) classifier.

Throughout the paper, we propose a novel framework for user intent detect. The aims of this study are two folds: to build a Korean dialogue corpus annotated with articulated and economical tagging sets, and to conduct a classification task with CNN on top of GloVe, nothing that CNN has been shown state-of-the-art performances on a text categorization task [12]. In order to fulfill this objective, we will briefly review previous studies of speech acts, WPM, and GloVe in Section 2. Section 3 will present the details of our framework for detecting users' intention including the architecture of a Convolutional Neural Network (CNN) classifier on top of GloVe. Section 4 will explain the result of our classifier. In Section 5, we will discuss main findings of our study and our future research.

This work was supported by the National Research Council of Science & Technology (NST) grant by the Korea government (MSIP) (No. CRC-15-07-KIER).

II. RELATED WORKS

A. Dialogue Act

Understanding users' speech intentions enables a dialogue system to act more adequately. Various tags expressing user intention force have been proposed as speech act tags or dialogue acts [4]. Tags are given to each utterance so that a corpus has been built with dialogue acts. Such an annotated corpus is subsequently able to be utilized as a train and test set for a classifier to successfully map between users' utterances and corresponding tags. Most of traditional frameworks for speech acts such as Dialog Act Markup in Several Layers (DAMSL) have adopted speech act categories [5]. Those categories are ideally devised for human-to-human dialogues and revised them to be capable of annotating human-to-robot dialogues. However, scrutinizing and analyzing the illocutionary force in terms of users' utterances is not necessarily enough to determine system's next responses and operations. For example, the table below shows the limit of the existing annotation scheme to capture genuine users' intention.

TABLE I. AN EXAMPLE OF TAGGED SENTENCES

	Transcription	**DAMSL scheme**	**System's ideal action**
1	What's going on tomorrow?	Info-Request	READ schedule on tomorrow
2	Would you read my schedule on tomorrow?	Open-option	
3	I need to check things to do tomorrow.	Assert	
4	I was wondering if I have plans for tomorrow.	Other-statement	
5	Read my tommorw's schedule.	Action-Directive	

The five sentences in table 1 which are annotated differently each other: *What's going on tomorrow?* is tagged as 'Info-Request', and *Would you read my schedule on tomorrow?* is annotated as 'Open-option'. In the similar way, 'Assert' is mapped into *I need to check things to do tomorrow*', *I was wondering if I have plans for tomorrow* is so as 'Other-statement', and *Read my tomorrow's schedule* is tagged as 'Action-directive'. Here, the most important thing to successfully achieve SDS is let system know what to do in the next turn: "Read schedule on tomorrow in calendar app." Tagging dialogue acts in the way of DAMSL yields an additional processing for a system to decide a reasonable action in response of users' saying. Rather, a theoretically economical tagging framework is required to be set up in order that a system can map a sentence of users' utterance into a signal that directly instructs a system for an ideal action. We will introduce a newly designed framework for annotating user's utterance in Section 3.

B. Wordpiece Model (WPM)

Since the acoustic recognizer converts acoustic stream of spoken utterances into a continuum of texts, the fundamental question arises which segmented (or tokenized) elements would be the most appropriate units to represent semantic meanings. WPM is suggested as one of segmentation model, which inserts boundaries that divide transcribed texts into *sub-word units*, which is proposed by Google Voice Search [9]. It was first suggested to resolve segmentation problems in Japanese and Korean, in which have difficulty in segmenting words since they have few or no space between morphemes or words. This model is generated using a data-driven approach to maximize the likelihood of forming a sub-word unit of the training data [9].

The inventory of WPM consists of sub-word units whose probability of occurrence exceeds a certain threshold in the following ways. First, the model sets up a word inventory consisting of Unicode characters (e.g. Hangul for Korean). Then, a model is constructed for the training data based on the initial inventory. Subsequently, a model generates a new unit by combining units in the current inventory which maximize the likelihood, and this process is continued when the likelihood of being a new unit falls below a predefined threshold.

Since the sub-word units in the inventory contain all the vocabularies in corpus, WPM is able to segment sentences without giving rise to out-of-vocabulary words (OOVs) problems [9]. This inventory is effective for language modeling due to its simplicity of generating a word inventory.

C. Global Vector (GloVe)

There are two major model families for learning vector space representations of words: global matrix factorization methods and local context window methods. The former is latent semantic analysis and the latter the skip-gram model [6]. Despite their successful applicability in diverse fields, it is currently observed that the two leading models have problems: global matrix factorization methods show relatively poor performance on a word analogy task and local context window methods do not involve corpus statistics.

Along these lines, Global Vector model is suggested as a solution to the problem. GloVe proposes a specific weighted least squares model, which is designed to train on global word-word co-occurrence counts. It represents each word w $\in V_W$ and each context c $\in V_C$ as *d*-dimensional vectors x and c as in the equation below: We use F(w,c) to denote the number of co-occurrences of pair (w,c).

$$\vec{w} \cdot \vec{c} + b_w + b_c = \log\big(F(w,c)\big) \forall\, (w,c) \in D \quad (1)$$

b_w and b_c (scalars) are word/context-specific biases, and at the same time are parameters to be learned in addition to w and c.

Training is performed on word-word co-occurrence statistics from a corpus, and the resulting representations showcase interesting linear sub-structures of the word vector space. Pennington et al. (2014) proposed a new weighted least squares regression model as in the equation (2), where V is the size of the vocabulary, $X \in R^{VxV}$ is a word co-occurrence matrix [7].

$$\text{f(x)} = \begin{cases} (x/x_{\max})^{\alpha} & if\ x < x_{max} \\ 1 & otherwise \end{cases} \quad (2)$$

X_{ij} is the frequency of word i co-occurring with word j. In the equation (3), f(x) is a weighted function [7]:

$$\mathrm{J} = \sum_{i,j=1}^{V} f\left(X_{ij}\right)\left(w_i^{\tilde{T}} w_j + b_i + \widetilde{b_j} - \log X_{ij}\right)^2 \ (3)$$

We developed Korean version of GloVe [8] and the details are presented in Section 3.

III. A Framework for User Intent Detect

A. Corpus Development

Since there had been no dialogue corpus in Korean, building a dialogue corpus with transcribed texts is the highest-priority task in our research. We set up a forum of conversation with 20 test subjects using a Wizard of Oz methodology; we created an environment where a test subject believes to interact with a computer system which is actually a hidden operator [10]. Specifically, a test participant may think he or she is communicating with a computer using a speech interface; while the participant's sayings are being secretly transmitted into the computer of an operator (i.e. "wizard") in another room, and the operator types proper response in the texts which are transformed into acoustic waves into the computer of the test subject. This methodology helps gain relevant data of human-to-robot dialogues for training a SDS.

In this situation, we asked each test subjects to perform 15 dialogues with accordance to the precise tasks on the topic of schedule management. The task consists of 4 main system actions: create (35%), read (43%), update (10%), and delete (12%) schedules. In each task, the details are stated so that the test participant know which element he or she should say to complete a certain dialogue such as start date, alert, event title, location, and so on. 300 dialogues were collected and the corpus contains 1093 sentences of users' utterances. Three graduate students majored in linguistics manually annotated the transcribed sentences with previously defined dialogue acts described in the following Section.

B. Annotation Framework

As mentioned in Section 2, a theoretically economical tagging framework is extremely in demand so as to create a simple and direct mapping between a sentence of user's utterance and a dialogue acts. To build an annotation set which is more appropriate for a system to respond naturally, we adopted the framework of DSTC2 [3] and revised dialogue tags which are effective enough to extract users' intention and are specialized enough to determine a system operation.

The table below summarizes our revised version of dialogue act set for user utterances. The dialogue act set is designed by reducing unnecessary several tagging layers and directly representing the user's intention in terms of the most appropriate response for a system in the next turn. For example, all the five sentences in Table 1 ('*What's going on tomorrow?*', '*Would you read my schedule on tomorrow?*', '*I need to check things to do tomorrow.*', '*I was wondering if I have plans for tomorrow*', '*Read my tomorrow's schedule.*') are tagged into a single dialogue act, according to our annotation scheme: inform {(system_action, read), (date, tomorrow)}. A system can achieve 2 goals at a time; it can both directly understand the meaning of those 5 users' utterances and prepare to operate READ SCHEDULE action.

TABLE II. Dialogue Act Sets for User's Utterances

	Act	**Slot**	**Definition**
1	ack	empty list	An acknowledgement e.g. "okay"
2	affirm	empty list	An affirmative reply to the system's previous utterance e.g. "yes"
3	bye	empty list	A goodbye message of the user indicating that the conversation is finished
4	negate	empty list	A negative statement or a denial of the system's previous utterance e.g. "no"
5	null	empty list	Something not understandable to the system; outside its domain e.g. "pineapple"
6	repeat	empty list	A request for the system to repeat what it just said e.g. "please repeat that"
7	reqalts	one pair (s, v)	Requesting for changing the existing value of the slot; asking for alternative suggestions of the given value of the slot e.g. "are there any others"
8	restart	empty list	Asking the system to restart from the beginning e.g. "let's start again"
9	thankyou	empty list	User thanking the system e.g. "thanks"
10	deny	one pair (s, v)	Informing that the user does not want the value v for a certain slot s. s must be an informable slot and v a possible value for s as specified in the ontology.
11	request	one pair ("slot", s)	Asking the system for the value of the requestable slot according to the ontology.
12	inform	one pair (s, v)	Notifying that the value of slot s to be accepted as v
13	hello	empty list	Greeting the system e.g. "hi"
14	reqmore	empty list	Asking the system if the user can request more information

Consequently, our framework is more effective and economical due to the fact there is no need for additional processing to determine a system operation. Furthermore, this tagging set is expected to applied in a wide variety of domains for detecting the intentions of users, since this framework consists of two types of objects; it thoroughly divides two roles *pragmatic intention* and *semantic information* into two categories: *dialogue act* and *slot-value pair*(s). When it comes to expand a topic of dialogues, a new ontology is only needed to superinduce additional semantic information to the primitive knowledge of a system, without adding new categories of dialogue act.

C. Korean version of GloVe

We built GloVe on our own Korean corpus which we constructed for this study [8]. 100 million sentences (668,284,389 tokens) were collected by a web crawler from internet bulletin boards of the Korean web sites. Here, GloVe initially took its basic units of the corpus by spacing between words (i.e. *spaced-word units*). Spaced-word units that appeared less than 100 times in the corpus were ignored, resulting in vocabularies of 366,940 million.

To test the reliability of Korean version of GloVe [8], an experiment is conducted for word analogies and word similarity in which a Glove, a continuous bag-of-words (CBOW) model as well as a skip-gram model are compared. For the word similarity task, we obtained the Pearson correlation coefficient of 0.3133 compared with the human judgement in GloVe, 0.2637 in CBOW and 0.2177 in SG. For the word analogy task, the overall accuracy rate of 67% in semantic and syntactic relations was obtained in GloVe, 66% in CBOW and 57% in SG [8]. Thus, GloVe is advantageous to calculate semantic similarity of Korean spaced-word units and further discover linear relationships among them. In the following section, we will compare two kinds of basic units of Korean GloVe: a *spaced*- and *sub-word unit*.

D. Tokenization

Before converting the texts into vector spaces using GLoVe, we preprocessed the basic units of Korean GloVe corpus with WPM, by segmenting *spaced-word units* to the extent to maximize the likelihood of *sub-word units* (i.e. wordpieces). Sentences are converted into following corresponding word sequences.

TABLE III. AN EXAMPLE OF WORDPIECE SEQUENCES

Data		Transciption	Annotation
(1)	Spaced-word units	*na nayil sanchayk hanun* I tomorrow a walk take *ke ilceng* ***tunglokhay.cwe*** thing schedule set.up.imperative “Set up a schedule titled ‘take a walk’ tomorrow.”	inform {(system_action, create), (date, tomorrow), (event_title, take a walk)}
	Sub-word units	*_na_ _nayil_ _sanchayk_ _hanun_* *_ke_ _ilceng_* ***_tunglok hay cwe_***	
(2)	Spaced-word units	***tunglokhay.cwulay*** set.up.hortative “Would you set up a schedule, please?”	inform {(system_action, create)}
	Sub-word units	***_tunglok hay cwu lay_***	
(3)	Spaced-word units	*ke ilceng* ***sakcey.cwe*** that schedule delete..hortative “Delete that schedule.”	inform {(system_action, delete)}
	Sub-word units	*_ke_ _ ilceng_* ***_ sakcey cwe_***	

Compared with the original data, it is profitable to reduce the number of types by preventing the inventory from containing every possible spaced-word units that agglutinate various verbal suffixes. According to the traditional tokenization approach, *tunglok*, *tunglokhay*, *tunglokhaycwe*, *tunglokhaycwulay*, and *sakcey*, *sakceyhaycwe* are regarded as 6 distinctive units in the inventory. As illustrated in the above examples in Table 3, the spaced-word unit *tunglokhaycwe* is broken into three wordpieces *_tunglok*, *hay* and *cwe_* as in the example (2), and *sakceyhaycwe* is also segmentied into three sub-word units *_sakcey*, *hay* and *cwe_* as shown in (3). Similarily, *tunglokhaycwulay* is also separated into four wordpieces *_tunglok*, *hay*, *cwu* and *lay_*. It means that the inventory retains sub-word units (e.g. *tunglok*, *sakcye*, *hay*, *cwu*, and *lay*) which are segmented by WPM, and these segements enable to deal with spaed-words with a variety of morphemes (e.g. *tunglok*, *tunglokhay*, *tunglokhaycwe*, *tunglokhaycwulay*, *sakcey*, *sakceyhay*, *sakceyhaycwe*, and *sakceyhaycwulay*). WPM has the advantage in the way that it enables tackling almost infinite words with the very light inventory. Thus, in case of Korean, it is more likely to segment spaced-words into lexical (e.g. *tunglok*, and *sakcye*) and functional (e.g. *hay*, *cwu*, and *lay*) morphemes (or sub-word units) using statistical methods only.

To extract features for a classification task which allocates suitable dialog act for each corresponding text, we reconstructed corpus for training GloVe with sub-word units that WPM segmented the text of original corpus of GloVe, by excluding vocabularies whose frequency of occurrence is less than 100. The table 4 compares two kinds of basic units of corpora for training Korean GloVe.

TABLE IV. SUMMARY OF BASIC UNITS FOR TRAINING GLOVE.

	Token	**Type**	**Vocabulary**
Spaced-word Units	668,284,389	23,675,186	366,940
Sub-word Units	**939,465,894**	**195,542**	**150,758**

Compared between the original and preprocessed corpora of Korean GloVe with regard to the size of type and vocabulary, it is observed that the number of types which segmented by WPM decreased significantly. The effectiveness of WPM in tokenizing the corpus of GloVe is proven without both expanding the inventory infinitely and using morphosyntactic information.

E. The architecture of CNN classifier

Within natural language processing, the sequence of word vectors has been involved as feature vectors in a classification task using deep learning [11]. Our approach is based on recent work by Kim. Y. (2014), which proposes CNN, which has shown notable performance on image classification, to involve text categorization [11]. CNN has been proven as a state-of-the-art model on several text classification tasks (e.g. sentiment analysis and question classification), despite of its simple architecture that needs little tuning [12]. Furthermore, this model is robust enough to deal with utterances consisting of a variable length of input text. To be specific, CNN provides an effective mechanism for a text categorization task which substitutes the role of Term Frequency – Inverse Document Frequency (TF-IDF) which only depends on words appeared at the surface level and fails to capture slot cue words. Thus, we generated CNN classifier to conduct a text categorization task for detecting users’ intention with following details.

We generate a multi-class CNN classifier consisting of a convolutional layer, a pooling layer and a fully connected layer [11], as illustrated in Figure 1. Let $w_i \in \mathbb{R}^k$ be the *k*-

dimensional word vector corresponding to the i-th word in the utterance of our feature maps [11]. A sentence u of length n is described as

$$u = w_1 \oplus w_2 \oplus \dots \oplus w_n \quad (4)$$

where $\oplus$ is the concatenation operator and w_i denotes each word consisting of a sentence. A feature map $c \in \mathbb{R}^{n-h+1}$ is obtained by a filter $w \in \mathbb{R}^{h \cdot k}$, which is applied to a window of h words to produce a new feature, followed by a rectified linear unit (ReLU) activation function [13]. To conduct a classification task, we use Korean GloVe with two different kinds of basic feature units (i.e. a *spaced-word* and a *sub-word unit*), and make use of them as feature vector alternatively.

The maximum length (n) of an utterance tokenized as a spaced-word unit is determined to be 12, and that as sub-word units is to be 15. We zero-pad remaining vector spaces for utterances which is shorter than n, without changing any hyper-parameters; we use filter windows (h) of 3, with 100 feature maps each. These values were obtained from a grid search on the SST-2 dev set [11].

To handle varying length of dialogue utterances, a pooling layer is used to take the maximum value of each feature:

$$\hat{c} = \max\{c_j\}, \qquad 0 \leq j < 100 \quad (5)$$

The resulting $\hat{c}$ is the most representative feature corresponding to each feature. These features form the penultimate layer, and are passed to a fully connected softmax layer for classification [13]. In the fully connected layer, a softmax activation function produces output of 14 classes which stands for the number of dialogue acts for users' utterances.

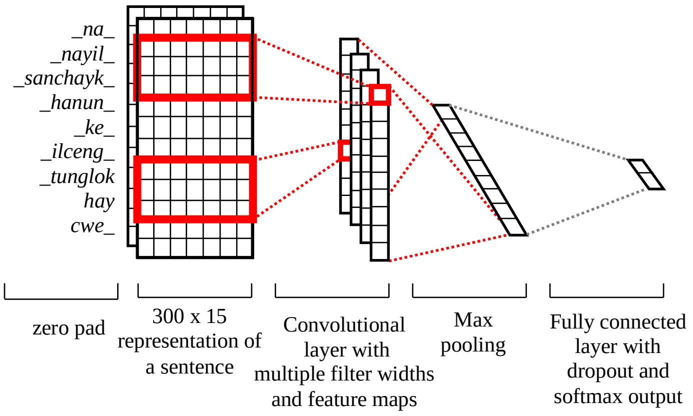
_na_
nayil
sanchayk
hanun
_ke_
ilceng
_tunglok
hay
cwe_
zero pad
300 x 15 representation of a sentence
Convolutional layer with multiple filter widths and feature maps
Max pooling
Fully connected layer with dropout and softmax output

Figure 1. The architecture of the CNN classifier.

IV. Result

To get an evaluation result of the classification task, we utilize 5-fold cross-validation with dividing the dataset into training (80%) and test (20%). The Table 5 below summarizes the accuracy rates in accordance with whether the basic unit of Korean GloVe is a spaced- or a sub-word unit.

TABLE V. Comparison of Accuracy Rate

	The basic units of Korean GloVe	
	Spaced-word	***Sub-word***
Train data	0.872	**0.914**
Test data	0.809	**0.853**

The classifier shows the accuracy of 85.2% in matching the correct dialog acts, which is higher than its counterpart with spaced-word units in the Korean. In both train and test data, it is observed that the sub-word units rather than spaced-word units contribute to achieve higher accuracy rate in the classification task.

V. Discussion and Future Research

This study aims at developing a novel approach to automatically acquire knowledge to detect user intents from the texts. We propose the generalized framework for User Intent Detect in the following ways. We first build a dialogue corpus in Korean and manually annotated with autonomously defined dialogue act sets. The annotation set is designed to map between transcribed texts and dialogue acts that let a system know what to do in the next turn. It is more economical than other traditional approach in that our tagging set removes intermediate processing to determine system actions. By establishing Korean version of GloVe, we can represent words in the vector space in an unsupervised way. Furthermore, it is observed that the statistical tokenizing termed WPM contributes to preprocess the corpus of GloVe and achieve high accuracy rate in the classification task to detect the users' intention.

We believe that our framework will lead to the CNN classifier that enables to learn and detect users' intention in other domain within the generalized annotation framework by using the simple tokenization method in an unsupervised way. In future research, we improve our classifier to decode semantic information that is conveyed in the form of slot-value pairs, and continue to establish multi-domain corpora in a variety of topics. In sum, the effectiveness of the proposed classifier model can be shown in different domains, indicating good generality and providing a reasonable direction for achieving SLU.

Acknowledgment

The authors would like to thank all members of Spoken Dialogue System Research Group, Sogang University for their contribution to the first step to construct a corpus and achieve Spoken Language Understanding in Korean dialogues.

References (will be revised)

[1] Wang, Y. Y., Deng, L., & Acero, A. (2005). Spoken language understanding. *IEEE Signal Processing Magazine*, *22*(5), pp.16-31.

[2] Chen, Y. N., Sun, M., Rudnicky, A. I., & Gershman, A. (2016, Mar.). Unsupervised user intent modeling by feature-enriched matrix

factorization. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6150-6154.

[3] Henderson, M., Thomson, B., & Williams, J. (2014, June). The second dialog state tracking challenge. In *15th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, vol. 263.

[4] Walker, M., & Passonneau, R. (2001, Mar.). DATE: a dialogue act tagging scheme for evaluation of spoken dialogue systems. In *Proceedings of the first international conference on Human language technology research*, pp. 1-8.

[5] Core, M. G., & Allen, J. (1997, Nov.). Coding dialogs with the DAMSL annotation scheme. In *AAAI fall symposium on communicative action in humans and machines*, vol. 56.

[6] Mikolov, T., Yih, W. T., & Zweig, G. (2013, June.). Linguistic Regularities in Continuous Space Word Representations. *In HLT-NAACL*, vol. 13, pp. 746-751.

[7] Pennington, J., Socher, R., & Manning, C. D. (2014, Oct.). Glove: Global Vectors for Word Representation. In *EMNLP*, vol. 14, pp. 1532-43.

[8] Yang, H., Lee, Y. I., Lee, H. J., Cho, S. W., & Koo, M. W. (2015). A Study on Word Vector Models for Representing Korean Semantic Information. *Journal of the Korean Society of Speech Sciences* vol, 7(4).

[9] Schuster, M., & Nakajima, K. (2012, Mar.). Japanese and korean voice search. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5149-5152.

[10] Rieser, V., & Lemon, O. (June, 2008.). Learning Effective Multimodal Dialogue Strategies from Wizard-of-Oz Data: Bootstrapping and Evaluation. In *ACL*, pp. 638-646.

[11] Kim, Y. (2014). Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.

[12] Johnson, R., & Zhang, T. (2014). Effective use of word order for text categorization with convolutional neural networks. *arXiv preprint arXiv:1412.1058*.

[13] Shi, H., Ushio, T., Endo, M., Yamagami, K., & Horii, N. (2016). Convolutional Neural Networks for Multi-topic Dialog State Tracking, In *the fourth dialog state tracking challenge (DSTC4) (IWSDS)*.