# Customization of an example-based dialog system with user data and distributed word representations

Eichi Seto*, Ryota Nishimura*, Norihide Kitaoka*
* Tokushima University, Tokushima, Japan
E-mail: ryota@nishimura.name Tel/Fax: +81-88-656-7259

***Abstract*—There is a need to develop spoken dialog systems which are capable of engaging in natural conversations with people, for example, the socially-isolated elderly. We propose an example-based dialog system featuring an adaptation method which customizes the dialog for each user. After retrieving user profile-related information from the Web, named entity extraction is applied to the retrieval results. Words with a high term frequency-inverse document frequency (TF-IDF) score are adopted as user related words. We then calculate the similarity between the selected user-related words and words in the existing example phrases of the dialog system. Cosine similarity between the distributed representations of the nouns is calculated using word2vec embedding. We then generate phrases adapted to the user by substituting user-related words for highly similar words in the original example phrases. Word2vec also has a special property which allows arithmetic operations such as plus and minus to be applied to distributed word representations. By applying operations to the words used in the original phrases, we were able to derive replacement words related to the user and insert them in the example phrases. We evaluated the naturalness of the generated phrases and found that the system could generate natural phrases.**

## I. Introduction

The performance of natural language dialog systems has been improved over time, and applications using these systems have become more popular [1]. Various types of dialog systems have been developed, such as a counseling dialog system [2] and a conversational knowledge teaching agent [3]. Recently, chat-like dialog systems are actively investigated. Such kind of dialog systems are demanded to be utilized for symbiotic machines such as humanoid robots. Reminiscence therapy systems also need such dialog technology[4]. However, when designing a chat-like spoken dialog system, there are various problems which need to be resolved in order to provide natural conversations. This study addresses response sentence generation, which is the generation of a response to a user's utterance, one of the key problems in developing such spoken dialog systems. In this study, we propose an example-based spoken dialog/chat system which is adapted to a specific user. Such systems as reminiscence therapy dialog systems is used by a specified user, and thus we can adopt user adaptation. Example-based spoken dialog systems can robustly respond to user utterances if their example database contains a wide enough variety of utterances. However, it is more difficult to generate personalized responses which are related to a specific user's interests or preferences, so our goal is to develop a database which is customized for each user. In [5], a domain adaptation for a sequence-to-sequence-based dialog system has been investigated. Enrichment of sequence-to-sequence-based dialog generaton using external memory has also been proposed [6]. Here, we propose a method of creating an example database using information from a user's profile acquired through a web search, and then using word2vec operations [7], [8], [9], [10], [11] to substitute user specific information into the example phrases.

## II. Example-based spoken dialog systems

An example-based spoken dialog system is a system that "talks" with a user using an example database consisting of pairs of input examples and the corresponding responses. Figure 1 shows examples of input and response pairs. Figure 2 shows the flow of an example-based dialog system. Example responses may correspond to multiple example inputs because various user utterances can have the same meaning, (e.g., "Do you know what time it is?", "Do you have the time?", "What time is it?"). By preparing multiple input examples the system can respond to a wider variety of user utterances. Likewise, multiple example responses can be chosen to correspond to the same example input, (e.g., if asked "How are you today?" the system can respond in different ways, such as "Fine, thanks" or "So-so"). By preparing multiple example responses we can provide conversations which are less monotonous.

Typical spoken dialog systems recognize a user's utterances using speech recognition software and then match the content of the recognized sentence with the system's input examples. Each input example is scored according to the number of matched content words, and after all of the input examples are scored the one with the highest score is selected. A score for each of the possible response sentences is then calculated. When a response sentence corresponds to multiple input sentences, the highest score among the corresponding input sentences is used for the score of the response sentence.

"Takemaru-kun" [12], [13], [14] is an example-based, spoken dialog chat system which has been successfully used with the public. The system contains prepared responses such as greetings, self-introductions and directions for navigating community centers and surrounding facilities. As various users talk to the system, the system logs the real-world dialogs. These logs are then transcribed and adopted as new examples, making the system more robust to an increasing variety of user utterances. However, our system does not adopt previous dialogs as examples because its responses are customized for a

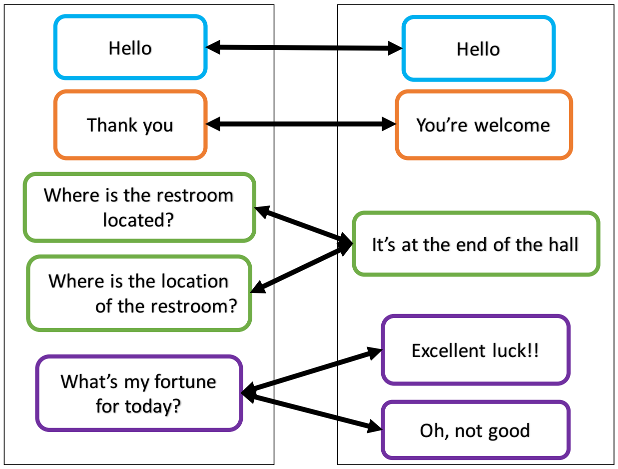


Fig. 1. Examples of input and response pairs in an example-based dialog system

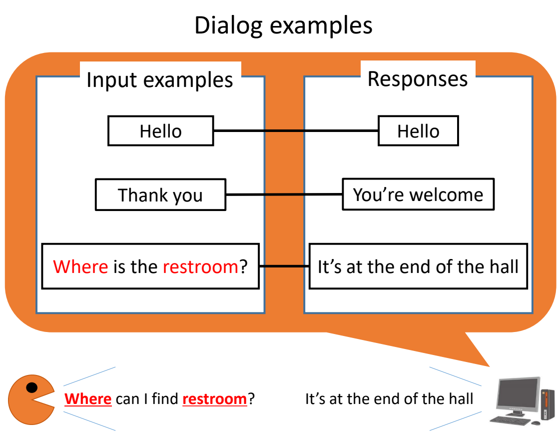


Fig. 2. System response selection in an example-based dialog system

specific user, and we do not anticipate an extremely wide range of possible utterances from one person. Instead, we customize the example database using limited information about the user.

There have been many studies on example-based spoken dialog systems. For example, a method to improve the robustness of the example-based dialog modeling framework using an agenda-based approach and n-best recognition hypotheses has been proposed [15]. A new probabilistic framework for spoken dialog management using a frame-based belief state representation has also been proposed [16]. These dialog management studies focus on selecting the correct response according to a given situation. In this study, we propose a method to adapt the example database itself to topics related to the user, with the goal of providing a natural, personalized chat interaction with that specific user. To develop a more intelligent system in the future, we will need to use additional information such as dialog histories and natural language analyses.

## III. Method of generating user-adapted examples

If example phrases similar to the user's utterances are included in the example database, an example- based spoken dialog system can provide natural responses. Thus, it is important to include example phrases likely to be employed by the user in the example database. Our system addresses this issue by generating user-adapted examples using profile information about the user.

### A. Acquiring words for user adaptation

In this study, we call words related to the user which are acquired to generate user-adapted examples "words for user adaptation". First, information about the user is retrieved from the Web using basic information about the user in queries. Named entity extraction using KNP [17], a Japanese syntactic analyzer, and morpheme analysis using JUMAN++ [18], a Japanese morphological analyzer, are then applied to the retrieval results. The named entity has a classification defined by IREX (Information Retrieval and Extraction eXercise). Table I shows an example of named entity classification and examples. In this study, we acquire words which fall into the categories of PERSON, LOCATION, ORGANIZATION and ARTIFACT. When conducting morpheme analysis using JUMAN++, the program acquires other words, known as "Artifact-Food", which are not included in the named entity's classification. We consider words that are collected in the named entity's classification to be important conversation topics, thus words acquired by named entity extraction and morpheme analysis are selected as candidates for words for user adaptation. Next, we calculate the TF-IDF (term frequency-inverse document frequency) score of the candidate words. TF-IDF is a numerical statistic that is intended to reflect the importance of a word within a document in a collection or corpus. In this study, the score for each word is obtained by normalizing the TF-IDF score for each search query. The following equation is used for normalization:

$$x^i_{score} = \frac{x^i - \mu}{\sigma} \tag{1}$$

In Equation (1), $x^i_{score}$ is the normalized value of word $i$ in word set $x$, $\mu$ is the average score and $\sigma$ represents the standard deviation. Words which exceed the threshold are considered to be related to the user and are selected as words for user adaptation. Figure 3 shows this procedure. Table II shows an example of a user profile. Table III shows examples of words selected for user adaptation which were acquired using the user's profile when the threshold for TF-IDF scores is set to 0.6.

### B. Adaptation of examples

Example phrases are adapted to the user using the words acquired as described above. We generate user- adapted examples by calculating similarity between words in the original examples and the acquired words using word2vec. Word2vec

TABLE I
NAMED ENTITY CLASSIFICATION

| Classification | Example |
|---|---|
| PERSON | Shohei Ohtani |
| LOCATION | United Kingdom, Yoshino River |
| ORGANIZATION | Tokushima University |
| DATE | today, April 2018 |
| TIME | 3:00 AM |
| MONEY | 500 yen |
| PERCENT | 90% |
| ARTIFACT | Nobel Prize |

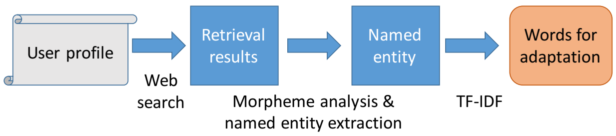


Fig. 3. Acquisition of words for user adaptation based on information retrieval

is a tool which reconstructs the linguistic contexts of words using vectors to represent words in a 200-dimensional vector space. This distributed representation of words reveals possible meanings of the words using word contexts. Therefore it is possible to calculate the similarity between words mathematically using cosine similarity. We adapt the existing examples in a database to particular users by replacing words in the examples with the words acquired for user adaptation. The procedure for generating user adapted examples is as follows. First, we acquire words for user adaptation and compare each word with the dialog examples. We pick a noun from an example contained in a dialog example database created through morphological analysis using Mecab, and then calculate the similarity between that noun and each word for user adaptation. If the similarity is higher than a threshold, we replace the word in the example database and generate a new example. Figure 4 shows this procedure.

### *C. Example generation using addition and subtraction of distributed word representations*

Distributed representations of words are generated using word2vec, which enables us to use vectors to calculate similarities between words. Word2vec also has a feature which allows the application of addition and subtraction operations to word meanings, for example the distributed representation obtained from the calculation $\overrightarrow{king} - \overrightarrow{man} + \overrightarrow{woman}$ is $\overrightarrow{queen}$. Our proposed system can generate user-adapted examples using this feature, which is helpful when we replace one of the words in an example associated with the name of a place or a facility with a user-adapted word. The new example may not make sense, so we have to replace another word using addition and subtraction on the distributed representation to generate a consistent user-adapted example. When "Tokyo" is replaced with the user adaptation word "Kyoto" in the sentence "Meiji Shrine in Tokyo is famous", the new sentence "Meiji Shrine in Kyoto is famous" is inconsistent. "Meiji Shrine" should be replaced by the name of a shrine located in Kyoto, such as "Shimogamo Shrine". To do this, we perform the calculation

TABLE II
EXAMPLE USER PROFILE

| | |
|---|---|
| Birthplace | Kyoto |
| Hobby | Travel (to Hokkaido) |
| Favorite foods | Steamed buns |
| Favorite celebrities | Yuzuru Hanyu |

$\overrightarrow{MeijiShrine} - \overrightarrow{Tokyo} + \overrightarrow{Kyoto}$ to obtain the names of shrines in Kyoto. Words located near the resulting vector become potential candidates. In this study, we focus on specific words as described in Section III-A. Figure 5 shows this procedure.

One problem with this method is that the results of the addition and subtraction of distributed word representations using word2vec may produce incorrect results due to a lack of accuracy. For this reason, it is necessary to select the correct word from a list of candidate words obtained from the calculation results. The correct word is selected using two values, similarity between the results of the addition and subtraction operation and similarity with the word for user adaptation. First, we obtained multiple candidates using only addition and subtraction, and then we compared the scores of each word obtained using the following formula to rescore them:

$$score = \lambda * similarity(v) + (1 - \lambda) * similarity(w) \quad (2)$$

In Equation (2), $similarity(v)$ represents cosine similarity with the results of addition and subtraction of replacement words, while $similarity(w)$ represents cosine similarity with the words for user adaptation. In this study, we set $\lambda$ to 0.3. We selected the word that got the highest score and generated adapted examples through replacement. Figure 6 shows this procedure.

## IV. EXPERIMENTAL EVALUATION

### *A. Experimental conditions*

We used Mecab [19] for morphological analysis and Mecab-ipadic-neologd [20] as the morpheme dictionary, since this dictionary is more robust when encountering new words than the conventional IPA Dictionary, and there was no need to perform consolidation of noun phrases into compound words. This allows our system to better analyze proper nouns such as the names of buildings and place names. In addition, we obtained semantic information about words using JUMAN++ in order to select the words to be used for addition and subtraction in distributed representations.

We retrieved information related to the user's profile using Google and selected the named entities which obtained normalized TF-IDF scores higher than 0.6 for user adaptation. TF represents the number of appearances in the search results, while IDF was calculated using Wikipedia data. Word2vec was used to calculate similarity and was trained using the Japanese version of Wikipedia dated July 1, 2017. Six sets of adapted examples were generated using the profiles of three people and two sets of dialog examples. The first set of dialog examples contained 100 manually created examples, and the

TABLE III
WORDS ACQUIRED FOR USER ADAPTATION

| | | |
|---|---|---|
| Kyoto Prefecture* | Hakodateyama* | Kinkakuji* |
| Arashiyama* | Kiyomizu Temple* | Gion District* |
| Toyako* | New Chitose Airport* | Kyoto* |
| Mao Asada | Yuzuru Hanyu | Noboribetsu* |

∗ place or sightseeing spot in Japan

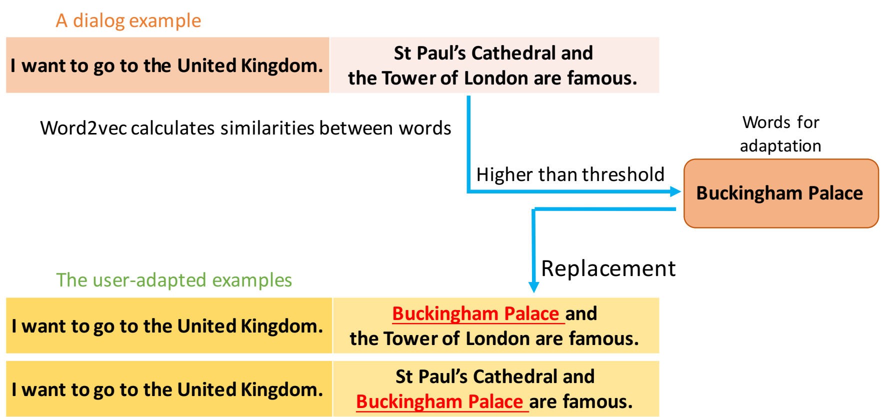


Fig. 4. System response selection in an example-based dialog system

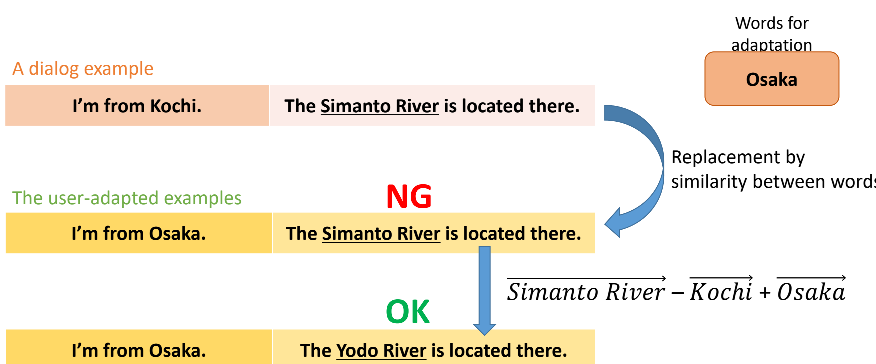


Fig. 5. Generate user-adapted examples using vector operation

second set included 162 examples selected from Twitter. We used 162 tweet-reply pairs for the Twitter examples, selected from 116,292 pairs acquired during October 2017.

To evaluate the performance of the proposed method, we examined the user-adapted examples it generated using each set of dialog examples and judged them to be natural or unnatural.

### B. Experimental results

We called the total number of user-adapted dialog examples generated in our experiments "#Generated", while the total number of generated examples determined to be natural was labelled "#Success". We calculated our success rate using the following formula:

$$Success\_rate(\%) = \frac{\#Success}{\#Generated} * 100 \tag{3}$$

*1) Experimental results using 100 manually created examples:* First, we replaced just one word in each dialog example, using only the calculated word2vec similarities. We also changed the similarity threshold used to decide which replacement words were selected. Table IV shows #Generated, #Success and Success_rate when using different thresholds. When we raised the threshold, the #Generated and Success_rate decreased. This is because there were many examples requiring addition and subtraction in order to obtain useable results, but we were only using word2vec similarity at this point. We then generated examples using addition

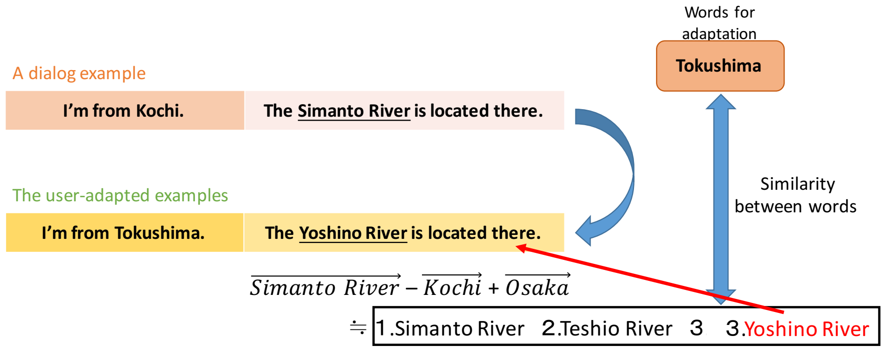


Fig. 6. Select a replacement word to generate the user-adapted example

TABLE IV
EXPERIMENTAL RESULTS FOR SELECTED THRESHOLDS USING MANUALLY GENERATED DIALOG EXAMPLES

| Similarity threshold | 0.70 | 0.73 | 0.75 | 0.78 |
|---|---|---|---|---|
| #Generated | 806 | 551 | 394 | 208 |
| #Success | 168 | 101 | 70 | 36 |
| Success rate | 20.8 | 18.3 | 17.7 | 17.3 |

TABLE V
EXPERIMENTAL RESULTS USING ADDITION AND SUBTRACTION ON MANUALLY GENERATED EXAMPLES

| Threshold | 0.73 |
|---|---|
| #Generated | 551 |
| #Success | 165 |
| Success rate | 31.9 |

| #Addition and Subtraction | 311 |
|---|---|
| #Success | 81 |
| #Failure | 230 |

TABLE VI
EXAMPLE OF ADAPTATION GENERATION USING VECTOR OPERATION WITH MANUALLY GENERATED EXAMPLES

| | |
|---|---|
| Original dialog | Tokyo is a nice place.<br>Did you go to Akihabara? |
| Words for user adaptation | Osaka |
| Replacement of similar words | Tokyo $\rightarrow$ Osaka |
| Addition and subtraction equation | $\overrightarrow{Akihabara} - \overrightarrow{Tokyo}$ $+\overrightarrow{Osaka}$ |
| Replacement of words by addition and subtraction | Akihabara $\rightarrow$ Umeda |
| Generated user adaptation | Osaka is a nice place.<br>Did you go to Umeda? |

and subtraction on the distributed representations, setting the threshold to 0.73. The replacement word candidates derived using addition and subtraction had similarity scores of 0.6 or more. We performed adaptation on examples with a $\lambda$ of 0.3 in Equation (2). Table V shows experimental results using addition and subtraction, the total number of examples using addition and subtraction and the number of successes and failures. A total of 311 dialog examples were generated using addition and subtraction on 551 dialog examples. The number of natural examples generated was 81. Thus, we generated 64 more natural dialog examples than when we created the examples without using addition and subtraction. Table VI shows user-adapted examples when using addition and subtraction on distributed word representations. Akihabara is a famous district in Tokyo, but the user's profile indicated he or she lives in Osaka. First, the system replaced Tokyo with Osaka, and then Akihabara was replaced with Umeda, a district in Osaka. As a result, we obtained a new, consistent and natural example.

*2) Experimental results using Twitter dialog examples:* We perform the experiment again under the same conditions as in the previous section, except we changed the set of test dialog examples used. First, we only replaced at most one word in each example using only word2vec similarities. Table VII shows #Generated, #Success and Success_rates when using different thresholds. When we raised the threshold, #Generated decreased but the success rate increased. We then generated examples using addition and subtraction on distributed representations, using a threshold of 0.73. The replacement word candidates generated using addition and subtraction had to have similarity scores of 0.6 or more. We performed adaptation of dialog examples which set $\lambda$ to 0.3 in Equation 2. Table VIII shows our experimental results when using addition and subtraction, the total number of examples using addition and subtraction and the number of successes and failures. 85 dialog examples were generated using addition and subtraction on 406 examples. The number of natural dialog examples when using addition and subtraction was 30, thus we generated 12 fewer natural examples than when the examples were generated without using addition and subtraction. Table IX shows an example of inconsistent user-adaption when using addition and subtraction on distributed word representations. First, the system replaced Nara with Fukui (two Japanese prefectures), but then Osaka was also incorrectly replaced with Fukui. Addition and subtraction were executed and the original words were replaced with semantically similar words, but the example input and response were connected in a negative manner ("Do you live in Osaka?" "No, I live in Fukui."). As a result, there were many examples which were judged to be unnatural. Although there were some difficult dialog examples

TABLE VII
EXPERIMENTAL RESULTS FOR SELECTED THRESHOLDS USING TWITTER GENERATED EXAMPLES

| Similarity threshold | 0.70 | 0.73 | 0.75 | 0.78 |
|---|---|---|---|---|
| #Generated | 626 | 406 | 289 | 139 |
| #Success | 280 | 196 | 147 | 72 |
| Success rate | 44.7 | 48.2 | 50.8 | 51.8 |

TABLE VIII
EXPERIMENTAL RESULTS USING ADDITION AND SUBTRACTION ON TWITTER GENERATED EXAMPLES

| Threshold | 0.73 |
|---|---|
| #Generated | 406 |
| #Success | 184 |
| Success rate | 45.3 |

| #Addition and Subtraction | 85 |
|---|---|
| #Success | 30 |
| #Failure | 55 |

which our proposed method was not able to adapt, we did have some success creating user adapted dialog examples using our proposed method.

## V. CONCLUSION

In this study we proposed a user adaptation technique for creating personalized sets of input and output dialog examples using user profile information to build a dialog system that could chat naturally with a human being. We generated user-adapted examples by calculating similarity between acquired words related to the user and the original words in the examples. We replaced words which were grammatically the same (nouns and proper nouns), whose similarity score was higher than a threshold. We were able to generate natural examples using addition and subtraction on distributed word representations. By selecting replacement words from candidates we were able to improve our success rate.

In our evaluation experiment, we generated adaptations of dialog examples created manually and of dialog examples taken from Twitter, using information from the profiles of three people. We were able to generate natural user-adapted dialog examples through replacement based on distributed representation. Future work includes increasing our success rate by improving the accuracy of the addition and subtraction operations on word representations.

TABLE IX
EXAMPLE OF ADAPTATION GENERATION USING VECTOR OPERATION WITH TWITTER GENERATED EXAMPLES

| | |
|---|---|
| Original dialog | Well, wait, do you live in Osaka? |
| | No, it is peaceful Nara. |
| Words for user adaptation | Fukui |
| Replacement of words by similarity | Nara $\rightarrow$ Fukui |
| Addition and subtraction equation | $\overrightarrow{Osaka} - \overrightarrow{Nara} + \overrightarrow{Fukui}$ |
| Replacement of words by addition and subtraction | Osaka $\rightarrow$ Fukui |
| Generated user adaptation | Well, wait, do you live in Fukui? |
| | No, I live in peaceful Fukui. |

## ACKNOWLEDGMENTS

This study received support from the Strategic Information and Communications R&D Promotion Program (SCOPE) of the Ministry of Internal Affairs and Communications (MIC) of Japan.

## REFERENCES

[1] C. Lee, S. Jung, S. Kim, and G. G. Lee, "Example-based dialog modeling for practical multi-domain dialog system," *Speech Communication*, vol. 51, no. 5, pp. 466–484, 2009.

[2] S. Han, K. Lee, D. Lee, and G. G. Lee, "Counseling dialog system with 5w1h extraction," in *Proceedings of the SIGDIAL 2013 Conference*, 2013, pp. 349–353.

[3] K. Lee, P. H. Seo, J. Choi, S. Koo, and G. G. Lee, "Conversational knowledge teaching agent that uses a knowledge base," *16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 2015.

[4] S. Ashida, "The effect of reminiscence music therapy sessions on changes in depressive symptoms in elderly persons with dementia," *Journal of music therapy*, vol. 37, no. 3, pp. 170–182, 2000.

[5] P.-J. Chen, I.-H. Hsu, Y.-Y. Huang, and H.-Y. Lee, "Mitigating the impact of speech recognition errors on chatbot using sequence-to-sequence model," in *Proceedings of 2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*. IEEE, 2017, pp. 497–503.

[6] K. Yu, Z. Zhao, X. Wu, H. Lin, and X. Liu, "Rich short text conversation using semantic-key-controlled sequence generation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 8, pp. 1359–1368, 2018.

[7] T. Mikolov, W.-t. Yih, and G. Zweig, "Linguistic regularities in continuous space word representations," in *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2013, pp. 746–751.

[8] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[9] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in Neural Information Processing Systems*, 2013, pp. 3111–3119.

[10] Y. Goldberg and O. Levy, "Word2vec explained: Deriving mikolov et al.'s negative-sampling word-embedding method," *arXiv preprint arXiv:1402.3722*, 2014.

[11] X. Rong, "Word2vec parameter learning explained," *arXiv preprint arXiv:1411.2738*, 2014.

[12] R. Nishimura, Y. Nishihara, R. Tsurumi, A. Lee, H. Saruwatari, and K. Shikano, "Deploying a speech-based information system as a research platform for speech recognition research in real environments," *Electronics and Communications in Japan (Part II: Electronics)*, vol. 88, no. 12, pp. 43–54, 2005.

[13] R. Nisimura, A. Lee, H. Saruwatari, and K. Shikano, "Public speech-oriented guidance system with adult and child discrimination capability," in *Proceeding of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'04), 2004.*, vol. 1. IEEE, 2004.

[14] T. Cincarek, H. Kawanami, R. Nisimura, A. Lee, H. Saruwatari, and K. Shikano, "Development, long-term operation and portability of a real-environment speech-oriented guidance system," *IEICE transactions on information and systems*, vol. 91, no. 3, pp. 576–587, 2008.

[15] C. Lee, S. Jung, and G. G. Lee, "Robust dialog management with n-best hypotheses using dialog examples and agenda," *Proceedings of ACL-08: HLT*, pp. 630–637, 2008.

[16] K. Kim, C. Lee, S. Jung, and G. G. Lee, "A frame-based probabilistic framework for spoken dialog management using dialog examples," in *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue of the Association for Computational Linguistics*. Association for Computational Linguistics, 2008, pp. 120–127.

[17] D. Kawahara and S. Kurohashi, "A fully-lexicalized probabilistic model for Japanese syntactic and case structure analysis," in *Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*. Association for Computational Linguistics, 2006, pp. 176–183.

[18] H. Morita, D. Kawahara, and S. Kurohashi, "Morphological analysis for unsegmented languages using recurrent neural network language model," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 2015, pp. 2292–2297.

[19] T. Kudo, K. Yamamoto, and Y. Matsumoto, "Applying conditional random fields to Japanese morphological analysis," in *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, 2004.

[20] T. Sato, "Mecab-ipadic-NEologd: neologism dictionary for MeCab," 2016, https://github.com/neologd/mecab-ipadic-neologd.