Chat Response Generation Based on Semantic Prediction Using Distributed Representations of Words

Kazuaki Furumai , Tetsuya Takiguchi and Yasuo Ariki

Abstract In this paper, we propose a chat response generation system using distributed expression of words by word2vec. With the conventional one-hot representation method, there was a problem that the model becomes complicated as the vocabulary increases, and only words that appear in the dialogue corpus can be handled. We address these problems by using word2vec and extend it to handle unknown words that did not appear in the conversation corpus. In a subjective evaluation experiment, we showed that various responses can be generated by estimating words by semantic prediction.

1 Introduction

In recent years, dialogue systems have been actively studied owing to advances of artificial intelligence. There is a rule base method for generating responses according to rules created beforehand in response generation of the dialogue system, but there is a problem that cost is required for various kinds of responses. The chat system in this research is not a presumption of a specific topic or task, but is called a non task oriented type which focuses on the dialog itself with humans. Unlike task-oriented systems, this system is required to deal with various topics. Therefore, it is necessary to use a method to generate a response sentence automatically instead of a rule base method.

As far as we know, word representation in dialog systems is mainly based on one-hot representation, but in chat, it is expected that the number of words handled

Kazuaki Furumai
Kobe University, Japan, e-mail: furumai@me.cs.scitec.kobe-u.ac.jp

Tetsuya Takiguchi
Kobe University, Japan e-mail: takigu@kobe-u.ac.jp

Yasuo Ariki
Kobe University, Japan e-mail: ariki@kobe-u.ac.jp

will be very large. However, when trying to deal with various topics, if one-hot representation is used, an increase in the number of dimensions of word vectors can not be avoided, and the model becomes complicated. In addition, it is not possible to generate response sentences other than words appearing in the corpus, and it is heavily dependent on corpus. Therefore, by using semantic representation vector of fixed dimension learned with some text data beforehand instead of one-hot representation, it is possible to prevent complexity of the model and also to deal with words that do not exist in the corpus. In this paper, we propose a method to generate response sentences by word prediction by Recurrent Neural Network, using distributed representation of word instead of one-hot representation.

2 A dialogue system by RNN encoder-decoder

Those that are often used as a method for automatically generating responses in dialog systems are RNNs as seen in the Neural Conversational model (Vinyals et al [8]) and the Neural Responding Machine (Shang et al. [7]). It receives sequence $X = (x_1, ..., x_{T_x})$ of input word vector and outputs sequence $Y = (y_1, ..., y_{T_y})$ of output word vector.

The hidden layer $h_{(t)}$ of RNN can be represented by $h_{(t)} = f(h_{(t-1)}, x_t)$. This model uses RNN for processing input word sequence X as Encoder and RNN for generating output word series Y as Decoder and hidden layer $h_{(T_x)}$ for $h_{(0)}$ in Decoder, and therefore is called RNN Encoder-Decoder. In this paper, we use the RNN Encoder-Decoder model.

3 Distributed representation of words

Distributed representation of words is based on the distribution hypothesis and represents words by low-dimensional real-valued vectors, for which word2vec proposed by Mikolov et al. [6] [4] [5] is mainstream. Using a one-hot representation makes it impossible to consider the relationship between words, whereas using a distributed representation makes it possible to perform operations such as (King - Man + Woman = Queen) that considers the meaning of words.

There are two learning methods of word2vec, CBOW (Continuous Bag-of-Words) and Skip-gram. It is pointed out that learning by skip-gram model shows better results in [4], so in this paper we use Skip-gram model to learn word2vec.

4 The proposed method

In this paper, we generate a response sentence using the distributed representation vector proposed by word2vec as input / output vector of RNN Encoder-Decoder. A schematic diagram of the model is shown in Fig. 1. First, the input word sequence is converted to the d_{word} dimension vector by word2vec learned beforehand and input to the Encoder. Next, let the hidden layer $h_{(T_x)}$ generated by Encoder be the initial state $h_{(0)}$ of the hidden layer of Decoder. The output vector of the Decoder can be treated as the semantic prediction vector $y_{semantic_t}$, and is the d_{word} dimensional vector whose element takes real number. When generating responce, this sentence prediction vector $y_{semantic_t}$ is used to output a response sentence with the word y_t having the highest cos similarity among the word vector set V created by word2vec as the corresponding word. When the number of vocabulary is N, and the word vector learned by word 2 vec is $W_k \in V(k = 1,\ldots,N)$, it can be expressed as

$$y_t = \underset{W_k}{\arg\max}\, cos(y_{semantic_t}, W_k)$$

When the correct word sentence is $T = (t_1, \ldots, t_{T_t})$, the loss function L used for learning is $L = \sum_i |t_i - y_{semantic_i}|$.

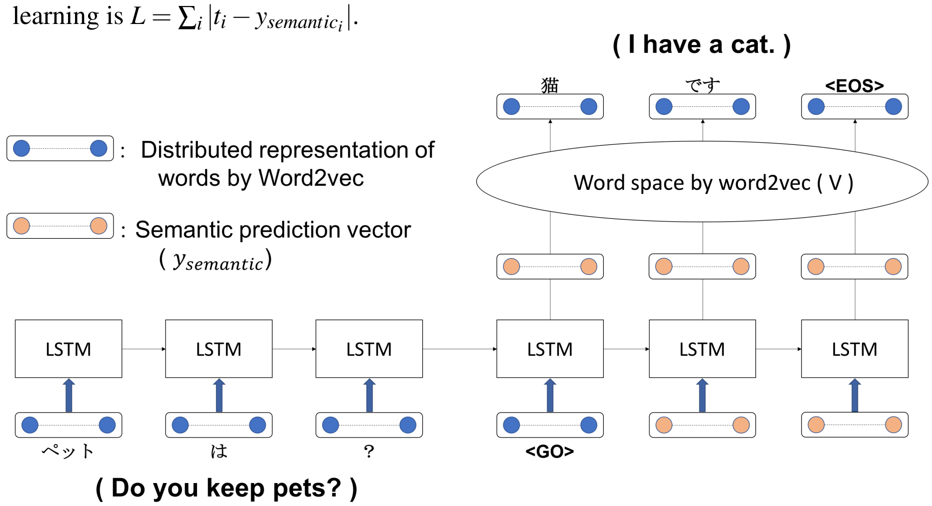

Fig. 1 Schematic diagram of the proposed method

5 Data sets

Since the data set that learns words and the data set that learns response sentence generation may be different, in this paper we prepared a dialogue corpus collected on Twitter and a data set created from Japanese Wikipedia articles. After shaping it into an appropriate format, we use a morphological analysis using MeCab [3] separately.

Since we do not consider the speaker character and the dialog histories in this research, we created a dialogue corpus with 360,000 pairs of Tweet / Reply pairs from Twitter. For learning of word2vec, in addition to the collected dialogue corpus, we used Japanese Wikipedia article (3G). We used the Twitter dialogue corpus for learning dialogue after learning word2vec using these data sets.

6 Experiment

6.1 Experimental conditions

The dimension number of word distributed representation by word2vec is $d_{word} = 128$. We excluded words with an occurrence count of 10 or less, and trained by skip-gram model. As a result, the number of vocabulary became about 200,000 words. About the RNN Encoder-Decoder, Long Short-Term Memory (LSTM) cell is used, and the number of units is 256, and the number of hidden layer is 3. Adam [2] is used as the optimization method at learning, and the learning rates are $\alpha = 0.0001, \beta_1 = 0.9, \beta_2 = 0.99$. $< GO >$ indicating the beginning of the sentence and $< EOS >$ indicating the end of the sentence were also learned as words with word2vec as special symbols. When generating a response sentence by the RNN Decoder, $< GO >$ is input as the first word, and a response is generated until $< EOS >$ is output. In addition, when the current cos similarity is 0.5 or less, or when the current cos similarity is 60% or less of the cos similarity of the output word at one previous time, the word is excluded.

In machine translation etc, BLEU [1] is used for evaluation, which calculates the score on the degree of matching of partial word strings. However, it is pointed out that there is a difference between evaluation by BLEU and manual evaluation . For example, there is no problem between languages having similar grammatical structures such as English and French, but differences arise tasks between languages with different the grammatical structures, such as English and Japanese. In the chat system, various combinations of input sentences and output sentences are considered, and it is a more complicated task, so it is conceivable that a difference is generated between BLEU evaluation and manual evaluation. Therefore, the following two evaluation indexes are prepared.

- Appropriateness : Degree of feeling that the system is reactive or understanding to input sentence
- Variety : Degree of performing various responses

Regarding variety, evaluation criteria are based on whether it is possible to make a response peculiar to the conversation, rather than a safe response. Response sentences were generated by 46 sentences which were collected from Twitter and not used for learning, and 5 point liker scale evaluations were conducted for each evaluation on each generated sentence by 10 participants.

6.2 Result

The experimental results are shown in Fig. 2. For each evaluation index, the results of averaging the respective evaluation values are shown. Compared with the one-hot representation method, the proposed method improved the Variety. From the viewpoint that synonyms and relationships between words can be considered in the proposed method, improvement of appropriateness was expected, but in fact there was little difference. However, there is an opinion that judgment of appropriateness was difficult when the meaning of the input sentence (user's utterance) selected from Twitter is difficult to understand. This means that improvement of the questionnaire is necessary. Regarding variety, the influence of the beam search [9] used in the conventional method can be considered, so that a close examination will be necessary.

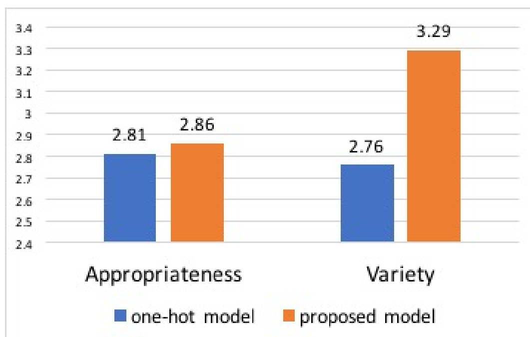

Fig. 2 Experimental results 5 point liker scale evaluation

An example of the generated response sentence is shown in Table 1. In one-hot model, since the words "Darth Vader" and "Guile" are not included in the training data, and they were treated as $< unk >$, the response has not changed. On the other hand, in the proposed method (word2vec model), each vector has been converted to another vector, and the response sentence has changed.

7 Conclusion

In this paper, we proposed a method to generate response sentences using distributed representation of words as input / output. We showed that variety evaluation is higher than response sentence generation by conventional one-hot model. Since,

Table 1 Example of reply sentence generation

input	output
初めまして〜お話しましょう (Nice to meet you. Let’s talk)	(W) ええよ！ (OK!) (O) ありがとうございます（ ´∀｀ ）(Thank you)
仲良くしてください！ (Let me be your friend.)	(W) ほんなら、何て呼びましょ? (Hey, what should I call u?) (O) なんて呼んでください（ ´∀｀ ）(Please call me.)
ガイル強いよね (Guile is strong, right?)	(W) 本当に...?(Sure?) (O) そうだった（￣ − ￣ ）(That’s right.)
ダース・ベイダー強いよね (Darth Vader is strong, right?)	(W) 本当に !!! (I totally agree!) (O) そうだった（￣ − ￣ ）(That’s right.)

(W) = word2vec model, (O) = one-hot model

in the current model, the one with the highest cos similarity is output and the process corresponding to beam search is not implemented, many response sentences with many grammatical errors are generated in many cases. In addition, even when words not included in the dialogue corpus appeared at the time of output generation, it often became ambiguous in terms of the whole sentence. In the future we will consider using seqGAN [10] and other models and aim to improve accuracy. In addition, since the Twitter corpus contains many noises, even a generated sentence that seems to be correct could have a bad evaluation. In the future, we need to think about a data set of multi-turn conversation which is less noisy and can also take dialog histories into account.

References

1. Doddington, G.: Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. Proc. of the Second International Conf. on HLT ’02 pp. 138–145 (2002)
2. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv:1412.6980 (2014)
3. KUDO, T.: Mecab : Yet another part-of-speech and morphological analyzer. http://mecab.sourceforge.net/ (2005)
4. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. arXiv:1301.3781 (2013)
5. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. Proc. of NIPS pp. 3111–3119 (2013)
6. Mikolov, T., Yih, S.W.t., Zweig, G.: Linguistic regularities in continuous space word representations. Proc. of NAACL-HLT 2013 pp. 746–751 (2013)
7. Shang, L., Lu, Z., Li, H.: Neural responding machine for short-text conversation. Proc. of ACL 2015 pp. 1577–1586 (2015)
8. Vinyals, O., Le, Q.: A neural conversational model. ICML Deep Learning Workshop (2015)
9. Wiseman, S., Rush, A.M.: Sequence-to-sequence learning as beam-search optimization. Proc. of EMNLP pp. 1296–1306 (2016)
10. Yu, L., Zhang, W., Wang, J., Yu, Y.: SeqGAN: Sequence generative adversarial nets with policy gradient. arXiv:1609.05473 (2016)