Generating Cafeteria Conversations with a Hypernetwork Dialogue Model

Jun-Hyuk Oh [1] Hyo-Sun Chun [1] Byoung-Tak Zhang[1,2]

[1]School of Computer Science and Engineering
[2]Cognitive Science and Brain Science Programs
Seoul National University, Seoul 151-744, Korea

{jhoh, hschun, btzhang}@bi.snu.ac.kr

Abstract. This paper introduces a data-driven dialogue system that generates the cashier's response to the customer's order. The proposed system learns dialogue rules from the corpus without prior knowledge which manage dialogue flows. The support vector machine combined with bag-of-words model is used to recognize and classify dialogue act (DA) of the customer. The hypernetwork dialogue model is used to manage dialogue flows and generate responses. The experimental results show that proposed system can generate appropriate dialogues without dialogue rules and the performance of dialogue generation is significantly improved as the data grows.

1 Introduction

These days, in the field of machine learning and natural language processing, many researchers are studying sentence generation [1], [2]. However, many of these models use a lot of prior knowledge such as a specific grammar to generate sentences. These approaches are not scalable because they are bounded by their specific language or a specific situation. For example, a method to generate sentences based on several Korean grammar rules cannot be applied to other language. There are also many researches about dialogue management (DM) [12-14]. This is a challenging problem because a model is required to not only recognize the speech act of a user and also generate appropriate dialogues. So, in order to reduce the complexity, most of works define a set of rules or dialogue flows with prior knowledge. In [12], for example, they provide a script language in which the dialogue generation rules can be coded. Thus, developers need to foresee all possible situations in order to code the dialogue rules. In other words, if a new situation occurs, developers should modify the rules manually. Therefore, instead of these approaches, data-driven approaches are more advisable so that many researches are in progress today [4], [8], [10].

This paper describes a data-driven dialogue system in cafeteria situations. The system plays a cashier's role by automatically recognizing a customer's dialogue act (DA) and responding to the customer. The corpus is composed of Korean dialogues between customers and cashiers. The system consists of two parts: DA classification and dialogue generation. In the DA classification part, Bag-of-words (BoW) model is used for feature extraction of spoken sentences, and a Support Vector Machine (SVM) classifier is used to classify them. In the dialogue generation part, a hypernetwork dialogue model learns the flow of DA sequences from the corpus and predicts the next DA in the given situation. If the predicted DA is one of cashier side DAs, an appropriate response is retrieved from the corpus corresponding to the predicted DA. Otherwise, the system requires the user to input a new spoken sentence of the customer side. The architecture of the system is illustrated in Fig. 1. The proposed method does not need prior knowledge such as a set of dialogue rules because the system learns the rules from the corpus. The experimental results show that it is possible to manage dialogues without prior dialogue rules based on a hypernetwork dialogue model.

The rest of this paper is organized as follows. In Section 2, DA classification is described. Dialogue generation based on a hypernetwork dialogue model is described in Section 3. In Section 4, experimental results are presented. Finally, we conclude with summary and future works in Section 5.

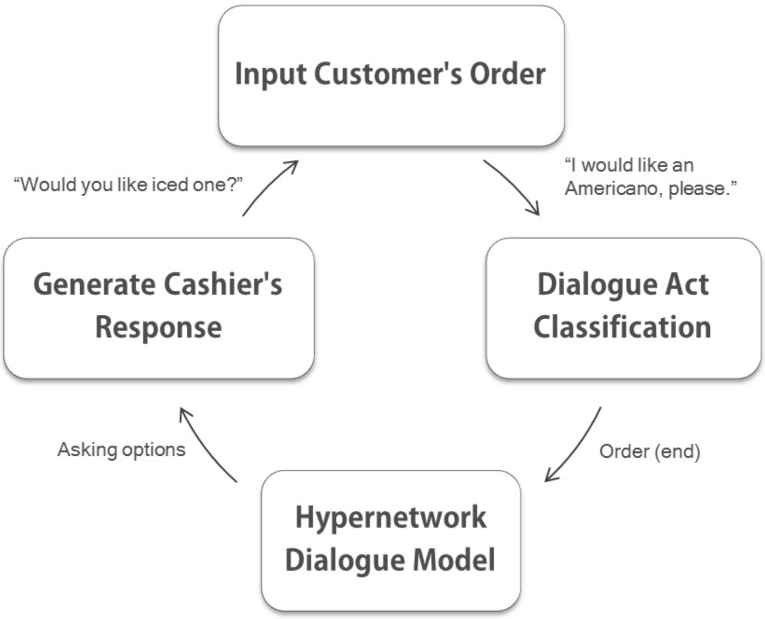
Input Customer's Order
"I would like an Americano, please."
Dialogue Act Classification
Order (end)
Hypernetwork Dialogue Model
Asking options
Generate Cashier's Response
"Would you like iced one?"

Fig. 1. Architecture of the dialogue system: First, a user inputs an order at the customer side. Next, the system recognizes DA of the input sentence. Then, a hypernetwork dialogue model predicts the next DA. If the predicted DA is one of cashier side DAs, an appropriate response is retrieved from the corpus.

2 Dialogue Act Classification

Table 1. Dialogue acts in cafeteria situations

DA	Role	Example
Greeting	Both	Hello.
Inducement to order	Cashier	Can I take your order?
Order (End)	Customer	I would like an Americano, please.
Order (Continue)	Customer	I would like a Café latte and
Payment information	Cashier	The total is 15,000 won.
Asking options	Cashier	Would you like iced one?
Deciding options	Customer	No. I would like an iced one.
Order Confirmation	Cashier	One Americano and two Café latte.
Question about menu	Customer	Is there café mocha?
Modification on order	Customer	I would like to change it to a Café latte.
Asking packing	Cashier	Here or to go?
Menu information	Cashier	Ice flake is 9,000 won.
Question about payment	Customer	Is it 8,000 won?
Sign request	Cashier	Sign on the screen, please.
Yes/No	Both	Yes.
Backchannel	Both	Yes.
Thanks	Cashier	Thanks.
Paying	Customer	Here you are.
Begin	-	(dummy for the system)
End	-	(dummy for the system)

Dialogue act (DA) is the meaning of a spoken sentence at the level of illocutionary force [9]. For DA classification, DAs in cafeteria situations are defined (Section 2.1). Then, BoW model for feature extraction and a SVM classifier for learning algorithm are used to classify a new sentence into pre-defined DA (Section 2.2).

2.1 Dialogue Acts in Cafeteria Situations

DAs in cafeteria situations are bounded [6]. Based on this idea, we define a set of DAs that commonly occur in cafeteria situations described in Table 1. DA 'Begin' and 'End' are dummies for the system. The first DA is always 'Begin', and the last one is 'End'. Note the difference between 'Yes/No' and 'Backchannel'. The spoken sentence 'Yes' is belongs to 'Yes/No' when it is used to respond to a question like 'Would you like iced one?' On the other hand, 'Yes' can also belong to 'Backchannel' when it is used to agree with the partner without meaning.

2.2 Prediction of Dialogue Acts

BoW model is used to extract feature of sentences and a SVM classifier is used to predict dialogue acts of given spoken sentences. In BoW model, a dictionary T is constructed from all of the distinct words in the corpus as follows:

$$T = \{w_1, w_2, \cdots, w_n\}$$

$w_1, w_2, \cdots, w_n$ are distinct words in the corpus. Based on the dictionary T, a sentence S is represented as a n-dimensional vector of word frequencies as follows:

$$< f(S, w_1), f(S, w_2), \cdots, f(S, w_n) > \in R^n$$

$$f(S, w_i) \equiv \text{the number of word } w_i \text{ in the sentence } S$$

After labeling DA for each sentence according to Table 1, a SVM with linear kernel is trained to predict DA of given sentences. Each training instance $< x_i, y_i > \in \mathrm{D}$ is represented as BoW model x_i and its DA y_i.

3 Dialogue Generation based on Hypernetwork Dialogue Model

Before generating dialogues, a hypernetwork dialogue model predicts the following DA (Section 3.1, 3.2). If the predicted DA is one of customer side DAs, the system requires the user to input an order and classify it into DA, which is discussed in Section 2. Otherwise (if predicted DA is one of cashier side DAs), the system generates an appropriate response by retrieving the most probable response from the corpus corresponding to the predicted DA (Section 3.3). This cycle is repeated until the predicted DA reaches 'End'.

3.1 Hypernetwork Model

Hypernetwork model is an extension of graphical model [3]. By the definition of hypergraph, an edge can connect more than two vertices, which is called a hyperedge. A hypernetwork model is represented as $\boldsymbol{H} = (\boldsymbol{X}, \boldsymbol{E}, \boldsymbol{W})$ where $\boldsymbol{X} = \{x_1, x_2, \cdots, x_n\}$, $\boldsymbol{E} = \{E_1, E_2, \cdots, E_{|\boldsymbol{E}|}\}$, and $\boldsymbol{W} = \{w_1, w_2, \cdots, w_{|\boldsymbol{E}|}\}$. $\boldsymbol{X}, \boldsymbol{E}, \boldsymbol{W}$ are sets of vertices, hyperedges, and weights, respectively. Each hyperedge is represented as $E_i = \left\{x_{i_1}, x_{i_2}, \cdots, x_{i_{|E_i|}}\right\}$ where $|E_i|$ is the cardinality of the hyperedge. A Hypernetwork model can be used as a probabilistic associative memory to store a data $\boldsymbol{D} = \{x^{(n)}\}_{n=1}^{N}$ where $x^{(n)}$ is n-th pattern to store. The energy of the hypernetwork is defined as follows:

$$\varepsilon\left(x^{(n)}; \boldsymbol{W}\right) = -\sum_{i=1}^{|\boldsymbol{E}|} w_{i_1 i_2 \cdots i_{|E_i|}} x_{i_1}^{(n)} x_{i_2}^{(n)} \cdots x_{i_{|E_i|}}^{(n)}$$

$x_{i_1}^{(n)} x_{i_2}^{(n)} \cdots x_{i_{|E_i|}}^{(n)}$ are vertices connected by the hyperedge $x^{(n)}$ and $w_{i_1 i_2 \cdots i_{|E_i|}}$ is the weight of the hyperedge. $\boldsymbol{W}$ is a set of weights of hyperedges and represents the parameters for the hypernetwork model. The probability of the data generated from the hypernetwork is given as Gibbs distribution:

$$\mathrm{P}\left(x^{(n)} \middle| \boldsymbol{W}\right) = \frac{1}{Z(\boldsymbol{W})} exp\{-\varepsilon(x^{(n)}; \boldsymbol{W})\}$$