Toward Lexical Acquisition during Dialogues through Implicit Confirmation for Closed-Domain Chatbots

Kohei Ono[1], Ryu Takeda[1], Eric Nichols[2], Mikio Nakano[2] and Kazunori Komatani[1]

[1] The Institute of Scientific and Industrial Research (ISIR), Osaka University
Ibaraki, Osaka 567-0047, Japan
ohno@ei.sanken.osaka-u.ac.jp
[2] Honda Research Institute Japan Co., Ltd.
Wako, Saitama 351-0188, Japan

Abstract. This paper proposes a lexical acquisition framework for a closed-domain chatbot. It learns the ontological categories of unknown terms in dialogues through implicit confirmation instead of using explicit questions which are too abrupt. Analysis of human participant responses to implicit confirmation requests confirmed the feasibility of our approach: the response expressions differed in a manner that indicated the correctness of an unknown term's predicted category. We thus propose a two-tiered method to predict unknown term categories that attempts to predict the most specific category possible and backs off to a more general category when the predicted category is not sufficiently confident. Direct evaluation showed that this two-tiered method makes correct category predictions 54.3% more often than previously proposed explicit confirmation method.

1 Introduction

Traditionally, dialogue systems are classified into two categories: task-oriented dialogue systems and non-task-oriented dialogue systems or chat-oriented dialogue systems. However, as chatting has been found to be effective in building *rapport* [3], some recent task-oriented systems also have chat functionality. For simplicity, we will call any system that can engage in chat-oriented dialogue a *chatbot*. Since building an open-domain chatbot that always generates appropriate utterances is still difficult [5], we think it is worth building a *closed-domain* chatbot, which tries to continue dialogues in a specific domain. We are now building one in the food and restaurants domain.

Although it is desirable for closed-domain chatbots to have comprehensive lexical knowledge in their domains, it is not realistic to prepare all of it in advance. Therefore, we must expect that users will use terms that are outside of the system's vocabulary[3], i.e., terms whose ontological categories the system does not know. In this case, if the system can acquire the term's category during dialogue, it would reduce the cost of manual knowledge base expansion. In this paper, we refer to acquiring the category of an unknown term as *lexical acquisition*. While there is a study for trying to acquire

[3] Here, we use *term* to mean an expression that is an entity in the knowledge base, and it may consist of multiple words.

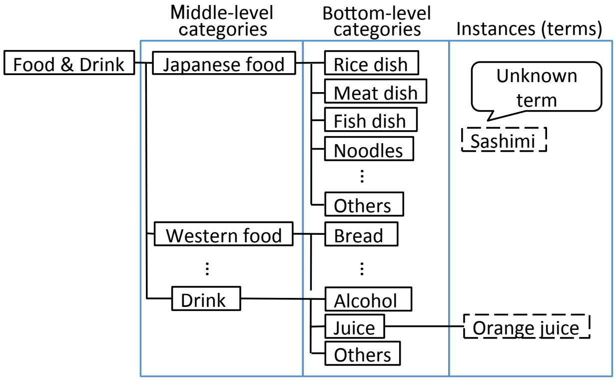

Fig. 1. Food-and-drink category hierarchy

relations among domain-dependent entities through dialogues [12], we here focus on lexical acquisition, which is required before such relations are obtained.

Previously, some methods for predicting the categories of unknown terms [10, 13] have been proposed. They acquire general categories of unknown terms, while our proposed method acquires more specific categories. Holzapfel et al. proposed a method using robot to acquire specific categories of unknown terms by iteratively asking questions [6]. For database search dialogue systems, Otsuka et al. proposed category acquisition methods that explicitly ask the user questions based on broad category prediction results [11]. Furthermore, Komatani et al. proposed a utility-based method for selecting appropriate questions [7] based on the results of category prediction. However, such explicit confirmation requests can degrade the user experience in chat-oriented dialogues, especially when the predicted category is incorrect.

In this paper, we propose a novel framework for acquiring the category of an unknown term in dialogue. It uses implicit confirmation, that is, it tries to acquire the term's category while pretending to know it. We think this category acquisition method is more suitable for chat-oriented dialogues than for task-oriented dialogues to which this method is also applicable, because repetitive explicit questions would be more troublesome. Although it does not guarantee category acquisition, it is sometimes used by humans in human-human dialogues. Our framework can be considered as an instance of implicitly supervised learning [1, 8] in that user responses to implicit confirmation requests are used as indicators for acquisition.

This paper is organized as follows. In Section 2, we explain our framework in detail. Then we explore the viability of our framework through two experiments. In Section 3, we analyze user responses to implicit confirmation requests and find that reply sentiment correlates with predicted category correctness. In Section 4, we find that our two-tiered approach predicts categories more effectively than previous approaches. In Section 5, we presents our conclusions.

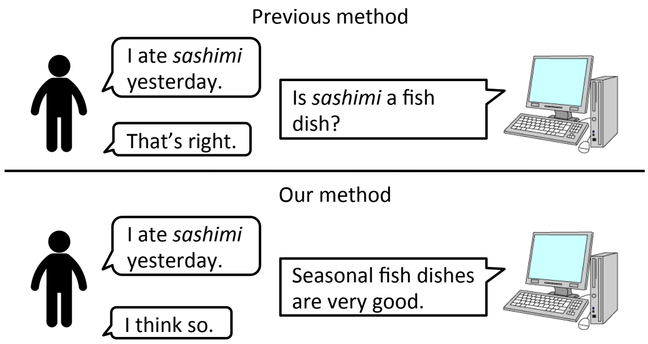

Fig. 2. Lexical acquisition through implicit confirmation

2 Framework for Lexical Acquisition

Our lexical acquisition framework requires an ontology, or category hierarchy. The category hierarchy used in this paper is shown in Figure 1. It organizes concepts into 3 levels: (i) *top-level* concepts such as `food-and-drink` and `restaurants`, (ii) *middle-level* concepts that are more specific but do not describe instances, and (iii) *bottom-level* concepts that describe instances of terms[4]. We assume that the top-level category is known using techniques like sequential labeling-based semantic slot extraction [4]. Our target is acquiring the category of the unknown term, which we assume belongs to one of the bottom-level categories.

Our proposed lexical acquisition process consists of the following three steps:

(1) Predict the bottom-level category of the unknown term. If the confidence for the predicted category is not high enough, back off to a middle-level category prediction.
(2) Generate an implicit confirmation request based on the predicted category.
(3) If a bottom-level category was predicted, determine if it is correct based on the user response to the implicit confirmation request.

Predicting a higher-level category in Step (1) is helpful because an implicit confirmation request based on an erroneous category prediction may cause a dialogue breakdown, which should be avoided. Instead, backing off to a higher-level category increases the possibility of the predicted category being correct. In this case, it is not possible to determine the bottom-level category, but we think this is better than a dialogue breakdown.

Figure 2 shows an example for the food-and-drink domain. Let us assume the system does not know the term *sashimi* (raw fish). When the user says "I ate *sashimi* yesterday," the system predicts that the bottom-level category of *sashimi* is *Fish dish*

[4] We do not deal with variation of expressions, that is, deciding if an newly encountered term refers to an existing instance in the hierarchy, though it is part of our future work.

Table 1. Intention of user responses to confirmation requests

Intention	Correct	Incorrect	Examples
Positive	19	4	"I think so."
Negative	3	15	"Is that so?" "No, I don't like it."
Indecisive	28	31	"I often cook it."
Total	50	50	

(Step (1)). Then, it generates an implicit confirmation request "Seasonal fish dishes are very good." while pretending to know that term (Step (2)). Finally, it interprets the user response "I think so," as positive and identifies *sashimi*'s category as *Fish dish*. If the confidence for the category *Fish dish* is lower than a set threshold in Step (1), the higher-level category *Japanese food* is predicted, and an utterance based on it is generated instead.

For comparison, the upper part of Figure 2 shows the previous method [7, 11], where the system asks explicit questions using only bottom-level category predictions. Our proposed framework improves upon theirs by using the entire category hierarchy to make higher-confidence category predictions and generate less intrusive implicit confirmation requests.

3 Analysis of User Responses to Implicit Confirmation Requests

In order for our framework to be viable, it must be possible to acquire the categories of unknown terms from the user responses to implicit confirmation requests. To confirm this, we asked ten participants to compose utterances for our dialogue system using ten different specified terms. Then, the system presented an implicit confirmation request for each utterance and collected the participant's response. We manually prepared ten implicit confirmation request sentences: five with the correct input term categories and five with incorrect ones. For example, we prepared sentences like "Rice dishes are filling, aren't they?", which is a correct confirmation request when the estimated category is *Rice dish*. The participants were asked to cooperate even when the system's confirmation requests were strange.

We manually classified the response intentions into *positive*, *negative*, and *indecisive* based on whether they agreed with the sentiment of the implicit confirmation request. We split the results into *correct* or *incorrect* based on the correctness of the term category in the implicit request (Table 1). While more than half of the replies were indecisive, we can see that positive user responses correspond to correct categories 83% of the time $(19/(19+4))$, while negative responses correspond to incorrect categories 83% of the time $(15/(3+15))$.

These results show that (1) response expressions change in a predictable manner based on whether or not the confirmation request contains the correct unknown term category, and (2) correct unknown term category predictions can be identified by those positive responses.

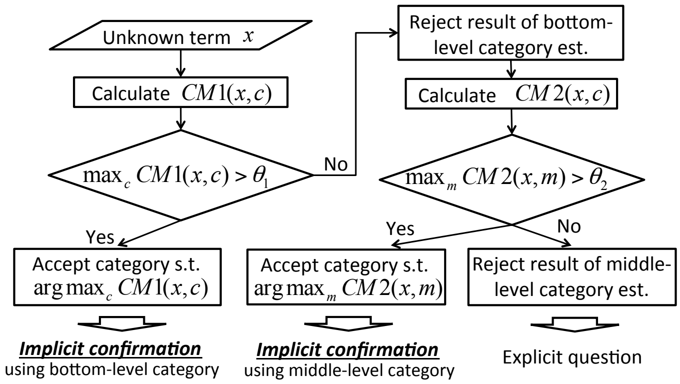

Fig. 3. Our scheme using category hierarchy

We leverage the latter finding to verify the predicted categories of unknown terms in our proposed framework. An example is shown below (*U* denotes the user and *S* the system).

U1: I had salty broiled mackerel for dinner.
S1: Seasonal fish dishes are good, aren't they?
U2: Yes, they are.

A rule-based approach was proposed to classify user responses to an explicit confirmations as positive or negative [9]. We may be able to use their approach in our framework.

Table 1 also shows that more than half of user responses were indecisive. In those cases, no category of the unknown term is acquired only by their surface expressions and this may be a disadvantage of lexical knowledge acquisition through implicit confirmation. In chatbots, however, we think it is more important to avoid abrupt explicit questions, even though the lexical knowledge acquisition is not always possible, to continue dialogues with users. As our future work, we will try to reduce such indecisive responses by changing contents of confirmation requests and by estimating user intention with dialogue context as well as their surface expressions.

4 Two-tiered Category Prediction

For Step (1) in Section 2, we propose a novel two-tiered method of unknown term category prediction that exploits the category hierarchy to improve category prediction and implicit confirmation requests. If the category prediction is clearly incorrect, the implicit confirmation request will likely confuse users, which is undesirable.

4.1 Category Prediction Approach

Our two-tiered category prediction consists of two steps as shown in Figure 3. First, we assume confidence measures (CMs), a kind of posterior probability, of each category given an unknown term. The steps are (1) calculating the CMs ($CM1$) of the bottom-level category irrespective of the middle-level category and thresholding by a parameter θ_1, and (2) calculating the CMs ($CM2$) of the middle-level category and thresholding by θ_2. At each step, we accept the category whose CM is the highest only when it exceeds a given threshold.

We calculated the CM of a middle-level category m by the sum of its child category CMs as follows:

$$CM2(x, m) = \sum_{c \in \text{child}(m)} CM1(x, c), \tag{1}$$

where x is an unknown term, and child(m) is middle-level category m's set of bottom-level categories. The CM is designed by considering the category hierarchy. This sum is expected to boost the CM of middle-level categories whose child-category CMs are relatively high but under the threshold.

4.2 Experiment and Discussion

We verify the effectiveness of our two-tiered category prediction approach by comparing its recall to the previous method [11].

The dataset used for evaluation contains 1,564 pairs of food or drink names and their bottom- and middle-level categories in the system's category hierarchy shown in Figure 1. The number of the bottom- and middle-level categories were 21 and 6, respectively. The number of child categories per middle-level category ranged from 1 to 7.

We used a) a maximum entropy (ME) model [2] for $CM1$ and b) a loss-function-based threshold adjustment method. The ME model's features are character n-gram ($n = 1, 2, 3$) and Japanese language's four character types, as used in a previous study [11]. The loss function is defined as $(\lambda \cdot FP(\theta) + FN(\theta))/N_\lambda$, where λ is a weight parameter, N_λ is a normalization factor, which is used for making uniform the range of values, and FP and FN are the false positive and false negative rates, respectively, with a threshold θ at each prediction level. The FP should be weighted because accepting an incorrect prediction causes erroneous confirmations that cause dialogue breakdown.

The evaluation procedure is as follows. First, we predicted the bottom-level categories by 10-fold cross validation for all terms in the dataset. Then, we determined θ_1 that minimizes the loss function using all the prediction results. Similarly, we predicted the middle-level categories for the rejected terms at the bottom-level and also determined θ_2 in the same way as for θ_1. Finally, we calculated the ratio of generating implicit confirmations according to the prediction results.

First, we adjusted the threshold parameters based on the relationship between the FP and FN rates. Figure 4 is a plot of the loss function at the bottom level. We set θ_1 to 0.7 so that it minimizes the loss function in which weights are given to FP, i.e., $\lambda = 2$ and 3. Threshold at the middle level θ_2 was set to 0.6 in the same way.

The category prediction results are given in Table 2. Predictions whose CMs exceeded the threshold are listed in the accepted column, while results which did not are listed in the rejected column. Accepted predictions are further divided by category

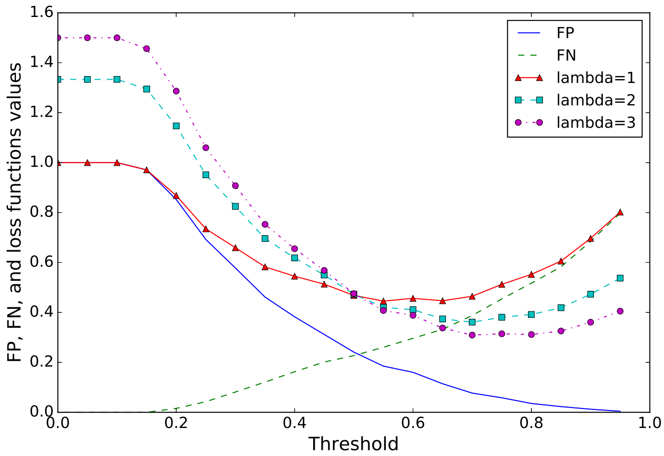

Fig. 4. Plot of the loss functions for determining θ_1

Table 2. Summary of unknown term category prediction results

	Accepted prediction results				Rejected prediction results	Total
	correct		incorrect			
	bottom-level	middle-level	bottom-level	middle-level		
# of terms	663	358	37	77	429	1564
Ratio	0.653		0.073		0.274	1.000

correctness and again by depth of prediction in the category hierarchy. Because the previous method [11] only makes bottom-level category predictions, its recall is 0.423 $(663/1564)$. However, our method also makes correct middle-level category predictions, so its recall is 0.653 $((663+358)/1564)$ - an improvement of 54.3% $(0.653/0.423 = 1.543)$. These results show that our two-tiered approach is more effective at predicting categories.

Furthermore, because in our framework better category predictions improve the quality of confirmation requests, our work is significant as a first step toward less intrusive dialogue systems that learn through user interactions.

5 Conclusion

In this paper, we proposed a novel framework for lexical acquisition through implicit confirmation to be used in closed-domain chatbots. We have shown that our approach is viable by two investigations: our analysis of user responses to implicit confirmation requests and direct evaluation of our two-tiered category prediction method.

Our next step is to develop and evaluate methods to determine the correctness of the predicted category from user responses to the implicit confirmation request. We will then incorporate this framework into a chatbot in the food and restaurant domain that we have been building and evaluate it in an in-the-wild user study.

Our framework assumes that sentence patterns for implicit confirmation requests are manually prepared. To reduce this effort is also among our future work.

References

1. Banerjee, S., Rudnicky, A.I.: Segmenting meetings into agenda items by extracting implicit supervision from human note-taking. In: Proceedings of the 2007 International Conference on Intelligent User Interfaces (IUI 2007). pp. 89–92 (2007)
2. Berger, A.L., Pietra, S.A.D., Pietra, V.J.D.: A maximum entropy approach to natural language processing. Computational Linguistics 22(1), 39–71 (1996)
3. Bickmore, T.W., Picard, R.W.: Establishing and maintaining long-term human-computer relationships. ACM Transactions on Computer-Human Interaction 12(22), 293–327 (2005)
4. Hahn, S., Dinarelli, M., Raymond, C., Lefèvre, F., Lehnen, P., Moschitti, R.D.M.A., Ney, H., Riccardi, G.: Comparing stochastic approaches to spoken language understanding in multiple languages. IEEE Transactions on Speech and Audio Processing 19(6), 1569–1583 (2011)
5. Higashinaka, R., Funakoshi, K., Araki, M., Tsukahara, H., Kobayashi, Y., Mizukami, M.: Towards taxonomy of errors in chat-oriented dialogue systems. In: Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL 2015). pp. 87–95 (2015)
6. Holzapfel, H., Neubig, D., Waibel, A.: A dialogue approach to learning object descriptions and semantic categories. Robotics and Autonomous Systems 56(11), 1004–1013 (2008)
7. Komatani, K., Otsuka, T., Sato, S., Nakano, M.: Question selection based on expected utility to acquire information through dialogue. In: Proceedings of the International Workshop on Spoken Dialogue Systems 2016 (IWSDS 2016) (2016)
8. Komatani, K., Rudnicky, A.I.: Predicting barge-in utterance errors by using implicitly supervised ASR accuracy and barge-in rate per user. In: Proceedings of the Joint conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing (ACL-IJCNLP 2009). pp. 89–92 (2009)
9. de Marneffe, M.C., Grimm, S., Potts, C.: Not a simple yes or no: Uncertainty in indirect answers. In: Proceedings of the 10th Annual Meeting of the Spacial Interest Group on Discourse and Dialogue (SIGDIAL 2009). pp. 136–143 (2009)
10. Meng, H., Ching, P.C., Chan, S.F., Wong, Y.F., Chan, C.C.: Isis: An adaptive, trilingual conversational system with interleaving interaction and delegation dialogs. ACM Transactions on Computer-Human Interaction 11(3), 268–299 (2004)
11. Otsuka, T., Komatani, K., Sato, S., Nakano, M.: Generating more specific questions for acquiring attributes of unknown concepts from users. In: Proceedings of the 14th Annual Meeting of the Spacial Interest Group on Discourse and Dialogue (SIGDIAL 2013). pp. 70–77 (2013)
12. Pappu, A., Rudnicky, A.I.: Learning situated knowledge bases through dialog. In: Proceedings of the 15th Interspeech. pp. 120–124 (2014)
13. Takahashi, Y., Dohsaka, K., Aikawa, K.: An efficient dialogue control method using decision tree-based estimation of out-of-vocabulary word attributes. In: Proceedings of the 7th International Conference on Spoken Language Processing (ICSLP-2002). pp. 813–816 (2002)