A Modules-Based, Task-Navigational Dialogue System

Motoki Yatsu Rafal Rzepka Kenji Araki
Graduate School of Information Science and Technology, Hokkaido University
{my, kabura, araki}@media.eng.hokudai.ac.jp

Abstract

There are needs of dialogue system that can act as a navigator to tasks that a user has but are uncertain of. We can realize smooth navigation to certain tasks, combining multiple non-task-oriented conversational agents with task-oriented ones. Methodology of treating both non-task-oriented and task-oriented modular dialogue systems in one integrated system has yet to be developed. The proposal method, which adopts FSM (finite-state machine) model, has each state for non-task-oriented and task-oriented dialogue, conditioned with a user's aim vector $\boldsymbol{a}$*: which has WordNet semantic similarity of user's utterance and keyword set of task domains as its component. Decision of a task domain is performed through evaluation of components of the vector on a state transition. Two dialogue examples which show the appropriateness of the task decision are presented.*

Key Words- Dialogue, Integrated Dialogue System, Human-Computer Interaction, domain selection, ontology

1. Introduction

We propose a dialogue system which can find user's need for assistance from a task-oriented dialogue system. Until the system find the user's need (*domain decision*), it can perform non-task-oriented dialogue, in which no definite task is assigned to the system.

1.1 Desire for communication

In some developed countries, it has been warned [1] that there will be more elderly people who live alone and have less communication opportunities with their neighbor people, which leads to desire for communication unsatisfied. From a worldwide view, the emergence of social media such as Twitter[1] and Facebook[2] also makes us estimate that people have unsatisfied desire to communicate.

A system was invented in the computer's first age as a substitute to a psychiatrist for a patient to talk to, known as ELIZA [2]. The technology was then improved to become systems that can imitate human's basic ability to converse with other social beings. We now call such systems chatterbots.

Not many researches had sought these objectives in methodological and multilingual perspectives until recently, but in this field we can notice a chatterbots which generates utterance in Japanese using linguistic knowledge in sentence end [3] and a statistical method [4], which applies word n-gram models in Japanese utterance generation in multiple ways.

On the other hand, dialogue agents which have multiple task domains were also developed, and are already implemented as automated online assistants, such as the online help center of IKEA[3].

1.2 Text-based dialogue interfaces

In works such as [5][6], the domain of dialogues were limited to task-oriented ones. Imai et al. [7] expanded it to non-task-oriented dialogues. However, the method does not accept a multi-turn task-oriented component dialogue system because the system selects an utterance from its components based on the scores calculated in each one turn.

The proposal method enables multi-turn task-oriented dialogue to be performed in the integrated system by adopting a state transition model.

1.3 Outline of this paper

In this paper, first we introduce the aim vector, then we overview the experimental system for the proposal method, and the method to calculate definiteness of the user's goal is reviewed. Then we evaluate two examples in which non-task-oriented dialogues were performed with domain decisions made successfully or not.

The turn differential $\Delta\boldsymbol{a}$ of aim vector $\boldsymbol{a}$ (see **2.1**) has a component which equals to an average of semantic similarity values (see **2.3**) of all the nouns in user utterance to nouns from all the keyword sets (see **2.4**) defined for each task-oriented component dialogue system.

The system judges if the degree $\|\boldsymbol{a}\|$ (see **2.1**) of the aim vector exceeds the threshold, and if it does, it transits to the task-oriented dialogue state after selecting the nearest domain (see **3.2**).

---

[1] http://twitter.com
[2] http://www.facebook.com
[3] http://www.ikea.com/

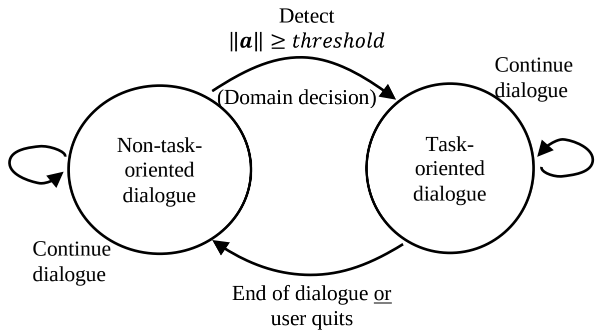

Figure 1. The system overview shown in a state diagram.

2. Core methods

In this chapter, we will discuss the main theories that govern the proposal method.

2.1 Aim vector and definiteness of user's aim

An *aim vector* $\boldsymbol{a}$ represents the orientation of user's current *aim*. In this paper, we define an aim as the user's goal with a certain orientation with which we can associate a task domain (see **3.3**) and with a measureable degree of definiteness.

The definiteness of the current user's aim can be expressed in Euclidean norm $\|\boldsymbol{a}\|$. In the discussions below, t as in $\boldsymbol{a}(t)$ denotes the number of present turn in a dialogue.

2.2 State Transition Model

The proposal method adopts a model of finite-state machine; in which one state transits to another if the transitional condition of a state is met. The system has two states: non-task-oriented dialogue state and task-oriented dialogue state. Figure 1 shows the state diagram. (See Chapter **3** for detailed description.)

2.3 Semantic Similarity in WordNet

From a user utterance, the system calculates a turn differential $\Delta\boldsymbol{a} = \boldsymbol{a}(t) - \boldsymbol{a}(t-1)$ of $\boldsymbol{a}$ that suggests the degree of interest by the user to each keyword set tied to tasks. Calculation is performed using the algorithm depicted in Figure 2. We used morphological analyzer MeCab[4] for extraction of nouns. In calculation, Leacock-Chodorow [8] semantic similarity is obtained:

$$sim(c_1, c_2) = \max\left[-\log\left(N_p/2D\right)\right] \qquad (1)$$

where N_p stands for the shortest path length between twowords c_1 and c_2, and D the depth of taxonomy. We used Natural Language Toolkit[5] running on Python and Japanese WordNet [9] to analyze Japanese utterance input and system outputs.

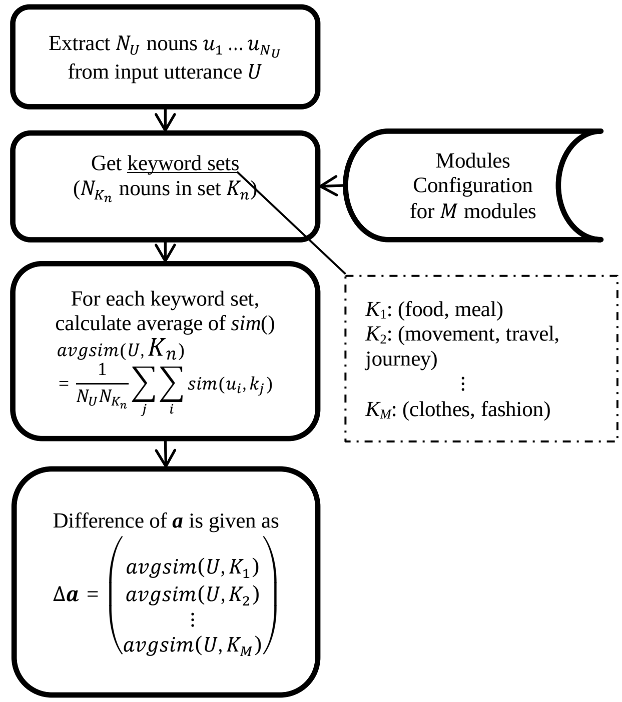

Figure 2. Chart of the algorithm which evaluates $\Delta\boldsymbol{a}$, the displacement of aim vector $\boldsymbol{a}$ and an example of keyword sets

2.4 Domains, keyword sets

A keyword set (an example shown in Figure 2) used in the proposal method includes a group of nouns related to a task domain. Each keyword set corresponds to a task domain which the task-oriented component dialogue system takes account of.

2.5 Modules

The component dialogue systems, whether task-oriented or non-task-oriented, of the proposal dialogue system are implemented modularly. The modules are configured using also a text file so that one can add an infinite number of modules without altering the main dialogue system program. For non-task-oriented dialogue, each module is modified based on an independent dialogue system and can input one line utterance and return one to the main system. To select a response, the weighted score sums of non-task-oriented dialogue modules are calculated using scripted evaluation methods (*inner modules*). We used the three methods from the work of Imai et al. [7] to evaluate the overall relevance of an input utterance to an output from a module.

---

[4] http://mecab.sourceforge.net

[5] http://www.nltk.org/

Usr >The lectures will begin tomorrow
Sys >Anyhow the lectures will begin, so *dashi* (noise)
Usr >so *dashi* what?
Sys >It is obvious that it is delicious *ta* (noise).
Usr >certainly it has some delicious taste, but it is about lectures in my university that I want to know.
Sys >But, it is unmistakable that fat is a delicious taste, and the body needs a moderate amount of it.
(part (1): from "Sys >Anyhow the lectures will begin" to "Sys >But, it is unmistakable…")
Usr >Do you know something about lectures in the university?
Sys >Exactly I know much about it *demo* (noise).
Usr >Lectures will be started in the university, so I need information about them.
Sys >You need it, is it university?
Usr >It's university.
Sys >On the contrary, some people who got 80% marks seems to have been rejected.
Usr > No, it's not about exams, it's on lectures. It might be better I'd ask the office instead.
Sys >I am sure it is high.
Usr >What do you mean by high?
Sys >Well.
(part (2): from "Sys >On the contrary…" to "Sys >Well.")
Usr >No information about lecture is on the Internet, so I have to go directly to the office. But it takes more than 2 hours by train…
Sys >It was that University is surely high.
Usr >I don't mean it stands on the high place. It takes a long time to move there because my house is far away from it.
Guide >From the conversation with the chat system, you seem interested in: *health, illness*.

Figure 3. A dialogue log between a user and the proposal system.
Japanese particles which can be regarded as noises are written with annotation. (*threshold* = 10.0)

The three methods were: computing Levenshtein distance between the output and a response in a man-to-man chat log, evaluation of grammaticality using N-gram probability, and comparison of word-end expressions.

3. The proposal method as a state machine

In this chapter, we will view the system's behavior as a finite-state machine.

3.1 Initial state

The user, who holds an aim with some degree of indefiniteness, starts a dialogue with the system. In our real world, the domain of the aim is oriented to a task domain or it is not to any task domain. In the latter case, the user engages in a non-task-oriented dialogue. Based on this observation the proposal system holds the non-task-oriented dialogue state.

The system responds to a user input with an utterance which scores the highest among the non-task-oriented modules (see **2.5**).

3.2 State transition and domain decision

If the norm $\|\boldsymbol{a}\|$ of cumulated vector $\boldsymbol{a}$ exceeds the threshold value, which suggests that the user's current interest to one of domains becomes definite, the non-task-oriented dialogue state transits to task oriented dialogue state. Before a state transition occurs, the system finds the largest component of $\boldsymbol{a}$, which is defined as related to (as described in **3.1**) the nearest domain keyword set. The nearest domain is applied to a task-oriented dialogue module most desirable to the user.

3.3 End of task-oriented dialogue

We can consider the end of task-oriented dialogue or user's halt from the dialogue as the condition of the state transition from the task-oriented state to the initial state.

4. Evaluation of the proposal system

In the subsections below, we will discuss what is needed for the proposal system to satisfy the demand of the main usage case. We show here 2 results from an ongoing Web experiment[6] with 5 keyword sets we chose randomly.

4.1 Erroneous decision

An example of domain decision is shown in Figure 3. This sample[7] was taken by one of us and the source dialogue text is in Japanese. In the dialogue part (1), the user shows that she is interested in the information about lectures given in her university and then asks the system for it. The unnatural word end of the system's utterance ("*dashi*") makes the user respond by reduplication of the chunks. The system returns a response, however, in a completely different context (with "*dashi*" interpreted as a noun), which delays the user's decision making.

In part (2) occurs another change of topic through an entire conversation. The user takes two extra turns conversing with the system in order to provide it with appropriate information.

In the end of the dialogue, the system shows the user a domain that is considered the most relevant (*i.e.* associated to the strongest component of $\boldsymbol{a}$) to the topic of the dialogue. However, we can consider this example as an erroneous decision since there are neither words of

[6] http://arakilab.media.eng.hokudai.ac.jp/~my/experiment.html
[7] All the dialogues shown in this paper are translated from Japanese originals, and any residual Japanese word or morpheme is italicized.

regions of body nor names of illness in the user's utterances in this dialogue example.

4.2 Correct decision

In Figure 4 is an example where a decision of a domain was correctly performed but in an earlier stage than the user had expected. In this case the user might have had other opinions about what she would like to do/talk against the computer. This example was chosen by a researcher not in the field of computational linguistics.

Usr >My job hunting begins next month.
Sys >Is warm fun?
Usr >I think it is better for me to buy a new business suit.
Sys >I don't understand better well.
Usr >Is there any shop that sells clothes cheap around here?
Sys >Cheap around here?
Usr >Yes, like a clothes corner in the department store.
Sys >Are talking about cheap?
Usr >Yes. About being cheap.
Sys >Continue about that.
Usr >Tell me about a store that sells clothes for men.
Guide >From the conversation with the chat system, you seem interested in: *clothes*.

Figure 4. A correct domain decision example.
(*threshold* = 5.0)

5. Discussion

The accidental changes that caused the erroneous decision shown in the dialogue of Figure 3 are supposedly due to no preservation of context. The main cause of no preservation of context is caused both for modules' internal reasons, which are innate in some types of component dialogue systems (e.g. which uses no turn-taking way of response-generation) thus cannot be fixed, and also caused by the system's criteria to switch modules we described in (**2.5**).

One way to deal with the problem may be to add a context inspector as an inner module (see **2.5**) and let the main system prefer an output that stays in the domain to which the previous utterance belongs.

On the other hand, we can conclude that the correct domain decision is caused mainly by user's disregard of utterance from the component dialogue systems. This assumption suggests that non-task-oriented dialogue state with only randomly generated responses are not necessarily effective to the user's domain selection. Also, the proposal system seems responsible for assisting the user on more stages rather than leaving entire dialogue to non-task-oriented modules, in order to preserve the context.

6. Conclusion and Future Work

In this paper, we observed how the proposal method which adopts FSM model is applied to certain usages. In the discussion we found issues such as no context preservation in dialogue, and innate problems in dialogue modules, and we mentioned a possible solution in the section. We must also consider what advantages modular architecture has in a dialogue system.

In this research, the system has a dialogue interface which deals with both the non-task-oriented and task-oriented dialogue, but not fully task-independent conversational capability, as shown in the experiment. In the future research, we aim to survey further in quantitative aspects, including threshold and error/success ratio.

References

[1] PK. Kam, "Empowering Elderly People: A Community Work Approach", *Community Development Journal*, Oxford University Press, 1996, pp. 230-240.

[2] J. Weizenbaum, "ELIZA—a computer program for the study of natural language communication between man and machine", *Communications of the ACM*, New York, NY, USA, 1966, pp. 36-45.

[3] S. Higuchi, R. Rzepka and K. Araki, "A Casual Conversation System Using Modality and Word Associations Retrieved from the Web", *Proc.of the 2008 EMNLP*, October 2008, pp. 382-390.

[4] M. Takahashi, R. Rzepka, and K. Araki, "A Performance Evaluation of Sentence Generation Method Using Web Search and Word N-gram Models", *NLP2010 Proceedings* (in Japanese), March 2010, pp. 391-394.

[5] N. Kanda, K. Komatani, et al, "Robust Domain Selection Using Dialogue History in Multi-domain Spoken Dialogue Systems(Speech Processing)", *Transactions of Information Processing Society of Japan 48(5)* (in Japanese), 2007-05-15, pp. 1980-1989.

[6] S. Ueno, I.R. Lane, and T. Kawahara, "Example-based training of dialogue planning incorporating user and situation models", *Proc. ICSLP*, 2004, pp.2837-2840.

[7] K. Imai, R. Rzepka, and K. Araki, "Performance Evaluation of Method for Selecting Best Replies from Candidate Sentences from Different Dialogue System", *IPSJ SIG Technical Report)* (in Japanese), 2010-NL-195(10), 2010, pp. 1-7.

[8] C. Leacock and M. Chodorow, "Combining local context and WordNet similarity for word sense identification", *An Electronic Lexical Database In WordNet: A Lexical Reference System and its Application*, The MIT Press, 1998, pp. 265–283.

[9] K. Kuroda, F. Bond and K. Torisawa, "Why Wikipedia needs to make Friends with WordNet" *in The 5th International Conference of the Global WordNet Association (GWC-2010)*, Mumbagi, 2010, pp 9-16.