Spoken Dialog Processing for Acquiring Taste and Knowledge

Arnaud JORDAN
Graduate School of Information Science and Technology
Hokkaido University
Sapporo, Japan
Email: arnaud@media.eng.hokudai.ac.jp

Kenji ARAKI
Graduate School of Information Science and Technology
Hokkaido University
Sapporo, Japan
Email: araki@media.eng.hokudai.ac.jp

***Abstract*—Recently many spoken agents have been developed, some of them use static rules to create a response which will make the user continue to speak, others try to get knowledge from the user to create more lively answers. Of course, both systems have different advantages and drawbacks. Static rules can be created very easily and give good result quickly. Nevertheless, we think Inductive Learning is the best method to create the most effective spoken agent because it makes the system much more human-like. A spoken agent has many different applications, but we focused on non-task oriented conversation.**

Inductive Learning lets the system evolve with the user input and progress slowly, but can progress without real limit and do not cost intensive human work. That is why, the new system that we created is based on Inductive Learning.

In this paper, we compare these two kinds of system and show that speaking with a system which has its own tastes and opinions can make the conversation more agreeable and lively.

***Keywords*-genetic algorithm, inductive learning, spoken dialogue processing, taste acquiring**

I. Introduction

Nowadays, we usually use many different kinds of spoken agents, they try to help or entertain us. Since Joseph Weizenbaum created ELIZA [1] in 1966 tons of methods have been developed and presented, each one has advantages and drawbacks. More recently, ELIZA system has been improved to create ALICE [2]. They are both using static rules and are non-task oriented.

Currently we are trying to create a flexible generic Spoken Dialog Agent which can be used in many different conditions. Like a human who can have many kinds of conversation, the new system must be able to adapt to user speaking too. There are many methods to get such system, however in our opinion, the most important ability is one's learning ability. That is why we used machine learning to create our new system. Human learn how to speak progressively while speaking together. So we aim to develop a system which learns while speaking with users.

In this paper, we compare two systems, one which uses static rule and can not answer questions with one which acquires knowledge and uses it to answer questions of the users. To compare the two systems we conduct an evaluation using the semantic differential method. Some subjects use the two systems and evaluate them.

Our original objective was to create a generic spoken dialog agent which can deal with many different situations. For examples it could be able to do common chatting, answer question, give some information or realize some task for the user. However, this paper's system focus on common chatting and question answering capacities only. We want to show that answering general question is very important to give a good impression to the user. Nevertheless, we do that without using static rule, only output generated using the user's knowledge. The system is used by many different users, in consequence the system learn knowledge from many different sources and while crossing those knowledge it can create his own specific tastes and opinions.

In the first part of the paper, we will explain the system's core and the most important capacities. Then, we will speak about the spoken agent's behavior. Finally, we present our experimental results, analyse them and conclude on the advantages of our system.

II. Outline

The system uses the same principles as GA-ILSD system [3], namely Inductive Learning with Genetic Algorithm, we also use and improve the Q-GA-ILSD[1] question treatment to replace interrogative pronouns into a meta-model. Furthermore, we use JUMAN [4] as Japanese Morphological Analysis System. JUMAN provide metadata about word type and also about word categories. We use both of this metadata to try to generate the best response as possible.

The main system's data flow is presented on Figure 1.

[1]Q-GA-ILSD adds the capability of answering questions to the GA-ILSD system. Basically, it changes interrogative pronouns into a meta-model to match the knowledge that has been acquired and in consequence be able to reply to questions.

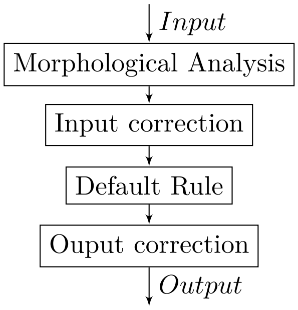

Figure 1. System's data flow

The input correction is basically a personal pronoun replacement, for example "I" is replaced by "You". If the user says "I am good", the system will understand "You are good". The output correction will be used to correct some time relative word, one day after we say "Today is warm", the output will be changed in "Yesterday was warm". However, this last correction is still under development.

The whole system is coded in Java, each rule is an object which requires an input stream and generates an output stream. Here, the input stream is the input sentence and the output stream is the response to that sentence. All rules we created, all classes we coded, can be considered as a spoken dialog agent framework.

A. *Main concepts*

In this section, we will speak about the concepts used by the system's main part, which are the basis of the system's data processing.

1) Learning: At the beginning, the system has no specific knowledge, we expect the system will evolve differently in function of the situation. If we provide too many static knowledge this will influence the evolution of the system. Therefore, the system has to get knowledge from the user's input, that is what we call learning. The system saves all user's inputs and tries to get the knowledge contained in it. For example, it saves the relations[2] between sentences and between words.

2) Statistics: While speaking and interacting with the user, the system will get many data, sometime the same data will be got several times. The more frequently a data is learned, more important it becomes. For example, if the system gets the input "I am a boy" two times and the input "I am a girl" one time, and then, when the user asks "What are you?" the output will be "I am a boy".

3) Evolution: In the same way as GA-ILSD algorithm, the system uses genetic algorithm to evolve and to delete wrong rules. To know if a rule is wrong or not, after using it, we analyse the user's reaction. If this reaction contains specific key word[3], the rule previously used is considered wrong. In other cases, the rule is considered correct, if a rule is considered wrong or correct its adaptive rate[4] will evolve in consequence.

4) Abstraction: When we speak to each other, sentences can be interpreted on many different levels. As you can see on Figure 2, a same sentence can be changed and used to generate many new outputs. We can add or remove data contained in the sentence, like only keeping some words and only using meta-data such as word type or word category. More the sentence contains precise data more the level is concrete, in an opposite manner, less data are precise more the level is abstracted.

今日は晴れです。[Input]
(Today the weather is clear.)
今日札幌で晴れです。[Adding context's local]
(Today the weather is clear in Sapporo.)
晴れです。[Removing time data]
(The weather is clear.)
形容詞 [Take word type]
(Adjective)

Figure 2. Example of sentence level

B. *Rule*

To generate output sentence we created a rule-based system. Each rule can get the input sentence and generate the output sentence if it is possible. However, to generate the most suitable output we needed to create many different kinds of rules. Those rules can be classified in four categories that will be explained in the next sections. The system uses all of these kinds of rules.

1) Simple rule: A simple rule is very simple as its name suggests. We can see their data flow on Figure 3. They contain two sentences; a match sentence and an output sentence. If the user's input sentence matches the rule's match sentence, the rule will reply its output sentence. In other cases, the rule will reply nothing.

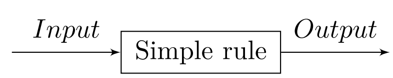

Figure 3. Simple rule data flow

お元気ですか。⇒ 元気です。
(How are you ? ⇒ I am fine.)

Figure 4. Example of a simple rule

[2]question/answer, sentence/response, etc...

[3]We use the same key word list as the GA-ILSD system
[4]Adaptive rate to user's dialog

On the Figure 4, "How are you ?" is the matched sentence and "I am fine" is the output sentence. The rule can reply only when the input sentence is "How are you ?".

2) Rule base: A rule base is a rule which contains other rules. In fact, it is just a list of rules. As we can see on Figure 5, when we ask the rule to generate an output, the rule will look for the most suitable rule to reply among all rules contained. If no rules match the input, the rule will not provide a response.

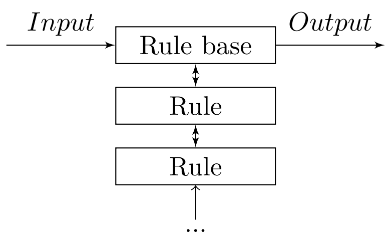

Figure 5. Rule base data flow

Default rule: The system's first rule which contains all other rules is called the default rule. It could be called the root rule too.

3) Abstract rule: An abstract rule contains a reference to one or several other rules. As shown on Figure 6, when the rule tries to generate an output, it will look for the knowledge contained in referenced rules to generate its own response. Unlike the rule base which directly uses other rule's output, the abstract rule amends the data using a specific strategy to generate a new response. Moreover, it can use the data contained in several rules to generate one output. For example, an abstract rule can combine two rules to generate a response. If the input is "What do you like?", the rule will search simple rules which match the input, find "I like cats" and "I like dogs", and generate the output "I like cats and dogs".

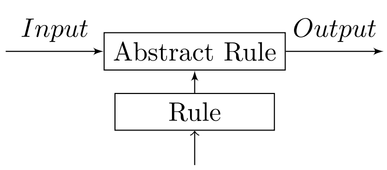

Figure 6. Abstract rule data flow

4) Recursive rule: A recursive rule changes the input sentence before sending it to another rule (cf. Figure 7). This rule adds or removes a part of the input sentence to ask another rule to reply to this new input. For example, if the user says "今日は晴れですね (Today is clear weather, isn't it?)", a recursive rule can remove not meaningful parts of the sentence to create a new input "今日晴れ (Today clear weather)".

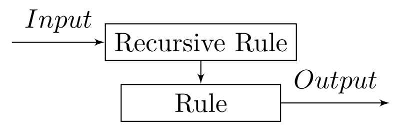

Figure 7. Recursive rule data flow

Example of recursive rule: One of the best example of recursive rule (cf. Figure 8) is the rule which changes interrogative pronouns into a meta-model. For example, "who" will be changed into the meta-model which matches all human names. There, we use a similar method as Q-GA-ILSD system. Then, this new sentence will be sent to the default rule to try to reply to the question using the new input.

友達は誰ですか。⇒ 友達は < 人名 > です。
(Who is your friend ?⇒<name> is your friend)

Figure 8. Example of a recursive rule

5) Adaptive rate: To choose the best rule to try to reply to the user on all possible ones, we calculate the adaptation rate of each rule. As a result, the highest adaptive rate is chosen. We can see the formula used on equation (1). This formula is inspired by GA-ILSD's adaptation rate. We added new variables to make the same formula able to handle any kinds of rules and to be able to hierarchize them.

Simple rule has a coefficient superior or equal to one and a reduction coefficient equal to one. However, recursive rule and rule base have a coefficient of one and a reduction coefficient inferior to one, the reason of these coefficients is that these rules do not generate an output, but use other rules which already have an adaptive rate so we only need to change it.

$$t_a = g * k * (\frac{nb_c}{nb_c + nb_i} + h) \tag{1}$$

t_a	Adaptation rate
g	Reduction coefficient
k	Coefficient
nb_c	Number of correct used
nb_i	Number of incorrect used
h	Number of learning

To generate a response we often had to use several different rules. The adaptive rate changes while combining different rules. We can see a general equation to calculate the final rule's adaptive rate through the equation (2).

$$t_a = \prod_{i=1}^{n-1} t_{a_i} \tag{2}$$

n	Number of rule used

III. Setting of the system

The system has been created to be as flexible as possible, as a consequence, many different settings are possible. In this paper we choose to focus on opinion and taste acquiring and to disable other capacities[5]. That makes the system more reactive and reduces the bug's number.

A. Used rules

As we explain in section II.B, the system uses many different rules. Moreover, the same rule can have many different instances. Concretely, we used about 30 different classes as rules. We list some of those rules below.

- *DefaultRule* is the first rule the system uses to reply an input sentence. DefaultRule is a kind of rule base.
- *ApologyRule* generates an apology such as "Sorry" if the input contains a key word like the sentence "You are wrong" which contains the key word "wrong".
- *DeclarationRuleBase* is a rule base which contains simple rules which just replies using the input sentence. When the user inputs a new sentence, the rule will generate a new simple rule to save it.
- *LinkRule* is a rule base which contains simple rules which replies using a sentence learned from the user. When the user inputs a new sentence, the rule will generate a new simple rule to save the link between the previous sentence and the new sentence.
  For example, if the system says "How are you?" and the user replies "I am fine", a simple rule which replies "I am fine" when the input is "How are you?" will be generated.
- *JiritsugoOnlyRule* is a recursive rule which removes not meaningful words[6] from the input sentence.
- *PronounRule* is a recursive rule which replaces interrogative pronouns by a meta-model.

All rules except the DefaultRule are encapsulated in a rule base. For example, as shown on Figure 9, the PronounRule is encapsulated in the DefaultRule and send amended input to the DefaultRule. PronounRule is used only one time because amended data contain no pronouns anymore, only pronouns' meta-model.

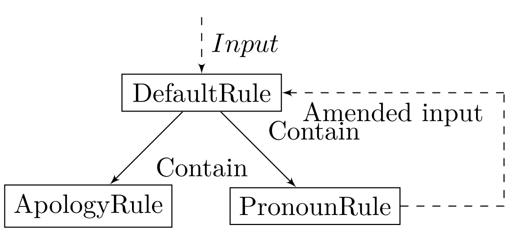

Figure 9. Example of rules arrangement

[5]task rule, sentence generation rule, etc...
[6]in Japanese grammar meaningful words are called Jiritsugo.

1) Number of rule: The starting system has 48 rules, however, if we input "Hello" the number of rules will increase to 53. At the end of the experiment the system contained about 2000 rules. Counting knowledge is very hard, however, we can consider each rule as atomic knowledge.

All new inputs are saved in one or several simple rules, which is why the number of rules increases more than one after each new input.

B. Behaviour

In the same way as a human, we can say that a spoken agent has its own behaviour. The behaviour means how the system will reply to some kinds of inputs, like question or declaration.

1) Aiduchi: If the system can not give a proper response we use *Aiduchi* to provide a response. *Aiduchi* are interjections showing that we understand and are listening what the speaker is saying. We can see on Figure 10 the using of *Aiduchi*.

猫は可愛いです。⇒ そうですね。
(Cats are cute. ⇒ That is right.)

Figure 10. Example of *Aiduchi*

2) Declaration: When the user enters a declarative sentence, the system tries to propose a question which matches this declaration, then if there is no one, it looks for a declaration. Finally, if there is no knowledge about the user's declaration the system just answers with an *Aiduchi* [5]. Figure 11 shows some of these cases. We are thinking about trying to add similar capability as Teruhisa's Guidance System [6] to give much more information to the user when it is possible.

晴れです。⇒ 晴れが好きですか。
(The weather is clear. ⇒ Do you like clear weather ?)
猫が好きです。⇒ 猫は可愛いですよ。
(I like cats. ⇒ Cats are cute.)
学校に行きます。⇒ そうですか。
(I go to school. ⇒ Really ?)

Figure 11. Response of a declaration

3) Question: If the input is a question, the system will try to answer it. If it can not, it will repeat the question to the user. This is a little unnatural behaviour. However, the objective of the system is to capture the user's feelings and opinions, therefore we have to ask the user to teach the system how to answer to the question. Take as example the conversation of Figure 12, the system reply "How are you?", but we can think this response is in fact similar to ”I do not know, how about you?”.

お元気ですか。⇒ お元気ですか。
(How are you ? ⇒ How are you ?)

Figure 12. Response to a question

IV. Experimentation

To prove the importance of having opinions and tastes to chat with a human, we consider to compare our new system with an ELIZA based system created in our laboratory[7].

We used semantic differential approach [7], [8] with same procedure as Uchida's paper [9]. We ask subjects to answer some questions, to test the two systems, and we get subject impression after tried each one. First system is randomly chosen and we only provide some basic information about the system and how to use it. We ask subjects to enter about thirties grammatically correct sentences in each system. After using one system the subject fills out the survey's form corresponding.

We also ask three introduction questions as below:

- Did you know spoken agent?
- Did you already use a spoken agent? If you did, which one?
- Do you prefer typing input sentence or use voice recognition?

At the end of the experiment, we conclude by asking what part of the dialog was the most impressive. We used this question's answer to get a more subjective impression's result about each system.

A. Results

14 subjects participate in the experiment.

As shown on Table I, only half of the subjects knew what was a spoken agent before the experiment. However, we think that many people already used a spoken agent, but did not realize it. For those who already used one, the name of Siri[8] and i コンシェル[9] (*iKonsheru*) was the most frequent.

Table I
Experimentation's result

Question	Yes (%)
Know spoken agent	50
Used spoken agent	36
Prefer voice input	50

Figure 13 shows a successful dialog's part which occurred several times during the experimentation.

[7]Accessible at: http://demo.media.eng.hokudai.ac.jp
[8]https://www.apple.com/jp/ios/siri/
[9]http://www.nttdocomo.co.jp/service/customize/iconcier/

お元気ですか。⇒ 元気です。
(How are you ? ⇒ I am fine ?)
好きな食べ物は何ですか？⇒ 味噌汁です。
(What food do you like ? ⇒ It is miso soup.)

Figure 13. A dialog part

Table II and Figure 14 show the results of the semantic differential form.

Table II
Experimentation's result

Adjective		System	ELIZA
Bad	Good	4.86	4.29
Odious	Cute	4.36	3.29
Boring	Interesting	4.50	4.57
Malicious	Kind	4.86	3.42
Idoit	Clever	4.57	4.14
Inconvenient	Usefull	3.79	3.29
Won't use	Want to use	4.14	4.29
Empty	Full	3.93	3.50
Tedium	Have many interest	5.43	5.43
Complicated	Simple	3.64	4.71
Machine-Like	Human-Like	3.43	3.36
Slow	Fast	4.57	5.86
Hard to be intimate	Easy to be intimate	4.14	4.07
Hard to understand	Easy to understand	4.86	4.29
Disappointed	Satisfied	4.86	4.36

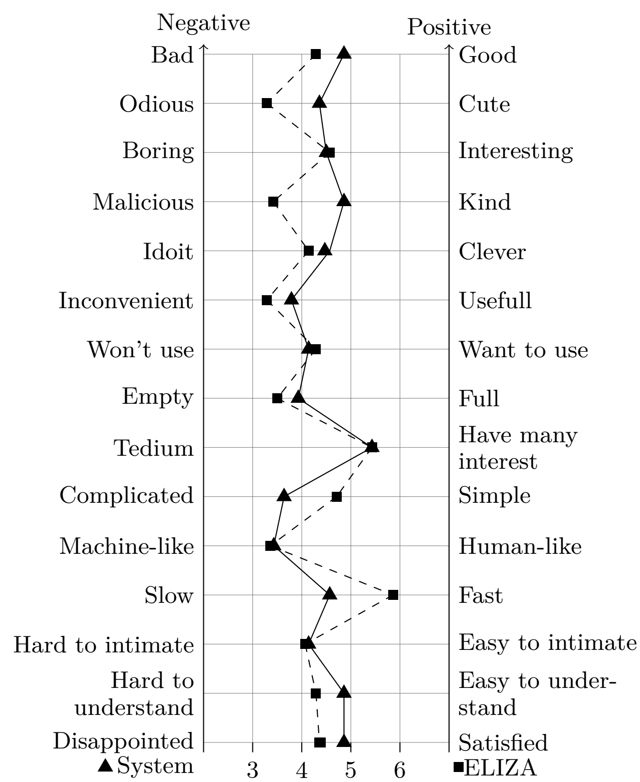

Figure 14. Semantic differential experimentation's results

B. Result's analyse

In general, the new system is a little bit better than a simple ELIZA system, but results are often very similar, many subjects prefer ELIZA system on some points. The new system's mean result is 4.40 and ELIZA's mean result is 4.19, but ELIZA got better results on two points. Moreover, five subjects give a better mean result to ELIZA.

The new system is still under development and some subject's speaking like dialect makes the system work in a bad way. Even if we ask users to use grammatically correct sentences and standard Japanese some inputs whose were supposed to be nicely answer just get an *Aiduchi* as response because no rule matches the user's input sentence due to some differences which did not change the sentence meaning.

If we watch in detail, we can see that new system is specially more "cute", "clever" and more "full", but not really better or wanted to be used more. We thought that a user feels that system was more powerful, but while speaking the system replies to question really well only a few times. In consequence, we think the system knowledge was not sufficient to get very good result, but potentially it can get better results if we teach him more thing. Moreover, even if the system gets the adequate knowledge if a question is constructed differently the system will not be able to match the correct rule. For example, "What food do you like ?" and "What are your liked food ?" have the same meaning, but different constructions, so the system will not understand it as an only one similar question.

We can notice that subjects who used non-standard Japanese encounter some problems with both systems especially due to morphological analysis error. However, the new system encounters more difficulties because if a user uses a non-usual sentence the system will not match with the knowledge it got until there and in an opposite manner, if the user teaches non-standard inputs, other users will not be able to access this knowledge too.

The ELIZA's system response is very fast and generating time does not change a lot in function of the input length. The new system tries to find the most adequate response to user's questions, in consequence question's answering can become longer and increase in function of the length of the input and of the quantity of acquired knowledge.

Every subject's dialog was quite different and impression too, but everyone enjoyed getting a proper response to some questions. However, the system still uses too many *Aiduchis* and becomes quickly annoying.

V. Conclusion

In conclusion, we had created a new spoken agent which learns taste and opinion from users to be able to reply many kinds of question. During the experimentation, we got some nice dialogs where the user could enjoy speaking with a spoken agent which looks clever. However, the system is not perfect at all, we still have to enhance it. The simplest way to increase the system response quality would be to get much more data, but this would make the system too slow to be used. We think about creating a better strategy to find the best response in the fastest way and to not have to check all system's database at all new input.

Anyway, we could show that being able to reply some easy questions let the user have better and more impressive conversation than a simple ELIZA system which replies to question with other questions. Most of subjects understand the new capacities of the system and tried to teach him some knowledge, but some subjects only ask questions without replies to the system's one. In consequence, we are thinking about adding other ways to get knowledge such as using social networking service.

In further research, we will experiment other system's behaviour to compare them and to find their advantages and drawbacks.

References

[1] J. Weizenbaum, "ELIZA - a computer program for the study of natural language communication between man and machine," *Communications of the ACM 9 (1)*, vol. 36–45, 1966.

[2] R. S. Wallace, "The anatomy of A.L.I.C.E." *Parsing the Turing Test*, vol. pp. 181-210, 2009.

[3] K. Araki and K. Tochinai, "Effectiveness of Natural Language Processing Method Using Inductive Learning," *Proceeidngs of the IASTED International Conference ARTIFICIAL INTELLIGENCE AND SOFT COMPUTING*, pp. 295–300, May 2001.

[4] "Japanese morphological analysis system JUMAN version 7," 2012, Kurohashi & Kawahara lab., Graduate School of Informatics, Kyoto University.

[5] M. Takeuchi, N. Kitaoka, and S. Nakagawa, "A spoken dialog system activating the natural response timing using prosodic and linguistic information for chat-like conversation," *IPSJ SIG Technical Report*, 2004.

[6] T. Misu, T. Kawahara, T. Shoji, and M. Minoh, "Speech-based interactive information guidance system using question-answering and information recommendation," *Journal of Information Processing*, 2007.

[7] D. R. HEISE, "The semantic differential and attitude research," *Attitude Measurement*, pp. 235–253, 1970.

[8] K. Anzou, K. Murata, and M. Numasaki, "Shakai Shinrigaku Kenkyuu Nyuumon," September 2009.

[9] Y. Uchida and K. Araki, "Evaluation of a System Noun Concepts Acquisition from Utterances about Images (SINCA) Using Daily Conversation Data," *Proceedings of NAACL-HLT*, pp. 65–68, 2009.