Comparison of two Knowledge Treatments for Questions Answering

Arnaud JORDAN† **Kenji ARAKI**†
†Graduate School of Information Science and Technology
Hokkaido University, Sapporo, Japan
{arnaud, araki}@media.eng.hokudai.ac.jp

Abstract

Recently, plenty of spoken agents have been proposed. In general, their objective is to be as human-like as possible. Some of them got very good results and are fun to use. In our laboratory, we also proposed several spoken agents with the main aim to speak naturally. Using spoken agents we can make many tasks easier and more enjoyable. In general, it is easier to speak than using a mouse or a keyboard to interact with a machine. However, is it really necessary for a machine to speak like a human? What kind of response is most suitable for the system? Those questions are really important and often not enough studied. For example, in our research we focus on question answering. Using a spoken agent, we can collect and share knowledge to reply to the questions from users. A spoken agent can acquire knowledge from some users and then share it to other users when they asked. In this case, what kinds of responses are most suitable? In this paper, we propose two new spoken agents which replies to questions in two different ways and compare them. One acquires knowledge, makes its own and uses it to reply like a human would. The other one acquires knowledge and shares it exactly as it learns it. This behavior is more machine-like. We got some interesting results which show that our systems' responses are useful in providing knowledge. And that most of users did not expect a machine to speak exactly like a human.

Keywords: Inductive Learning, spoken dialogue processing, question answering, natural language processing, multi-user system

1 Introduction

In recent years, many spoken agents have been developed. They use many different methods, but their objective is often the same. It is to be able to speak like a human. In our laboratory, we also proposed some new spoken agent like GA-ILSD [1] which aims to deal with non-task oriented dialog. Furthermore, we added question answering capability to the system and get a result's increase [2].

Since, there are many ways to use a spoken agent, we have to adapt it to the situation where it is used [3]. We think that the users do not expect to speak to a machine like to a human, but expect to get their wishes to be suited. In consequence, speaking in a human way is maybe not as important as we generally think.

In our current research, we focus on the user's question answering process and we aim to develop a new system dedicated to question answering using a rule based spoken agent. However, firstly, we have to look for the best way for a machine to reply to the user's question. We imagine two ways to answer questions and implement them in two new spoken agents. One which acquires knowledge, makes it own to reply to the questions and one which acquires knowledge and transfers it directly to reply to the user's questions. Both of those systems use the same framework as Spoken Dialog Processing for Acquiring Taste and Knowledge System [2]. In addition, the two systems have been developed for Japanese speaker.

Our research objective was to compare these two kinds of question answering and to get the user's impressions about them. We think those two kinds of answering have their own utilities, but they can be more useful in a specific situation. For example, while using at home or at work, one of the systems is maybe more suitable.

In the first part of this paper, we explain both systems' outlines. Then, we compare the two system's answers helped with experimentation we carried out. Finally, we conclude about differences of these two kinds of question answering.

2 Outline

We used the same base as Spoken Dialog Processing for Acquiring Taste and Knowledge System as a framework to create our two new systems. This framework is rule based and uses Inductive Learning to generate a response to the user's input. In consequence, the framework already acquires knowledge from users automatically. Nevertheless, we had to change how the system generates responses using this knowledge to change its comportment. Concretely, we changed rules which are used to generate the responses to the questions.

In addition, the framework does not use genetic algorithm, but uses selection and feedback similarly. We think genetic algorithm's crossover and mutation generates too many non-meaningful sentences especially for question's answering.

Figure 1 shows the input's treatment. All users' inputs are transferred to shared rules which can be used by any user's rule. In consequence, each user's knowledge is gathering up, and then can be used to generate a response.

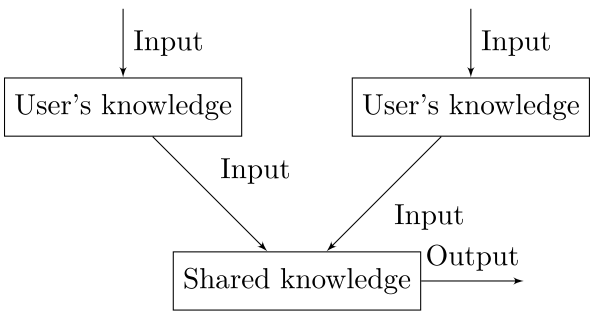

Figure 1. Treatment of user's input

We can see an example of a structure of the rules on Figure 2.

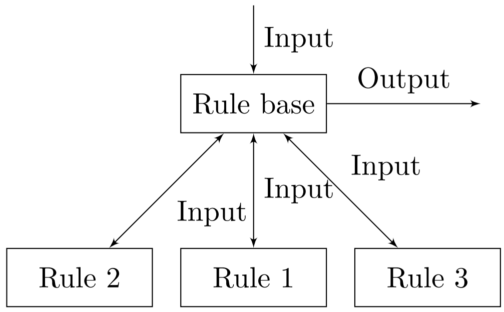

Figure 2. Example of a structure of the rules

New acquired knowledge are saved as a new rule[1] which is referenced in a rule base. When a user inputs a question, the rule base checks all rules it contains to find a rule which can generate a response.

The system acquired knowledge from each user while speaking with him. We call knowledge the input sentence as well as the link between a sentence and its response. For example, the system acquires link between "Ogenki desuka? (Are you OK?)" and its response "Genki desu (I am OK)". Moreover, the system used multi-level sentence, it means the system generates many different sentence's models from the same input, for example the sentence "Ogenki desuka? (Are you OK?)" can be simplified in "Genki? (OK?)".

Now we take a question as example, if the user input the sentence "Nani ga Suki desuka? (What do you like?)", the input will be firstly send to a rule base which checks other rules if they can generate the response. One of these rules is a rule which replaces interrogative pronouns by a meta-model, in our example the pronoun "Nani (What)" is replaced by a meta-model which matches all the common nouns. Then this rule asks the system to reply the new sentence "<Futsuu Meishi> ga Suki desuka? (<Common Noun> do you like?)", some rules match the new input and generate response such as "Neko ga Suki desu (I like cats)" and "Inu ga Suki desu (I like dogs)". These two replies are caught by the rule which generated the new input containing the meta-model. In consequence, the rule has two possible responses "Neko ga Suki desu (I like cats)" and "Inu ga Suki desu (I like dogs)" both of them are cor-

[1]For example: Rule 1, Rule 2, Rule 3

rect and can be used as a response. The first system we created simply choose the first response it get if there was no variable to prefer to use one more than the other. In this paper, we created two new systems which main difference is the choice of the adequate response.

The system also acquires links between words. For example, it acquires a link between "you" and "I", "you" and "am". If we acquire links on only one couple of sentences, the result has no meaning and cannot be used for anything. However, if we acquire links from all new inputs we can get exploitable statistics.

We can see on Table 1 that shows a small example of the result we got. We only care about the words of the same type to avoid too many links with some articles such as "a". Those links can have many meanings like synonym, antonym or meronym. However, the system just deals with links and not with their type.

Table 1. Word's link count

Word 1	Word 2	Count
I	you	5
car	street	3
car	cat	1
dog	cat	3

Using these links, we can change the user's input sentence to generate more responses. For example, if the user asks "Neko ga Suki desuka? (Do you like cats?)", the system can generate the new input "Inu ga Suki desuka? (Do you like dogs?)" and reply to this new input using "Inu ga Suki desu (I like dogs)". In consequence, when we ask "Neko ga Suki desuka? (Do you like cats?)", the system replies "Inu ga Suki desu (I like dogs)" which is a natural response in our opinion.

3 Two systems

To realize this research we modify our previous system to create two new conversational agents which have two different behaviors.

3.1 Spoken Dialog System with Acquiring Knowledge Function for Multi-User

Spoken Dialog **S**ystem with **A**cquiring Knowledge Function for **M**ulti-**U**ser[2,3] has been developed to be used by several people to become useful. It acquires knowledge from all users, and then uses the most often acquired knowledge to generate the response.

Figure 3 shows an example of dialog for SAMU.

> U: Amerika no Daitouryou ha Dare desuka?
> (Who is the president of the USA?)
> S: Amerika no Daitouryou ha Obama desu.
> (Obama is the president of the USA.)

Figure 3. Example of a dialog for SAMU

If we watch this example, we can feel that the system's response is quite human-like. The response is syntactically and grammatically correct, moreover this response has been learnt from a human who replied to the same question like that. In the response, we do not know who taught that knowledge to the system, but the system reply like if it used its own knowledge. The system chooses to reply "Obama is the president of the USA." because until now most of the people taught him this knowledge. However, maybe, some people said "Bush is the president of the USA.".

We think that behavior is the same as a human who chooses in general the knowledge he got the most often until now to reply to a question. However, if the president of the USA changes, the system will continue to reply the name of the last president until more people teach it the new name of the president or until someone tells the system it is wrong. In that case, using feedback the system can avoid wrong response quickly.

[2]Shorted by SAMU in this paper.

[3]Available on : `http://demo.media.eng.hokudai.ac.jp`

3.2 Spoken Dialog System with Acquiring Knowledge Function for Each User

Spoken Dialog **S**ystem with **A**cquiring Knowledge Function for **E**ach **U**ser[4,2] has to be used by several people too. It acquires knowledge from each user and uses all knowledge it got from all users to generate the response.

Figure 4 shows an example of dialog for SAEU.

U: Amerika no Daitouryou ha Dare desuka?
(Who is the president of the USA?)
S: Amerika no Daitouryou ha Obama desu to Hanako-san to Keita-san ga Iimashita. Amerika no Daitouryou ga Busshu desu to Takurou-san ga Iimashita.
(Hanako and Keita said that Obama is the president of the USA. Takuro said that Bush is the president of the USA.)

Figure 4. Example of a dialog for SAEU

Hanako, Takuro and Keita spoke with the system, all of them taught a knowledge which can be used to reply to "Who is the president of the USA?". Hanako and Keita taught the same knowledge to the system, "Obama is the president of the USA." and Takuro taught "Bush is the president of the USA.". The system does not look for the best response and just give all knowledge to the user. Each user's input is considered as a different knowledge even if several users said exactly the same sentence.

We think this behavior is a little more machine-like. In general, the human tries to avoid useless or wrong part of speech. If we ask someone "Who is the president of the USA?", he will certainly only reply the name of the current president and not cites from whom he learnt that name.

[4]Shorted by SAEU in this paper.

We think this system is not a simple spoken agent, but a knowledge sharing system using natural language. When you work in a group it is very hard to transfer information to other people in a smooth way. We think this kind of system can be a funnier way to exchange knowledge.

For example in a group of few people, if we want to know if someone like curry we can ask the system "Do you like curry?" and the system will teach us what thinks each member of the group.

Many users did not expect to chat with a machine like with a human, but they just expect to receive the information they want. In consequence, even if a system is more machine-like, we think it can be more useful than a human-like system. Most of the machines are seen as a tool to get objective results. If we have a simple question it is very easy to ask someone who is near you to get a response. However, in this case the response is subjective and may be wrong. SAEU's response is as complete as possible and lets the user choose which knowledge he believes or ignores.

4 Experimentation

Our experimentation focuses on the evaluation of both our systems' question answering. We do not study other parts of speech such as declarative sentence or greetings. We also compare our results to a baseline, ELIZA [4] system's response.

4.1 Preparation

To obtain a set of responses generated by our systems we asked some subjects to chat with both systems. First, they gave some knowledge to the systems and then they questioned the systems.

12 people participate in this first phase. We ask them to teach and to ask some knowledge to the systems about a specific topic to increase the systems' answering rate.

Table 3 shows information about people who participates to this work.

Table 2. Subject's information

	Number
Subject	12
Male	7
Female	5
Student	8
Worker	4
Average age	24.64

Each participant taught about 30 knowledge in the system, therefore at the end of this phase the system collected about 360 sentences containing knowledge. However, each sentence is used to generate several rules, in consequence at the end of the experimentation the system contained more than 1,000 rules.

To generate our baseline samples, we selected 20 questions randomly from our system's dialog and generated responses using an ELIZA like system created in our laboratory[5].

We also generated a set of human answers. We ask five people to reply to 20 questions chosen randomly in the same dialog corpus. We asked them to answer using the same specific topics used during this preparatory phase.

Since, humans adapt their way of speaking when they speak to a machine [5], we wanted to only use machine oriented questions. That is why; we used the same set of questions to create the baseline and human answers.

We can see examples of sample we got for SAMU, SAEU, ELIZA and for a human on Figure 5, 6, 7 and 8.

U: Suki na Tabemono de ha Nani desuka?
(What food do you like?)
S: Watashi ha Shūkurīmu ga Suki desu.
(I like cream puff.)

Figure 5. A dialog sample for SAMU

[5]Available on : `http://demo.media.eng.hokudai.ac.jp`

U: Nani wo Tabemasuka?
(What do you eat?)
S: Donburi wo Tabemashita. to Hiroki-san ga Iimashita.
(Hiroki said he ate Donburi.)

Figure 6. A dialog sample for SAEU

U: Shokuji wo Shimasuka?
(Do you have meal?)
S: Shokuji desuka?
(Meal ?)

Figure 7. A dialog sample for ELIZA

U: Karē wo Tsukuru no ha Suki desuka?
(Do you like cook curry?)
S: Futsuu kana.
(So-so.)

Figure 8. A dialog sample for a human

4.2 Evaluation

We used a semantic differential approach [6] [7] to compare our two new systems' responses and the baseline. We use the same set of adjectives as our paper [2] to facilitate future comparisons. Moreover, we also asked several questions to know for example if the answer was natural or not and which system the subject would choose to use.

4.2.1 Experimentation's stream

The subjects have to evaluate 20 answers randomly chosen from the set of samples on each system, (5 questions' answer for each system). We asked them to evaluate the response as a system response and not as a human one. After filling out a form, the new sample is displayed. The list of adjectives contained in the form is randomly sorted and the subject's previous answers are cleared.

We also ask the subjects to choose if the answer is human-like or not, if the answer have meaning or is unnatural or is natural, and if the answer contains knowledge or not.

At the end of the experimentation we explain three systems' behaviors more in detail. Then we ask the subjects the three questions below.

- Which system do you want to use?
- Which system do you want to buy?
- Which system is useful?

4.2.2 Experimentation interface

To make sure all samples and all forms are randomly organized we created a simple interface using the Swing Java GUI widget toolkit [8]. The subjects use it to fill out the form.

You can see a screenshot of the interface we used on Figure 9.

Figure 9. Experimentation interface

4.2.3 Subjects

In this second phase of the experiment, we ask subjects who did not participate in the first part to evaluate previously generated outputs.

Table 3. Subject's information

	Number
Subject	10
Male	6
Female	4
Student	7
Worker	3
Average age	25.50

4.3 Results

Table 4 and Figure 10 show the results of the semantic differential questionnaires. We also calculate the mean result of all adjectives and the mean result of all adjectives excluding the Machine-like/Human-like adjective. We do not draw the human's answer results on Figure 10 to keep it easy to read. We analyze those results in the next section.

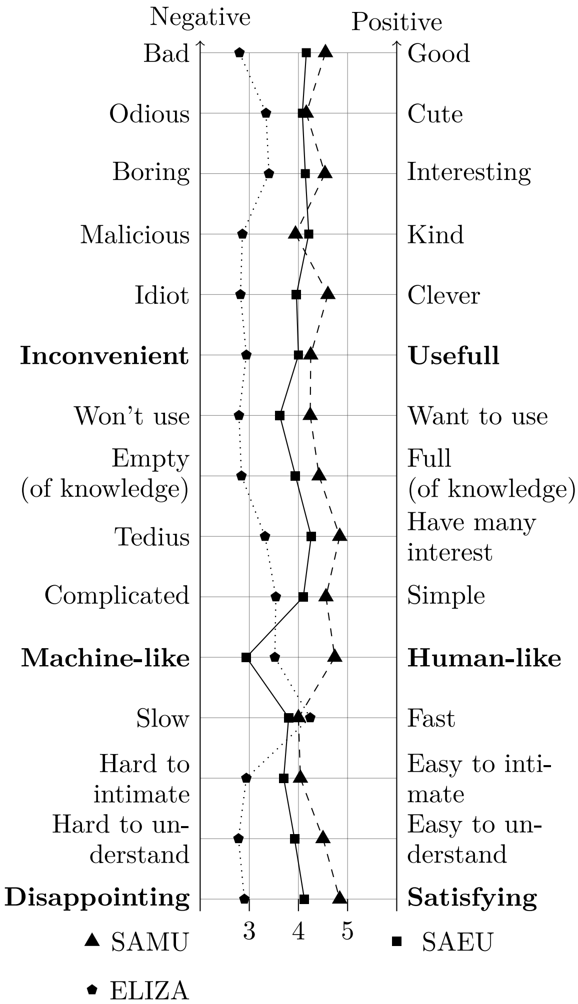

Figure 10. Semantic differential experimentation's results

Table 5 shows the result of the answer's evaluation.

Table 6 shows the results of three questions we asked to the subjects to know what system they would prefer to use.

Table 4. Experimentation's result

Adjective		SAMU	SAEU	ELIZA	Human
Bad	Good	4.55	4.16	2.80	5.39
Odious	Cute	4.16	4.08	3.34	4.60
Boring	Interesting	4.54	4.14	3.40	4.78
Malicious	Kind	3.94	4.21	2.86	4.52
Idiot	Clever	4.60	3.96	2.82	5.02
Inconvenient	**Usefull**	**4.25**	**4.00**	2.94	4.80
Won't use	Want to use	4.24	3.62	2.79	4.78
Empty (of knowledge)	Full (of knowledge)	4.42	3.94	2.84	4.57
Tedius	Have many interest	4.84	4.26	3.32	4.66
Complicated	Simple	4.56	4.10	3.54	4.84
Machine-like	**Human-like**	**4.74**	**2.94**	**3.52**	5.44
Slow	Fast	4.00	3.80	4.24	4.12
Hard to be intimate	Easy to be intimate	4.04	3.70	2.94	4.99
Hard to understand	Easy to understand	4.50	3.92	2.78	5.18
Disappointing	**Satisfying**	**4.84**	**4.12**	2.90	4.92
Mean		**4.41**	**3.93**	3.13	4.84
Mean (Machine-like/Human-like excluded)		**4.34**	**4.00**	3.15	4.76

Table 5. Evaluation of the answers

	SAMU[%]	SAEU[%]	ELIZA[%]	Human[%]
Human-like	82.0	18.0	52.0	88.0
Having knowledge	90.0	94.0	22.0	90.0
Natural	78.0	32.0	36.0	94.0
Unnatural	16.0	64.0	28.0	6.0
No meaning	6.0	4.0	36.0	0.0

Table 6. Final questions' result

	SAMU [%]	SAEU [%]	ELIZA [%]
Which system you want to use?	80.0	10.0	10.0
Which system you want to buy?	60.0	40.0	0.0
Which system is useful?	50.0	50.0	0.0

4.4 Result's analyse

First, we can confirm that SAMU's responses are human-like (82.0%), natural (78.0%), contain knowledge (90.0%), and that SAEU's responses are more machine-like (82.0%) and a little unnatural (64.0%). However, the subjects understand them (96.0%) and think they contain knowledge (94.0%). The ELIZA's responses are quite human-like (52.0%), but often the subjects did not understand them (36.0%) or thought they were unnatural (28.0%) and in general they contained no knowledge (88.0%).

Second, we confirm that our two new systems give a better impression than an ELIZA like system[6]. Even if ELIZA's answers are quite human-like, the users expect a machine to give them knowledge and not only to reply with a natural response. In addition, both systems' results are quite similar. Even if SAEU answer looks more machine-like (the difference is 1.80 points), the other parts of the impression are just a little worse (the mean difference is 0.34 points). In consequence, we can confirm that human-like answers are maybe not that important for a

[6] SAMU and SAEU means are 4.41 and 3.93 instead of 3.10 for the ELIZA system

machine, even if responses are more machine-like the final impression become just a little worse. We think most people understand that a system does not speak like a human and do not really care about that. The most important thing for them is to get an understandable response which contains the knowledge they expect.

The three last questions' results give us similar results as the semantic differential experiment's results. In general, the subjects want to use SAMU (80.0%), we think it is because it is more human-like. However, if we ask them which system they want to buy and which system they think is useful SAEU get results close to SAMU, we think that is maybe because SAEU looks more objective or precise.

5 Conclusion

In this paper, we checked that we can use the same framework to create many different dialog agents which have different behaviors very easily with minimal time cost. Moreover, when we enhance the framework all systems using it will be enhanced at the same time.

The results from both of the systems are better than an ELIZA system and give a good impression to the user. As expected, SAEU was more machine-like, but that did not really annoyed the subjects. In general, most of the users did not expect a machine to reply like a human and forgave a machine that is not natural if the response was understandable and contained the expected knowledge.

In future research, we think we will enhance the SAEU system to create a spoken dialog agent dedicated to knowledge sharing. Since, the SAEU system keeps and shares all the knowledge it acquired and it does not select a specific knowledge or change knowledge to reply to the user question, we think it could be a natural way to share knowledge. In addition, even if the response is a little unnatural, that may not be a too big drawback.

Acknowledgment

We would like to thank firstly all Araki's laboratory people who help us during the system development. In addition, we thank all subjects who accept to evaluate the system for our experimentation. Finally, we want to thank all readers to be interested in our research and to read this paper.

References

[1] Kenji Araki and Michitomo Kuroda. Generality of spoken dialogue system using SeGa-IL for different languages. *Proceeding of the IASTED International Conference COMPUTATIONAL INTELLIGENCE*, pages 70–75, 2006.

[2] Arnaud Jordan and Kenji Araki. Spoken dialog processing for acquiring taste and knowledge. *Proceedings of PACLING2013*, 2013.

[3] Sebastian Varges, Fuliang Weng, and Heather Pon-Barry. Interactive question answering and constraint relaxation in spoken dialogue systems. In *Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue*, pages 28–35. Association for Computational Linguistics, 2009.

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

[5] Joakim Gustafson, Anette Larsson, Rolf Carlson, and K Hellman. How do system questions influence lexical choices in user answers? pages 22–25, 1997.

[6] Kyoshi Anzou, Kouji Murata, and Makoto Numasaki. *Shakai Shinrigaku Kenkyuu Nyuumon*. September 2009. (in Japanese).

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

[8] Oracle. Jdk 6 swing (java foundation classes (jfc))-related apis and developer guides, 2013. [Online; accessed 21-September-2013].