The State of the Art in Dealing with User Answers

Staffan Larsson

Talkamatic AB and University of Gothenburg
Gothenburg, Sweden
`staffan@talkamatic.se, sl@ling.gu.se`

1 Introduction

This paper takes a look at how dialogue systems from three major players (Apple, Google, and Microsoft) on the market actually deal with some conversational behaviours frequently encountered in human-human dialogue[1]. The work presented here builds on the "Trindi Tick-list" (Bos et al., 1999).

2 Over-answering

What happens if the user gives more information than the system asked for? To keep it simple, we assume that the extra information given is still relevant to the task.

U: Make a call
S: OK, who do you want to call?
U: John
S: What number?
U: Mobile
S: OK, placing the call.

The dialogue above does feel a bit tedious; why not speed it up a little by giving more than one bit of information at a time?

U: Make a call
S: OK, who do you want to call?
U: John's mobile
S: OK, placing the call.

Note that S, when responding to U's question about who to call, also provides the desired kind of number (mobile).

At least two of the systems we tested do not seem to handle over-answering very well. Here's how one of them (Google Now) responds:

U: Make a call

S: OK, who do you want to call?
U: Fred's mobile
S: Try repeating the contact name or say Cancel

On the positive side, the two other systems we tested (Siri and Cortana) deal well with over-answering, and behave as expected.

3 Other-answering

We will now look at the related phenomena of other-answering. What if S had been mentally focusing on the number type rather than the person to call, and blurted out "mobile" in response to U's questions about who to call? A graceful continuation by the system would be to repeat the question about who to call.

U: Make a call
S: OK, who do you want to call?
U: Mobile
S: OK, mobile. Who do you want to call?

It turns out that Google Now has trouble with getting an answer to another question than the one they asked.

S: Who do you want to call?
U: Mobile
S: Say the name of the contact you want to call, or say Cancel

And here is the Cortana response:

S: Who would you like to call?
U: Mobile
S: I couldn't find that contact.

Basically, it seems that all these systems rely heavily on the assumption that the user will provide the information asked for. Whatever occurs after the question will be interpreted as an answer to it, even if it is not an appropriate answer (a name

---

[1]Some of the results presented here have previously been published in blog post format on the Talkamatic blog, `http://www.talkamatic.se/category/blog/`.

	Google Now	Siri	Cortana
Over-answering	misunderstood	+	+
Other-answering	misunderstood	+	ignored
Embedded revision	ignored	halts	halts

Table 1: Dialogue behaviours related to answering questions in three dialogue systems

in the phonebook), and even if it is a perfectly appropriate answer to a question that the system is just about to ask (about number type).

However, one system (Siri) does deal with other-answering in the way we outlined above. As an example, we use the email domain[2].

U: Send an email
S: To whom shall I send it?
U: It's about the trip
S: OK, I updated the subject [showing "Trip" as subject]
S: Ready to send it. To whom shall I send it?

As we can see, Siri also explicitly acknowledges understanding and integrating the "other" information.

4 Answer revision

By *answer revision*, we mean cases where a resolved question is given a new answer intended to replace the previous one. The desired behaviour from the system in such cases is that the old answer should be retracted and replaced by the new answer. Here is an example:

U: Call Jane
S: OK, Jane. What number?
U: No, Fred
S: OK, Fred. What number?
U: Mobile
S: Calling Freds mobile.

Siri ignores bare revisions altogether:

U: Call Jane
S: Which phone number for Jane?
U: Fred
S: Which phone number for Jane?

Embedded corrections starting with "no" result in the dialogue grinding to a halt:

U: Call Jane
S: Which phone number for Jane?
U: No, Fred
S: OK. (nothing happens)

Our second system (Google Now) does not seem to handle any kind of answer revision, regardless of form:

U: Call Jane
S: Home or mobile? [showing "Jane"]
U: Fred / No, Fred
S: Home or mobile? [still showing "Jane"]
U: mobile
S: Calling Jane's mobile.

Cortana ignores bare revisions, similar to Google Now. For embedded revisions, Cortana seems to misunderstand and get the "no" but not revised answer, thus ending up asking for a name that the user has already provided:

U: Call John
S: What number? Mobile or work?
U: No, Peter
S: Sure, who do you want to call?

5 Conclusions and future work

Our results are summarized in Table 1. Google Now does not do very well, which indicates that perhaps it is not intended as a full dialogue system. Cortana is also not very successful. Siri does quite well, but there is still room for improvement. As existing systems are improved and new systems appear on the market, investigations such as the one presented here need to be continually revised.

References

Johan Bos, Staffan Larsson, I Lewin, C Matheson, and D Milward. 1999. Survey of existing interactive systems. Technical Report D1.3, TRINDI (Task Oriented Instructional Dialogue) project.

---

[2]The reason for using the email domain is that in the phone domain, if the user does not respond to the question about who to call, Siri assumes that the person to call is the user her/himself. We leave a discussion of such behaviours for another paper.