CS442 Mobile Computing, Networking & Applications
Essay #9
Kite: Building Conversational Bots from Mobile Apps

20160042 Inyong Koo

This paper introduces *KITE*, an intuitive system for bootstrapping task-oriented bots using existing applications. I was interested in natural language processing and always wanted to learn how chatting bots like Apple's Siri, Samsung's Bixby work. I had some research on speech recognition[1], but this is my first time to learn about the slot-filling approach and the actual process of composing a necessary conversation.

The two main modules of *KITE* are task model extraction and question/answer generation modules. I think the novelty of this paper lies in the task model extraction module, which analyzes existing GUI based application and configures the logical structure of chat. I was impressed with how it comprehends the context and extracts the slots and intents. Especially, inspecting interaction traces was a reasonable, effection approach.

However, it leaves some concerns with applicability. The system needs to access user traces (or log data.) And it appears like the application should follow a specific convention. I wonder if the system can construct a task model from non-linear interaction flow with hard-coded UI components.

This lead to my proposal. The authors claimed that the slot-filling approach requires significant developer effort. For instance, the hand-designed control structures for each task is one thing. However, *KITE* does not resolve this limitation completely. It relies on existing hand-designed GUI structures. *KITE* is an auto-formator from an interface model to another. Of course, designing a GUI is more straightforward than constructing a chatbot control structure. But can't there be a better expression of interface structure? I think it would be cool to generate a simple GUI implementation and a chatbot from a simple modular diagram we can see in Figure 6 of the paper.

Nevertheless, I expressed some concerns about the task model extraction module, but the results appear optimistic. Testing with 25 apps, some of which so famous even I recognized, *KITE* has successfully extracted intents and slots for the target task with acceptable precision and recall. Since we are not entirely relying on *KITE* and developers will manually relieve errors.

Moving on to the question/answer generation module, I should probably study more about the seq2seq learning problems and neural sequence transduction models. I think the paper explained only briefly about the principles. I'm eager to learn about both the rule-based and neural network approach in NLP, and I will look into referenced papers later.

Meanwhile, Acquiring data from Twitter was an insightful approach. It was interesting to see how the authors create a multi-domain 5 million conversations dataset from what we call 'big data.' I have a question though; I searched a little about the BLEU score, and it says the score ranges from 0 to 1. I wonder how the multi-domain A2Q BLEU score can be 1.62.

Overall, I think this paper presents insightful, novel research. I am new to this field, and I learned a lot. I found the presentation video from the author, and I'd like to see the actual web implementation and the created bot.

[1] This paper doesn't cover speech recognition though.