Towards Learned Feedback for Enhancing Trust in Information Seeking Dialogue for Radiologists

www.dfki.de/RadSpeech/

Daniel Sonntag
German Research Center for Artificial Intelligence, Saarbrücken, Germany
sonntag@dfki.de

Introduction

We present a new approach to equip a multimodal QA system for radiologist with some form of self-knowledge about the expected dialogue processing behaviour and the results themselves for enhancing the trust in the system.

Many databases attached to a knowledge retrieval system, which the QA engines address (we use a Web Service based access to **Linked Data** operating with SPARQL queries and the HTTP protocol), are not available under special circumstances or deliver results only after several minutes of processing time for special questions. This situation changes constantly and is hardly predictable. This means **the radiologist** is constantly unsure about the query success and how much time it will take to get a "trustworthy" answer.

Our Approach

The availability of explanation capabilities can address the majority of trust concerns identified by the user.

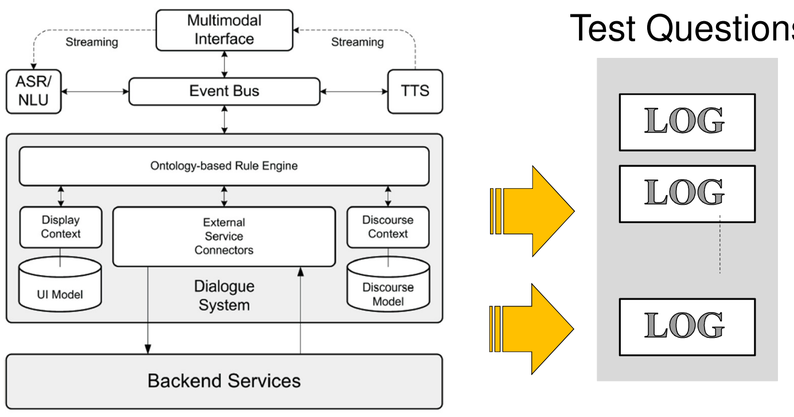
Learned models are used to provide feedback of the QA process, i.e., what the system is doing and delivers as results. The resulting automatic feedback behaviour should enhance the user's trust in the system.

With the help of **association rules**, we should be able to predict empty results and answer times, and classify queries for the probability of success according to query features and specific access and quality properties of the answer services.

We focus on the semi-automatic procedure:

1. The operationalisation algorithm must be run to mine the current data sets. The resulting rules are combined with the manually created rule set.
2. A dialogue system expert selects additional rules he finds useful, although he is not an expert of the radiology domain (diagnosing the cause of (dis)satisfaction, misunderstanding, and expected or unexpected behaviour).
3. According to the symptoms encountered in the pre-selected predictive model, the action rules (dialogue moves that give feedback to enhance trust) can be updated.

Transaction Set / Logging Framework



Speech Dialogue System

The generic framework follows a programming model which eases the interface to external third-party components (e.g., the automatic speech recognizer (ASR), natural language understanding (NLU), or synthesis component (TTS)).

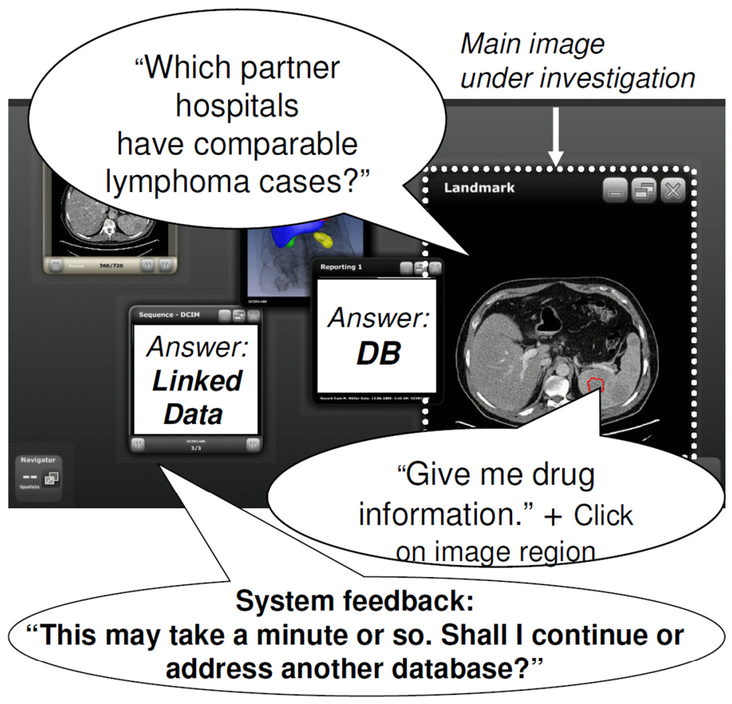


Figure 1. Speech-based dialogue and touchscreen interface.

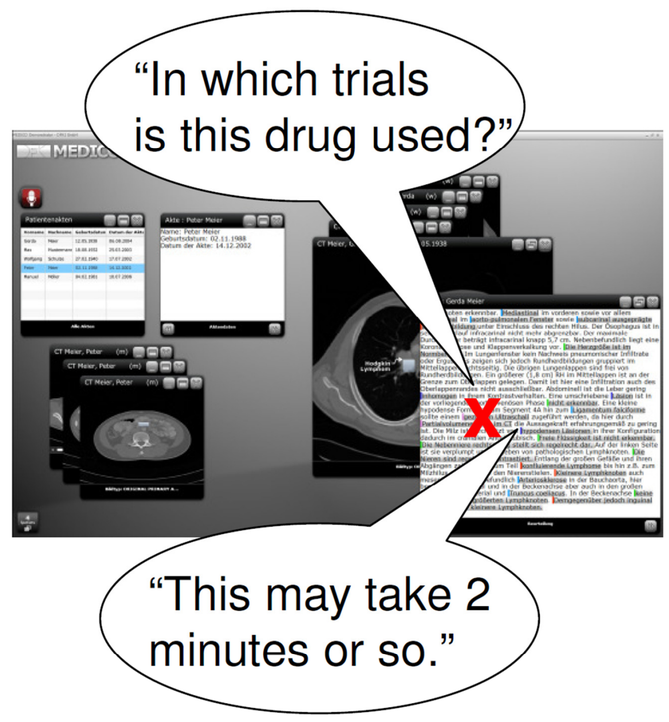


Figure 2. Touchscreen interaction and speech dialogue for additional drug information in a patient's textual finding.

Evaluation

Are there any precise characteristics and features in the ontological dialogue manager assertions (information state) that will cause the medical QA dialogue to fail or succeed?"

```
Supp  Conf  Lift
0.180 0.644 0.964
QUERY_NO_FOCUS => TURN_SLOW

0.139 0.780 1.117
QUERY_FOCUS(DRUG#Lymph) QUERY_SEMANTIC => TURN_SLOW

0.056 0.832 1.424
QUERY_FOCUS(DISEASE) QUERY_SEMANTIC => TURN_NORMAL
```

- Ten human test users (5 radiologists, 5 medical students) reported that especially the long response times for some queries (if longer than 15 seconds) were perceived as being much shorter when the question feedback was adequate. (ANOVA test and is significant at alpha = 0.05.)

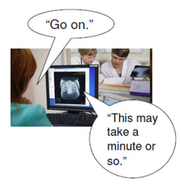
Second test: specialised test set where at least 50% of the example questions should fire one of the abovementioned feedback behaviours.

- The adapted system received a much higher overall score due to better ratings on the user evaluation of the user questions:
  (a) "**The error messages are helpful and enhance trust in the system.**";
  (b) "**The pauses between question input and answer output seem to be short.**" These results are again significant at alpha = 0.05.

This means we can semi-automatically learn models to improve the question feedback and trust in the multimodal QA system for radiologists. A drawback is that our recall for the initiation of answer feedback was low. (This means that we missed many situations where a system-initiative feedback would have been appropriate.)

Multimodal Dialogue

1. U: "Show me the CTs, last examination, patient XY." (retrieval stage)
2. S: Shows corresponding patient CT study picture series.
3. U: "Show me the internal organs: lungs, liver, then spleen."
4. S: Shows patient images according to referral record.
5. U: "Annotate with lymph node enhancement (+ pointing gesture on region)"; so lymphoblastic (expert finding)."
6. S: "Region has been annotated."
7. U: "Give me drug information (of this region)." **(A pattern based on the question focus and the SPARQL query concepts could be mined).**
8. S: "This may take a minute or so. Shall I continue or address another database?" **(2000ms)**
9. U: Confirms procedure
10. S: "Five corresponding details found in drugbase while using Radlex search terms." **(33000ms)**



The radiologist switches to the differential diagnosis of the suspicious case, before the next organ (liver) is examined, the image annotations can be completed, and the medication is prescribed.

Supported by Federal Ministry of Economics and Technology

This research is supported by the German Federal Ministry of Economics and Technology under the grant number 01MQ07016 (THESEUS).

The responsibility for this publication lies with the author.