KTH Computer Science
and Communication

Opposition for Bachelor's thesis in Computer Science

Thesis compiled by
Gustaf Lindstedt
Martin Runelöv

Title of thesis:
Context modeling using a common sense database

Opponent:
Johannes Linder

This review is based upon the Opposition protocol for Master's projects from CSC at KTH.

This Bachelor's thesis aimed to investigate how well common sense databases can be used to create a general purpose chatbot capable of mimicking human behavior. A prototype chatbot using the common sense database ConceptNet was built in order to evaluate the usefulness in context modeling. The title of the report is well-adjusted to fit the subject and clearly describes the content of the report, except perhaps a missing "Chatbot"-word (since the report investigates context modeling in the context of chatbots).

The first chapter is a general introduction to the subject which very naturally leads the reader to the problem statement; the text starts by introducing the wide area of chatbots and tests to evaluate performance and then narrows the the text down to context modeling and, finally, common sense databases. Furthermore, the problem statement is defined in detail by three questions and supports the purpose of the report.
However, two things could be defined more concretely:

1. The construction of a chatbot prototype is mentioned rather vaguely in the problem statement and could be rephrased more clearly.
2. How the chatbot prototype will be evaluated is not clear in the problem statement and could be explained more concretely.

Otherwise, the authors properly motivate their choice of method of tackling the problem statement (a literature study in addition to some type of subjective test run of the prototype) by arguing for the wide area of the problem domain as well as the abstract nature of measuring conversation quality.

The next chapter presents a wide background to the subject of chatbots and different approaches in building them. Since part of the report's goal is to compare methods for building chatbots it is relevant to present a wide overview of different chatbots in the background. The only critique that can be found here is to place section 2.3 regarding ontologies before section 2.2 where it is assumed the reader is aware of the meaning of ontologies.

The Analysis of the report is exhaustive and interesting, discussing factors and angles to consider when modeling context in conversations. However, no theory or previous research in the field of context modeling is referenced in relation to the authors own ideas, something that would greatly increase the credibility of their arguments.

The implementation is explained in a perfect level of detail in section 4.1. When the Result section is presented in the same chapter, however, it is again not exactly defined how the results were obtained (how the prototype was evaluated) and some discussion tends to float around the presentation of the results rather than clearly presenting them. An improvement could be to formally define the evaluation procedure of the prototype and then try and move the discussion of the results to the pure Discussion section.

The discussion held in chapter 5 is exhaustive and a lot of different angles to improve performance is brought to light. It would raise credibility to link the discussion to previous research or theory in the field to back up the arguments regarding improvements.
The conclusions seem credible and deducible from the results obtained, that common sense databases as of yet lack in size and quality. These results regarding limitations of the current state of common sense databases could be newsworthy for researchers in the field.
Lastly, the bibliography feels credible and relevant, with a lot of research papers and journals in the field of Context Modeling and Artificial Intelligence.

The stronger features of the report is the birds-eye overview of the subject the reader receives from the background section as well as the implementation description which was described in exactly the right amount of detail. The most notable weaknesses would be the somewhat loosely defined method of evaluation the results were obtained from and the lack of comparison between the authors conclusions and previous research in the discussion.

Questions to author:

1. What techniques does one use to construct well formed sentences in chatbots?
2. Which one was the most limiting factor in your prototype: The NLP complexity or the lack in quality in common sense databases?
3. Do you think the lack of 'world state'-knowledge in your prototype had noticeable impact on your results?