Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce   Patricia Schmidtova
Název práce   Theatre play generation
Rok odevzdání   2022
Studijní program   Informatika   **Studijní obor**   Matematická lingvistika

Autor posudku   Prof. Dr. Lucie Flek   **Role**   Oponent
Pracoviště   University of Marburg,
Dept of Mathematics and Computer Science,
Language Technologies Group

Text posudku:

The master thesis explores the newly opening possibility of generating a full theatre play with large pretrained language models. These models have been recently shown to possess impressive capabilities based on the learnt word probabilities, for example answering open questions, completing short narratives or generating short chit-chat dialogs. The novel research problem addressed in this thesis is therefore not unrealistic, albeit very ambitious.

The author. Ms Patricia Schmidtova, focuses on the technology based on the GPT-2 language model variants, as these are more transparent and easier to operate than the better performing, but partially monopolized 2022 industry models, while still yielding a competitive performance. She introduces and patiently explains to the reader multiple challenges that occur when aiming at producing and evaluating a long creative piece such as a theatre play script, and shows numerous major weaknesses of current systems. She further proposes, implements and evaluates multiple heuristics to improve these, mostly based on a cretive use of the state of the art natural language inference approaches, combined with hierarchical generation strategies, state-of-the-art decoding methods, and character orchestration heuristics.

 The thesis shows research creativity as well as expertise, is engaging and well written in a scientific style, exceeding the typical master student project requirements in scope, detail and structure, and is written in a clear, scientific style. The individual experiments were very well described, in particular the choice of various experimental setup options are very well explained and grounded in contemporary literature, as well as clarifying why other popular alternatives were unsuitable for a given scenario. The experiments are excellently related to each other and to the research questions. The research contributions of this thesis are major, from creating a new unique data set, over conducting extensive evaluations of the fine-tuned models on a large collection of existing datasets, to introducing new evaluation metrics and creative generation improvement heuristics.

The scope of the literature review is excellent, encompassing a very broad research area, but still clearly showing both an in-depth understanding of the current technologies and the familiarity with the state-of-the-art literature in the creative narrative generation field.

The methods used in the experimental chapters are very reasonable, i.e. appropriately chosen, while still novel enough to be interesting for the research community. The results obtained are

very promising and useful for future researchers in the area. In addition, the end to end evaluation with both the computer science students and the theatre professionals yields a unique comparison of the expectations and perceived challenges of the system. Even though some of the results might be disappointing for the linguistics community (the largest generic pretrained model still often being ranked better than smaller, but more complex approaches), overall the methods proposed are exciting and likely to be further used when the computational limitations allow them to be combined with the current largest models.

Some questions to the methodology:
- Couldn't the repetition problem, which was hindering the NLI entailment usage, be addressed as a preprocessing step, e.g. with an SBERT similarity? Generated utterances could be rejected if high similarity, otherwise still accepted if high entailment, couldn't they?
- The author mentions that the beam search didn't improve the results. We have an opposite experience from most of our short generation exercises (task-oriented chatbots). Could you elaborate on why this wasn't the case here and how it was assessed?

As can be seen from the above assessment, Ms Schmidtova worked on the assigned task thoroughly with very good results. Her experiments were scientifically conducted, critically evaluated, and iteratively improved. The writing is very clear and comprehensive, and the research questions were clearly presented and answered in detail.

Práci doporučuji k obhajobě.

Práci navrhuji na zvláštní ocenění.

Pokud práci navrhujete na zvláštní ocenění (cena děkana apod.), prosím uveďte zde stručné zdůvodnění (vzniklé publikace, významnost tématu, inovativnost práce apod.).

The thesis addressed a new research problem - an automated theatre play generation, and showcases the complexity of this topic in impressive breadth and depth using state of the art technologies. The thesis is very well written in scientific English, answers the research questions thoroughly in a set of meaningful comparative experiments, and makes the area accessible to newcomers interested in the field. The evaluation in collaboration with the theatre artists is unique as well, and yielded multiple publications already.

Datum 15. August 2022   **Podpis**