

Master Thesis Review

Faculty of Mathematics and Physics, Charles University

Thesis Author	Ekaterina Garanina		
Thesis Title	Table-to-Text Generation via Logical Forms		
Submission Year	2024		
Study Program	Computer Science		
Branch of Study	Language Technologies and Computational Linguistics		
Review Author	Ondřej Dušek	Role	Supervisor
Department	Institute of Formal and Applied Linguistics		

Review Text:

Thesis Topic The focus of Ekaterina Garanina’s master’s thesis is table-to-text generation including mathematical and logical operations. Solving this task is crucial for the automatic production of informative summaries or statements describing data tables. Within the current research context in data-to-text generation, this topic is extremely timely, as documented by many related works from recent years. Not only that, this thesis targets a specific gap in the current research: the problem of selecting relevant content for the output. Current research typically assumes pre-selected parts of the table on the input. The chosen approach is a pipeline combining finetuned T5 language models (LMs) with symbolic operations. First, multiple viable logical operations templates are produced by a LM based on the table statistics. Then, all templates are filled by correct information, selected using explicit logical operations over the table. Finally, another T5 LM generates the final natural-language statements. This approach is novel and very appropriate for remedying another major problem of current neural data-to-text systems in general: preserving factual accuracy of the generated statements.

Contents Summary The implementation mainly focuses on the LogicNLG dataset, which is used for primary evaluation, while parts of the system are trained on the related Logic2Text dataset. The evaluation is working with the BLEU n-gram overlap metric, following prior works, but also uses trained neural factual consistency metrics TaPas and TaPEX, as well as human evaluation on a sample of the outputs. The author first reimplemented a baseline end-to-end generation model (which assumes preselected table columns) using a T5 LM, and she went further to improve the model’s input textual representation of the table data. The main implemented model is, however, a full three-step pipeline, with the first and last step based on finetuned T5 LMs and the middle step based on explicit symbolic operations:

1. A T5 using a carefully chosen data input format representing the table and basic statistics produces logical form templates, which summarize interesting facts from the table and include various logical operations (from a set defined by the LogicNLG/Logic2Text datasets).
2. The templates are exhaustively filled in using a rule-based module, based on the explicit symbolic notation of the templates. This means that all the necessary mathematical and logical operations are symbolic and guaranteed to be accurate, which is supported by specific consistency checks. I would like to note that this step represented a substantial implementation workload.
3. The last step uses a filled template and produces a corresponding sentence, again using a finetuned T5 LM.

The overall system is able to generate multiple factual statements from the dataset. These can be then presented to the user as a whole, in a varied subset, or as a single randomly selected item.

All the implemented steps are first evaluated individually, to support the final model choice for use in the pipeline. Finally, the whole pipeline is evaluated end-to-end. All the evaluations are carefully compared to reasonable baselines. The end-to-end evaluation is further supported by a comparison to multiple prior works on the dataset. Special care is taken to resolve discrepancies in metric implementation and data formats, which goes way beyond the usual customs of today’s NLP literature.

The results are very positive overall, starting with the change of baseline input format which resulted in a factual correctness improvement. The main pipeline model shows greatly improved factual correctness, compared to both the internal as well as multiple external baselines, reaching new state of the art in correctness metrics on

LogicNLG. While BLEU did not improve, this was partially caused by producing valid statements not reflected by the references, partially due to slight formatting differences (the thesis reverted some disfluency-inducing preprocessing steps on the LogicNLG dataset). The human evaluation confirms much higher factual correctness than the baseline, with slightly decreased interestingness. However, unlike factual correctness, interestingness showed to be rather subjective, with substantial disagreements between annotators. A further analysis of the automatic metrics generally used for LogicNLG showed their problems and limits, underlining the importance of the human evaluation.

Text Structuring The text is structured into 8 numbered chapters. Chapter 1 is a short introduction describing the problem at hand and the main research questions. Chapter 2 gives a brief background on the basic techniques for the task. Chapter 3 gives a short overview of alternative related datasets, standard metrics and previous approaches for the task. The short Chapter 4 describes in detail the datasets used in the experiments, including some of their issues (such as the preprocessing mentioned above). Chapter 5 gives an account of the baseline end-to-end LM, including the input formatting improvements. Chapter 6 represents the main implementation work of the thesis: It gives a detailed account of the whole pipeline system and its individual steps, including intrinsic evaluations and discussions on model selection. End-to-end evaluation of the whole pipeline, comparison to baselines, human evaluation and important discussions are included in Chapter 7. Chapter 8 concludes the thesis with a short informative summary.

The whole text is well organized and written in excellent English. It is a little dense at times due to the complex subject, but stays well understandable.

Work Progress Works on the started in early 2023 and continued up until the end of summer, in joint supervision with Prof. Gertjan van Noord from Rijksuniversiteit Groningen. The three of us held regular meetings, where the author always came prepared and showed her extremely efficient approach to the matter. She stayed highly motivated throughout the works, worked very diligently and independently, offered creative solutions and dealt smoothly with setbacks she encountered. Any significant issues of mine or the co-supervisor were resolved successfully before the thesis submission.

In addition to her work on her thesis, the author also got involved in works on a related paper on data visualization, prepared by PhD students under my supervision, where she only confirmed her creativity and efficiency.

Overall evaluation I believe that Ekaterina Garanina's thesis is outstanding, especially considering the extent and depth of her research work on this very challenging and very timely topic. She was able to develop a well-working solution to the entire problem of table-to-text generation, including the content selection step, which was generally ignored in previous research. The result is of a sufficient quality for a publication at a top-class conference in the field.

I am very happy with the final thesis, which I believe fulfills all the requirements and standards of Charles University to be accepted. I do not have any specific questions for the defense.

I recommend that the thesis be defended.

I nominate the thesis for a special award.

I believe that the thesis deserves to be nominated for the Dean's award, due to the timely topic, novel solution, and state-of-the-art results, as detailed above.

Prague, 31 January 2024

Signature: