

**Critique Report on the Paper: “Condition  
Aware and Revise Transformer for Question  
Answering [3]”**  
*Sumaiya Tabassum Nimi*

## Background Information and Purpose of the Research

Since for the automated question answering platforms, tangential information obtained from external sources like pertinent databases are important for providing precise replies, there were research works conducted in the corresponding domain for utilizing these information while generating the responses, assuming that the obtained information were free from any error. But in real life applications, these information often tend to be erroneous or even absent at times. In order to alleviate the issues arising due to these defective conditions, in the paper [3], a novel Transformer[2]-based model was proposed for selecting responses to users' queries that not only directly incorporated the relevant external information into the model for finding the model output but also revised the information based on the conversations with the users' whenever necessary and selected appropriate output based on the revised information.

## Methods Proposed

The goal of this work was to select an appropriate response from the set of candidate responses based on both conversation with the user and a set of condition obtained from external sources, that might be revised based on their original values and the conversation history. Towards this end, in the paper [3], novel architecture called CAR-transformer was proposed that selected the responses undergoing the following steps.

- Step-1: A dialogue encoder module transformed the conversation history and the users' queries into a series of vector representations.
- Step-2: Based on the users' queries, condition values from external sources were accumulated and a conditions reviser module generated revised condition values as one-hot vector representations based on the collected original values of the conditions and the the encoded representations of the conversation with the users generated in Step-1.
- Step-3: Then a conditions encoder module generated vector embedding representing the conditions using the one-hot vector representations generated in Step-2. The Conditions encoder module was basically a fully connected neural network.
- Step-4: Finally, a MLP classifier model selected the appropriate response using concatenation of the outputs generated in Step-1 and Step-3 as input.

## Strength of the Work

- Incorporation of conditions obtained in the form of categorical values from external sources like databases into end-to-end trainable deep learning models so that these directly influence the generated network outputs.

- Novel inclusion of a condition reviser module in the classification module to correct the inaccuracies resulting from the erroneous external condition values based on the conversations with the users, so that more pertinent responses could be selected by taking into account both the users' queries and the revised condition values.
- Experimental evaluations on both a custom dataset and a publicly available dataset done on both automatic and manual metrics demonstrated the superiority of the proposed architecture compared to the models proposed in the prior research works.
- It was shown that the proposed condition reviser module worked by verifying that the distributions of the revised condition values were closer to that of the true condition values.

## Limitations of the Work

- In Table-7, performance on the bAbI dataset [1] was evaluated to further solidify the effectiveness of the proposed architecture. But the results were compared with those produced by only two other models. Comparison with some more models, like was done in case of the custom dataset in Table-4 would have further solidified the claim of superiority.
- Description of the custom dataset used for evaluation seemed limited. It was not clear if each output response entailed exactly one of the 35 candidate responses. If exactly one was chosen, it could be seen from Figure-4 that the categories were not uniformly distributed, some of them were very frequent while some were really scant. Hence it seemed there should be an evident bias towards the most frequent categories. However, there was no discussion on any measure taken for alleviating the resulting bias.
- The proposed model incorporated a large number of representations generated by the constituent modules. Hence it seemed this model would not be suitable for deploying in memory-constrained edge devices.
- Ablation studies should have been done regarding the effect of the values of the hyperparameters like  $\eta$ , length of the generated representations, duration of conversation history taken into consideration etc. on the end results other than reporting results on some fixed values, so that the choices of the values of these hyperparameters were justified.
- The exact architectures of the neural network models used in different modules were should have been explicitly discussed so that the results could be reproduced.

## Questions Unanswered

- How many conditions could be revised at a time by the condition reviser module? That depended on the training actually, hence another question naturally arose, in the custom dataset, how many conditions were revised per sample?
- It was not clear if the external database was updated with the revision(s) generated by the condition reviser module or if it was solely for the purpose of selecting the appropriate response for the users.
- The applicability of the proposed approach on security-sensitive applications like banking transactions was not discussed. And it seemed that since this approach involved querying the database, some doubts would be raised by the stakeholders regarding deployment of designed QA platform into these security-sensitive applications.
- The experimented condition values in the custom dataset were all categorical. It was not clear if the scope of the condition reviser module could be extended to continuous values, like amount or number of products ordered also.
- Effects of length of conversation history should have been discussed so that custom dataset could be constructed for training the proposed model for obtaining the best results.

## Suggested Future Studies

- The paper presented a novel idea of suggesting edits to database based on conversations with the users. Architectures proposed in the paper for implementing this idea could be adopted for enabling these capabilities into many application scenarios other than the discussed scenario of e-commerce. For example, e-health applications can be developed that updates patients' conditions based on their conversation, so that remote monitoring of time-to-time health updates can be done by the healthcare providers.

## References

- [1] Chaitanya K Joshi, Fei Mi, and Boi Faltings. Personalization in goal-oriented dialog. *arXiv preprint arXiv:1706.07503*, 2017.
- [2] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.
- [3] Xinyan Zhao, Feng Xiao, Haoming Zhong, Jun Yao, and Huanhuan Chen. Condition aware and revise transformer for question answering. In *Proceedings of The Web Conference 2020*, pages 2377–2387, 2020.