REVIEW OF AN ATTENTION-BASED DEEP NET FOR LEARNING TO RANK

What is this paper about

A new attention-based deep neural networks are proposed to propose learning2rank problems. Specifically, multiple embeddings and a new attention mechanism are proposed to model interaction between the queries and search results. Three datasets are used to validate it's effectiveness. Results show the model's superiority.

What strength does this paper have

- multiple embeddings are used to reduce randomness of model
- attention mechanism are used to learning2rank problems.
- adequate experiment

Weaknesses and questions:

- There are many standard data set for learning to rank problem, such as: Microsoft Learning to Rank Dataset and Yahoo Learning to Rank Challenge, while the authors construct dataset theirselves to experiment instead of using these standard datasets.
- In the datasets you build, there are only two levels of search results, "related" and "not related". It's obvious that rank problem becomes easier. I prefer the performence when more related levels are exist in search result.
- It's worth be discussed whether your work does make sense. Multiple embeddings and complexity attention mechanism are used in your model, only to gain tiny improvement. It looks like superiority of the model may originate from complexity of the model, there isn't any essential innovation.

Typos, Grammar, and Style

quotation: The topics can be categorized into 7 superclasses: religion, computer, **"for sale,"** cars, sports, science, and politics.

REVIEW OF DATA-DISTORTION GUIDED SELF-DISTILLATION FOR DEEP NEURAL NETWORKS

What is this paper about

In order to reduce memory and time consumption while preserving performance, the authors propose a self-distillation mechanism which can transfer features between different distortion versions of the same training data , without assistance of other intermediate network. experiments show that model gains the better performance.

What strength does this paper have

- MMD metric is used to reduce the discrepency between global representation vectors of two-branch distorted versions data
- KL divergency is used to gain consistent posterior distributions vector between two-branch distorted versions data
- abundant experiments validate the effectiveness of self-distillation mechanism

Weaknesses and questions:

- Have you tried other data-distorted methods? Did you try a number of data-distorted methods and these two work best? In MNIST dataset, one number has even more handwritten versions, or data-distorted versions. Whether your self-distillation mechanism can gain improvement on the dataset?
- The function of your model eventually learns the mirror immutability etc. There are many ways to achieve this goal, and you just apply MMD metric and KL divergency mechanism, so there isn't any essential innovation.
- Wide coverage is used to write about your MMD metric and KL divergency mechanism, actually these are very simple and do not need to be detailed. What’s more, the optimization section can be substitute by a simply phrase of “ backpropagation algorithm”. All of these make the paper lack substantive content.