MULTIMODAL CHAIN OF THOUGHT

Selected Topics in DS 2024 Course

Author
Vasily Ivanov
2024

1 Project status report

Here I would comment on quality of the report.

- Main problem is described clearly but main method is described kinda vague
  Setting and literature are pointing in the right direction although It feels like I miss the core idea of chain-of-thought reasoning itself, though I captured it in general I still don't understand details of the logic behind this approach. Seems like it can be described as probability model or Bayesian network, to make things clearer. As far as I get it, this is just strategy of inference or postprocessing of the outputs aimed to get better results for the consequent reasoning.
  However later in the text details of the experiments can be obtained.
- Everything becomes alot clearer when you read later in text description of the model's pipeline and how LLM and representation encoders are connected.
- Main metrics is supposed to be accuracy but this information is not highlighted enough and is hidden in the text
- Multiple time it is pointed to the Appendix A if you want to reach for the details, but appendix A is missing and probably final version of the report didn't compile successfully or something like that.
- It seems like metrics are not optimal for the capacity of the model and I guess it happened because of the time pressure. I guess it is bad for scientific paper in the good journal but for student's works seems okay.
- Direction of thought and experiments seems valid and consequent to me. It would be interesting to read full version of these experiments.
- Could find which exactly LLM model was used.

In my eyes general structure is good and especially the abstract that gives short description of the paper in a solid and mature way. Literature links looks relevant and help to grasp some understanding even without diving too far. LLava-like approach looks like the good idea in general and appropriate for the task.
Data description could be a bit more structured and short for the sake of clearness.
Would like to read more about experiments and results but it is not a paper in nips so I guess this is kinda excessive from me.
Also it would be really nice if colleagues used smaller LLM and I would be able to run this thing on single or 2 GPUs.

2 Repository

- The repo looks clear and structured in terms of code, although lacks some important things for the high level of quality.

- README.md gives clear sequence of steps although pipeline is not automated enough and requires alot of hand actions. It would be nice to have a docker container or singularity container for that matter in order to make environment easier to build...

- Python lib requirements and versions are missing.

- Hardware requirements are non trivial and missing. Multiple A100 GPUS are needed in order to train the model, probable 6 or so.

In general it would be really nice if environment and requirements were clear and easy to setup. Data could be combined and put into single file, uploaded to the cloud storage for example.

3 Repeatability

Seems like procedure of training requires running LLM on the GPU and it needs huge VRAM capacity in order to run experiments. I was able to recreate environment but it seems like I can't in reasonable time to find proper hardware to recreate these experiments. I was able to run inference on single a100 GPU and it worked.

4 Conclusions

To me it seems that with a bit of polish that would be a really solid quality project, maybe even of minor paper level.