Project Progress Report

Team 18

Solving Crossword Puzzles using Loopy Belief Propogation and SweepClip

1. Project Idea

There is no change in the core idea of our project. The objective remains to develop a system capable of solving New York Times-style crossword puzzles using advanced techniques in natural language processing and probabilistic graphical models (used in LBP).

2. Methodology

Our approach involves two key stages:

1. **Candidate Answer Generation:** We generated a fine-tuned version of the GPT-4o mini model to generate plausible candidate answers for each clue. Initially, the fine-tuning code we found online was outdated. Therefore, we spent time updating and adapting the code to fine-tune the GPT-4o mini effectively.
2. **Answer Selection using Graph-based Inference:** After generating the list of candidate answers, we construct a factor graph that represents the constraints of the crossword puzzle (e.g., intersecting letters must match). We apply *Loopy Belief Propagation (LBP)* and the *SweepClip* technique to infer the most consistent set of answers across the entire puzzle.

3. Current Progress

- The fine-tuning of GPT-4o mini for crossword clue-answer pairs has been completed.
- A working implementation of Loopy Belief Propagation for crosswords has been integrated with the fine-tuned version.
- We are currently achieving suboptimal results in terms of letter and word accuracy.

4. Challenges Faced

- Updating the fine-tuning code for GPT-4o mini took longer than anticipated due to deprecation of certain APIs and changes in model interfaces.

- OpenAI no longer provides free fine tuning methodology which we figured out in later stages of our project development.
- Accuracy metrics are currently below expectations, indicating the need for better clue modeling and belief propagation tuning.

5. Future Plan

Our immediate focus is on improving the accuracy of our solution. We aim to:

- Enhance candidate generation by improving prompt engineering and possibly using ensemble approaches.
- Tune parameters of the LBP algorithm for better inference.
- Explore additional graph pruning and constraint-based techniques to improve the belief propagation step.
- Conduct systematic evaluations to identify the key bottlenecks in accuracy.
- Currently we have used only got 40 mini for fine turning our model, however we are planning to use other model such as lemma, claude etc as well.

6. Conclusion

The project is progressing steadily. While the foundational pipeline is in place, improving accuracy is our next milestone.