УДК 004.912

РАЗРАБОТКА ИНСТРУМЕНТА АНАЛИЗА ИНФОРМАЦИОННЫХ КОНТЕКСТОВ И ИССЛЕДОВАНИЕ МЕТОДОВ, НАПРАВЛЕННЫХ НА УСКОРЕНИЕ И ПОВЫШЕНИЕ КАЧЕСТВА ТЕКСТОВОЙ РАЗМЕТКИ

Новицкий И.А. (ИТМО)
Научный руководитель – кандидат технических наук, доцент Ефимова В.А.
(ИТМО)

Введение.

В поисковом агенте Alice AI LLM Search на основе больших языковых моделей ответы формируются на основе информационных контекстов — релевантных запросу фрагментов текстов из интернет-источников, полученных с помощью поиска. Для повышения качества генерируемых ответов выполняется разметка достоверности фрагментов; далее она используется для обучения и улучшения модели подтверждённости, которая классифицирует фрагменты как 0/1/2 («не подтверждён» / «частично подтверждён» / «подтверждён»). Класс «частично подтверждён» соответствует ситуации, когда факты во фрагменте совпадают с источниками более чем на 50%. Практика проверки и контроля качества разметки больших языковых моделей рассматривается как необходимое условие получения надёжных данных для обучения моделей. Основная проблема — высокая трудоёмкость ручной проверки: в среднем разметка одного фрагмента занимает 7–8 минут, из которых 3–4 минуты уходит на поиск подтверждающих источников и около 4 минут — на проверку фактов. Дополнительно разметчики нередко используют стороннюю языковую модель для поиска по контекстам, но ожидание ответа занимает 2–5 минут и также снижает скорость работы.

Основная часть.

Работа направлена на разработку инструмента-помощника, встроенного в процесс разметки, который снижает долю ручного поиска и упрощает проверку подтверждённости утверждений фрагмента по информационным контекстам. Актуальность обусловлена тем, что текущая цепочка обработки на сторонней языковой модели (deepseek-v3.1-terminus) не оптимизирована под задачу: генерация занимает 2–5 минут, удовлетворённость качеством ответов составляет около 50%, порядка 30% ответов содержат вымышленные факты, а соответствие выдаваемых фрагментов заданному контексту не превышает 70%. Для ускорения предлагается использовать более компактную модель, полученную методом переноса знаний. Для повышения качества предусматривается систематическая оценка корректности и «привязки» ответа к контекстам, включая схему, где одна языковая модель оценивает другую, а также сравнение по автоматическим показателям качества. Дополнительно рассматривается применение методов обучения с подкреплением (в том числе подход Cross Entropy Reinforcement Learning) для улучшения поведения модели в задаче поиска и представления подтверждающих фактов.

Выводы.

Ожидается сокращение времени разметки одного фрагмента с 7–8 до 3–4 минут и повышение качества разметки (F1/точность/полнота) с текущих значений около 0.6 до уровня выше 0.85 при росте удовлетворённости исполнителей выше 70%. Также целевым является повышение точности соответствия выдаваемых фрагментов заданному контексту с ~70% до ~90%. Достижимость целей предлагается подтверждать сравнением показателей «до/после» внедрения на реальных задачах разметки

Список использованных источников:

1. Boix-Adserà E. Towards a theory of model distillation // arXiv. – 2024. – arXiv:2403.09053.
2. Cheng X., Mayya R., Sedoc J. To Err Is Human; To Annotate, SILICON? Reducing Measurement Error in LLM Annotation // arXiv. – 2025. – arXiv:2412.14461.
3. Pangakis N., Wolken S., Fasching N. Automated Annotation with Generative AI Requires Validation // arXiv. – 2023. – arXiv:2306.00176.
4. Lee J., Shin J., Cho B. Evaluation of Large Language Models: Review of Metrics, Applications, and Methodologies // Preprints. – 2025.
5. He H., Shi X., Mueller J., Sheng Z., Li M., Karypis G. Distiller: A Systematic Study of Model Distillation Methods in Natural Language Processing // arXiv. – 2021. – arXiv:2109.11105.