

Отзыв научного руководителя на выпускную квалификационную работу
Пределаиной Анастасии Игоревны
на тему: «Нейросетевые методы выделения сочинительных связей»

Уровень образования: бакалавриат
Направление: 02.03.01 «Математика и компьютерные науки»
Основная образовательная программа: СВ.5152.2019 «Математика, алгоритмы и анализ данных»

Выделение сочинительных связей из текстов на естественных языках (coordination analysis, CA) — ценный этап предобработки предложений, к примеру, перед решением задач извлечения информации и перед иными методами анализа связи частей предложения, перечислений — и так далее. В ряде работ демонстрируется, что применение CA для преобразования предложений в более простые позволяет увеличить качество работы целевых методов.

Лучших показателей качества достигает метод извлечения сочинительных связей на основе Iterative Grid Labeling (IGL). Это в первую очередь один из лучших способов решения задачи OpenIE (open information extraction — выделение предикатов и их аргументов без использования “типизированных” меток в обучающей выборке). С момента публикации IGL (работа об OpenIE6, 2020) появились методы для OpenIE, скорость работы которых выше (например, DetIE в 2022). Для выделения же сочинительных связей нового быстрого метода с тех пор разработано не было.

Работа Анастасии Игоревны Пределаиной посвящена заполнению этого “пробела”: апробированы и оценены способы переосмысления архитектуры DetIE (BERT + однопроходное предсказание) применительно к задаче выделения сочинительных связей. В частности, проанализированы виды ошибок модели, на основе этого анализа предложены многочисленные варианты улучшений, приведены положительные и отрицательные результаты в пределах доступных вычислительных мощностей.

В тексте ВКР

- даны необходимые введение и обзор подходов к задаче, включая подробное описание базовой архитектуры DetIE,
- мотивация, цель и схемы экспериментов изложены ясно,
- в точности воспроизведены оценки качества, достигнутые IGL-CA в соответствующей работе, что позволило убедиться в честности сравнений,
- подготовлен анализ ошибок с пояснительными примерами, что хорошо продемонстрировало разницу в трудности между задачей OpenIE, для которой была разработана DetIE, и задачей выделения сочинительных связей, что уже само по себе ценный результат,
- на основе анализа предложено множество изменений и дополнений к архитектуре, пред- и пост-обработке данных, и измерена ценность каждой из модификаций релевантными метриками качества,
- найдено улучшение работы DetIE-CA (базовой модели в работе Анастасии Игоревны) путём применения аналога метода “лучевого поиска”, при этом

качество метода сопоставимо с IGL-CA (bert-base) и превосходит его по скорости более, чем в три раза,

- получен также ряд и отрицательных результатов, что также ценно, так как сужает дальнейший возможный поиск новых решений,
- повествование выстроено логично, текст описания квалификационной работы имеет достаточный уровень подробности и стилистически выверен.

Работа соискательницей над проектом выполнена самостоятельно и продолжалась непрерывно с августа 2022 года. Все новшества, кроме предложенных научным руководителем изначально при постановке задачи, а также технологический подход к подробному протоколированию результатов с использованием соответствующих современных информационных систем — также заслуга Анастасии Игоревны, что демонстрирует понимание процессов в проектах с “уклоном” в науки о данных, умение работать с данными и архитектурами современных искусственных нейронных сетей.

Стоит отметить и то, что Анастасия Игоревна легко разбиралась в необходимых научных статьях по теме, самостоятельно находила релевантную литературу, а также быстро при необходимости разбиралась в постороннем программном коде (например, для воспроизведения чужих результатов для сравнений).

На основании изложенного, на взгляд научного руководителя, работа достойна оценки “отлично”, а Анастасия Игоревна Пределина заслуживает присвоения степени бакалавра.

Рекомендуется расширить работу экспериментами с модификацией модели на основе BERT-Large (для убедительного перебора гиперпараметров с моделью такого размера потребуется более мощное оборудование) и подготовить результаты к публикации в научном издании или представлению на релевантной конференции.

Кандидат физико-математических наук
Доцент факультета математики и компьютерных наук СПбГУ
Авдюшенко Александр Юрьевич



6 июня 2023 года