

Отзыв на выпускную квалификационную работу

«Оптимизация алгоритмов рекомендательной системы»

студента бакалавриата по направлению 01.03.02 «Прикладная математика и информатика»

Папернюка Александра Олеговича

В выпускной квалификационной работе А.О. Папернюка исследуется задача построения рекомендательной системы. Данный тип систем вошел в нашу жизнь незаметно, но присутствует в ней весьма устойчиво. С деятельностью данной продукции так или иначе знакомы абсолютно все пользователи различных ресурсов сети Интернет, даже не всегда этот факт осознающие: стоит хотя бы один раз заказать в сети товары из любой категории, подобрать для прослушивания интересующий трек, скачать книгу или фильм, запросить новости в конкретной категории, – и контекстная реклама, поисковики, базы данных тут же напомнят о спектре Ваших интересов, выбирая их более или менее точно и предлагая новые из спектра, который связывает с Вашими запросами. На рынке работает огромное количество систем, которые более или менее точно определяют спектр интересов пользователей, охватывают большие или меньшие по объему данные базы, медленнее или быстрее обрабатывают информацию. Из всего выше сказанного вытекает актуальность выбранной студентом тематики исследования: только оптимизация по выбранным критериям и постоянное совершенствование уже существующих алгоритмов может помочь стать новым системам конкурентно способными.

Вопрос интересовал Александра ещё во время обучения на третьем курсе, при выборе темы научного исследования. Он изучал работу уже существующих алгоритмов рекомендательных систем и проводил их сравнительный анализ по быстродействию, информационному охвату, эффективности использования той или иной метрики. Это помогло ему во время прохождения практики в текущем учебном году в отделе музыки компании ООО «Мейл.Ру», целью которой было создать и реализовать в виде готового программного пакета новую модель, которая для каждого пользователя рекомендует наиболее подходящие музыкальные треки таким образом, чтобы они были подобраны исключительно в соответствии с предпочтениями клиента. С задачей Александр справился, о чём свидетельствуют благоприятный отзыв куратора его практики.

В работе Папернюка А.О. анализируется специфика таких подходов, как коллаборативная фильтрация, алгоритмы, основанные на контенте и гибридные системы. В ВКР рассмотрено несколько подходов к реализации алгоритмов коллаборативной фильтрации, для каждого из которых применены метрики, по-разному характеризующие рекомендуемые списки аудиозаписей. Алгоритмы были реализованы в виде программного кода на языке Python в открытом приложении Jupyter Notebook. С помощью построенного программного продукта были протестированы системы и проведен сравнительный анализ эффективности построенных систем, отобраны наилучшие. В дальнейшем на их основе планируется строить гибридные системы. Статистический анализ

полученных результатов демонстрирует несмещенность оценок используемых метрик.

Объем ВКР составляет 46 страниц, включая список литературы (21 наименование) и приложения, содержащие программный код модулей, реализующих описанные алгоритмы и методы подсчета метрик. Автоматический анализ текста на наличие заимствований в системе blackboard выявил в тексте 86,7% оригинальных фрагментов, что свидетельствует о высокой степени оригинальности представленной работы. Часть результатов, полученных в ВКР, была представлена дипломантом на конференции «Процессы управления и устойчивость» (CPS'21). Статья, подготовленная по результатам этого выступления, находится в данный момент в печати.

Недостатки ВКР могут быть сформулированы следующим образом:

- Обзор литературы по заданной тематике можно было бы значительно расширить с учетом того, какое внимание уделяется разработке рекомендательных систем в настоящее время;
- Работа алгоритмов, построенных в проведенном исследовании, апробирована на двух предоставленных компанией «МЕЙЛ.РУ» наборах данных (датированных концом сентября 2020 и концом октября 2020), однако отсутствие апробации на других наборах оставляет открытым вопрос об универсальности построенных алгоритмов;
- В части текста, относящейся к Постановке задачи, были частично анонсированы результаты, относящиеся, скорее, к выводам из проведенного исследования: такие формулировки как «были подобраны 3 метрики», «были посчитаны метрики на тестовых данных», «был основан исключительно на предпочтениях каждого клиента» стилистически более уместны для подведения итогов и заключения.

К безусловным плюсам работы относятся глубинный анализ механизма работы уже используемых на практике систем, модернизация существующих алгоритмов за счет модификации их архитектуры и внедрения функций ранжирования, ранее применимых только для анализа текстов, а также возможность дальнейшего развития созданных дипломантом систем.

Приведенные замечания не влияют на общее положительное впечатление от работы. На основе всего вышесказанного считаю, что выпускная квалификационная работа А.О. Папернюка заслуживает оценки «отлично».

Научный руководитель,
стар. преп. кафедры МТИСР
факультета ПМ-ПУ СПбГУ,
к. ф.-м. н.



С.Ш. Кумачева