

Отзыв о квалификационной работе бакалавра Сулягиной Анастасии.

В рамках квалификационной работы Анастасия исследовал существующие методы персонализации новостных лент пользователей в социальных сетях на примере модельных и реальных данных, предоставленных сетью ОК.

Выбранная автором тема, безусловно, интересна как с практической, так и с теоретической точки зрения. Информационная нагрузка, генерируемая социальными сетями на пользователя, в большинстве случаев значительно превосходит тот объем, который человек в состоянии обработать самостоятельно. При этом новостная лента играет ключевую роль в поддержании социальных связей и может значительно влиять на эмоциональное состояние пользователя. Следовательно, и сами социальные сети, и их пользователи заинтересованы в разработке и внедрении алгоритмов персонализации, вытягивающих на верх наиболее значимый для конкретного пользователя контент.

В работе Анастасии приводится анализ существующих работ, посвященных как общим методам ранжирования, так и специфики их применения в социальных сетях. В обзоре рассмотрены разные подходы к ранжированию (точечные, парные, списочные), методы машинного обучения (линейные модели, нейронные сети, решающие деревья и леса), а также метрики качества ранжирования (NDCG, MAP и т.д.).

В экспериментальной части проводится сравнение нескольких недавно разработанных подходов (ListMLE, ListNet, LambdaMART), классических подходов к ранжированию (логистическая регрессия, RankSVM), а также нескольких «наивных» подходов (сортировка по времени и соотношению клики/показы). Кроме того, эффективность работы разных моделей оценивалась для различных сегментов пользователей, в зависимости от количества данных для обучения.

Результат сравнения, однако, показал достаточно слабое преимущество методов машинного обучения над более простыми методами, что, по всей видимости, вызвано излишней увлеченностью автора методами и слабо проработанными признаками, а так же практически полностью отсутствующим анализом предоставленных данных классическими методами статистики. Это и является основным недостатком работы.

Кроме того среди недостатков следует отметить, что в процессе сравнительного анализа было сильно ограничено число пользователей, для которых проверялось качество результатов, что уменьшает их статистическую значимость. Анализ значимости признаков в полученных моделях так же проведен должным образом не был.

В процессе работы Анастасия, с одной стороны, показала высокий уровень самостоятельности, способность решать возникающие сложности. С другой стороны, излишняя самонадеянность и недостаток коммуникации с научным руководителем на начальных этапах негативно сказались на итоговом результате.

Рекомендую оценить квалификационную работу Сулягиной Анастасии на «хорошо» и допустить её к защите.

научный руководитель,
кандидат физико-математических наук
Бугайченко Дмитрий Юрьевич

