

РЕЦЕНЗИЯ
на выпускную квалификационную работу обучающегося СПбГУ
Марулина Дмитрия Алексеевича
по теме «Методы Фильтрации объявлений по аренде недвижимости»

Выпускная квалификационная работа Марулина Дмитрия Алексеевича посвящена вопросам фильтрации и делению на группы объявлений по аренде недвижимости.

В начале работы приводится постановка задачи, обзор литературы, обзор существующих решений и описание сути задачи.

В первой главе проводится анализ предметной области, а именно приводятся примеры объявлений, составленных частниками и агентствами.

Вторая глава состоит из теоретического упоминания метода фильтрации. Здесь рассказывается про метод распределения Дирихле, указываются основные шаги метода сэмплирования по Гиббсу, приводятся общие принципы перплексии.

К сожалению, работа читается не очень легко, постоянно возникают вопросы «а что это, а что то?»», но объяснения возникают существенно позже.

К стилистическим замечаниям можно отнести несогласованные местами предложения, например с. 14 (в описании алгоритма LDA) «Основная задачей этого является алгоритма – нахождение скрытых тем, с помощью которых можно раскрыть значение текста, который.», с. 25 «При всех проведенных экспериментах с байесовским классификатором, добиться точности выше 60%.» -- что хотел сказать автор в этом предложении? Ни перед предложением, ни после нет указания того, что это цель, поэтому непонятно наличие инфинитива в предложении.

По существу работы можно сказать следующее: с одной стороны есть впечатление, что проделано достаточно большое исследование, а с другой стороны остается множество неясностей, в частности:

- 1) Непонятно каким образом можно по значению перплексии определить ошиблась модель или сработала верно. Ведь мы не можем контролировать работу системы на каждом шаге. Машинное обучение предполагает, что система сама учится и это происходит в автоматическом режиме. А пока получается, что мы сами знаем, что этот результат не подходит, поэтому модель отбрасывается.
- 2) Не описан алгоритм автоматического определения «удачности» модели.
- 3) В пункте 2.5 приведена схема алгоритма. Однако не ясны принципы работы блоков «Определение порогов», «Оценка порогов» и «Оценка модели».
- 4) Непонятно как выглядит выборка объявлений, как загружается выборка в разработанное программное обеспечение, каким образом определяются пороги. В основном в работе указаны общие фразы, что что-то делается, но как именно не ясно.
- 5) В последней главе внимание уделено разработке приложения, однако ни описания функций, ни ссылки на github не присутствует. Поэтому не понятно, в каком состоянии находится разработка.

Мне, как потенциальному пользователю такой системы тематика крайне интересна, и я, к слову, в настоящее время (июнь), как раз воспользовался бы такой системой для фильтрации объявлений, но из изложения настоящей работы мне не стало понятно, как результатами описываемого исследования реально на практике можно воспользоваться.

С учетом вышесказанного, считаю, что при ответе на все замечания и при возникновении у комиссии ощущения, что всё понятно, работу можно оценить на «хорошо».

Рецензент,
к.ф.-м.н., главный специалист
Главного Управления
по организации работы с персоналом СПбГУ,
Главный администратор CRIS системы PURE СПбГУ



Лепихин Т.А.