

РЕЦЕНЗИЯ
на выпускную квалификационную работу
«Анализ тональности русскоязычных комментариев в социальных
медиа»

автор: Бакытбек Нурсултан

Работа посвящена разработке метода для определения тональности русскоязычных текстов-комментариев к видеозаписям. Предложено решение, базирующееся на использовании методов машинного обучения с учителем. Автором использованы наивный байесовский классификатор и метод опорных векторов.

Вызывает недоумение использование в качестве примера текстового документа с комментариями, содержащими ненормативную лексику. Была ли в этом такая экстренная необходимость? Если бы было указано, что этот текст имеет определенную конкретную эмоциональную окраску, связанную с наличием данного конкретного слова в комментарии, но здесь же приведено просто так, без смысла, без должной необходимости. Использование ненормативной лексики не этично в выпускной работе бакалавра Санкт-Петербургского государственного университета.

Проверить достоверность полученных результатов не представилось возможным, поскольку отсутствует инструкция по использованию разработанного программного обеспечения и попытки запуска разработанного автором программного обеспечения завершились неудачей («File labeledTrainData.tsv does not exist»).

Автором указано, что для обучающей выборки было выбрано 85% комментариев, так как при обучающей выборке в 50% алгоритмы продемонстрировали «очень низкое качество работы». Логично предположить, что если бы выбрано было 95% комментариев, то точность была бы еще выше. Также возникает вопрос: а будет ли применим результат обучения для классификации комментариев от других видео (по той же тематике)? Ответа на данный вопрос в работе не найдено.

Проведенная «ручная экспертиза» не вызывает доверия. Сказано, что в экспертизе участвовало 3 человека и анализ файлов в проекте позволяет сделать вывод, что экспертизу проводили по независимым наборам комментариев. Если бы одни и те же комментарии были отправлены на экспертизу независимо трем различным «экспертам», то сравнивая их оценки, можно было бы тогда говорить о какой-то экспертной оценке тональности, а так это не более чем субъективное мнение знакомых. Вот, например, непонятно по какой причине комментарий «Отец и сын» оценен как положительный, «Такая ВАТА(((((((» тоже экспертом отмечен как позитивный (хотя наличие «грустных смайлов» должно бы наводить на мысль о негативности отзыва). Попытаться осуществить обучение классификаторов по таким экспертным оценкам скорее всего и не должны давать хороших результатов.

В соответствии с указанными недостатками и руководствуясь критериями оценивания выпускной квалификационной работы, считаю, что выпускная работа в целом может быть оценена на «удовлетворительно».

Рецензент к.ф.-м.н., доцент кафедры КТС



Погожев С. В.
29 мая 2017 г.