

## РЕЦЕНЗИЯ

на выпускную квалификационную работу магистра  
«Метод программного построения векторной модели неформализованного текста  
на естественном языке на основе морфемного синтаксического анализатора»  
выполненную студенткой гр. 23544/2  
Санкт-Петербургского политехнического университета Петра Великого  
Коваленко Татьяной Викторовной

Выпускная квалификационная работа Коваленко Т.В. посвящена актуальной проблеме обработки и анализа текстов методами машинного обучения. Целью работы является улучшение существующих подходов к классификации текстов методами машинного обучения путём применения усовершенствованной с помощью морфемного синтаксического анализатора векторной модели текста.

Выпускная квалификационная работа магистра состоит из введения, четырех глав, заключения и списка использованных источников.

Во введении автор рассматривает актуальность проблемы обработки и анализа текстов, в частности задачи классификации текстов. Обосновывает необходимость улучшения существующих методов построения векторных моделей текстов.

В первой главе Коваленко Т.В. описывает существующие подходы к решению задачи классификации, дает обзор методов построения векторных моделей текстов, используемых в решении задачи классификации с применением машинного обучения, проводит анализ существующих алгоритмов морфемного разбора слов. Согласно всем выявленным недостаткам существующих подходов сформулированы задачи по улучшению методов построения векторных моделей текстов путём разработки алгоритма разбиения слов на морфемы, применении его для построения морфемной векторной модели текста, построения классификатора текста – нейронной сети и применении морфемной векторной модели для обучения классификатора.

Во второй главе описываются предлагаемый метод построения векторной модели текста на основе морфемного синтаксического анализатора; подготовка корпуса русскоязычных текстов и базы обучающих текстов; придуманные и разработанные алгоритмы разбиения слов на морфемы; предлагаемая архитектура нейронной сети, используемая в классификаторе текстов.

В третьей главе приведены детали реализации предлагаемых методов и алгоритмов.

В четвертой главе представлены полученные результаты применения морфемной векторной модели текстов для решения задачи классификации текстов. Оценено качество разработанных алгоритмов разбиения слов на морфемы,

проведен сравнительный анализ с результатами классификации с применением словной векторной модели текста, сделаны соответствующие выводы.

Работа выполнена технически грамотно, предлагаемый метод описан четко и ясно.

В тоже время, в работе замечены следующие недостатки:

1. Полученные результаты только частично соответствуют ожидаемым, так как качество классификации с применением морфемной модели текста оказалось не лучше качества словной модели.
2. Оценка качества классификации проводилось на достаточно малом наборе данных, что может значительно влиять на полученные результаты.
3. Сравнительный анализ производился только со словными векторными моделями текстов. Хочется увидеть сравнение с существующими морфемными моделями.

Вопросы по работе:

1. Будут ли описанные результаты исследований применяться на практике?
2. Были ли в достаточной степени исследованы существующие алгоритмы разбиения слов на морфемы?

Указанные недостатки не влияют на общее положительное впечатление от работы.

Выпускная квалификационная работа Коваленко Т.В. по теме «Метод программного построения векторной модели неформализованного текста на естественном языке на основе морфемного синтаксического анализатора» соответствует требованиям, предъявляемым к выпускным квалификационным работам и заслуживает оценки «отлично».

Рецензент,  
ООО «Оупен Текст Технолоджи»,  
Инженер-консультант

В. А. Подлесов



31.05.2018