

ОТЗЫВ
на автореферат диссертации
Кудинова Михаила Сергеевича
«Статистическое моделирование русского языка с помощью нейронных сетей»
на соискание учёной степени кандидата технических наук
по специальности 05.13.17 – «Теоретические основы информатики»

Актуальность темы диссертационной работы не вызывает сомнений, особенно в контексте современных направлений исследований в области информационных технологий и на фоне редкой встречаемости отечественных научных работ по данной теме. Выбранная автором тема будет интересна специалистам в области построения систем интеллектуальной обработки текстовых и речевых данных на русском языке, который относится к группе флективных языков. Данное свойство русского языка заметно затрудняет построение качественных языковых моделей, использующихся, в частности, в системах автоматического распознавания речи.

Основная трудность моделирования флективных языков заключается в отсутствии достаточных объёмов обучающих текстовых корпусов для статистически значимого покрытия всех возможных n-грамм. В связи с этим автором диссертационной работы предложен подход раздельного предсказания лемм и морфологических признаков, реализованный на аппарате нейронных сетей. Данный подход, кроме снижения требования к объёму обучающего корпуса, позволяет генерировать на выходе словоформы, которых могло не быть в обучающем корпусе. Таким образом автоматически расширяется словарь языковой модели. Также для снижения числа гипотез распознавания, автором предложена гибридная статистическая языковая модель на рекуррентной нейронной сети и тематическом разложении левого контекста. Предложенные автором подходы обладают научной новизной и перспективой дальнейшего развития в задаче построения языковых моделей флективных языков.

Автор демонстрирует владение темой и умение использовать современные наработки и технологии в научных исследованиях. Результаты проведенных экспериментов заслуживают доверия, поскольку были проведены на достаточно представительных речевом и текстовом корпусах.

В качестве замечаний можно отметить, что в тексте автореферата не упоминается модификация рекуррентной нейронной сети (LSTM – Long Short Term Memory), способная учитывать долговременные зависимости и которая, благодаря этому своему свойству, широко используется в задаче построения языковых моделей. Также к недостаткам можно отнести то, что автор решает задачу предсказания морфологических признаков слова, опираясь только на левый контекст, тогда как правый контекст также может служить источником информации для более точного предсказания. Возможно, это обусловлено архитектурой увязки с системой распознавания речи.

Несмотря на указанные замечания, на основе автореферата можно сделать следующий вывод. Автором проделаны необходимые и важные исследования, обладающие научной новизной и имеющие практическое значение. Считаю, что работа соответствует требованиям ВАК, а ее автор, Кудинов Михаил Сергеевич заслуживает присуждения ученой степени кандидата технических наук по специальности 05.13.17 – Теоретические основы информатики.

Кушнир Дмитрий Алексеевич,
к.т.н., руководитель группы речевых технологий
ООО «ДСС-Лаб»

Подпись _____
Заведующий _____



Начальник отдела кадров
- Н.А. Буренина