

ОТЗЫВ

на автореферат диссертации Ватолина Алексея Сергеевича на тему «Обучение и оценивание мультязычных нейросетевых моделей семантического векторного представления научных текстов», представленной на соискание ученой степени кандидата технических наук по специальности 2.3.8 «Информатика и информационные процессы».

Диссертационная работа А.С. Ватолина посвящена актуальной теме. В условиях экспоненциального роста объемов научных публикаций разработка эффективных инструментов для семантического поиска, анализа и верификации информации становится важной задачей для развития науки. Работа предлагает комплексное решение для создания и оценки специализированных мультязычных моделей для русскоязычного и англоязычного научного домена, что вносит значительный вклад в данную область исследования.

Научный вклад автора состоит в том, что им решен ряд научных и практических задач. К наиболее существенным результатам можно отнести:

1. Разработку семейства компактных и эффективных двуязычных моделей (SciRus-tiny и SciRus-small) для векторного представления научных текстов. Предложенные данные для обучения и двухэтапная методология, сочетающая маскированное языковое моделирование и контрастивное дообучение на парах «заголовок-аннотация», доказала свою высокую эффективность, позволив достичь качества, сопоставимого с более крупными англоязычными аналогами.
2. Создание и апробацию первого мультизадачного русско-английского эталонного набора данных RuSciBench для оценки качества моделей на научных текстах. Данный инструментарий, включающий задачи классификации, регрессии и поиска, заполняет существующий пробел в средствах оценки для русскоязычной научной среды и, благодаря интеграции в международную систему оценки МТЕВ, способствует повышению воспроизводимости и стандартизации исследований в данной области.
3. Предложение оригинальной полуавтоматической методики формирования наборов данных для верификации научных фактов на русском языке. На основе этой методики создан первый в своем роде русскоязычный набор данных для оценки RuSciFact, который позволяет оценивать способность моделей к логическому выводу.

Практическая значимость работы подтверждается успешным внедрением модели SciRus-tiny в информационно-поисковую систему «нейропоиск» на крупнейшем российском научном портале eLibrary.ru, что упрощает анализ научной информации для широкого круга исследователей. Разработка и публикация открытых наборов данных и моделей также имеет большую практическую ценность, поскольку стимулирует дальнейшие исследования и разработки в научном сообществе.

Автореферат последователен, структурирован, отличается научным стилем и логичностью изложения. Содержание автореферата и публикаций в полной мере отражают ключевые выводы диссертации.

В качестве замечания, носящего рекомендательный характер, можно отметить, что было бы интересно увидеть анализ применимости разработанных моделей для других, смежных с научной, областей (например, техническая документация, патенты). Однако данное замечание не снижает общей высокой оценки проделанной работы.

В заключение, диссертационная работа Ватолина Алексея Сергеевича представляет собой законченное научное исследование, содержит решение важной научной задачи, обладает научной новизной и практической ценностью. Работа соответствует всем требованиям, предъявляемым ВАК РФ к кандидатским диссертациям, а ее автор, Ватолин Алексей Сергеевич, безусловно, заслуживает присуждения ученой степени кандидата технических наук по специальности 2.3.8 «Информатика и информационные процессы».

Доктор биологических наук,

Кандидат технических наук,

Ведущий научный сотрудник лаборатории исследования биомеханических систем

Степанян Иван Викторович

Электронная почта: neurocomp.pro@gmail.com


11.11.2025

Федеральное государственное бюджетное учреждение науки

Институт машиноведения им. А.А. Благонравова Российской академии наук

101000, Москва, Малый Харитоньевский переулок, д.4

Телефон: 8 (495) 628-87-30

Электронная почта: info@imash.ru

Подпись Степаняна И. В. заверяю



И. В. Степанян