

Отзыв научного руководителя

о диссертационной работе Потапенко Анны Александровны «Семантические векторные представления текста на основе вероятностного тематического моделирования», представленной на соискание ученой степени кандидата физико-математических наук по специальности 05.13.17 — «Теоретические основы информатики»

Вероятностное тематическое моделирование — одно из направлений обработки естественного языка, активно развивающееся с конца 90-х годов. Тематическая модель выявляет тематику коллекции текстовых документов, определяя, к каким темам относится каждый документ, и какие слова образуют каждую тему.

А. А. Потапенко начала исследования в области тематического моделирования под моим научным руководством в 2011 году, будучи студенткой 3-го курса кафедры Математических методов прогнозирования ВМК МГУ. Закончив с отличием специалитет МГУ в 2014 году, она продолжила исследования в аспирантуре НИУ «Высшая школа экономики».

С первой же поставленной задачей — построить обобщённый алгоритм обучения для нескольких известных тематических моделей — А. А. Потапенко справилась отлично, полностью самостоятельно выполнив реализацию и все вычислительные эксперименты. Эти наработки легли в основу третьей главы диссертации. Последующие работы А. А. Потапенко внесли существенный вклад в развитие нового подхода в тематическом моделировании — аддитивной регуляризации (ARTM). Ею впервые была показана возможность улучшения одновременно нескольких критериев качества модели путём комбинирования регуляризаторов и выработаны приёмы последовательного применения регуляризаторов, успешно используемые при решении многих прикладных задач в рамках ARTM. Эти результаты вошли в четвёртую главу диссертации. В пятой главе А. А. Потапенко объединила тематическое моделирование с идеями дистрибутивной семантики, что позволило ей впервые предложить методы построения векторных представлений слов, одновременно обладающие свойствами интерпретируемости и разреженности наряду с возможностью решать задачи семантической близости и учитывать мультимодальные данные.

А. А. Потапенко проделала большой объём как теоретической, так и экспериментальной работы. Обзорные разделы диссертации представляют собой прекрасное систематическое введение в современное состояние исследований по тематическому моделированию и векторным представлениям текста. Выступая на многочисленных научных конференциях и семинарах в России и за рубежом, она

