

ОТЗЫВ

на автореферат диссертации Карповича Сергея Николаевича «Математическое и программное обеспечение вероятностного тематического моделирования потока текстовых документов», представленной к защите на соискание ученой степени кандидата технических наук по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей».

В автореферате диссертации С.Н. Карповича достаточно полно раскрывается актуальность диссертационного исследования, ее цель и исследовательские задачи. Выполненная работа посвящена важному для решения практических задач обработки текстов на естественном языке вопросу повышения эффективности существующих алгоритмов классификации и кластеризации текстовых документов. Автором разработан и обоснован новый метод построения вероятностных тематических моделей на основе размеченных данных, алгоритм многозначной классификации текстовых документов, метод определения тем новых в вероятностной тематической модели слов. Предложенные методы повышают эффективность существующих методов. Предлагаемый алгоритм расширяет возможности применения вероятностного тематического моделирования для решения практических задач многозначной классификации текстовых документов.

Новизна полученных результатов состоит в разработке нового метода построения динамической вероятностной тематической модели, в которой использован разработанный алгоритм многозначной классификации текстовых документов и метод определения тем «нового слова» с помощью произведения Адамара. Разработанные методы и алгоритм важны для решения практических задач анализа коллекций и потоков текстовых документов.

Судя по автореферату, диссертационная работа хорошо структурирована и представляет собой законченное научное исследование. Результаты, выносимые на защиту, прошли апробацию на конференциях различного уровня и опубликованы в 3 журналах, входящих в список ВАК.

К недостаткам работы стоит отнести следующее:

1. Недостаточно полно отражены существующие направления исследований в области вероятностного тематического моделирования.

2. Очень кратко описаны сценарии применения вероятностного тематического моделирования в практических задачах информационного поиска и рекомендательных системах.
3. На стр. 15 ошибочно указано, что D_2 – документы корпуса для обучения. На мой взгляд, должно быть сказано, что D_2 – это документы корпуса для тестирования алгоритма.

Третий отмеченный недостаток, скорее всего, является опiskой, а первые два недостатка, по-видимому, связаны с ограниченностью объёма автореферата и принципиально не влияют на общую оценку диссертационной работы Карповича С.Н., которая имеет существенную практическую и теоретическую значимость, отвечает требованиям п. 9 Положения о присуждении ученых степеней, утвержденного постановлением Правительства Российской Федерации от 24 сентября 2013 № 842, предъявляемым к кандидатским диссертациям, а ее автор заслуживает присуждения ученой степени кандидата технических наук по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей».

Зав. лабораторией искусственного интеллекта
Федерального государственного бюджетного
учреждения науки Институт систем информатики
им. А.П. Ершова Сибирского отделения
Российской академии наук
к. т. н.

О.А. Загорюлько
10 октября 2017 г.

Адрес: 630090, Новосибирск, проспект Лаврентьева, 6

Телефон: (383) 3328359, +7 9137852655

E-mail: zagor@iis.nsk.su