

**ОТЗЫВ НА АВТОРЕФЕРАТ ДИССЕРТАЦИИ**  
**Кузьмина Арсентия Александровича**  
на тему «Иерархическая классификация коллекций документов»,  
представленной на соискание ученой степени  
кандидата физико-математических наук  
по специальности 05.13.17 – «Теоретические основы информатики»

Диссертация Кузьмина А.А. посвящена разработке и исследованию методов иерархической классификации текстовых документов. В работе рассматриваются коллекции текстов, в которых для некоторого подмножества документов задана экспертная кластеризация в виде дерева тем. Для классификации неразмеченной части коллекции предлагается иерархическая взвешенная функция сходства, на базе которой строится вероятностная модель классов документов коллекции. С помощью вариационного байесовского вывода строится оценка совместного апостериорного распределения классов документов и параметров предложенной модели, позволяющая ранжировать экспертные кластеры в порядке убывания их релевантности заданному документу. Параметры и гиперпараметры предложенной модели оптимизируются с помощью EM алгоритма. Предложенный подход позволяет настраивать параметры модели и классифицировать неразмеченные документы при небольшом размере обучающей выборки. При сравнении с известными аналогами предложенный подход показал сравнимые и более высокие результаты на исследованных наборах данных.

По автореферату имеются следующие замечания:

1. Следовало бы указать время работы и объем использованной алгоритмами памяти в вычислительном эксперименте, а также сравнить их асимптотику с полученными теоретическими оценками.
2. Для сравнения с другими алгоритмами следует также использовать общедоступные наборы данных, используемые другими исследователями, такие как веб-каталоги (например, данные Open Directory Project).
3. Стоило бы привести сравнение работы алгоритма (стр. 13) при различных значениях параметров априорных распределений параметров и гиперпараметров (формулы 14, 15).

Несмотря на указанные замечания, считаю, что диссертационная работа «Иерархическая классификация коллекций документов» соответствует требованиям ВАК, предъявляемым к кандидатским диссертациям, а ее автор – Кузьмин Арсентий Александрович – заслуживает присвоения ученой степени кандидата физико-математических наук по специальности 05.13.17 – теоретические основы информатики.

Доцент кафедры Алгоритмы и технологии  
программирования МФТИ (ГУ)  
к.ф.-м.н.

Хританков Антон Сергеевич

Адрес: Россия, 141701, Московская область,  
г. Долгопрудный, Институтский пер., 9.  
Тел: +7 (495) 408-45-54  
E-mail: anton.khritanov@gmail.com, atp@phystech.edu



ЗАВЕРЯЮ  
УЧЕНЫЙ СЕКТОР



Золотухина

Хританков