УДК 004.852

# Использование алгоритма аддитивной регуляризации тематических моделей для анализа текстов англоязычных песен разных декад

А.В. Ларионова (Университет ИТМО, Санкт-Петербург)
Л.Д. Орлова (Университет ИТМО, Санкт-Петербург)
С.Ю. Чанова (Университет ИТМО, Санкт-Петербург)
Научный руководитель – к.т.н. Н.Ф. Гусарова
(Университет ИТМО, Санкт-Петербург)

**Введение.** Тематическое моделирование – это технология статистического анализа текстов для автоматического выявления тематики в больших коллекциях документов. Одной из самых известных тематических моделей является модель вероятностного латентно-семантического анализа (PLSA), но основная проблема заключается в том, что задача тематического моделирования имеет множество решений, а PLSA выбирает только одно из них. Тем самым, варианты для выбора лучшего решения просто не предоставляются.

**Основная часть.** ARTM решает проблему неединственности и неустойчивости, накладывая дополнительные ограничения на модель. Регуляризация служит для задания желаемых свойств решения с помощью критериев-регуляризаторов. Аддитивная регуляризация позволяет задавать несколько критериев одновременно. Предположительно, использование ARTM может улучшить качество кластеризации текстов, повысить точность поиска, различность тем. Для проверки этого предположения было реализовано тематическое моделирование англоязычных песен разных декад с целью проведения сравнительного анализа содержания песен и изменения спектра тем с течением времени.

**Выводы.** Проведенные эксперименты показывают, что тематическая модель с аддитивной регуляризацией показывает лучшие значения критериев качества модели и обеспечивает лучшую интерпретируемость тем, чем модель без регуляризаторов.

Авторы: Ларионова А.В., Орлова Л.Д., Чанова С.Ю.
Научный руководитель: Гусарова Н.Ф.