

Автоматическая настройка параметров АРТМ под широкий класс задач

Е. В. Иванова¹, С. Н. Матвеева², Т. А. Голубева³, А. В. Трусов⁴,
В. В. Черноног⁵, М. В. Царицын⁶,

Аннотация: В работе рассматривается задача тематического моделирования. Тематическое моделирование нашло своё применение в таких областях как машинное обучение и обработка естественного языка. В работе используется широко известная библиотека BigARTM, использование которой требует настройки большого числа параметров. В работе рассматривается возможность нахождения универсального набора значений параметров для широкого класса задач. Для нахождения этих значений, используется метод $\langle \dots \rangle$. Для оценки качества используется критерий $\langle \dots \rangle$. Полученная, с помощью подобранных коэффициентов, ошибка не больше чем на $\langle X \rangle\%$ больше чем в "локально лучших моделях".

Ключевые слова: тематическое моделирование, ARTM, BigARTM, глобальная оптимизация.

1 Введение

BigARTM (сокр. англ. additive regularization for topic modeling) — библиотека для тематического моделирования на основе аддитивной регуляризации тематических моделей, разрабатываемая А. Фреем (МФТИ), К. Воронцовым (ВЦ РАН) и М. Апишевым (МГУ).

¹Московский физико-технический институт, trusov.av@phystech.edu

²Московский физико-технический институт, matveeva.sn@phystech.edu

³Московский физико-технический институт, golubeva.ta@phystech.edu,

⁴Московский физико-технический институт, trusov.av@phystech.edu,

⁵Московский физико-технический институт, chernonog.vv@phystech.edu,

⁶Московский физико-технический институт, tsaritsyn.mv@phystech.edu,