

Автоматическая настройка параметров BigARTM под широкий класс задач.*

Гришанов А. В.¹, Булатов В. Г.¹, Воронцов К. В.¹

grishanov.av@phystech.ru, bt.uytya@gmail.com, vokov@forecsys.ru

¹ Московский физико-технический институт

Открытая библиотека BigARTM позволяет строить тематические модели, используя широкий класс возможных регуляризаторов. При этом задача настройки коэффициентов оказывается весьма сложной. В данной статье мы оптимизируем перебор параметров, используя механизм относительных коэффициентов регуляризации и автоматический выбор N-грамм. Проверяется гипотеза о том, что существует универсальный набор относительных коэффициентов регуляризации, дающий «достаточно хорошие» результаты на широком классе задач. Для эксперимента использовались наборы данных Victorian Era Authorship Attribution Data Set, 20 Newsgroups, МКБ-10. Модель с подобранными коэффициентами имеет качество не более чем на $\langle X \rangle\%$ хуже «локально лучших моделей»

Ключевые слова: тематическое моделирование, аддитивная регуляризация тематических моделей, PLSA, BigARTM.

1 Введение

TODO

2 Название раздела

TODO

*Задачу поставил: Булатов В. Г. Консультант: Воронцов К. В.