

Автоматическая настройка параметров BigARTM под широкий класс задач

Гришанов А. В.

Московский физико-технический институт
Факультет управления и прикладной математики
Кафедра «Интеллектуальные системы»

Задачу поставил д.ф.-м.н., К. В. Воронцов
Консультант Виктор Булатов

Москва,
2019 г.

Автоматизация настройки параметров BigARTM

Проблема

Настройка параметров BigARTM требует работы эксперта. Требуется автоматизировать этот процесс.

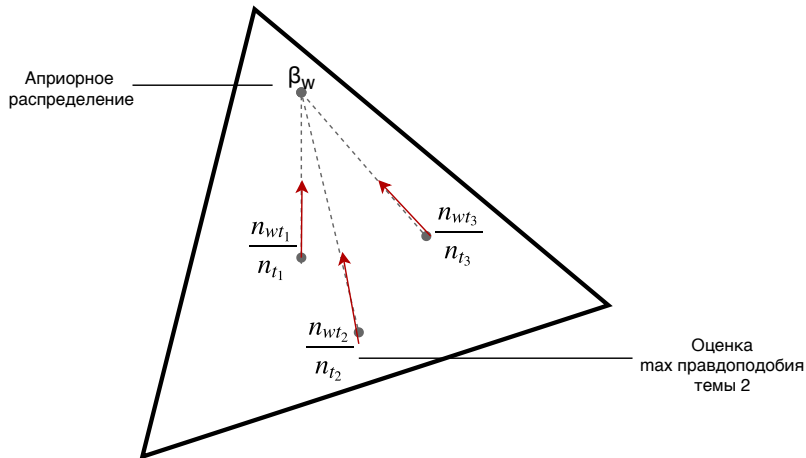
Цель работы

Проверить гипотезу о существовании конфигураций, хорошо работающих на широком классе задач.

Метод решения

Предлагается использовать относительные коэффициенты регуляризации.

Относительные коэффициенты регуляризации



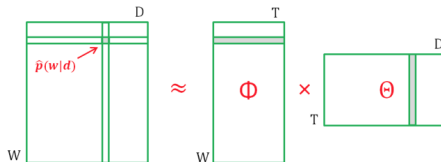
- Описание подхода ARTM
Konstantin Vorontsov, Anna Potapenko. Additive Regularization of Topic Models.
- Относительные коэффициенты регуляризации
Дойков Н.В. Адаптивная регуляризация вероятностных тематических моделей.
- Модель PLSA
David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. Journal of machine Learning research, 2003.

Задача тематического моделирования

$d \in D$ — документы, $w \in W$ — слова, $t \in T$ — темы

$$p(w|d) = \sum_{t \in T} p(w|t)p(t|d) = \sum_{t \in T} \varphi_{wt}\theta_{td}$$

Ставится задача $F \approx \Phi\Theta$



Пришли к следующей модели:

PLSA

$$L(\Phi, \Theta) = \sum_{d \in D} \sum_{w \in d} n_{dw} \sum_{t \in T} \phi_{wt} \theta_{td} \rightarrow \max_{\Phi, \Theta} \quad (1)$$

Разложение матрицы F в произведение матриц Φ и Θ не единственно. В частности, для любой невырожденной матрицы S размера $T \times T$ верно, что $F = (\Phi S)(S^{-1}\Theta)$.

При наложении на модель дополнительных требований (регуляризаторов $R_i(\Phi, \Theta)$) получим:

Аддитивная регуляризация тематических моделей

$$L(\Phi, \Theta) + \sum_{i=1}^n \tau_i R_i(\Phi, \Theta) \rightarrow \max_{\Phi, \Theta} \quad (2)$$

Переход от абсолютных τ к относительным

Формула М-шага, сглаживающего или разреживающего φ_{wt} :

$$\varphi_{wt} = \text{norm}_{w \in W}(n_{wt} + \tau), \quad \text{norm}(x_t) = \frac{\max\{x_t, 0\}}{\sum_{s \in T} \max\{x_s, 0\}} \quad (3)$$

Проведём репараметризацию. Пусть $\beta_w = \frac{1}{|W|}$ — равномерное распределение.

$$\varphi_{wt} = \text{norm}_{w \in W}(n_{wt} + \tau) = \frac{n_{wt} + \tau}{\sum_{w \in W} n_{wt} + \tau} = \frac{n_{wt} + \tau}{n_t + \tau|W|} \quad (4)$$

Представим это в виде выпуклой комбинации $\frac{n_{wt}}{n_t}$ и $\frac{1}{|W|}$.

$$\frac{n_{wt} + \tau}{n_t + \tau|W|} = (1 - \lambda) \frac{n_{wt}}{n_t} + \lambda \frac{1}{|W|} \Rightarrow \tau = \frac{n_t \lambda}{(1 - \lambda)|W|} \quad (5)$$

Значит, сглаживание Фи можно трактовать, как нахождение компромисса между $\varphi_{wt} = \frac{n_{wt}}{n_t}$ и $\varphi_{wt} = \frac{1}{|W|}$.

- Рассмотрим набор датасетов $\{\mathcal{D}_{ex}, \mathcal{D}_{in}\}$, где \mathcal{D}_{ex} имеют внешний критерий качества, а \mathcal{D}_{in} — только внутренние.
- Необходимо проверить гипотезу о том, что существуют общие коэффициенты регуляризации $\tau_{general}$, которые **не хуже** чем PLSA и **лучше** PLSA по нескольким критериям.
- Для каждого из первых найдём лучшие параметры, затем будем искать общие.
- В конце проверим выполнение гипотезы на всех данных.

Результаты эксперимента

Фиксируем следующие относительных коэффициенты:
декоррелирование — 0.04; разреживание тем в документах — 0.1; разреживание слов в темах — 0.2

20news groups

	perplexity	Φ sparsity	Θ sparsity
PLSA	2580	0.882	0.001
BigARTM (with relative regularizers)	2560	0.900	0.860

NIPS

	perplexity	Φ sparsity	Θ sparsity
PLSA	1000	0.800	0.890
BigARTM (with relative regularizers)	995	0.850	0.920