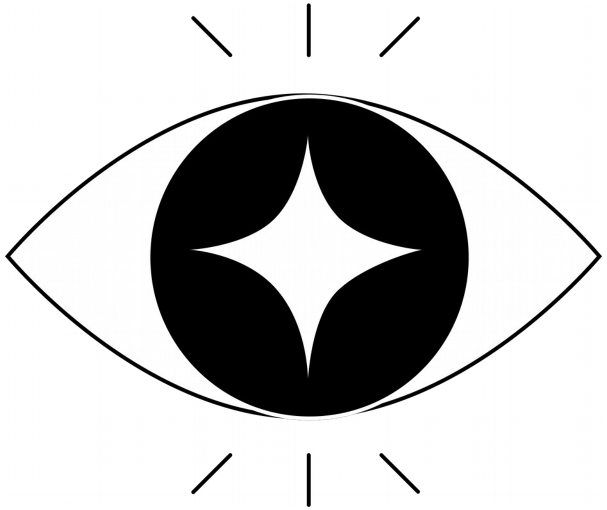
Тематическое моделирование LDA

На этом уроке

- Модель LDA;
- Практика.



Модель LDA

Более старый и популярный подход к регуляризации тематических моделей — модель LDA (латентное размещение Дирихле);

Для регуляризации использует априорное распределение Дирихле и параметры Φ и Θ;

Предположи $\alpha \in \mathbb{R}^{|T|}$ и $\beta \in \mathbb{R}^{|W|}$ толбцы ϕ_t и Θ_d генерируются их распределений Дирихле с параметрами

Априорное распределение на параметры

Предположим, что вектор-столбцы ϕ_t генерируются из распределения Дирихле с параметрами $\beta \in \mathbb{R}^{|T|}$:

$$\mathrm{Dir}(\phi_t|\beta) = \frac{\Gamma(\beta_0)}{\prod_{w\in W}\Gamma(\beta_w)} \prod_{w\in W} \phi_{wt}^{\beta_w - 1}$$

$$\phi_{wt} > 0, \qquad \beta_0 = \sum_{w\in W} \beta_w, \qquad \beta_w > 0$$

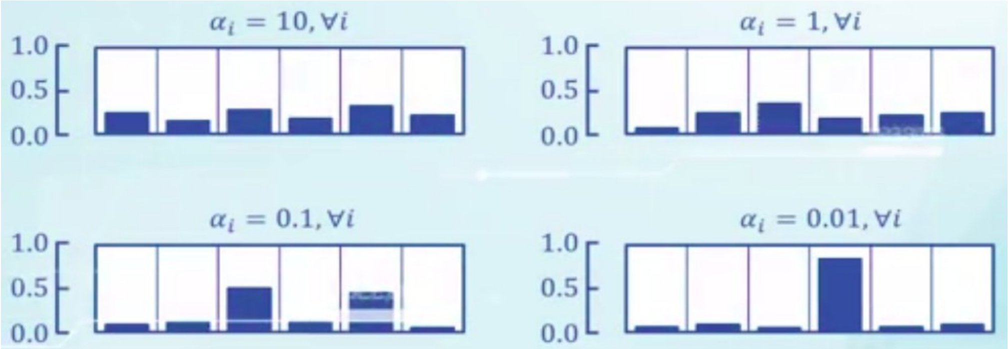
Предположим, что вектор-столбцы Θ_d генерируются из распределения Дирихле с параметрами $\alpha \in \mathbb{R}^{|W|}$:

$$\mathrm{Dir}(\theta_d|\alpha) = \frac{\Gamma(\alpha_0)}{\prod_{t\in T}\Gamma(\alpha_t)} \prod_{t\in T} \theta_{td}^{\alpha_t - 1}$$

$$\theta_{td} > 0, \qquad \alpha_0 = \sum_{t\in T} \alpha_t, \qquad \alpha_t > 0$$

Почему именно распределение Дирихле

- Распределение Дирихле позволяет порождать разрежённые векторы;
- У него есть параметры, которые позволяют управлять степенью разрежённости;



- Оно является сопряжённым к мультиномиальному, что упрощает байесовский вывод в случае задачи тематического моделирования;
- Серьёзных лингвистических обоснований для использования распределения Дирихле нет.

Обучение модели LDA

Обучать модель LDA можно разными способами:

- Вариационный вывод (Variational Bayes);
- Сэмплирование Гиббса (Gibbs Sampling);
- Максимум апостериорной вероятности (MAP).

VB и GS основываются на байесовском выводе. Используется сложный математический аппарат.

Параметры Φ и Θ не ищутся напрямую. Вместо этого строятся распределения на всевозможные значения параметров.

Это более сложная задача, которая в случае тематических моделей кажется избыточной. Итоговым результатом всё равно являются не сами распределения, а их точечные оценки.

Обучение модели LDA

Обучать модель LDA можно разными способами:

- Вариационный вывод (Variational Bayes);
- Сэмплирование Гиббса (Gibbs Sampling);
- Максимум апостериорной вероятности (MAP).

MAP, как и метод максимального правдоподобия, строит точечную оценку параметров Φ и Θ.
Покажем, что этот подход приводит к модели, похожей на ARTM с регуляризатором сглаживания/разреживания.

MAP для модели LDA

Применим к апостериорному распределению на параметры формулу Байеса:

$$p(\Phi, \Theta | D, \alpha, \beta) = \frac{p(D|\Phi, \Theta)\, p(\Phi, \Theta|\alpha, \beta)}{p(D|\alpha, \beta)} \propto p(D|\Phi, \Theta)\, p(\Phi, \Theta|\alpha, \beta)$$

Прологарифмируем результат и распишем:

$$\ln\bigl(p(D|\Phi, \Theta)\, p(\Phi, \Theta|\alpha, \beta)\bigr) =$$

$$= \ln\left(\prod_{i=1}^{n} p(d_i, w_i|\Phi, \Theta)\, p(\Phi|\beta)\, p(\Theta|\alpha)\right) =$$

$$= \ln\left(\prod_{d \in D}\prod_{w \in d} p(d, w|\Phi, \Theta)^{n_{dw}} \prod_{t \in T} \mathrm{Dir}(\phi_t|\beta) \prod_{d \in D} \mathrm{Dir}(\theta_d|\alpha)\right)$$

MAP для модели LDA

Прологарифмируем результат и распишем:

$$\ln\big(p(D|\Phi,\Theta)\,p(\Phi,\Theta|\alpha,\beta)\big) =$$

$$= \ln\left(\prod_{d\in D}\prod_{w\in d} p(d,w|\Phi,\Theta)^{n_{dw}} \prod_{t\in T}\mathrm{Dir}(\phi_t|\beta) \prod_{d\in D}\mathrm{Dir}(\theta_d|\alpha)\right) =$$

$$= \left.\sum_{d\in D}\sum_{w\in d} n_{dw}\ln\sum_{t\in T}\phi_{wt}\theta_{td} + \quad\right\}\ \text{Правдоподобие}$$

$$\underbrace{+\sum_{t\in T}\sum_{w\in W}\ln{\phi_{wt}}^{\beta_w-1} + \sum_{d\in D}\sum_{t\in T}\ln{\theta_{td}}^{\alpha_t-1}}_{\text{Регуляризатор}} \to \max_{\Phi,\Theta}$$

MAP для модели LDA

Регуляризатор априорного распределения Дирихле:

$$R(\Phi, \Theta) = \sum_{t\in T} \sum_{w\in W} (\beta_w - 1) \ln \phi_{wt} + \sum_{d\in D} \sum_{t\in T} (\alpha_t - 1) \ln \theta_{td}$$

Применим общую теорему ARTM и получим формулы M-шага EM-алгоритма:

$$\phi_{wt} = \underset{w\in W}{\operatorname{norm}}(n_{wt} + \beta_w - 1)$$

$$\theta_{td} = \underset{t\in T}{\operatorname{norm}}(n_{td} + \alpha_t - 1)$$

В зависимости от значений гиперпараметров происходит либо сглаживание (α_t, $\beta_w > 1$), либо разреживание ($0 < \beta_w$, $\alpha_t < 1$).

Модификации модели LDA

В научной литературе десятками описываются разные специализированные варианты LDA:

- Модели для классификации;
- Мультиязычные модели;
- Динамические модели;
- Иерархические модели.
- …

Каждая обычно представляет собой базовую модель и некоторый регуляризатор для решения целевой задачи. Байесовский подход приводит к сложности комбинирования нескольких требований в одной модели. ARTM обходит это ограничение и позволяет собирать в одной модели любой число регуляризаторов без дополнительного математического вывода.

Основные выводы

- Классическая математическая LDA вводит априорное распределение Дирихле на параметры модели;
- LDA можно обучать с помощью методов байесовского вывода и MAP;
- Байесовский вывод для тематических моделей оказывается достаточно трудоёмким и негибким;
- MAP строит точечные оценки параметров, как и метод максимума правдоподобия;
- MAP приводит к EM-алгоритму, похожему на ARTM с подходящим регуляризатором.