Вероятностные тематические модели
Лекция 8. Мультимодальные ARTM

К. В. Воронцов
vokov@forecsys.ru

Этот курс доступен на странице вики-ресурса
http://www.MachineLearning.ru/wiki
«Вероятностные тематические модели (курс лекций, К.В.Воронцов)»

ВМК МГУ • весна 2017

Содержание

Напоминание. Мультимодальная ARTM

W^m — словарь токенов m-й модальности, $m \in M$
$W = W^1 \sqcup \cdots \sqcup W^M$ — объединённый словарь всех модальностей

Максимизация суммы log правдоподобий с регуляризацией:

$$\sum_{m \in M} \tau_m \sum_{d \in D} \sum_{w \in W^m} n_{dw} \ln \sum_t \phi_{wt} \theta_{td} + R(\Phi, \Theta) \to \max_{\Phi, \Theta}$$

EM-алгоритм: метод простой итерации для системы уравнений

$$\text{E-шаг:} \quad p_{tdw} = \underset{t \in T}{\text{norm}}\left(\phi_{wt} \theta_{td}\right)$$

$$\text{M-шаг:} \quad \begin{cases} \phi_{wt} = \underset{w \in W^m}{\text{norm}}\left(\sum_{d \in D} \tau_{m(w)} n_{dw} p_{tdw} + \phi_{wt} \frac{\partial R}{\partial \phi_{wt}} \right) \\ \theta_{td} = \underset{t \in T}{\text{norm}}\left(\sum_{w \in d} \tau_{m(w)} n_{dw} p_{tdw} + \theta_{td} \frac{\partial R}{\partial \theta_{td}} \right) \end{cases}$$

Параллельные и сравнимые корпуса текстов

Parallel — точный перевод (с выравниванием предложений), пример: EuroParl, протоколы европарламента, 21 язык.

Comparable — не перевод, а пересказ на другом языке, пример: Википедия.

W^ℓ — словарь языка ℓ из множества языков L.

Модель ML-P (MultiLingual Parallel)

- каждый язык — отдельная модальность
- $\theta_{td} = p(t|d)$ общее для всех связных документов $d = \bigsqcup\limits_{\ell \in L} d^\ell$

Дополнительные данные — двуязычные словари:

- $\Pi_k(w) \subset W^k$ — все переводы слова $w \in W^\ell$ в языке k

---

I. Vulić, W. De Smet, J. Tang, M.-F. Moens. Probabilistic topic modeling in multilingual settings: an overview of its methodology and applications. 2015

Пример тем. Мультиязычная модель Википедии

216 175 русско-английских пар статей. Языки — модальности.
Первые 10 слов и их вероятности $p(w|t)$ в %:

Тема №68				Тема №79			
research	4.56	институт	6.03	goals	4.48	матч	6.02
technology	3.14	университет	3.35	league	3.99	игрок	5.56
engineering	2.63	программа	3.17	club	3.76	сборная	4.51
institute	2.37	учебный	2.75	season	3.49	фк	3.25
science	1.97	технический	2.70	scored	2.72	против	3.20
program	1.60	технология	2.30	cup	2.57	клуб	3.14
education	1.44	научный	1.76	goal	2.48	футболист	2.67
campus	1.43	исследование	1.67	apps	1.74	гол	2.65
management	1.38	наука	1.64	debut	1.69	забивать	2.53
programs	1.36	образование	1.47	match	1.67	команда	2.14

Асессор оценил 396 тем из 400 как хорошо интерпретируемые.

Vorontsov, Frei, Apishev, Romov, Suvorova. BigARTM: Open Source Library for Regularized Multimodal Topic Modeling of Large Collections. AIST-2015.

Пример тем. Мультиязычная модель Википедии

216 175 русско-английских пар статей. Языки — модальности.
Первые 10 слов и их вероятности $p(w|t)$ в %:

Тема №88				Тема №251			
opera	7.36	опера	7.82	windows	8.00	windows	6.05
conductor	1.69	оперный	3.13	microsoft	4.03	microsoft	3.76
orchestra	1.14	дирижер	2.82	server	2.93	версия	1.86
wagner	0.97	певец	1.65	software	1.38	приложение	1.86
soprano	0.78	певица	1.51	user	1.03	сервер	1.63
performance	0.78	театр	1.14	security	0.92	server	1.54
mozart	0.74	партия	1.05	mitchell	0.82	программный	1.08
sang	0.70	сопрано	0.97	oracle	0.82	пользователь	1.04
singing	0.69	вагнер	0.90	enterprise	0.78	обеспечение	1.02
operas	0.68	оркестр	0.82	users	0.78	система	0.96

Асессор оценил 396 тем из 400 как хорошо интерпретируемые.

Vorontsov, Frei, Apishev, Romov, Suvorova. BigARTM: Open Source Library for Regularized Multimodal Topic Modeling of Large Collections. AIST-2015.

Регуляризация по двуязычным словарям. Модель ML-TD

Гипотеза. Если $u \in \Pi_k(w)$, то тематика слов w и u близка:

$$\mathrm{KL}\big(\hat{p}(t|u) \,\big\|\, p(t|w)\big) \to \min,$$

где $\hat{p}(t|u) = \frac{n_{ut}}{n_u}$, $\quad p(t|w) = p(w|t)\frac{p(t)}{p(w)} = \phi_{wt}\frac{n_t}{n_w}$.

Модель ML-TD (MultiLingual Translation Dictionary)

$$R(\Phi) = \tau \sum_{\ell,k \in L} \sum_{w \in W^\ell} \sum_{u \in \Pi_k(w)} \sum_{t \in T} n_{ut} \ln \phi_{wt} \;\to\; \max_{\Phi}.$$

Недостатки. Модель ML-TD не учитывает два обстоятельства:

- тематику омонимов сближать не нужно,
- слово может иметь разные переводы в разных темах.

Дударенко М. А. Регуляризация многоязычных тематических моделей // Вычислительные методы и программирование. 2015. Т. 16. С. 26–36.

Матрица вероятностей переводов. Модель ML-TDP

Гипотеза. Переводы слов зависят от тем: $\pi^{k\ell}_{uwt} = p(u|w,t)$, темы согласуются в разных языках через переводы слов:

$$\mathrm{KL}\big(\hat{p}(u|t) \,\big\|\, p(u|t)\big) \to \min;$$

$\hat{p}(u|t) = \frac{n_{ut}}{n_t}$ — частотная оценка по модальности (языку) k,
$p(u|t)$ — модель темы t в языке k по языку ℓ:

$$p(u|t) = \sum_{w \in \Pi_\ell(u)} p(u|w,t)p(w|t) = \sum_{w \in \Pi_\ell(u)} \pi^{k\ell}_{uwt}\phi_{wt}.$$

Модель ML-TDP (MultiLingual Translation Dictionary Probability)

$$R(\Phi,\Pi) = \tau \sum_{\ell,k \in L} \sum_{u \in W^k} \sum_{t \in T} n_{ut} \ln \sum_{w \in \Pi_\ell(u)} \pi^{k\ell}_{uwt}\phi_{wt} \to \max_{\Phi,\Pi}.$$

Дударенко М. А. Регуляризация многоязычных тематических моделей // Вычислительные методы и программирование. 2015. Т. 16. С. 26–36.

Формулы M-шага для моделей ML-TD и ML-TDP

ML-TD (MultiLingual Translation Dictionary):

$$\phi_{wt} = \underset{w \in W^\ell}{\text{norm}} \left(n_{wt} + \tau \sum_{k \in L \setminus \ell} \sum_{u \in \Pi_k(w)} n_{ut} \right)$$

ML-TDP (MultiLingual Translation Dictionary Probability):

$$\phi_{wt} = \underset{w \in W^\ell}{\text{norm}} \left(n_{wt} + \tau \sum_{k \in L \setminus \ell} \sum_{u \in \Pi_k(w)} \pi^{\ell k}_{wut} n_{ut} \right)$$

$$\pi^{k\ell}_{uwt} = \underset{u \in W^k}{\text{norm}} \left(\pi^{\ell k}_{wut} n_{ut} \right)$$

Смысл регуляризации:

условные вероятности $\phi_{wt} = p(w|t)$ согласуются с их частотными оценками по словам других языков

Тематические переводы слов $\pi^{k\ell}_{uwt} = p(u|w,t)$

Темы, в которых $p(\text{«sum»}|\text{«сумма»},t) > 0.9$

Тема №6		Тема №12		Тема №20	
множество	set	математика	triangle	вектор	vector
пространство	space	треугольник	square	координата	coordinate
группа	point	теорема	number	пространство	field
точка	left	точка	point	преобразование	tensor
элемент	limit	математический	theorem	базис	transform
функция	symmetry	угол	angle	тензор	basis
предел	function	координата	mathematics	сила	space
отображение	open	экономика	real	векторный	force
симметрия	property	число	theory	точка	rotation
открытый	topology	квадрат	geometry	система	thermometer

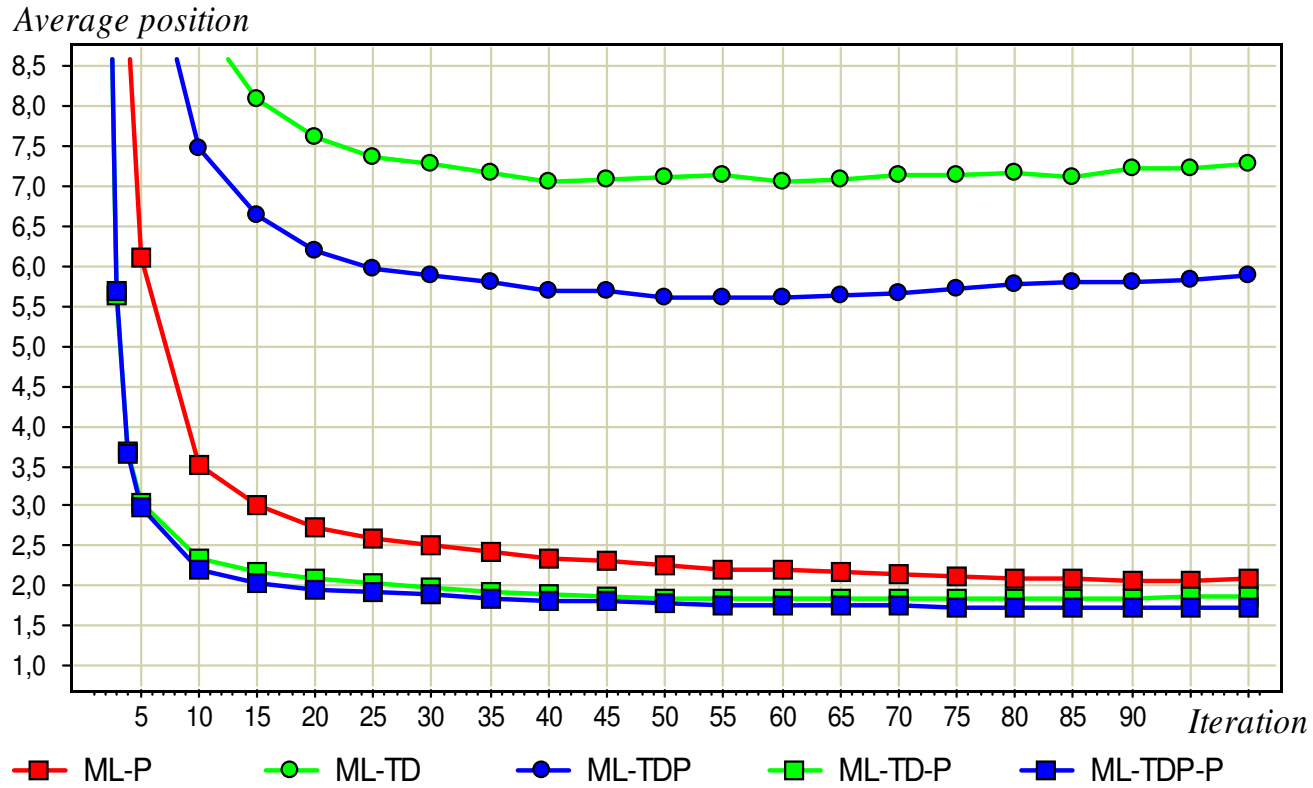
Темы, в которых $p(\text{«total»}|\text{«сумма»},t) > 0.9$

Тема №5		Тема №19		Тема №22	
орбита	space	программный	software	игра	game
аппарат	nasum	версия	version	видеосигнал	character
космический	orbit	работа	news	игрок	video
земля	instrument	компания	company	фильм	player
поверхность	earth	анонимный	work	головоломка	series
солнечный	surface	примечание	note	серия	puzzle
станция	solar	терминатор	release	качество	movie
запуск	system	журнал	support	шахматы	jason
система	landing	рей	terminator	джейсон	world
атмосфера	camera	персонаж	anonymous	буква	chess

Кросс-язычный поиск: ищем документ по его переводу

Wiki: $|D| = 586$, категория «Математика», $|T| = 100$,
$|W^{\text{рус}}| = 19\,305$, $|W^{\text{eng}}| = 23\,413$, переводов 82 642 пар.

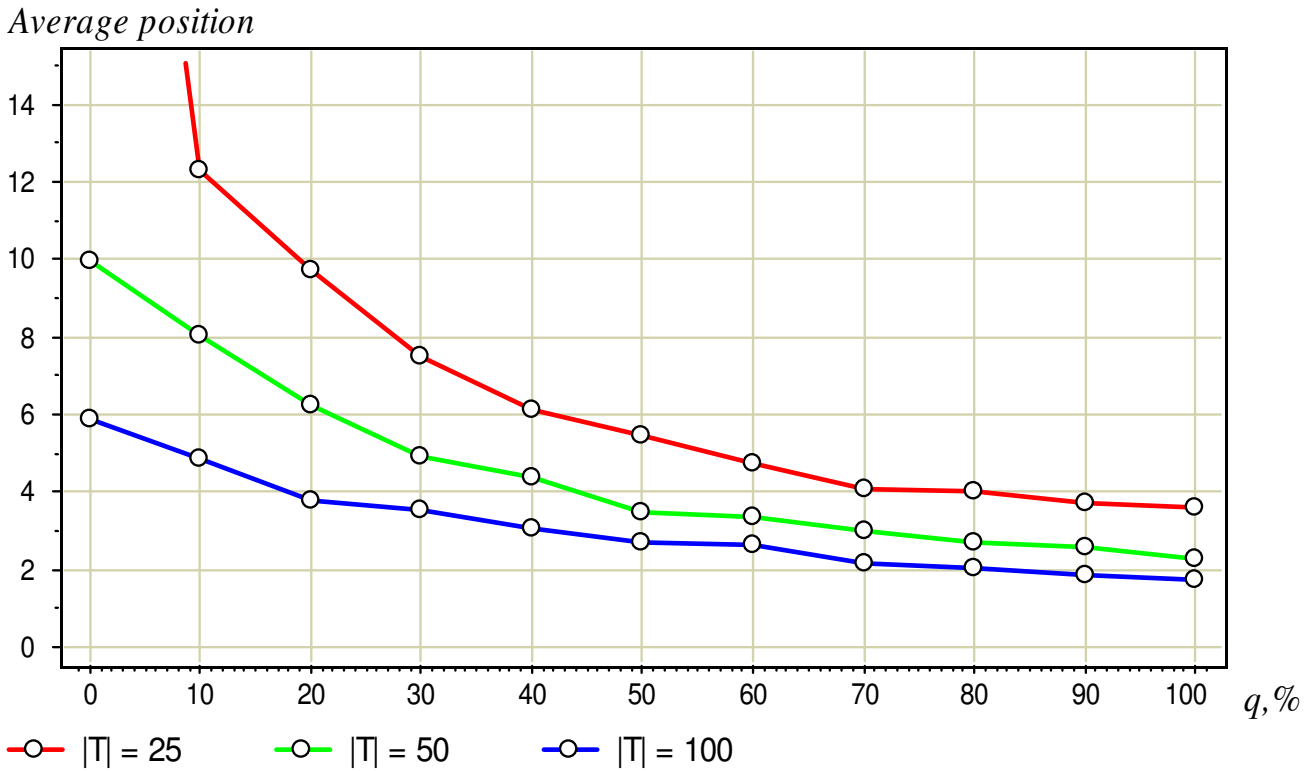
Качество поиска — средняя позиция перевода в выдаче:



Кросс-язычный поиск: ищем документ по его переводу

Wiki: $|D| = 586$, категория «Математика», $|T| = 25, 50, 100$, $|W^{\text{рус}}| = 19\,305$, $|W^{\text{eng}}| = 23\,413$, переводов 82 642 пар.

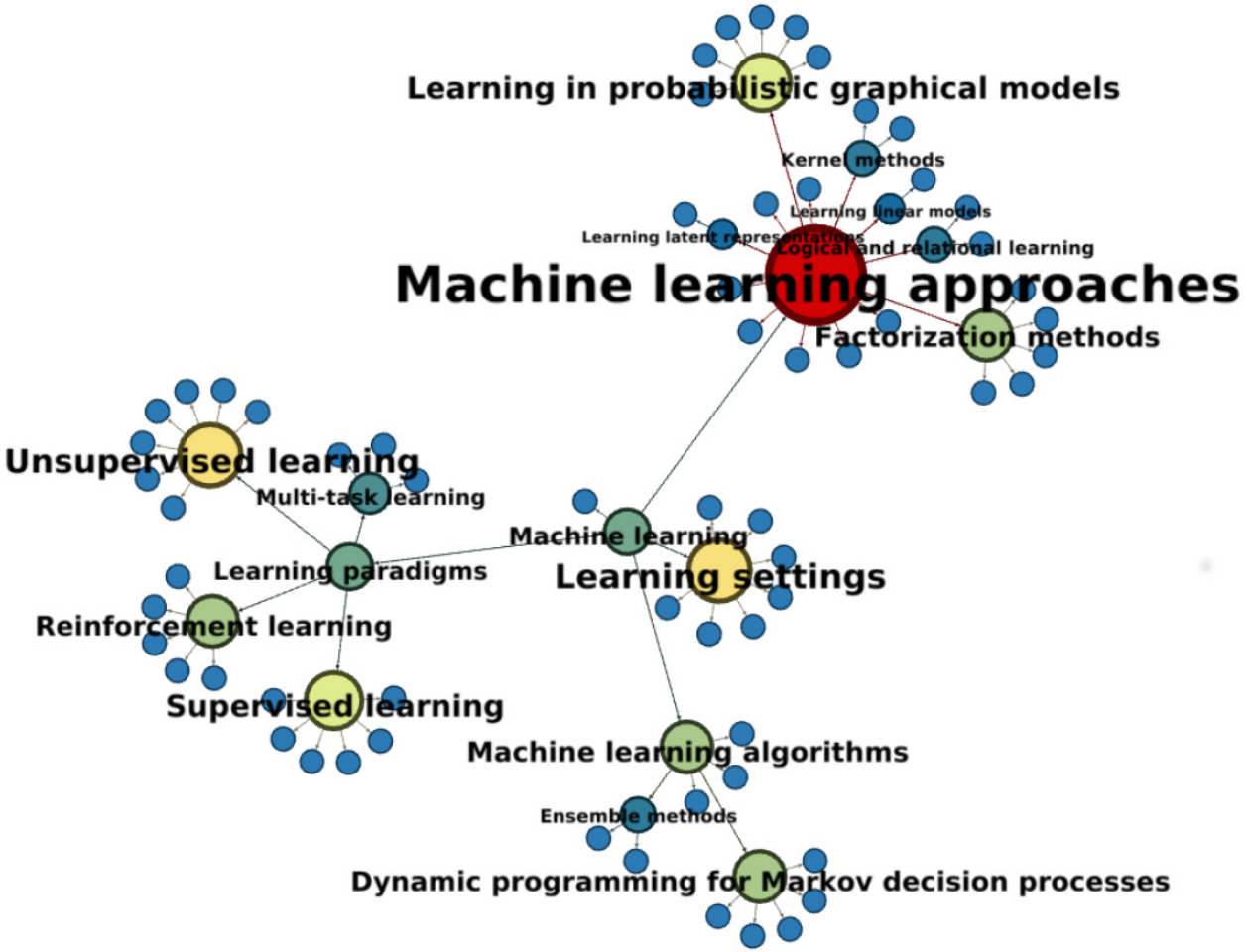
Зависимость средней позиции перевода в выдаче
от числа тем $|T|$ и доли q параллельных текстов в коллекции:



Резюме по мультиязычным моделям

- Главное чудо: для построения мультиязычных тем достаточно иметь сравнимые корпуса.
- Сравнимая коллекция является более сильным источником многоязычной информации, чем словарь переводов (!)
- Модель с вероятностями переводов — самая сильная
- Не обязательно, чтобы все документы имели параллельные
- Главное применение — по запросу на одном языке ищем:
  — тексты на другом языке — *кросс-язычный поиск*,
  — тексты на всех языках — *мульти-язычный поиск*.
- Применение в статистическом машинном переводе:
  выбор варианта перевода согласно тематике документа.

Пример тематической иерархии



Georgeta Bordea. Domain adaptive extraction of topical hierarchies for Expertise Mining. 2013.

Иерархические тематические модели

- структура иерархии: дерево / многодольный граф
- направление: снизу вверх / сверху вниз / одновременно
- наращивание: поверхинное / послойное

Открытые проблемы:

- "Despite recent activity in the field of HPTMs, determining the hierarchical model that best fits a given data set, in terms of the structure and size of the learned hierarchy, still remains a challenging task and an open issue."
- "The evaluation of hierarchical PTMs is also an open issue."

Zavitsanos E., Paliouras G., Vouros G. A. Non-Parametric Estimation of Topic Hierarchies from Texts with Hierarchical Dirichlet Processes. 2011.

Регуляризатор Ф: родительские темы как псевдо-документы

Шаг 1. Строим модель с небольшим числом тем.

Шаг *k*. Пусть модель с множеством тем T уже построена. Строим множество дочерних тем S (subtopics), $|S| > |T|$.

Родительские темы приближаются смесями дочерних тем:

$$\sum_{t\in T} n_t \,\mathrm{KL}_w\Big(p(w|t) \,\Big\|\, \sum_{s\in S} p(w|s)p(s|t)\Big) \;\to\; \min_{\Phi,\tilde{\Psi}},$$

где $p(s|t) = \tilde{\psi}_{st}$, $\tilde{\Psi} = (\tilde{\psi}_{st})_{S\times T}$ — *матрица связей*, .

Родительская $\Phi^p \approx \Phi\tilde{\Psi}$, отсюда регуляризатор матрицы Φ:

$$R(\Phi,\tilde{\Psi}) = \tau \sum_{t\in T}\sum_{w\in W} n_{wt} \ln \sum_{s\in S} \phi_{ws}\tilde{\psi}_{st} \;\to\; \max.$$

Родительские темы t — «документы» с частотами слов n_{wt}.

Регуляризатор Θ: родительские темы как модальность

Шаг 1. Строим модель с небольшим числом тем.

Шаг k. Пусть модель с множеством тем T уже построена. Строим множество дочерних тем S (subtopics), $|S| > |T|$.

Родительские темы приближаются смесями дочерних тем:

$$\sum_{d\in D} n_d \,\mathrm{KL}_t\Big(p(t|d) \,\Big\|\, \sum_{s\in S} p(t|s)p(s|d)\Big) \;\to\; \min_{\Theta,\Psi},$$

где $\psi_{ts} = p(t|s)$, $\Psi = (\psi_{ts})_{T\times S}$ — матрица связей.

Родительская $\Theta^p \approx \Psi\Theta$, отсюда регуляризатор матрицы Θ:

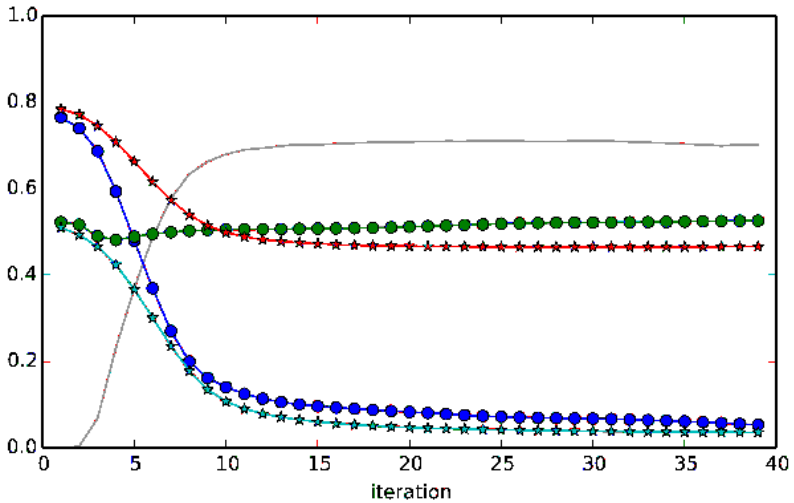
$$R(\Theta,\Psi) = \tau \sum_{d\in D}\sum_{t\in T} n_{td} \ln \sum_{s\in S} \psi_{ts}\theta_{sd} \;\to\; \max.$$

Родительские темы t — модальность с частотами токенов n_{td}.

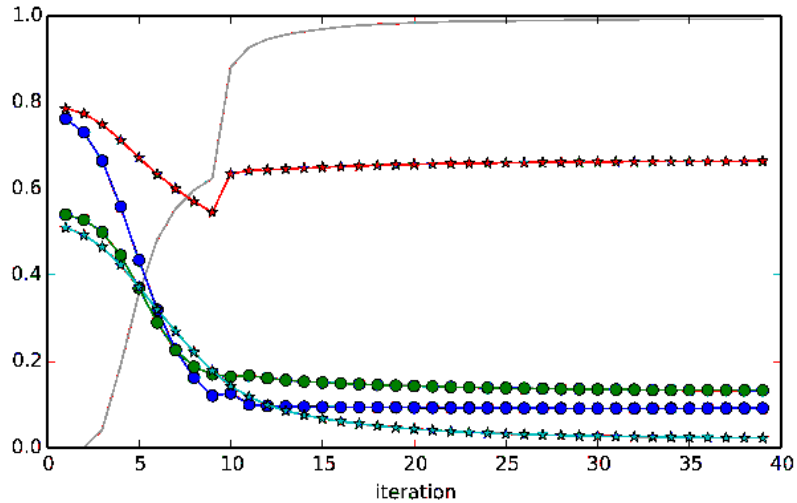
Эксперимент на коллекции ММРО-ИОИ

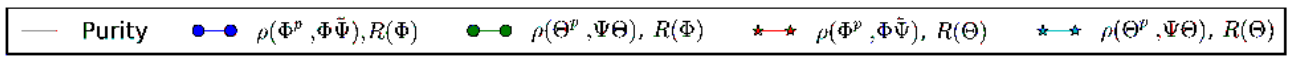
Среднее расстояние Хеллингера $\rho(\Phi^p, \Phi\tilde{\Psi})$ и $\rho(\Theta^p, \Psi\Theta)$ для регуляризаторов Φ и Θ при переходе между уровнями $1 \to 2$:

Разреживание Φ с 1-й итерации



Разреживание Φ с 10-й итерации





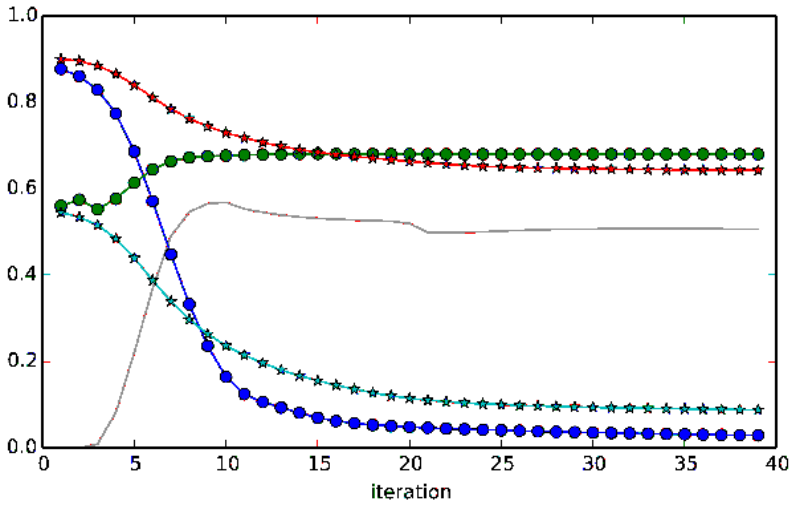
Вывод. Регуляризатор Θ плохо приближает Φ^p.

Chirkova N. A., Vorontsov K. V. Additive regularization for hierarchical multimodal topic modeling // JMLDA, 2016.

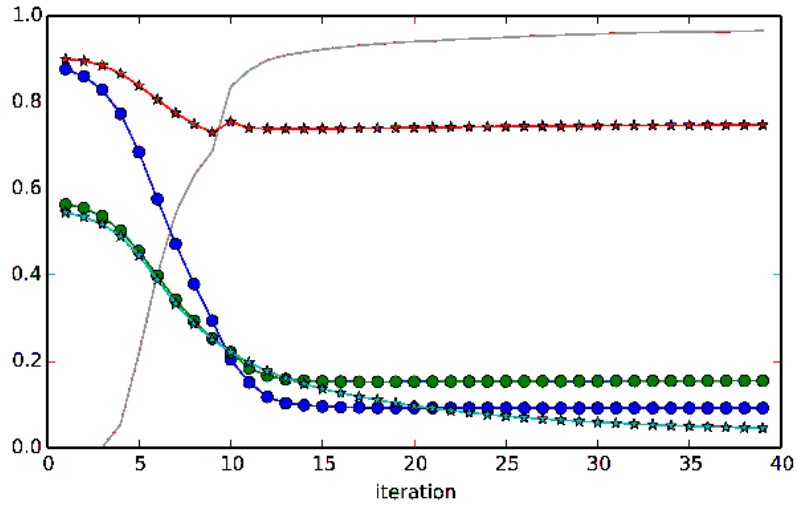
Эксперимент на коллекции ММРО-ИОИ

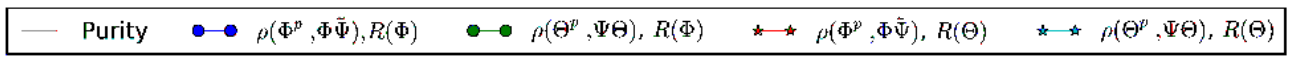
Среднее расстояние Хеллингера $\rho(\Phi^p, \Phi\tilde{\Psi})$ и $\rho(\Theta^p, \Psi\Theta)$ для регуляризаторов Φ и Θ при переходе между уровнями $2 \to 3$:

Разреживание Φ с 1-й итерации



Разреживание Φ с 10-й итерации





Вывод. Регуляризатор Θ плохо приближает Φ^p.

Chirkova N. A., Vorontsov K. V. Additive regularization for hierarchical multimodal topic modeling // JMLDA, 2016.

Выводы

- Регуляризатор Φ приближает $\Phi^p \approx \Phi\tilde{\Psi}$ и $\Theta^p \approx \Psi\Theta$.
- Регуляризатор Θ приближает только $\Theta^p \approx \Psi\Theta$.
- Максимальное разреживание $\psi_{ts} \in \{0, 1\}$ даёт иерархию–дерево.
- Нельзя допускать вырождения $\psi_{ts} = p(t|s) \equiv 0$.

Дальнейшие задачи:

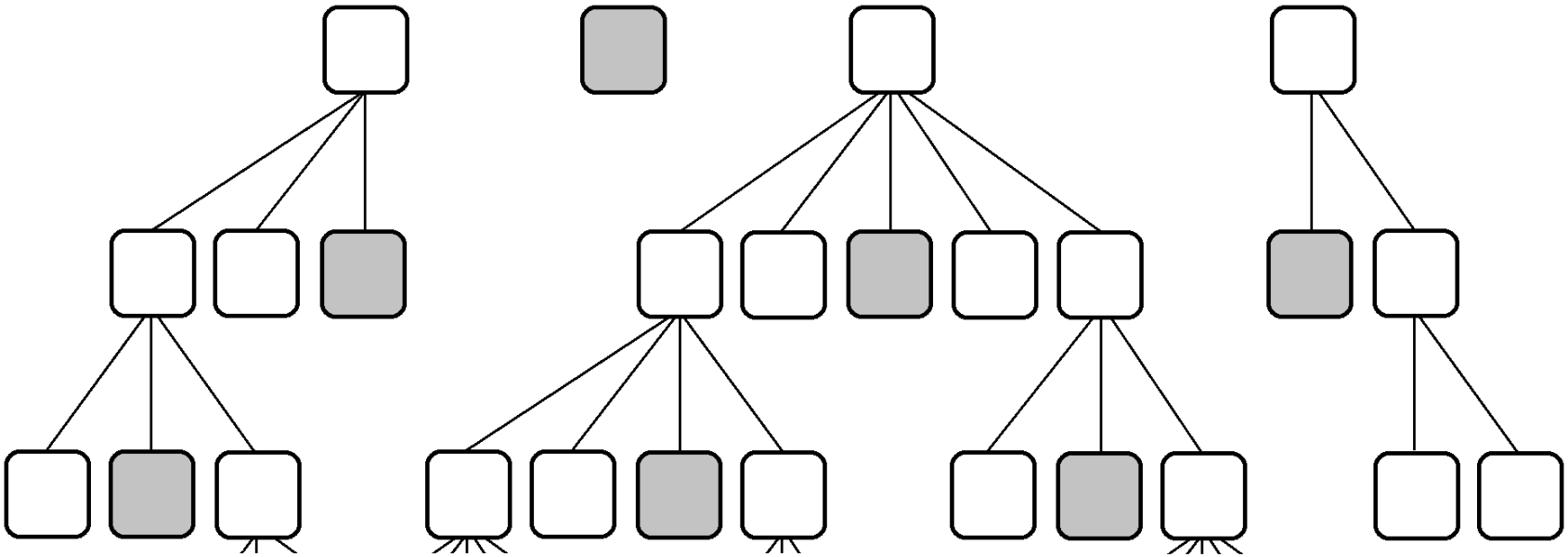
- Согласованная регуляризация: $\tilde{\psi}_{st} p(t) = \psi_{ts} p(s)$

$$\tau_1 \sum_{t,w} n_{wt} \ln \sum_s \phi_{ws}\psi_{ts}\frac{n_s}{n_t} + \tau_2 \sum_{d,t} n_{td} \ln \sum_s \psi_{ts}\theta_{sd} \;\to\; \max_{\Phi,\Psi,\Theta}$$

- Наращивание уровня для заданного подмножества $T' \subseteq T$
- Критерий неоднородности темы для включения её в T'
- Иерархии с темами различной глубины

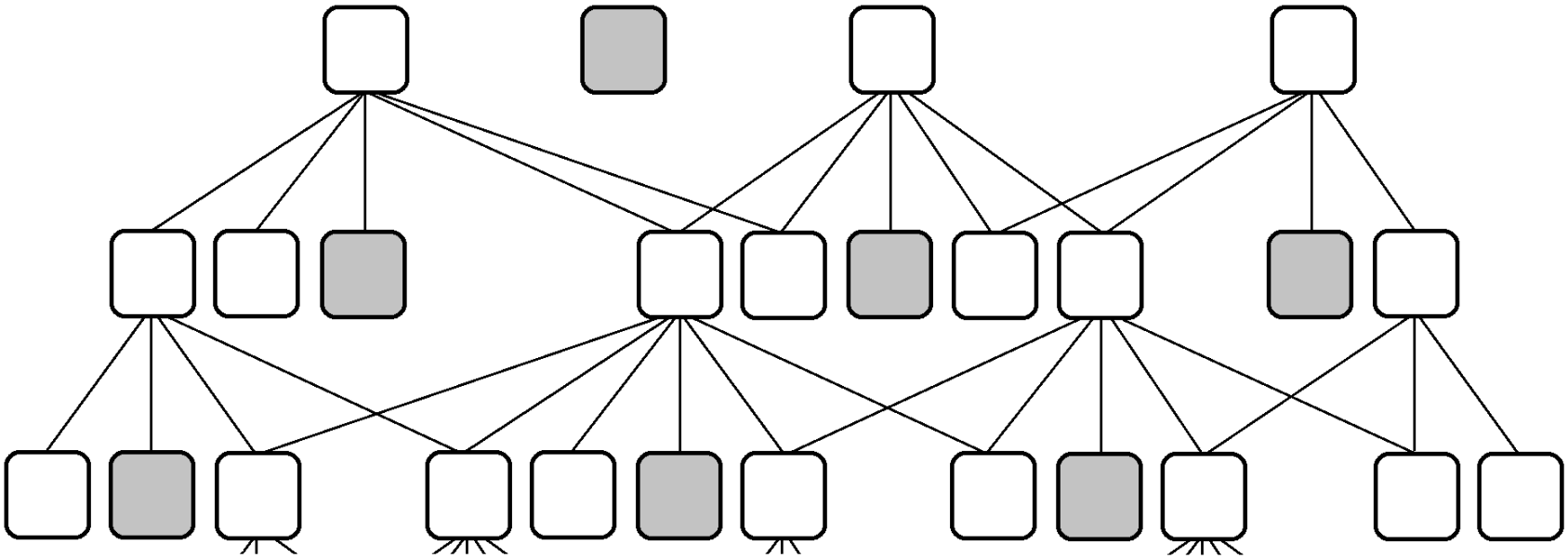
Иерархии с темами различной глубины

- На каждом уровне расщепляются не все темы (допускается вырожденность: $p(s|t) \equiv 0$ для некоторых t)
- Расщепляемая тема может иметь дочернюю фоновую, в которой собирается общая лексика родительской темы
- При максимальном разреживании $p(t|s) \in \{0, 1\}$ иерархия является деревом (корень не показан)



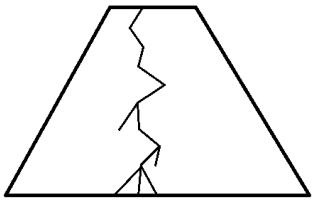
Иерархии с темами различной глубины

- На каждом уровне расщепляются не все темы (допускается вырожденность: $p(s|t) \equiv 0$ для некоторых t)
- Расщепляемая тема может иметь дочернюю фоновую, в которой собирается общая лексика родительской темы
- При умеренном разреживании $p(t|s)$ у вершины может быть несколько родителей (корень не показан)

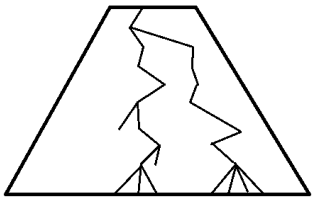


Иерархии с темами различной глубины

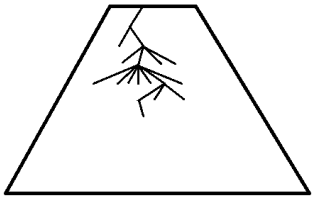
След документа в тематической иерархии определяет степень его специализации, назначение, аудиторию



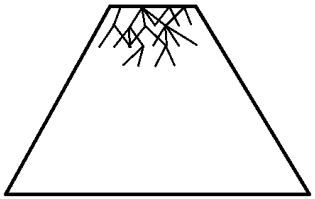
узко специализированный,
для профессионалов



междисциплинарное исследование,
для профессионалов



обзорный,
для ознакомления с предметной областью

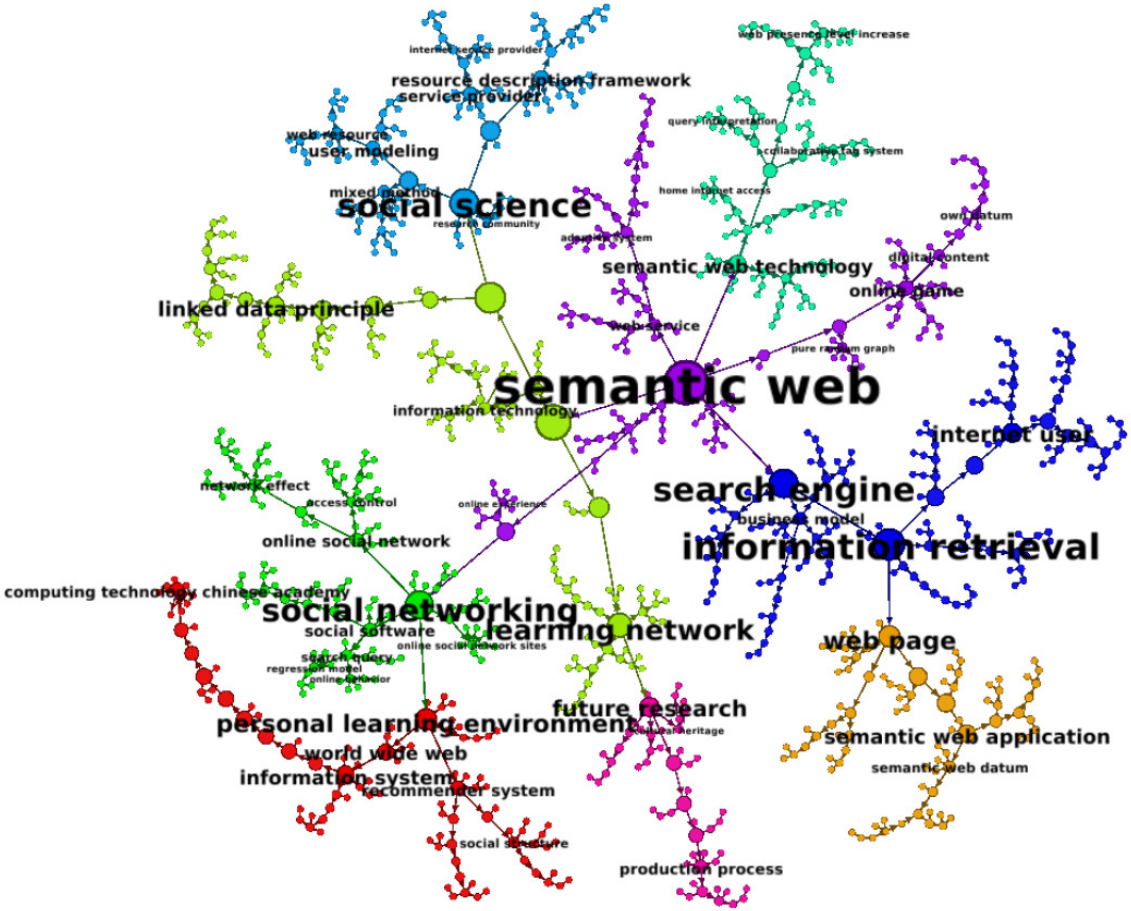


популярный или энциклопедиченский,
для расширения кругозора

Способы оценивания качества тематических иерархий

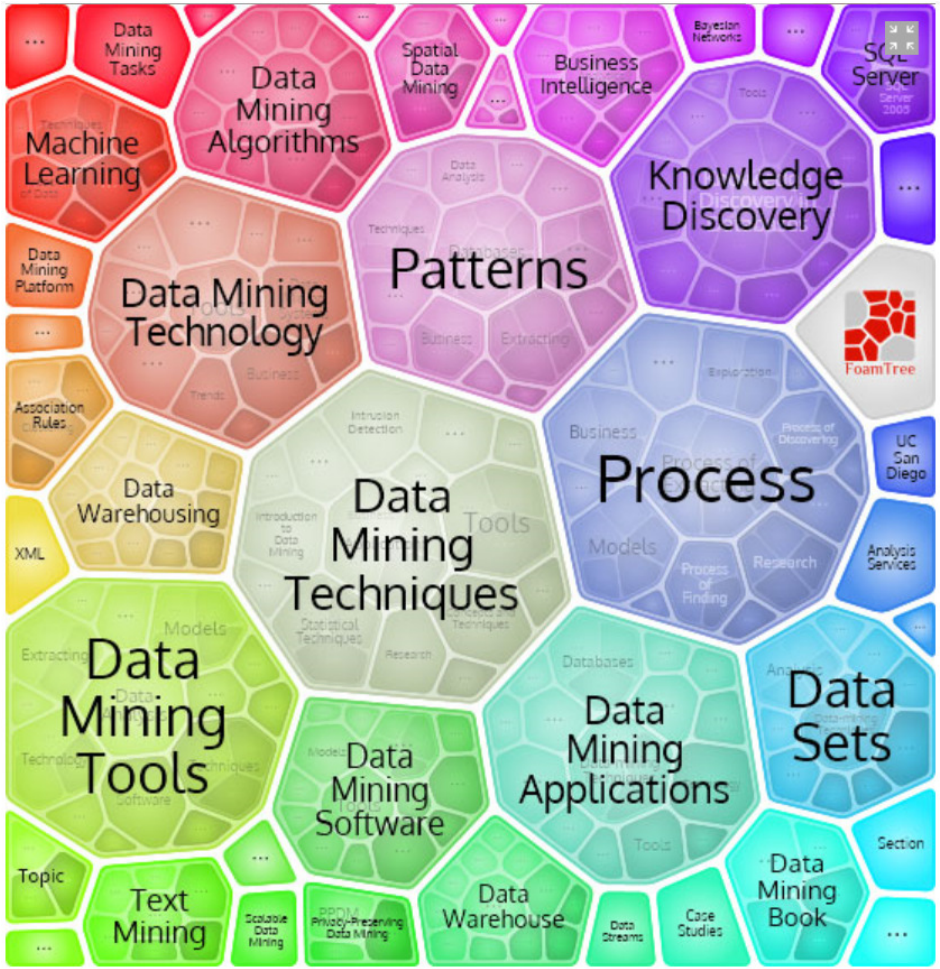
- *Перплексия* или правдоподобие: приводит ли постепенное дробление тем к более точному разложению
- *Устойчивость*: получаются ли схожие иерархии при различных начальных условиях
- *Полезность*: сколько шагов делает пользователь, чтобы найти документ по иерархии
- *Метод интрузий*: правильно ли асессоры определяют чужую тему, внедрённую в список дочерних тем
- *Сравнение с «золотым стандартом»*: насколько иерархия похожа на имеющуюся категоризацию документов

Визуализация древовидных иерархий



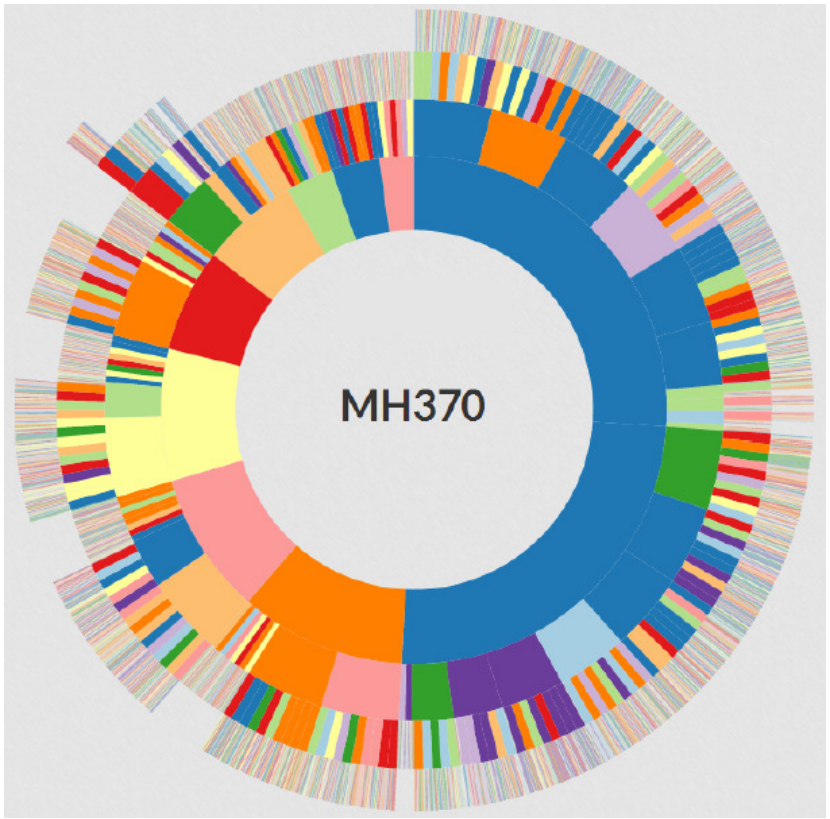
Georgeta Bordea. Domain adaptive extraction of topical hierarchies for Expertise Mining. 2013.

Визуализация древовидных иерархий в проекте FoamTree



https://carrotsearch.com/foamtree-overview

Визуализация древовидных иерархий



Smith A., Hawes T., Myers M.. Hiérarchie: interactive visualization for hierarchical topic models. Workshop on Interactive Language Learning, Visualization, and Interfaces, ACL, 2014.

Порождающая модальность

Основные предположения:

- C — порождающая модальность (категории, авторы, ...)
- $D \times W \times T \times C$ — дискретное вероятностное пространство
- коллекция — i.i.d. выборка $(d_i, w_i, t_i, c_i)_{i=1}^n \sim p(d, w, t, c)$
- d_i, w_i — наблюдаемые, темы t_i — скрытые
- два предположения об условной независимости:
  $p(w|d,t) = p(w|t), \;\; p(t|c,d) = p(t|c)$

Вероятностная модель порождения документа d:

$$p(w|d) = \sum_{t \in T} p(w|t) \sum_{c \in C} p(t|c)\, p(c|d) = \sum_{t \in T} \phi_{wt} \sum_{c \in C} \psi_{tc} \pi_{cd}$$

- $\phi_{wt} \equiv p(w|t)$ — распределение терминов в темах
- $\psi_{tc} \equiv p(t|c)$ — распределение тем в категориях
- $\pi_{cd} \equiv p(c|d)$ — распределение категорий в документах

ARTM для трёх-матричных разложений ΦΨΠ

Максимизация log правдоподобия с регуляризатором R:

$$\sum_{d\in D}\sum_{w\in d} n_{dw} \ln \sum_{t\in T}\sum_{c\in C} \phi_{wt}\psi_{tc}\pi_{cd} + R(\Phi,\Psi,\Pi) \;\to\; \max_{\Phi,\Psi,\Pi};$$

EM-алгоритм: метод простой итерации для системы уравнений

E-шаг:
$$p_{tcdw} \equiv p(t,c|d,w) = \underset{(t,c)\in T\times C}{\operatorname{norm}}\big(\phi_{wt}\psi_{tc}\pi_{cd}\big);$$

M-шаг:
$$\begin{cases} \phi_{wt} = \underset{w\in W^m}{\operatorname{norm}}\left(n_{wt} + \phi_{wt}\dfrac{\partial R}{\partial \phi_{wt}}\right); & n_{wt} = \sum\limits_{d,c} n_{dw}p_{tcdw} \\ \psi_{tc} = \underset{t\in T}{\operatorname{norm}}\left(n_{tc} + \psi_{tc}\dfrac{\partial R}{\partial \psi_{tc}}\right); & n_{tc} = \sum\limits_{d,w} n_{dw}p_{tcdw} \\ \pi_{cd} = \underset{c\in C}{\operatorname{norm}}\left(n_{cd} + \pi_{cd}\dfrac{\partial R}{\partial \pi_{cd}}\right); & n_{cd} = \sum\limits_{w,t} n_{dw}p_{tcdw} \end{cases}$$

Автор-тематическая модель (Author-topic model)

$C_d \subset C$ — множество порождающих категорий документа d

- Если $\pi_{cd} = \frac{1}{|C_d|}\left[c \in C_d\right]$, вклады авторов равны, то матрица Π фиксирована, EM-алгоритм на Π отдыхает :)
- Если $\pi_{cd} = 0, \quad c \notin C_d$, вклады авторов определяет модель, фиксирована структура разреженности матрицы Π, EM-алгоритм определяет только ненулевые элементы.
- Если множество C_d задано неточно или частично:

$$R(\Pi) = \sum_{d \in D} \sum_{c \in C_d} \ln \pi_{cd} \to \max$$

- Если множества C_d неизвестны, но Π разрежена:

$$R(\Pi) = -\sum_{d \in D} \sum_{c \in C} \ln \pi_{cd} \to \max$$

M. Rosen-Zvi, T. Griffiths, M. Steyvers, P. Smyth. The author-topic model for authors and documents. 2004.

Транзакционные данные

Выборка может содержать не только пары (d, w), но также тройки, ..., n-ки элементов разных модальностей.

Примеры:

- **Данные социальной сети:**
  (d, u, w) — в блоге d пользователь u записал слово w
- **Данные сети интернет-рекламы:**
  (u, d, b) — пользователь u кликнул рекламное объявление b на веб-странице d
- **Данные рекомендательной системы:**
  (u, f, s) — пользователь u оценил фильм f в ситуативном контексте s

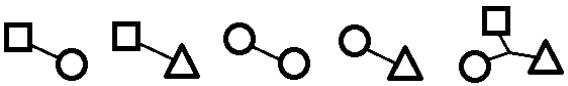
Хотим объяснить наблюдаемую выборку рёбер гиперграфа латентными тематическими профилями его вершин.

Тематическая модель гиперграфа: определения и обозначения

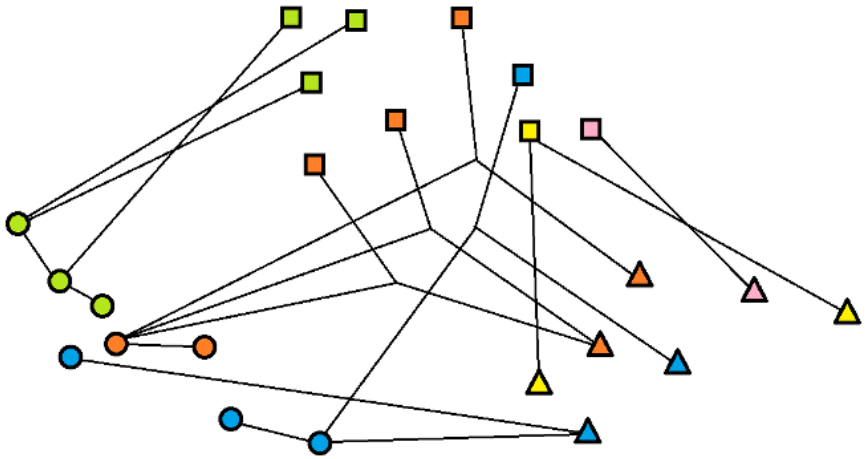
$\Gamma = \langle V, E \rangle$ — ориентированный гиперграф.
$V = V^1 \sqcup \cdots \sqcup V^M$ — разбиение вершин по модальностям

M — множество модальностей:

K — множество типов рёбер:



T — множество тем:



X^k — наблюдаемая выборка транзакций — рёбер типа k
ребро (d, x): вершина-контейнер $d \in V$ и вершины $x \subset V$,
n_{dx} — число вхождений ребра (d, x) в выборку X^k
$p_k(d, x)$ — неизвестное распределение на рёбрах типа k

Тематическая модель гиперграфа

Вероятностная тематическая модель рёбер типа k:

$$p_k(x|d) = \sum_{t \in T} \theta_{td} \prod_{v \in x} \phi_{kvt},$$

$\theta_{td} = p(t|d)$ — тематика контейнера не зависит от типа ребра k
$\phi_{kvt} = p_k(v|t)$ — для модальности v в теме t на рёбрах типа k

Задача максимизации log правдоподобия:

$$\sum_{k \in K} \tau_k \sum_{(d,x) \in X^k} n_{dx} \ln \sum_{t \in T} \theta_{td} \prod_{v \in x} \phi_{kvt} \to \max_{\Phi, \Theta},$$

$$\phi_{kvt} \geqslant 0, \quad \sum_{v \in V^m} \phi_{kvt} = 1; \qquad \theta_{td} \geqslant 0, \quad \sum_{t \in T} \theta_{td} = 1;$$

где $\tau_k > 0$ — веса типов рёбер.

EM-алгоритм для гиперграфовой ARTM

Задача максимизации регуляризованного правдоподобия:

$$\sum_{k\in K}\tau_k \sum_{(d,x)\in X^k} n_{dx} \ln \sum_{t\in T} \theta_{td} \prod_{v\in x} \phi_{kvt} \;+\; R(\Phi,\Theta) \to \max_{\Phi,\Theta}.$$

EM-алгоритм: метод простой итерации для системы уравнений со вспомогательными переменными $p_{ktdx} = p_k(t|d,x)$:

$$\text{E-шаг:}\quad p_{ktdx} = \underset{t\in T}{\text{norm}}\left(\theta_{td}\prod_{v\in x}\phi_{kvt}\right)$$

$$\text{M-шаг:}\quad \begin{cases} \phi_{kvt} = \underset{v\in V^m}{\text{norm}}\left(\sum_{(d,x)} [v\in x]\tau_k n_{dx} p_{ktdx} + \phi_{kvt}\frac{\partial R}{\partial \phi_{kvt}}\right) \\ \theta_{td} = \underset{t\in T}{\text{norm}}\left(\sum_{k\in K}\sum_{(d,x)} \tau_k n_{dx} p_{ktdx} + \theta_{td}\frac{\partial R}{\partial \theta_{td}}\right) \end{cases}$$

Резюме

- Модальности — мощное обобщение ARTM для учёта разнородных исходных данных
- Иерархические послойные модели реализуются с помощью модальностей или псевдо-документов
- Следующий шаг обобщения ARTM — гиперграфовые тематические модели для транзакционных данных
- Трёх-матричные и гиперграфовые модели пока не реализованы в BigARTM