Вероятностные тематические модели
Лекция 7. 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 2016

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: коллекция EuroParl, 21 язык
Comparable — не перевод, а пересказ на другом языке,
пример: Википедия

L — множество языков, каждый язык — модальность
W^ℓ — словарь языка ℓ

Дополнительные данные в мультиязычных коллекциях:

- двуязычные словари: $V_{\ell k} \subset W^\ell \times W^k$,
  $(w, u) \in V_{\ell k}$ — слово $u \in W^k$ является переводом $w \in W^\ell$,
  $\Pi_k(w)$ — множество всех переводов $w \in W^\ell$ в языке k
- выравнивание параллельных текстов по предложениям

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

Регуляризация по двуязычным словарям. Модель ML-TD

Гипотеза. Если $(w,u) \in V_{\ell k}$, то тематика слов w и u близка:

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

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

Модель ML-TD (MultiLingual Translation Dictionary)
Регуляризатор матрицы Φ:

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

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: $\pi_{uwt} = p(u|w,t)$, 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,

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

Модель ML-TDP (MultiLingual Translation Dictionary Probability)

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

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):

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

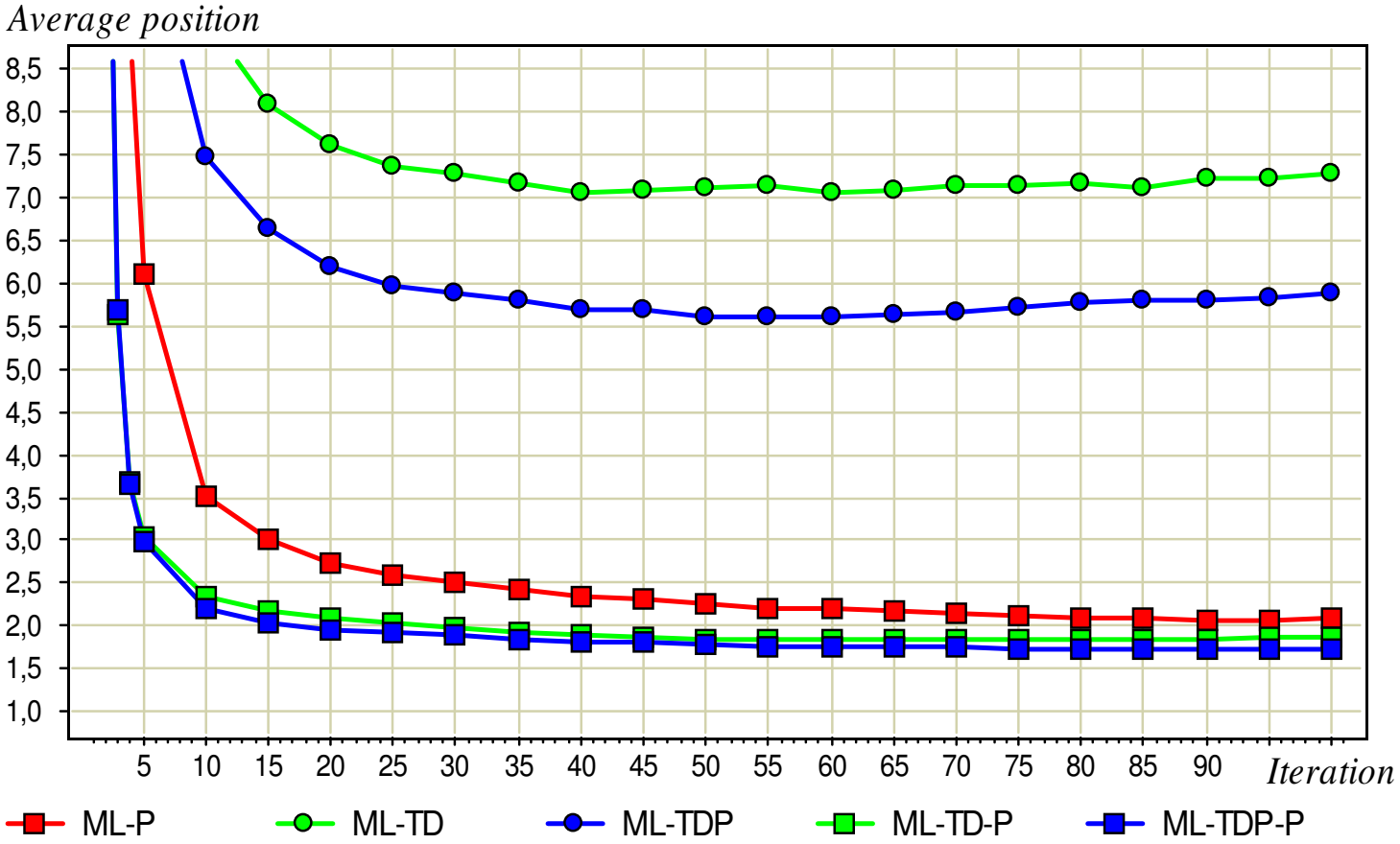
$$\pi_{uwt} = \underset{u \in W^k}{\text{norm}} \left(\pi_{wut} n_{ut} \right)$$

Смысл регуляризации:

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

Кросс-язычный поиск: ищем документ по его переводу

Качество поиска — средняя позиция перевода в выдаче
Wiki: $|D| = 586$, $|W^{\text{rus}}| = 19\,305$, $|W^{\text{eng}}| = 23\,413$, $|V| = 82\,642$



Пример. Переводы «сумма→sum» и «сумма→total»

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

Биграммная тематическая модель

n_{dvw} — частота пары слов «vw» в документе d
$\phi^v_{wt} = p(w|v,t)$ — распределение слов после слова v в теме t

Модель BTM (Bigram Topic Model):

$$\sum_{d\in D}\sum_{vw\in d} n_{dvw} \ln \sum_{t\in T} \phi^v_{wt}\theta_{td} \to \max_{\Phi,\Theta}$$

Это мультимодальная модель:
$M = W$, каждому слову v соответствует отдельная модальность,
$W^v = W$ — все слова, которые могут следовать за v.

Недостатки биграммной модели BTM:

- все пары соседних слов образуют биграммы;
- модель не описывает отдельные слова (униграммы);
- общее число токенов $O(|W|^2)$.

Hanna Wallach. Topic modeling: beyond bag-of-words // ICML 2006

Объединение униграмм и биграмм в одной модели

Модель TNG (Topical n-grams):

$$\sum_{d\in D}\sum_{vw\in d} n_{dvw} \ln \sum_{t\in T} \underbrace{\big(x_{vwt}\phi^{v}_{wt} + (1-x_{vwt})\phi_{wt}\big)}_{p(w|v,t)} \theta_{td} \to \max_{\Phi,\Theta}$$

$x_{vwt} = \mathrm{P}($пара слов «vw» является биграммой в теме $t)$.

Частные случаи:

- $x_{vwt} = x_{vt}$ — матрица параметров в модели TNG.
- $x_{vwt} \equiv 1$ — модель BTM;
- $x_{vwt} =$ [пара слов «vw» из словаря биграмм];

Xuerui Wang, Andrew McCallum, Xing Wei. Topical *N*-Grams: Phrase and Topic Discovery, with an Application to Information Retrieval. 2007.

Мультимодальная мультиграммная ARTM

W^n — словари n-грамм, отфильтрованные по трём критериям:
1) наличие подчинительных синтаксических связей;
2) статистическая значимость коллокации $\text{PMI}(vw) = \ln \frac{P_{vw}}{P_v P_w}$;
3) высокая тематичность $\text{KL}\left(\frac{1}{|T|} \,\middle\|\, p(t|vw)\right)$.

Связь с моделью TNG.
При $x_{vwt} = \lambda[vw \in W^2]$ сумма log-правдоподобий модальностей является оценкой снизу для log-правдоподобия модели TNG:

$$\sum_{d\in D}\sum_{vw\in d} n_{dvw} \ln \sum_{t\in T}\left(x_{vwt}\phi^v_{wt} + (1 - x_{vwt})\phi_{wt}\right)\theta_{td} \geqslant$$

$$\sum_{d\in D}\sum_{vw\in d} n_{dvw} \ln\left(\lambda \sum_{t\in T} \phi^v_{wt}\theta_{td} + (1-\lambda)\sum_{t\in T}\phi_{wt}\theta_{td}\right) \geqslant$$

$$\lambda \sum_{d,vw} n_{dvw} \ln \sum_{t\in T} \phi^v_{wt}\theta_{td} + (1-\lambda)\sum_{d,w} n_{dw} \ln \sum_{t\in T}\phi_{wt}\theta_{td} \to \max_{\Phi,\Theta}$$

Биграммы радикально улучшают интерпретируемость тем

Коллекция 1000 статей конференций ММРО, ИОИ на русском

распознавание образов в биоинформатике		теория вычислительной сложности	
unigrams	bigrams	unigrams	bigrams
объект	задача распознавания	задача	разделять множества
задача	множество мотивов	множество	конечное множество
множество	система масок	подмножество	условие задачи
мотив	вторичная структура	условие	задача о покрытии
разрешимость	структура белка	класс	покрытие множества
выборка	распознавание вторичной	решение	сильный смысл
маска	состояние объекта	конечный	разделяющий комитет
распознавание	обучающая выборка	число	минимальный аффинный
информативность	оценка информативности	аффинный	аффинный комитет
состояние	множество объектов	случай	аффинный разделяющий
закономерность	разрешимость задачи	покрытие	общее положение
система	критерий разрешимости	общий	множество точек
структура	информативность мотива	пространство	случай задачи
значение	первичная структура	схема	общий случай
регулярность	тупиковое множество	комитет	задача MASC

Сергей Стенин. Мультиграммные аддитивно регуляризованные тематические модели // Магистерская диссертация, МФТИ, 2015.

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 ФΨП

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

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

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 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:

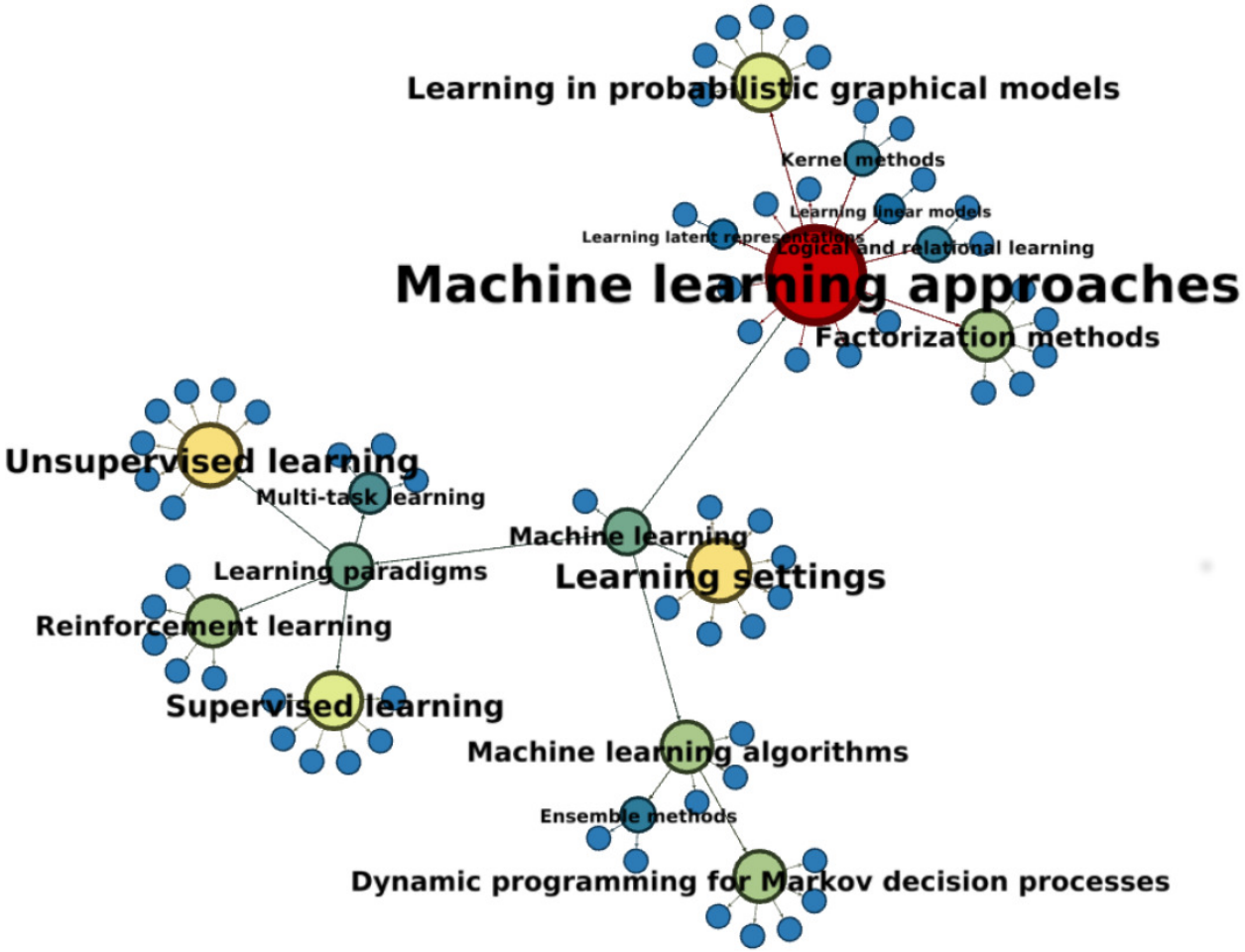
$$R(\Pi) = \sum_{d \in D} \sum_{c \in C_d} \ln \pi_{cd} \to \max$$

- 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.

Пример тематической иерархии



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

Иерархические тематические модели

- структура иерархии: дерево / многодольный граф
- стратегия: восходящая / нисходящая / одновременная
- 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:

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

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

$\Phi^p \approx \Phi\Psi$, отсюда регуляризатор матрицы Φ:

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

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:

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

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

$\Theta^p \approx \tilde{\Psi}\Theta$, отсюда регуляризатор матрицы Θ:

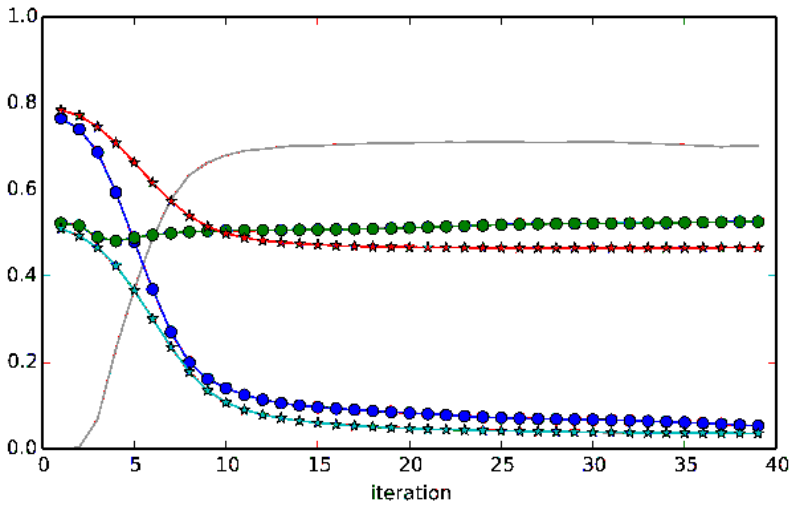
$$R(\Theta, \tilde{\Psi}) = \tau \sum_{d\in D}\sum_{t\in T} n_{td} \ln \sum_{s\in S} \tilde{\psi}_{ts}\theta_{sd} \;\to\; \max.$$

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

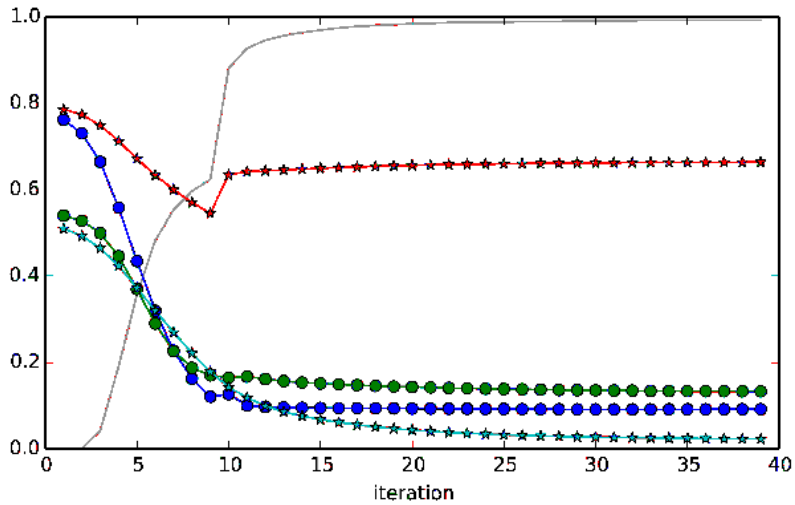
Эксперимент на коллекции ММРО-ИОИ

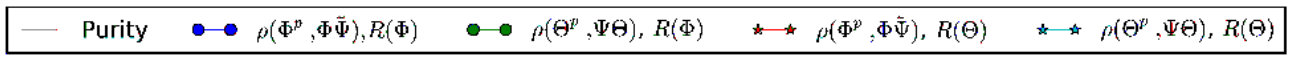
Качество аппроксимации матриц Φ и Θ родительского уровня матричными разложениями дочернего уровня:
зависимости среднего расстояния Хеллингера от номера итерации при переходе между 1-м и 2-м уровнями.

Разреживание Φ с 1-й итерации



Разреживание Φ с 10-й итерации

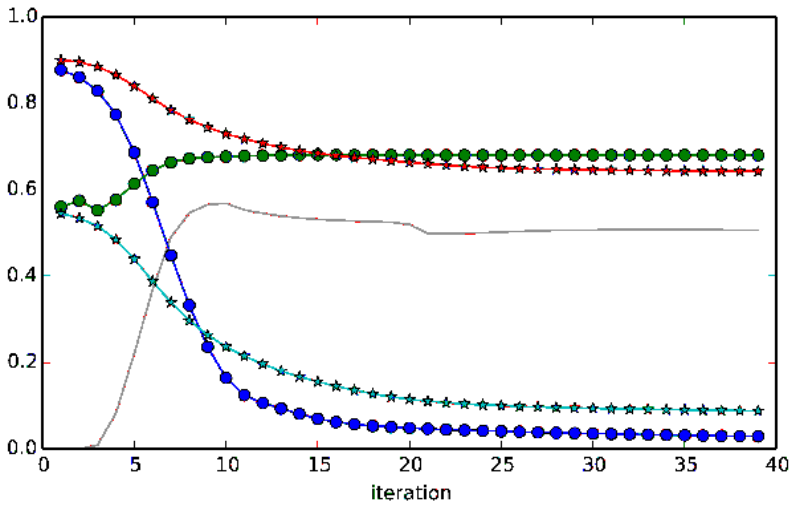




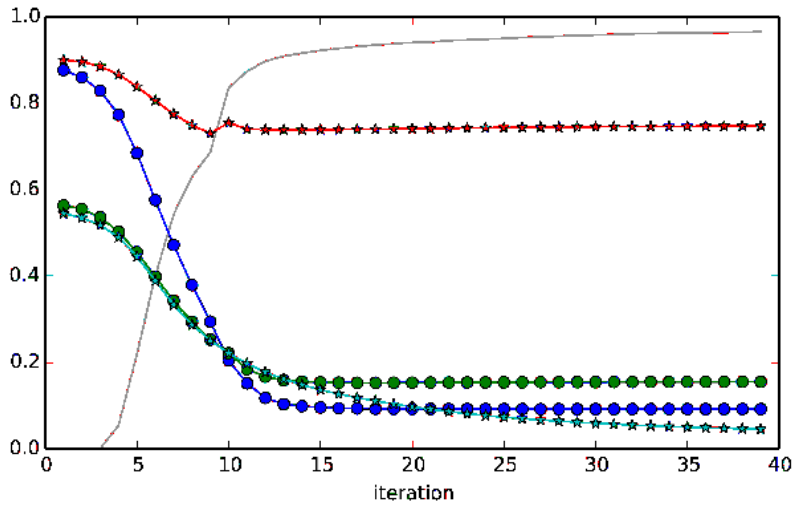
Эксперимент на коллекции ММРО-ИОИ

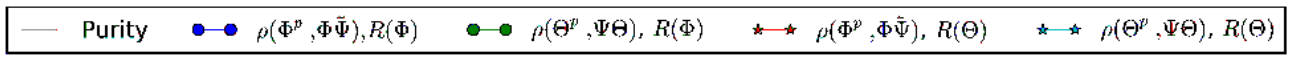
Качество аппроксимации матриц Φ и Θ родительского уровня матричными разложениями дочернего уровня: зависимости среднего расстояния Хеллингера от номера итерации при переходе между 2-м и 3-м уровнями.

Разреживание Φ с 1-й итерации



Разреживание Φ с 10-й итерации





Выводы

- Регуляризатор Φ приближает $\Phi^p \approx \Phi\Psi$ и $\Theta^p \approx \tilde{\Psi}\Theta$.
- Регуляризатор Θ приближает только $\Theta^p \approx \tilde{\Psi}\Theta$.
- Разреживание $\Psi = (\psi_{st})$ даёт иерархию, близкую к дереву;
- при этом надо следить за невырожденностью $p(t|s)$.

Дальнейшие задачи:

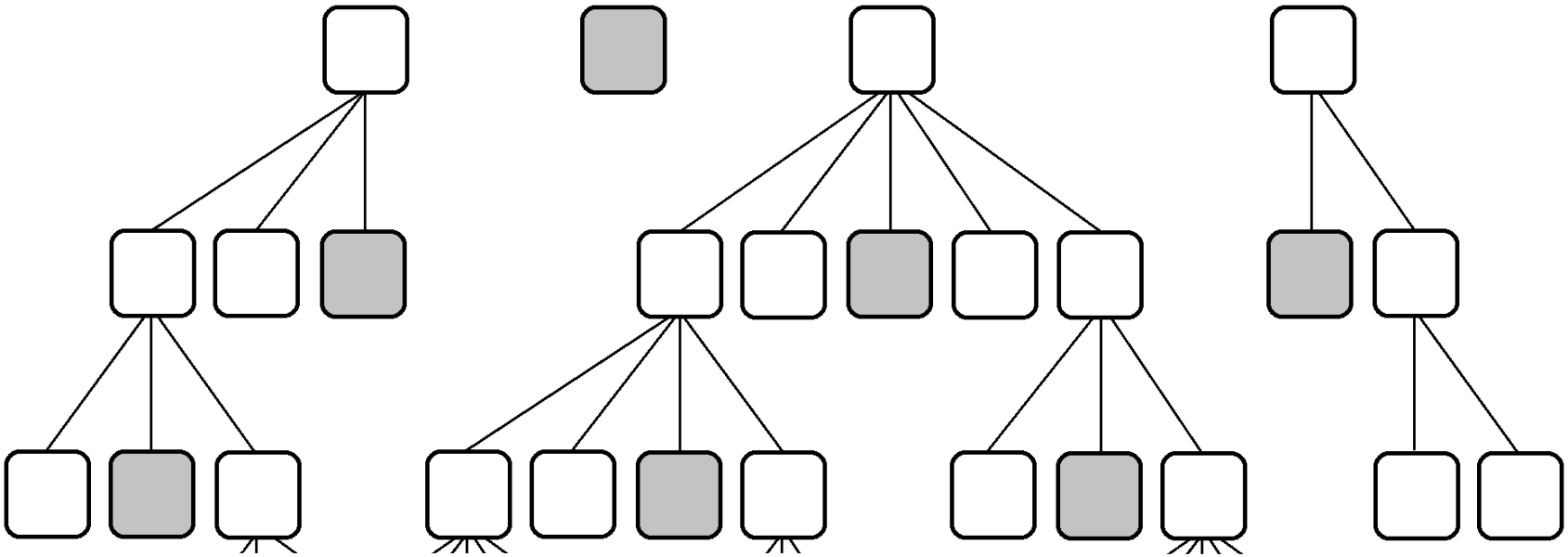
- Согласованная регуляризация: $\psi_{st} p(t) = \tilde{\psi}_{ts} p(s)$

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

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

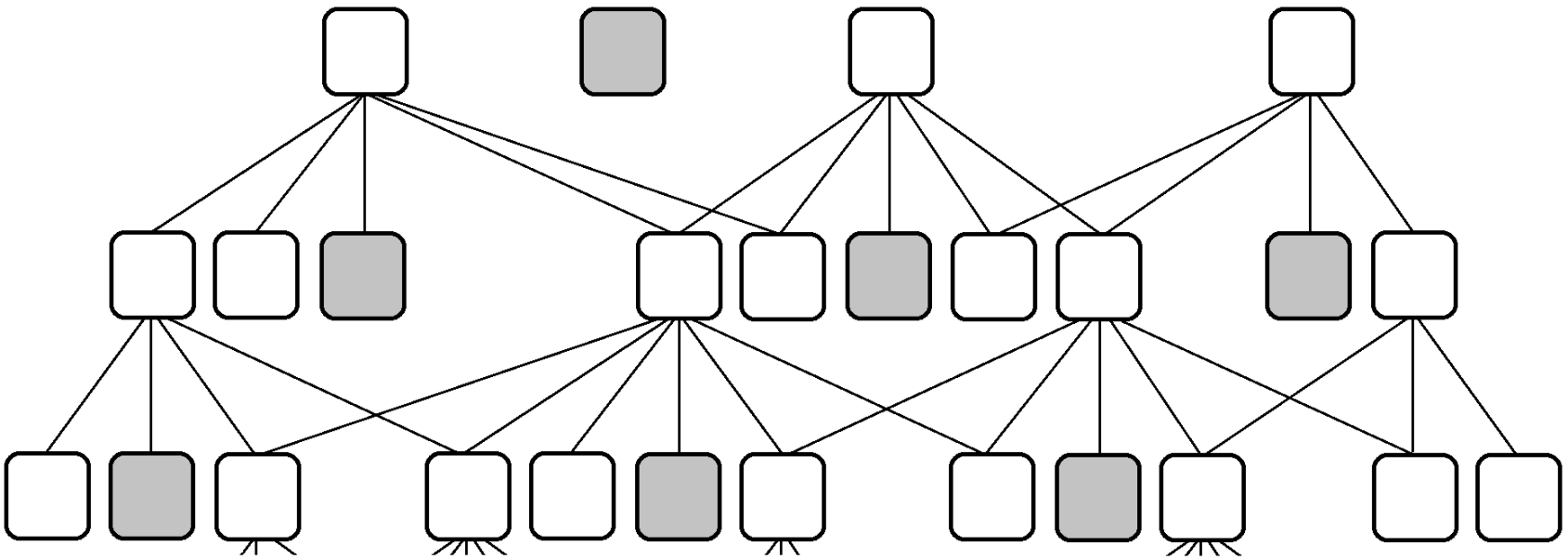
Иерархии с темами различной глубины

- 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)



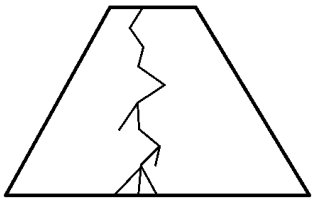
Иерархии с темами различной глубины

- 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)

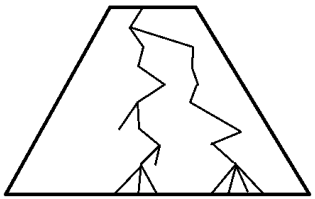


Иерархии с темами различной глубины

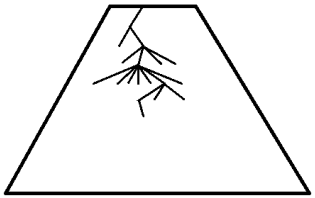
След документа в тематической иерархии определяет степень его специализации, назначение, аудиторию



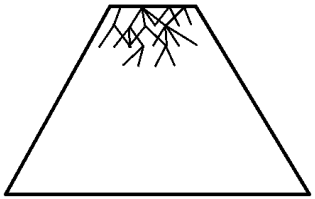
узко специализированный,
для профессионалов



междисциплинарное исследование,
для профессионалов



обзорный,
для ознакомления с предметной областью

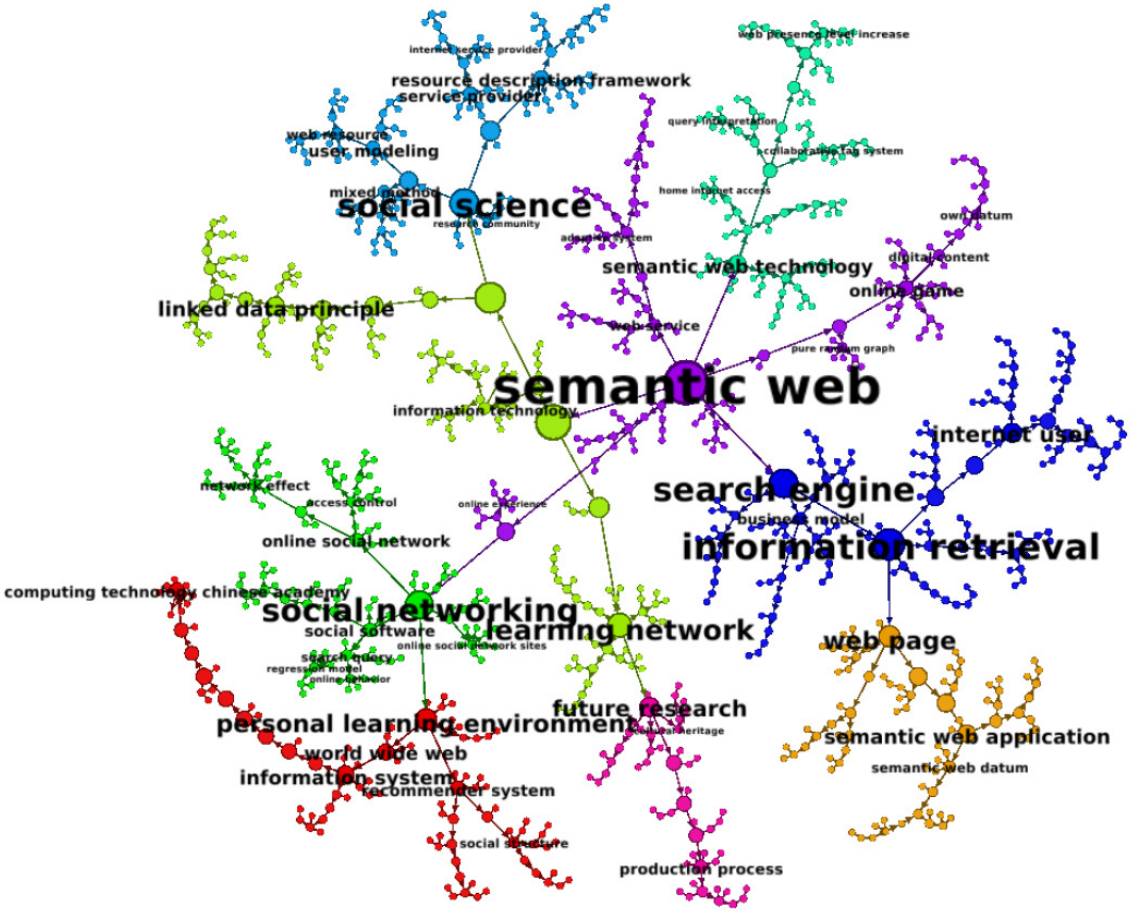


популярный или энциклопедиченский,
для расширения кругозора

Способы оценивания качества тематических иерархий

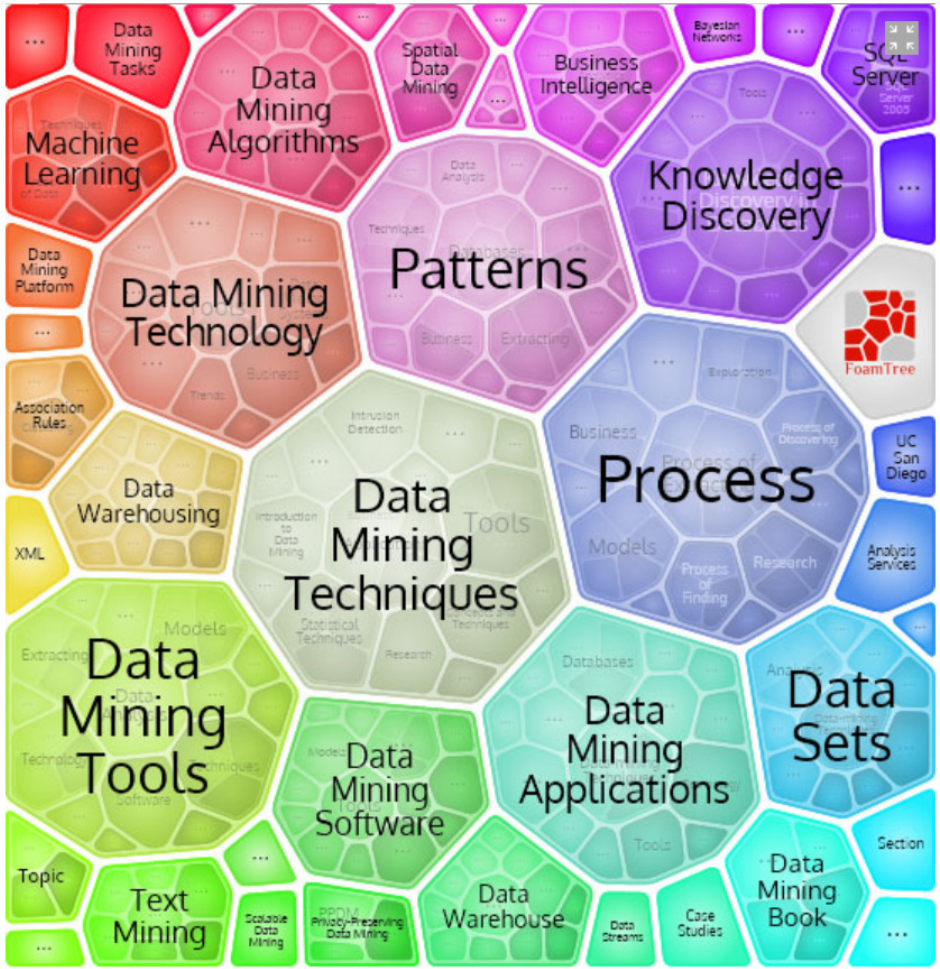
- *Перплексия* или правдоподобие: приводит ли постепенное дробление тем к более точному разложению
- *Устойчивость*: получаются ли схожие иерархии при различных начальных условиях
- *Метод интрузий*: правильно ли асессоры определяют чужую тему, внедрённую в список дочерних тем
- *Проверка адекватности*: правильно ли разделяются документы из двух разнородных коллекций
- *Сравнение с «золотым стандартом»*: насколько иерархия похожа на имеющуюся категоризацию документов

Визуализация древовидных иерархий



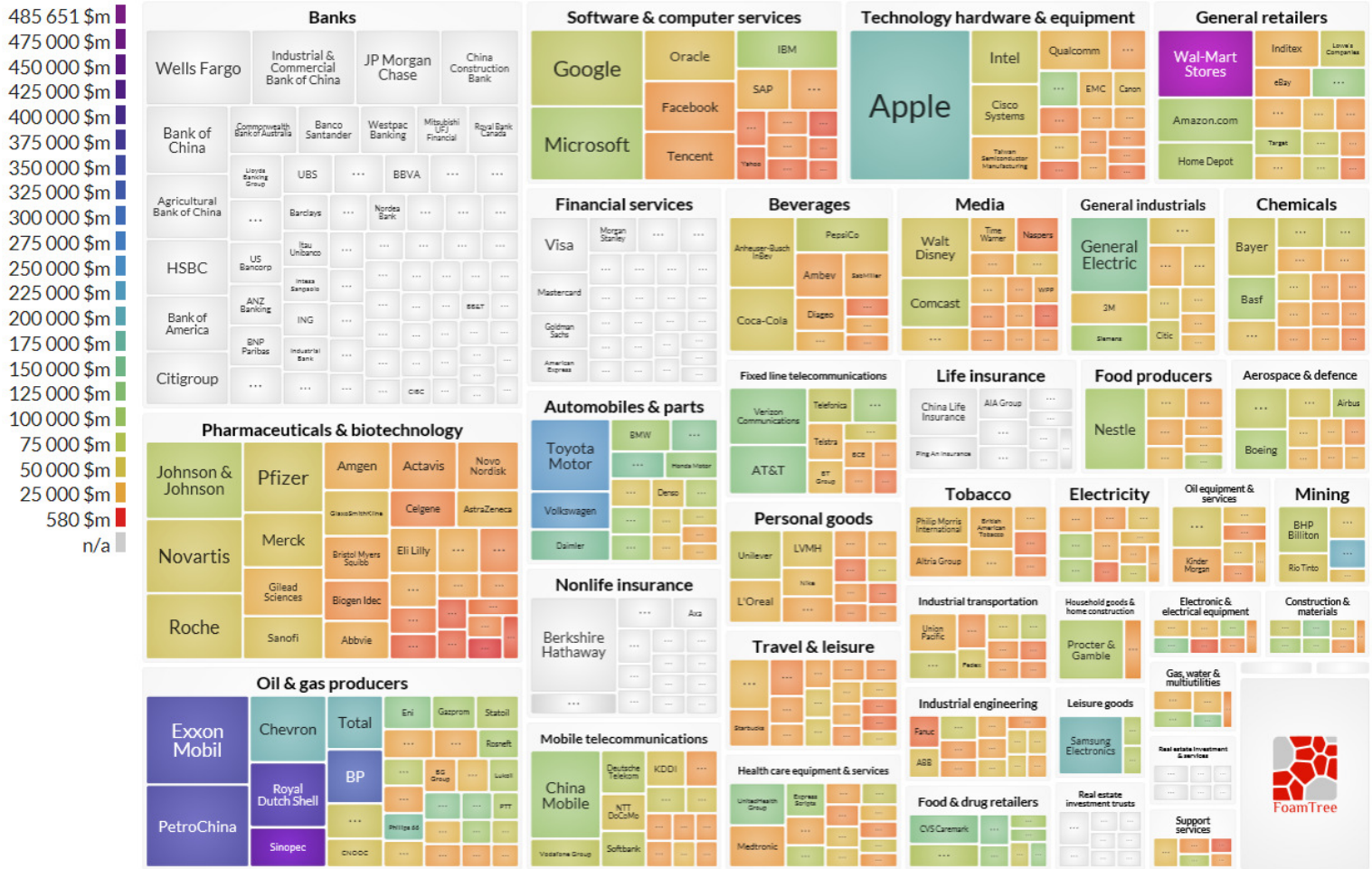
Georgeta Bordea. Domain adaptive extraction of topical hierarchies for Expertise Mining. 2013.

Визуализация древовидных иерархий в проекте FoamTree



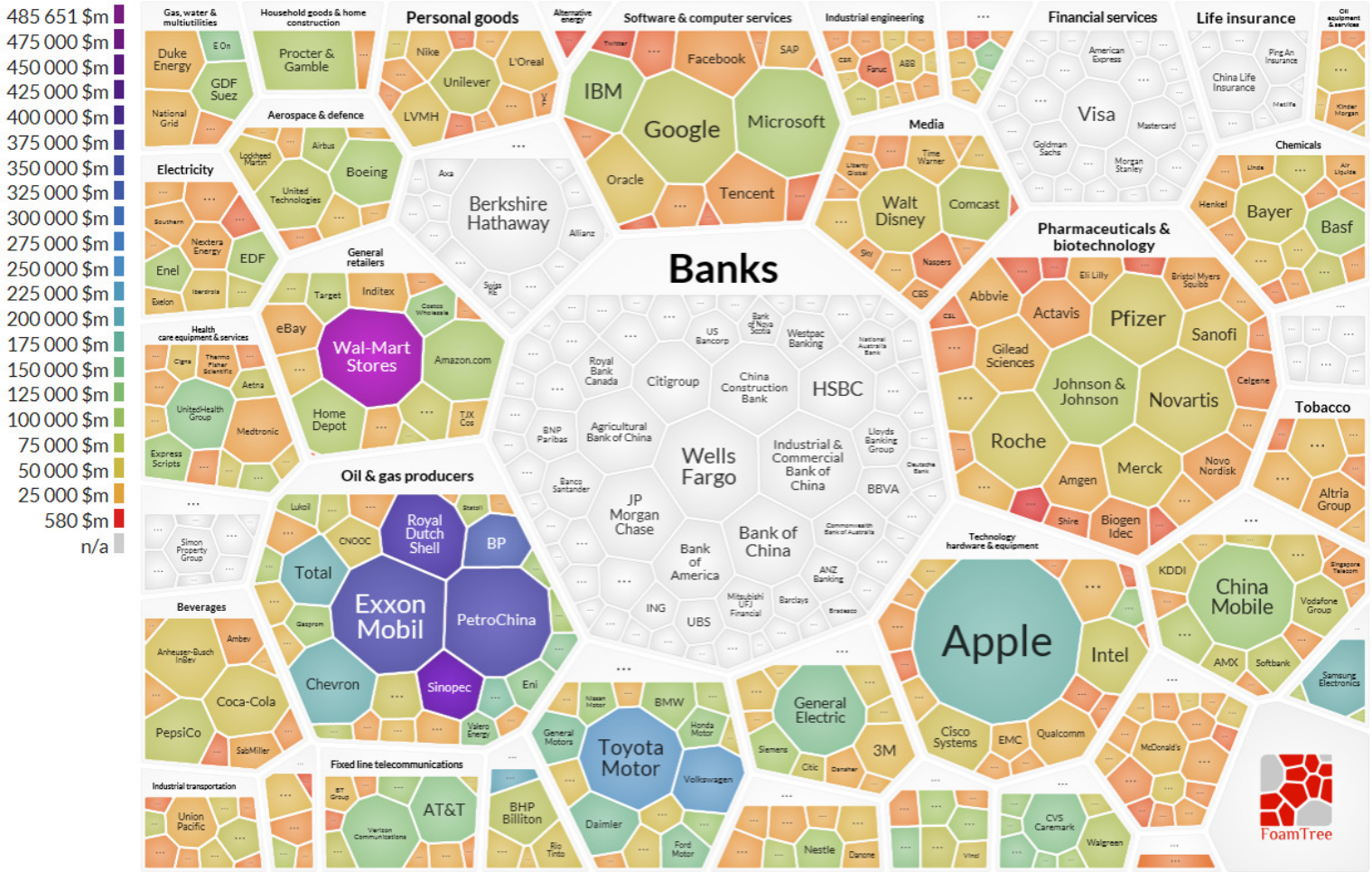
https://carrotsearch.com/foamtree-overview

Демо-пример FoamTree: рейтинг крупнейших компаний



https://get.carrotsearch.com/foamtree/demo/demos

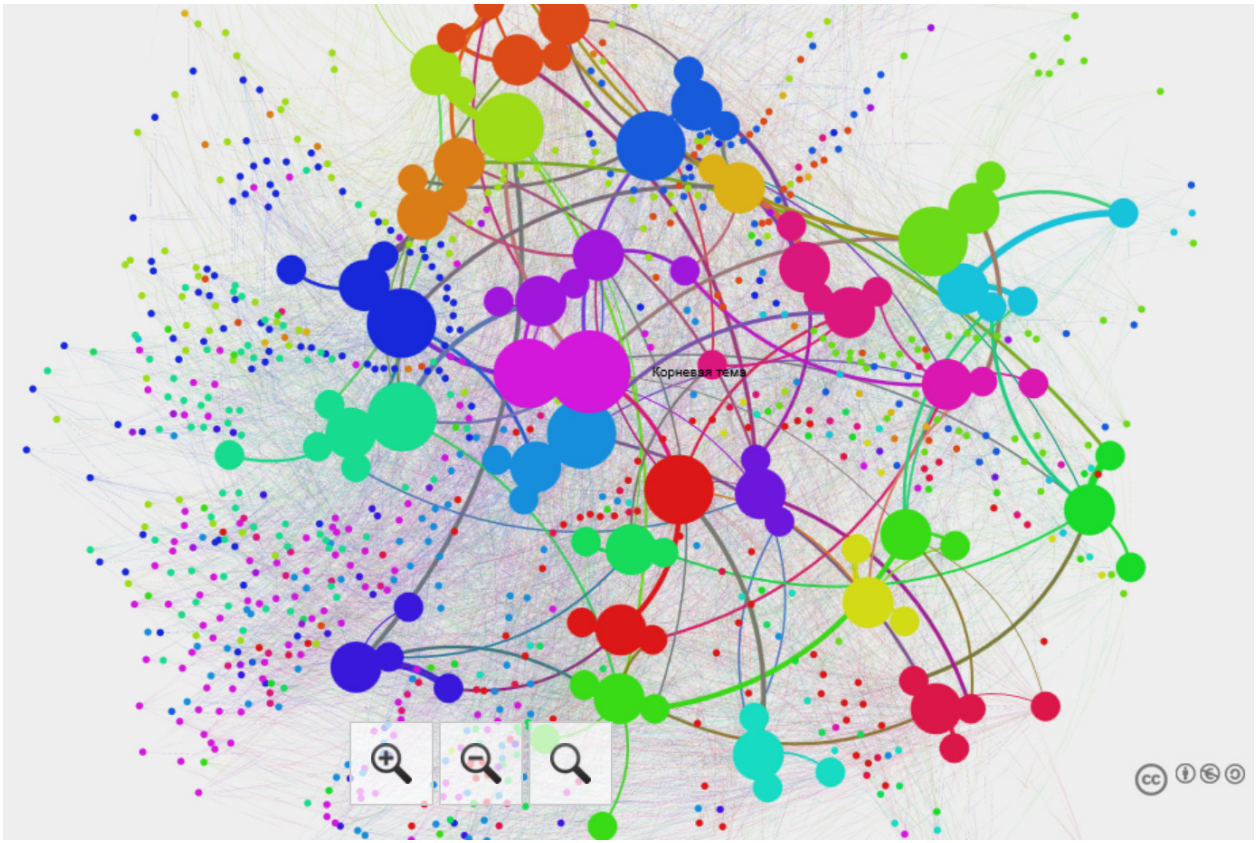
Демо-пример FoamTree: рейтинг крупнейших компаний



https://get.carrotsearch.com/foamtree/demo/demos

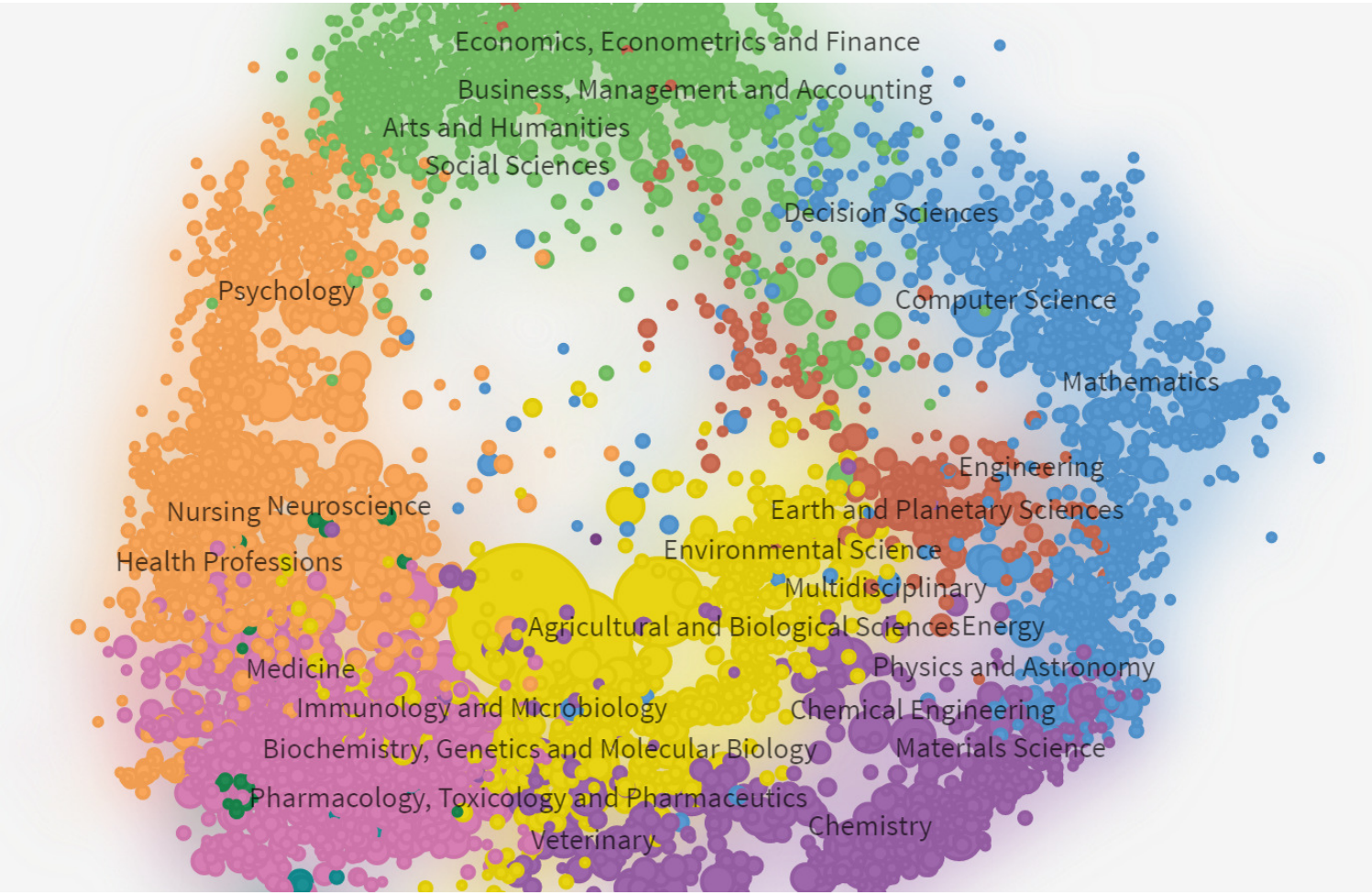
Визуализация не-древовидной иерархии

Коллекция ММРО/ИОИ: $|D| = 865$, $|W| = 42\,000$ n-грамм
7 регуляризаторов на каждом из трёх уровней иерархии



http://explore-mmro.ru

Scopus Scimago: визуализация структуры научных знаний



http://www.scimagojr.com/shapeofscience

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

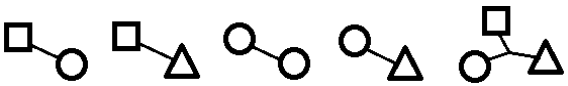
Хотим объяснить наблюдаемую выборку рёбер гиперграфа латентными тематическими профилями его вершин.

Тематическая модель гиперграфа: определения и обозначения

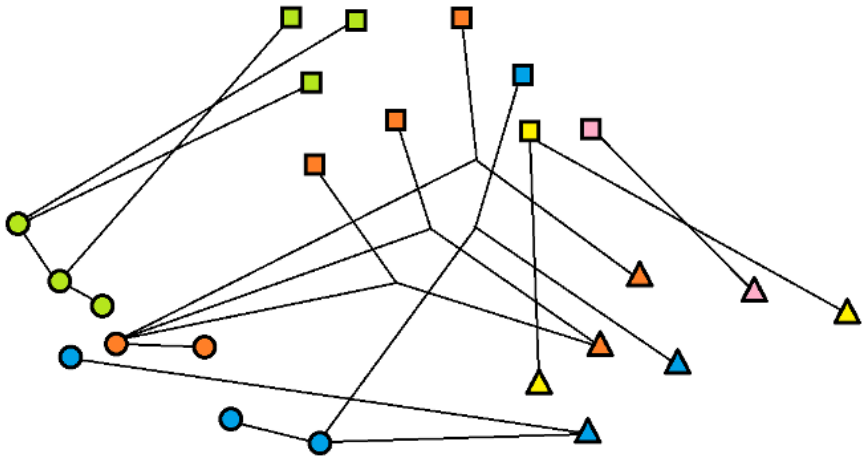
$\Gamma = \langle V, E \rangle$ — ориентированный гиперграф.
$V = V^1 \sqcup \cdots \sqcup V^M$ — разбиение вершин по модальностям

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 $p_{ktdx} = p_k(t|d,x)$:

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

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

Резюме

- Модальности — мощное обобщение ARTM для учёта разнообразных 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