

Средневзвешенная когерентность как мера интерпретируемости тематической модели

Жгутов Кирилл

Московский физико-технический институт

Курс: Автоматизация научных исследований
(практика, В. В. Стрижов)/Группа 125

Научный руководитель: доктор физико-математических наук
К. В. Воронцов

Консультант: В. А. Алексеев

2024

Цели исследования

Цель исследования: создать автоматическую метрику интерпретируемости тем тематической модели

Задача: исследовать поведение средневзвешенной когерентности для различных параметров

Задача тематического моделирования

Дано: коллекция текстовых документов

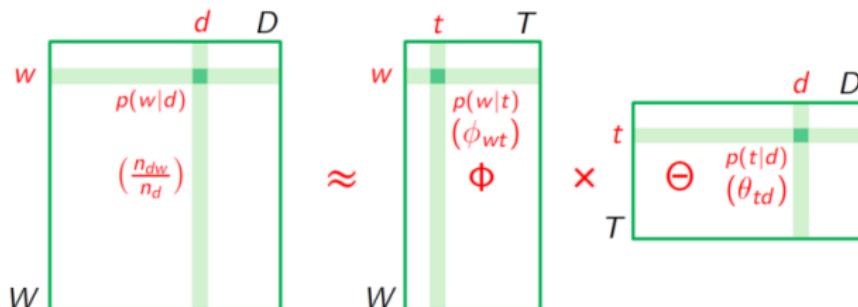
- ▶ n_{dw} - количество вхождений слова w в документ d
- ▶ n_d - количество слов в документ d
- ▶ $p(w|d) = \frac{n_{dw}}{n_d}$

Найти: параметры тематической модели

- ▶ $\phi_{wt} = p(w|t)$ вероятности слова w в каждой теме t
- ▶ $\theta_{td} = p(t|d)$ вероятности темы t в каждом документе d

$$p(w|d) = \sum_{t \in T} p(w|t)p(t|d)$$

Это задача стохастического матричного разложения



Интерпретируемость тематической модели

topic_0		topic_7	
token		token	
фильм	0.038650	год	0.014330
роль	0.012110	космический	0.011500
год	0.011740	который	0.011340
хороший	0.010600	дорога	0.006440
режиссёр	0.010200	сообщать	0.006110

Интерпретируемая и не интерпретируемая темы.

Когерентность¹

Когерентность (согласованность) темы t по k топовым словам

$$coh_t = \frac{1}{k(k-1)} \sum_{w_i, w_j \in argtop_k \phi_{wt}} PMI(w_i, w_j),$$
$$PMI(u, v) = \ln \frac{p(u, v)}{p(u)p(v)}$$

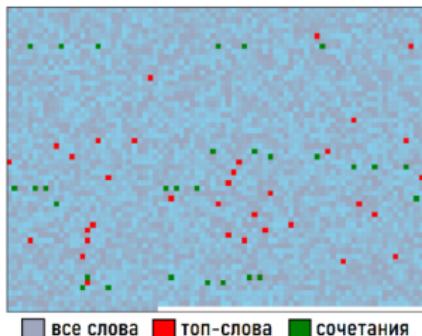
Gold-standard средняя корреляция Спирмена между оценками разных экспертов.

Resource	Method	Median	Mean
WordNet	HSO	0.15	0.59
	JCN	-0.20	0.19
	LCh	-0.31	-0.15
	LESK	<u>0.53</u>	<u>0.53</u>
	LIN	0.09	0.28
	PATH	0.29	0.12
	RES	0.57	0.66
	VECTOR	-0.08	0.27
Wikipedia	WuP	0.41	0.26
	RACO	0.62	0.69
	MiW	0.68	0.70
	DOCSIM	0.59	0.60
	PMI	0.74	0.77
Google	TITLES	<u>0.51</u>	
	LOGHITS		-0.19
Gold-standard	IAA	0.82	0.78

¹Newman D., Lau J.H., Grieser K., Baldwin T. Automatic evaluation of topic coherence // Human Language Technologies, HLT-2010, Pp. 100-108.

Когерентность

Классическая когерентность имеет недостаток: она учитывает лишь небольшое подмножество слов.



Напротив, если предположить существование суперсимметрии, то введение новых **частиц** приводит как раз к такому объединению. Оказывается, что суперсимметрия не только обеспечивает объединение **взаимодействий**, но и стабилизирует объединённую теорию, в которой присутствуют два совершенно разных масштаба: масштаб **масс обычных частиц** (порядка 100 **масс протона**) и масштаб великого объединения (порядка 10^{16} **масс протона**). Последний масштаб уже близок к так называемому **планковскому** масштабу, равному **обратной ньютонаской константе тяготения**, что составляет порядка 10^{19} **масс протона**. На этом масштабе мы ожидаем проявление **эффектов квантовой гравитации**. В этом моменте нас **ожидает** приятный **сюрприз**. Дело в том, что гравитация всегда стояла несколько особняком по отношению к остальным **взаимодействиям**. **Переносчик гравитации, гравитон, имеет спин 2**, в то время как **переносчики остальных взаимодействий имеют спин 1**. Однако **суперсимметрия перемешивает спины**.

first **top words** of **topic 3**: физика with **top 10** in bold: **частица, электрон, кварк, атом, энергия, вселенная, фотон, физика, физик, эксперимент, масса, теория, свет, симметрия, протон, эйнштейн, нейтрино, вещество, квантовый, ускоритель, детектор, волна, эффект, свойство, спин, гравитация, материя, адрон, поль, частота**

Средневзвешенная когерентность

$$\text{coh}_{t_0} = \frac{\sum_{u,v} \text{rel}_{t_0}(u, v) \text{coh}(u, v)}{\sum_{u,v} \text{rel}_{t_0}(u, v)}$$

$\text{coh}(u, v)$ сочетаемость пары слов u, v в текстах,

$\text{rel}_t(u, v)$ релевантность слов u и v теме t , в частности,

$\text{rel}_t(u, v) = [\phi_{ut}, \phi_{vt} > \text{top}_k \phi_{wt}]$, $\text{coh}(u, v) = PMI$

- когерентность Ньюмана

Альтернативный золотой стандарт

1. Вместо оценки экспертами каждой темы по топ словам предлагается выделять тематические цепочки

В исследованиях мы действительно можем находить корреляции между стилем вождения и особенностями личности. Например, склонные к экстраверсии водители могут больше отвлекаться на внешние факторы и стимулы внешней среды и в этом отношении представляют большую опасность. В свою очередь, люди, которым требуется большее количество психических ресурсов, для того чтобы справиться с тревогой, будут вести себя осторожнее в условиях трафика. Вместе с тем есть и обратная сторона: та же характеристика интроверсии за счет высокого уровня тревожности приводит к чрезмерной осторожности. Для таких водителей характерен крадущийся тип вождения, что будет влиять на общее тревожное поведение всех участников трафика.

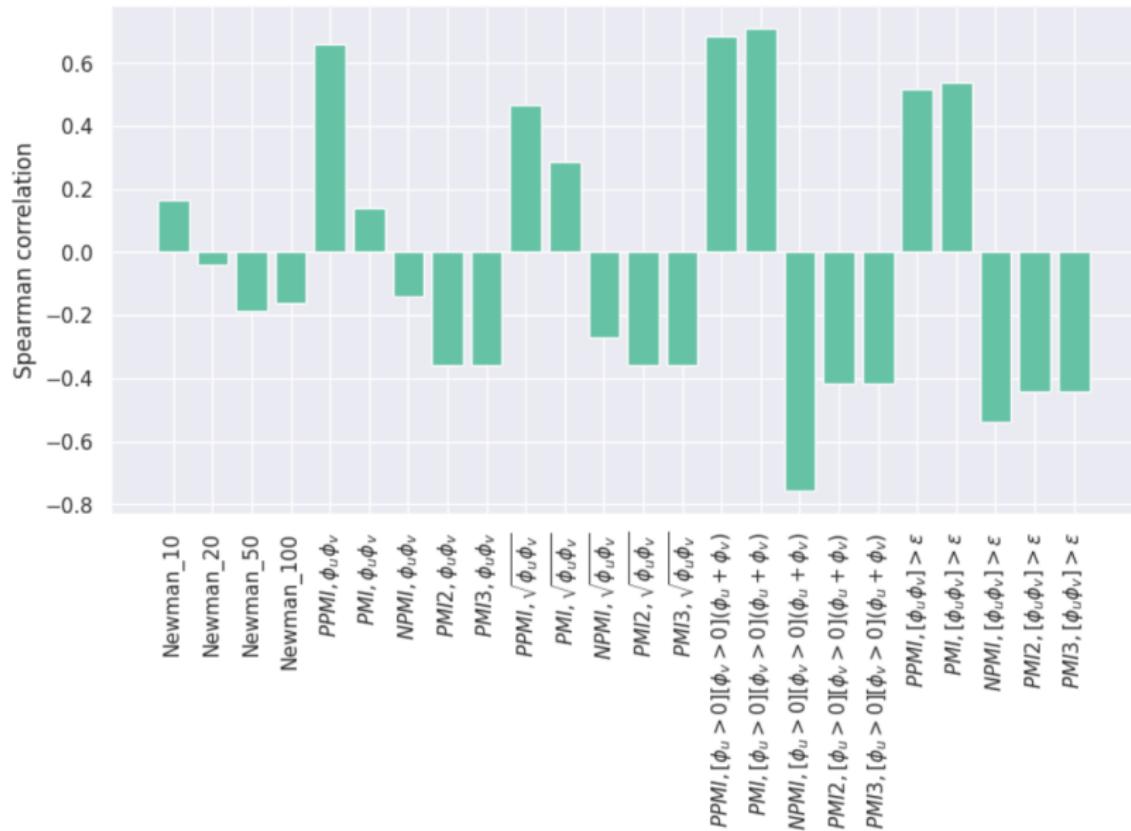
2. Вместо корреляции Спирмена с оценками экспертов, корреляция с согласованностью с цепочками слов

$$\text{cons}_{t_0} = \text{mean}_{C_{di} \in C(t_0)} p(t|C_{di})$$

C - множество цепочек

$C(t) = \{C_{di} \mid t = \text{argmax}_t p(t|C_{di})\}$ - множество цепочек, согласованных с темой t

Результаты эксперимента



Список лиетратуры

1. Newman D., Lau J.H., Grieser K., Baldwin T. Automatic evaluation of topic coherence // Human Language Technologies, HLT-2010, Pp. 100-108.
2. Vasily A Alekseev, Vladimir G Bulatov, and Konstantin V Vorontsov. Intra-text coherence as a measure of topic models' interpretability. In Komp'juternaja Lingvistika i Intellektual'nye Tehnologii, pages 1–13, 2018
3. Константин Воронцов. Вероятностное тематическое моделирование: теория регуляризации artm и библиотека с исходным кодом bigartm. 2023.
4. Воронцов К. В. "Оценивание качества тематических моделей"(из курса лекций "Вероятностные тематические модели" How humans interpret topic models. Advances in neural information processing systems, 22, 2009.

Итоги

- ▶ Предложены новые методы оценки интерпретируемости
- ▶ Показана их более высокая корелированность с согласованностью с цепочками слов