Сегментация и суммаризация текстов

К. В. Воронцов
vokov@forecsys.ru

Этот курс доступен на странице вики-ресурса
http://www.MachineLearning.ru/wiki
«Математические методы анализа текстов
(курс лекций) / осень 2019»

МФТИ — ФИЦ ИУ РАН • 20 ноября 2019

Содержание

Методы сегментации TextTiling, TopicTiling

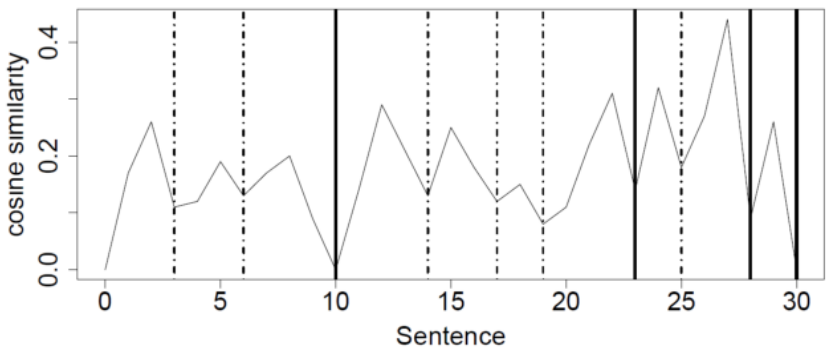
$(s_j)_{j=1}^{k_d}$ — последовательность предложений документа d

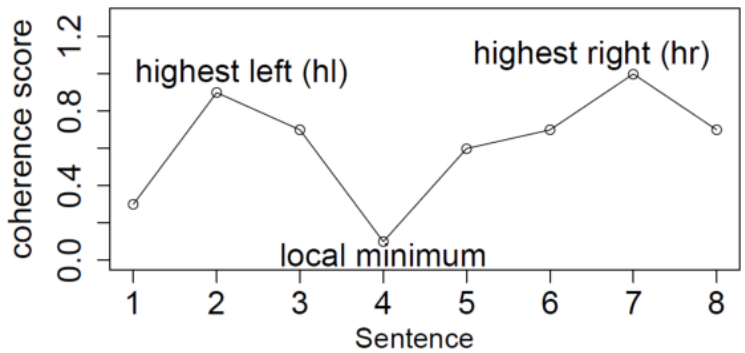
$v_s[t] = \frac{1}{|s|} \sum_{w \in s} v_w[t]$ — векторное представление предложения s

$v_w[t]$ — эмбеддинги слов (word2vec, тематические $p(t|d,w)$ и т.п.)

$c_j = \cos(v_{j-1}, v_j)$ — *coherence score*, оценка близости соседних предложений (чем глубже провал, тем чётче граница)

$d_j = \frac{1}{2}(\mathrm{hl}_j + \mathrm{hr}_j - 2c_j)$ — *depth score*, оценка глубины провала





Martin Riedl, Chris Biemann. Text Segmentation with Topic Models. 2012.

Эвристики для TopicTiling

Эвристики для определения числа сегментов:

- заданное число провалов с наибольшей глубиной d_j
- провалы с глубиной более $\text{avr}\{d_j\} + \delta\,\text{stdev}\{d_j\}$, $\delta = 0{,}5..1{,}2$

Дополнительные эвристики и параметры:

- filter: игнорировать короткие предложения (менее 5 слов)
- игнорировать стоп-слова
- подбирать число предложений слева и справа от j

Эвристики для тематической сегментации:

- использовать фоновые темы и игнорировать их в v_j
- использовать $p(t|d,w)$ или $\arg\max_t p(t|d,w)$
- подбирать число итераций
- подбирать параметры $|T|$, α, β в модели LDA

Martin Riedl, Chris Biemann. Text Segmentation with Topic Models. 2012.

Измерение качества сегментации

Базовые методы сегментации по векторам $p(w|s_j)$ и $p(t|s_j)$

- TT и TT-LDA — TextTiling (Hearst, 1997)
- C99 и C99-LDA — кластеризация предложений (Choi, 2000)

Коллекции для сравнения методов сегментации:

- *Choi dataset*: синтетический корпус, 700 документов по 10 сегментов, нарезанных из «Brown corpus»
- *Galley dataset*: синтетический корпус, 500 документов по 4–22 сегментов, нарезанных из «WSJ corpus»

Метрики для сравнения методов сегментации:

- Precision/Recall не учитывают границы между сегментами
- P_k (Beeferman et al., 1997)
- WD, WindowDiff (Pevzner and Hearst, 2002)

Martin Riedl, Chris Biemann. Text Segmentation with Topic Models. 2012.

Метрики для сравнения методов сегментации

Все метрики основаны на сравнении с идеальной сегментацией, т.н. «золотым стандартом» (gold standard).

- P_k (Beeferman et al., 1997) — чем меньше, тем лучше:
  $B_i = [\text{словопозиции } i \text{ и } i+k-1 \text{ лежат в одном сегменте}]$
  B_i^0 — то же самое для идеальной сегментации
  P_k — доля позиций, для которых $B_i \neq B_i^0$
- WD, WindowDiff (Pevzner and Hearst, 2002)
  $C_i = (\text{число сегментов между позициями } i \text{ и } i+k-1)$
  C_i^0 — то же самое для идеальной сегментации
  WD — доля позиций, для которых $C_i \neq C_i^0$

Doug Beeferman, Adam Berger, John Lafferty. Statistical models for text segmentation. 1999.

Lev Pevzner, Marti Hearst. A critique and improvement of an evaluation metric for text segmentation. 2002.

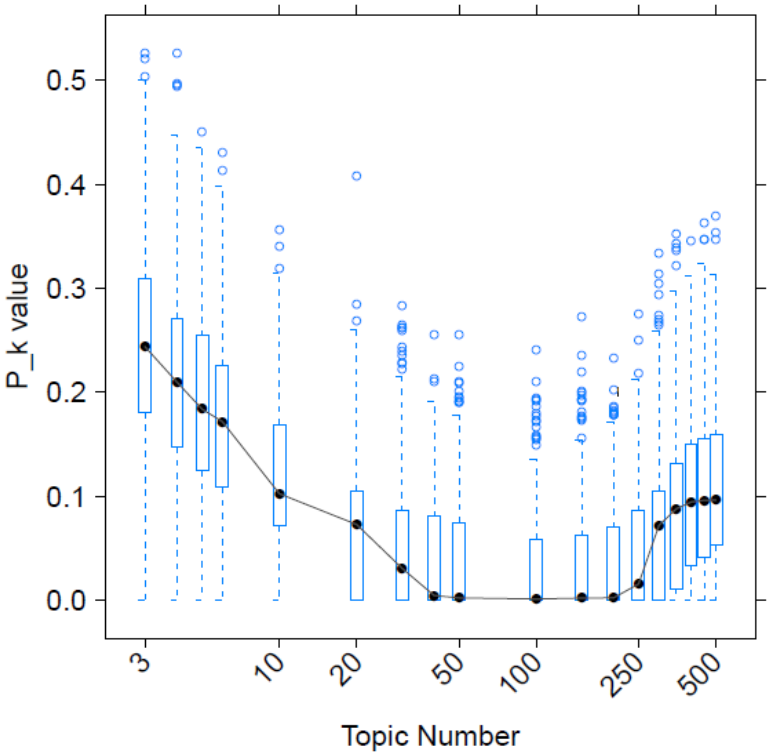
Результаты сравнения методов сегментации (Choi dataset)

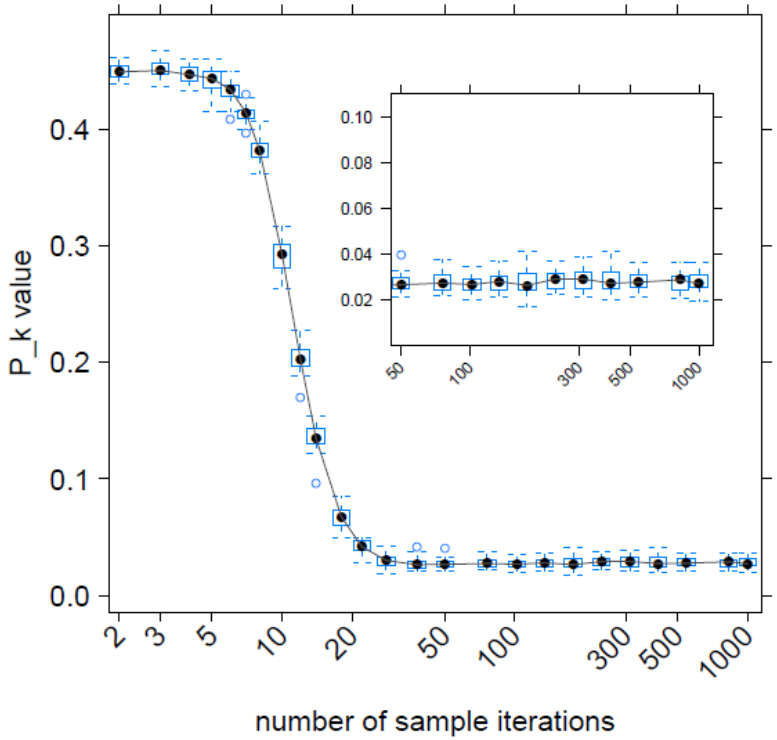
Method	Segments provided		Segments unprovided	
	P_k	WD	P_k	WD
C99	11.20	12.07	12.73	14.57
C99LDA	4.16	4.89	8.69	10.52
TT	44.48	47.11	49.51	66.16
TTLDA	1.85	2.10	16.41	21.40
TopicTiling	2.65	3.02	4.12	5.75
TopicTiling (filtered)	**1.50**	**1.72**	**3.24**	**4.58**

- Тематические модели лучше
- Лидирует TopicTiling с фильтрацией коротких предложений
- «Segments provided» — число сегментов известно (на реальных данных это нереалистичное предположение)

Martin Riedl, Chris Biemann. Text Segmentation with Topic Models. 2012.

Зависимости P_k ($k = 6$) от параметров модели

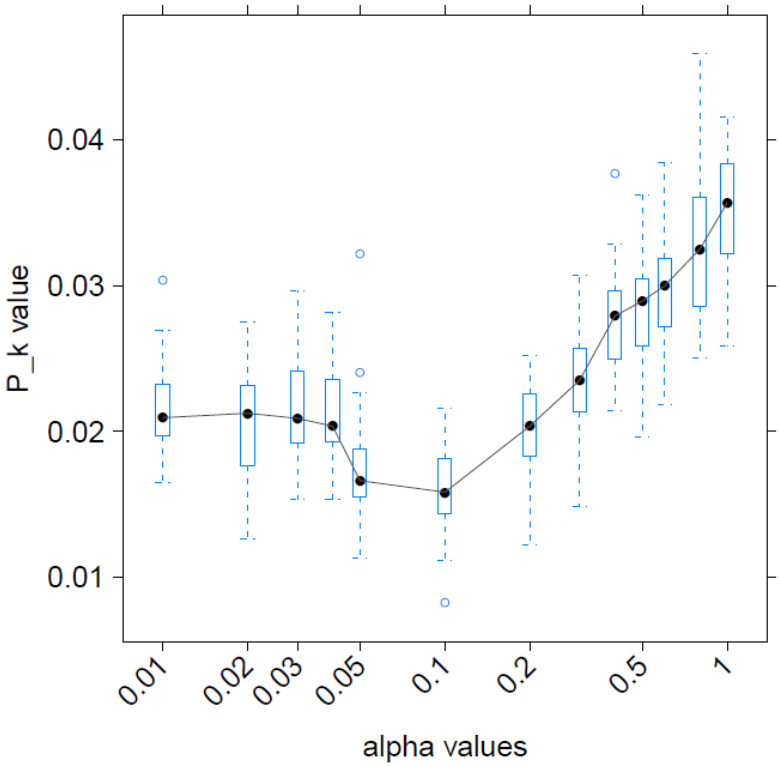


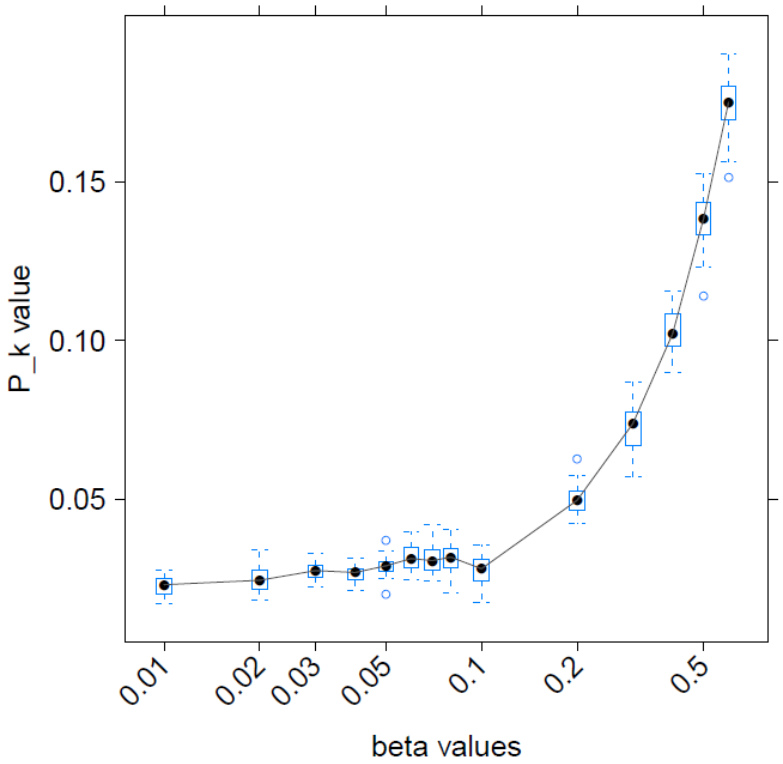


- Качество сегментации сильно зависит от $|T|$
- оптимальный диапазон $|T| = 50..150$ достаточно широк
- при $|T| = 100$ сходимость за 20–30 итераций

Martin Riedl, Chris Biemann. Text Segmentation with Topic Models. 2012.

Зависимости P_k ($k = 6$) от параметров α, β модели LDA

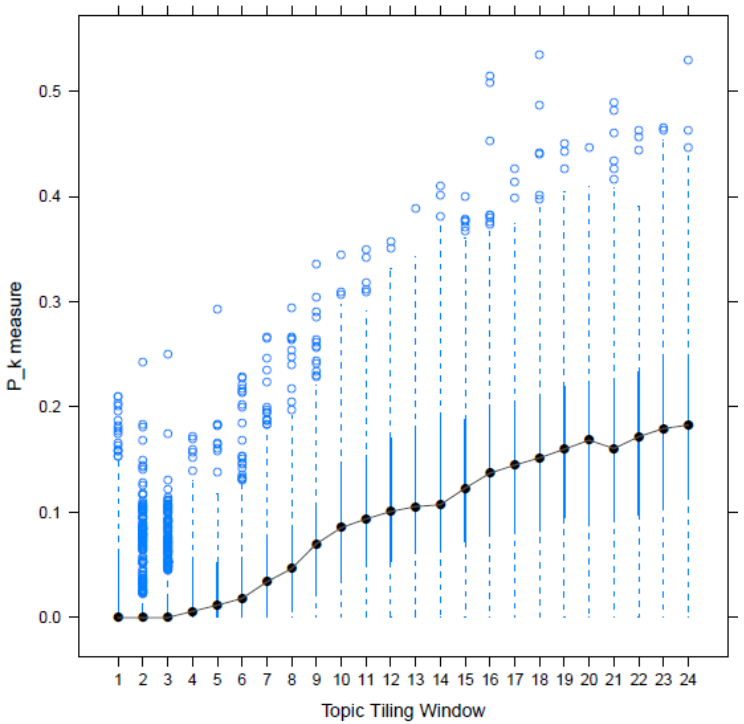




- Разреживать надо, но матрицу Θ — не слишком сильно
- параметры α, β менее критичны, чем число тем

Martin Riedl, Chris Biemann. Text Segmentation with Topic Models. 2012.

Зависимость P_k ($k = 6$) от ширины окна w (window)



фиксированное число сегментов:

seg. size	3-5		6-8		9-11		3-11	
	P_k	WD	P_k	WD	P_k	WD	P_k	WD
d=false,w=1	2.71	3.00	3.64	4.14	5.90	7.05	3.81	4.32
d=true,w=1	3.71	4.16	1.97	2.23	2.42	2.92	2.00	2.30
d=false,w=2	1.46	1.51	1.05	1.20	1.13	1.31	1.00	1.15
d=true,w=2	**1.24**	**1.27**	**0.76**	**0.85**	**0.56**	**0.71**	**0.95**	**1.08**
d=false,w=5	2.78	3.04	1.71	2.11	4.47	4.76	3.80	4.46
d=true,w=5	2.34	2.65	1.17	1.35	4.39	4.56	3.20	3.54

определяемое число сегментов:

seg. size	3-5		6-8		9-11		3-11	
	P_k	WD	P_k	WD	P_k	WD	P_k	WD
d=false,w=1	**2.39**	**2.45**	4.09	5.85	9.20	15.44	4.87	6.74
d=true,w=1	3.54	3.59	1.98	2.57	3.01	5.15	2.04	2.62
d=false,w=2	15.53	15.55	0.79	0.88	1.98	3.23	1.03	1.36
d=true,w=2	14.65	14.69	**0.62**	**0.62**	**0.67**	**0.88**	**0.66**	**0.78**
d=false,w=5	21.47	21.62	16.30	16.30	6.01	6.14	14.31	14.65
d=true,w=5	21.57	21.67	17.24	17.24	6.44	6.44	15.51	15.74

- Оптимальная ширина окна $w = 2$–3 предложения
- «d=true»: усреднение $\arg\max_t p(t|d, w)$ по каждому w
- Почему они не догадались использовать $p(t|d, w)$?

Martin Riedl, Chris Biemann. Text Segmentation with Topic Models. 2012.

Эксперименты на более реалистичных данных Galley's WSJ

фиксированное число сегментов:

Parameters	All words		Filtered	
	P_k	WD	P_k	WD
d=false,w=1	37.31	43.20	37.01	43.26
d=true,w=1	35.31	41.27	33.52	39.86
d=false,w=2	22.76	28.69	21.35	27.28
d=true,w=2	21.79	27.35	19.75	25.42
d=false,w=5	14.29	19.89	12.90	18.87
d=true,w=5	**13.59**	**19.61**	**11.89**	**17.41**
d=false,w=10	14.08	22.60	14.09	22.22
d=true,w=10	13.61	21.00	13.48	20.59

определяемое число сегментов:

Parameters	All words		Filtered	
	P_k	WD	P_k	WD
d=false,w=1	53.07	72.78	52.63	72.66
d=true,w=1	53.42	74.12	51.84	72.57
d=false,w=2	46.68	65.01	44.81	63.09
d=true,w=2	46.08	64.41	43.54	61.18
d=false,w=5	30.68	43.73	28.31	40.36
d=true,w=5	28.29	38.90	26.96	36.98
d=false,w=10	19.93	32.98	18.29	29.29
d=true,w=10	**17.50**	**26.36**	**16.32**	**24.75**

- Качество сегментации сильно зависит от коллекции
- Определять число сегментов стало труднее
- Окно пришлось расширить до $w = 5$–10 предложений
- Здесь «filtered» — учитывать только существительные, прилагательные и глаголы — помогает, но не сильно

Martin Riedl, Chris Biemann. Text Segmentation with Topic Models. 2012.

Задача суммаризации (аннотирования, реферирования) текста

Автоматическая суммаризация — *краткий* текст, построенный по одному или нескольким документам и *наиболее полно* передающий их содержание.

Полуавтоматическая — HAMS, human aided machine summarization

Основные типы задач суммаризации:

- *one-document* — на входе один документ $d \in D$
- *multi-document* — на входе набор документов $D' \subseteq D$
- ⊕ *topic* — на входе набор сегментов темы $p(d, s|t)$

Основные подходы к суммаризации:

- *extractive* — выбор некоторых предложений целиком
- *abstractive* — генерация текста на естественном языке

H.P.Luhn. The automatic creation of literature abstracts. 1958.
Juan-Manuel Torres-Moreno. Automatic Text Summarization. 2014.

Основные этапы выборочной (extractive) суммаризации

1. Внутреннее представление текста
   - граф / кластеризация / тематизация предложений в тексте
   - вычисление важности и других признаков предложений
2. Оценивание полезности (ранжирование) предложений
3. Отбор предложений для реферата
   - оптимизация критериев информативности и различности
   - оптимизация последовательности предложений
   - учёт целей и особенностей прикладной задачи (новости/статьи/веб-страницы/посты/мэйлы)

D.Das, A.Martins. A survey on automatic text summarization. 2007.

A.Nenkova, K.McKeown. A survey of text summarization techniques. 2012.

Yogita Desai, Prakash Rokade. Multi Document Summarization: Approaches and Future Scope. 2015.

Mahak Gambhir, Vishal Gupta. Recent automatic text summarization techniques: a survey. 2016.

TextRank — аналог ссылочного ранжирования PageRank

Текст — граф предложений. Предложение $s \in S$ тем важнее,
— чем больше других предложений c, похожих на s,
— чем важнее предложения c, похожие на s,
— чем меньше других предложений, на которые c также похоже.

Вероятность попасть в s, случайно блуждая по похожим:

$$\mathrm{PR}(s) = (1-\delta) + \delta \sum_{c \in S_s^{in}} \frac{\mathrm{PR}(c)}{|S_c^{out}|},$$

$S_s^{in} \subset S$ — множество предложений c, похожих на s,
$S_c^{out} \subset S$ — множество предложений, на которые похоже c,
$\delta = 0.85$ — вероятность продолжать блуждания (damping factor)

Sergey Brin, Lawrence Page. The Anatomy of a Large-Scale Hypertextual Web Search Engine. 1998.

Rada Mihalcea, Paul Tarau. TextRank: Bringing Order into Text. EMNLP-2004.

Определение сходства предложений

- Доля общих слов в двух предложениях
- Доля общих слов, за исключением слов общей лексики
- Доля общих n-грамм в двух предложениях
- Сходство векторных представлений двух предложений
- Сходство тематических распределений двух предложений

Другое применение TextRank — *извлечение ключевых слов* (keyword extraction).

В этом случае близость между совами (n-граммами) определяется по частоте их сочетаемости в окне ширины h

Покрытие терминологии и тематики документа

S_d — множество предложений документа d
$a \subset S_d$ — искомая суммаризация

Покрытие терминологии документа (lexicon coverage):

$$\text{WCov}(a) = \text{KL}\big(p(w|d) \| p(w|a)\big) \to \min_{a \subset S_d}$$

Покрытие тематики документа (topic coverage):

$$\text{TCov}(a) = \text{KL}\big(p(t|d) \| p(t|a)\big) \to \min_{a \subset S_d}$$

Избыточность суммаризации (redundancy):

$$\text{Red}(a) = \sum_{s,s' \in a} B_{ss'} \to \min_{a \subset S_d}, \qquad B_{ss'} = \text{sim}\big(p(w|s), p(w|s')\big),$$

где sim — одна из мер сходства: cos, JS, Jaccard и т.п.

Marina Litvak, Natalia Vanetik, Chunlei Liu, Lemin Xiao, Onur Savas. Improving Summarization Quality with Topic Modeling. 2015.

Задача многокритериальной дискретной оптимизации

Метод релаксации: вместо $a \subset S_d$ ищем $\pi_s = p(s|a)$, где $s \in S_d$.
В релаксированной задаче:

$$p(w|a) = \sum_{s \in d} p(w|s)p(s|a) = \sum_{s \in d} \frac{n_{ws}}{n_s} \pi_s$$

$$p(t|a) = \sum_{s \in d} p(t|s)p(s|a) = \sum_{s \in d} \theta_{ts} \pi_s$$

Сумма трёх критериев $\mathrm{WCov}(a) + \tau_1 \mathrm{TCov}(a) + \tau_2 \mathrm{Red}(a)$:

$$\sum_{w \in d} n_{dw} \ln \sum_{s \in d} \frac{n_{ws}}{n_s} \pi_s + \tau_1 \sum_{t \in T} \theta_{td} \ln \sum_{s \in d} \theta_{ts} \pi_s - \tau_2 \sum_{s,s' \in d} B_{ss'} \pi_s \pi_{s'} \to \max_{\{\pi\}}$$

Максимизация покрытия — это максимизация правдоподобия!

Можно добавить регуляризатор разреживания:

$$R(\pi) = -\tau_3 \sum_{s \in S_d} \ln \pi_s \to \max_{\{\pi\}}$$

Оценка полезности предложений

Дополнительные признаки для отбора предложений:

- *SumBasic* — средняя частота слов, исключая стоп-слова
- *Centriod* — средний TF-IDF слов, превышающий порог
- *LexicalChain* — число слов сильных лексических цепочек
- *ImpactBased* — число слов из ссылающихся контекстов
- *TopicBased* — число слов из запроса пользователя

Стратегии отбора предложений:

- по одному top-предложению от каждой из top-тем
- поощрять выбор соседних предложений
- штрафовать предложения с анафорой и эллипсисом

A.Nenkova, K.McKeown. A survey of text summarization techniques. 2012.

Тематическая модель предложений для суммаризации

S_d — множество предложений документа d;
n_{sw} — частота терма w в предложении s;
n_s — длина предложения s.

Отбор предложений для суммаризации: $p(s|t) \to \max\limits_{s \in S_d}$

Тематическая модель сегментированного текста:

$$p(w|d) = \sum_{s \in S_d} p(w|s) \sum_{t \in T} p(s|t)p(t|d) = \sum_{s \in S_d} p_{ws} \sum_{t \in T} \psi_{st}\theta_{td}$$

где $p_{ws} \equiv p(w|s) = \frac{n_{ws}}{n_s}$ — частота терма w в предложении s.

Вместо ϕ_{wt} нельзя взять $p(w|t) = \sum\limits_{d \in D} \sum\limits_{s \in S_d} p_{ws}\psi_{st}$. Почему?

Dingding Wang, Shenghuo Zhu, Tao Li, Yihong Gong. Multi-document summarization using sentence-based topic models // ACL-IJCNLP 2009.

BSTM — Bayesian Sentence-based Topic Models

Критерий максимума регуляризованного правдоподобия:

$$\sum_{d\in D}\sum_{w\in d} n_{dw} \ln \sum_{s\in S_d} p_{ws} \sum_{t\in T} \psi_{st}\theta_{td} \to \max_{\Phi,\Theta}$$

- Авторы утверждают, что модель переходит в обычную $p(w|d) = \sum_t \phi_{wt}\theta_{td}$, если предложение $\equiv$ слово
- Это не так, т. к. предложения уникальны: $S_d \cap S_{d'} = \emptyset$
- Модель разваливается на независимые модели документов (Litvak, 2015) такую LDA строят явно, это тоже работает!
- Но это не будет работать для multi-document summarization!
- А то, что модель «Bayesian», вообще не имеет значения ;)

Dingding Wang, Shenghuo Zhu, Tao Li, Yihong Gong. Multi-document summarization using sentence-based topic models // ACL-IJCNLP 2009.

Идея обобщения для много-документной суммаризации

Критерий максимума регуляризованного правдоподобия:

$$\sum_{d,w} n_{dw} \ln \sum_{t\in T} \phi_{wt}\theta_{td} + \tau \sum_{d,w} n_{dw} \ln \sum_{s\in S_d} p_{ws} \sum_{t\in T} \psi_{st}\theta_{td} + R \to \max_{\Phi,\Psi,\Theta}$$

EM-алгоритм: метод простой итерации для системы уравнений

$$\text{E-шаг:} \quad \begin{cases} p_{tdw} \equiv p(t|d,w) = \underset{t\in T}{\text{norm}}\left(\phi_{wt}\theta_{td}\right) \\ p_{stdw} \equiv p(s,t|d,w) = \underset{s,t\in S_d\times T}{\text{norm}}\left(p_{ws}\psi_{st}\theta_{td}\right) \end{cases}$$

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

ROUGE: Recall-Oriented Understudy for Gisting Evaluation

$r \in R$ — множество рефератов, написанных людьми
s — суммаризация, построенная системой
Чем больше, тем лучше — для всех метрик семейства ROUGE

Доля n-грамм из рефератов, вошедших в суммаризацию s:

$$\text{ROUGE-}n(s) = \frac{\sum\limits_{r \in R} \sum\limits_{w} [w \in s][w \in r]}{\sum\limits_{r \in R} \sum\limits_{w} [w \in r]}$$

Доля n-грамм из самого близкого реферата, вошедших в s:

$$\text{ROUGE-}n_{\text{multi}}(s) = \max_{r \in R} \frac{\sum\limits_{w} [w \in s][w \in r]}{\sum\limits_{w} [w \in r]}$$

Chin-Yew Lin. ROUGE: A package for automatic evaluation of summaries. 2004.

ROUGE: Recall-Oriented Understudy for Gisting Evaluation

$r \in R$ — множество рефератов, написанных людьми
s — суммаризация, построенная системой
Чем больше, тем лучше — для всех метрик семейства ROUGE

ROUGE-L(s) максимальная общая подпоследовательность s, r
ROUGE-W(s) штрафует за пропуски в подпоследовательности
ROUGE-S(s) аналог ROUGE-2(s) для биграмм с пропусками
ROUGE-SU-m(s) для биграмм с пропусками не длиннее m

$\mathrm{JS}\big(p(w|s), p(w|R)\big)$ — лучше всего коррелирует с экспертными оценками качества суммаризации (Lin, 2006).

Готовые пакеты для вычисления метрик: pyRouge и др.

Chin-Yew Lin. ROUGE: A package for automatic evaluation of summaries. 2004.

Chin-Yew Lin, Guihong Cao, Jianfeng Gao, Jian-Yun Nie.
An Information-Theoretic Approach to Automatic Evaluation of Summaries. 2006.

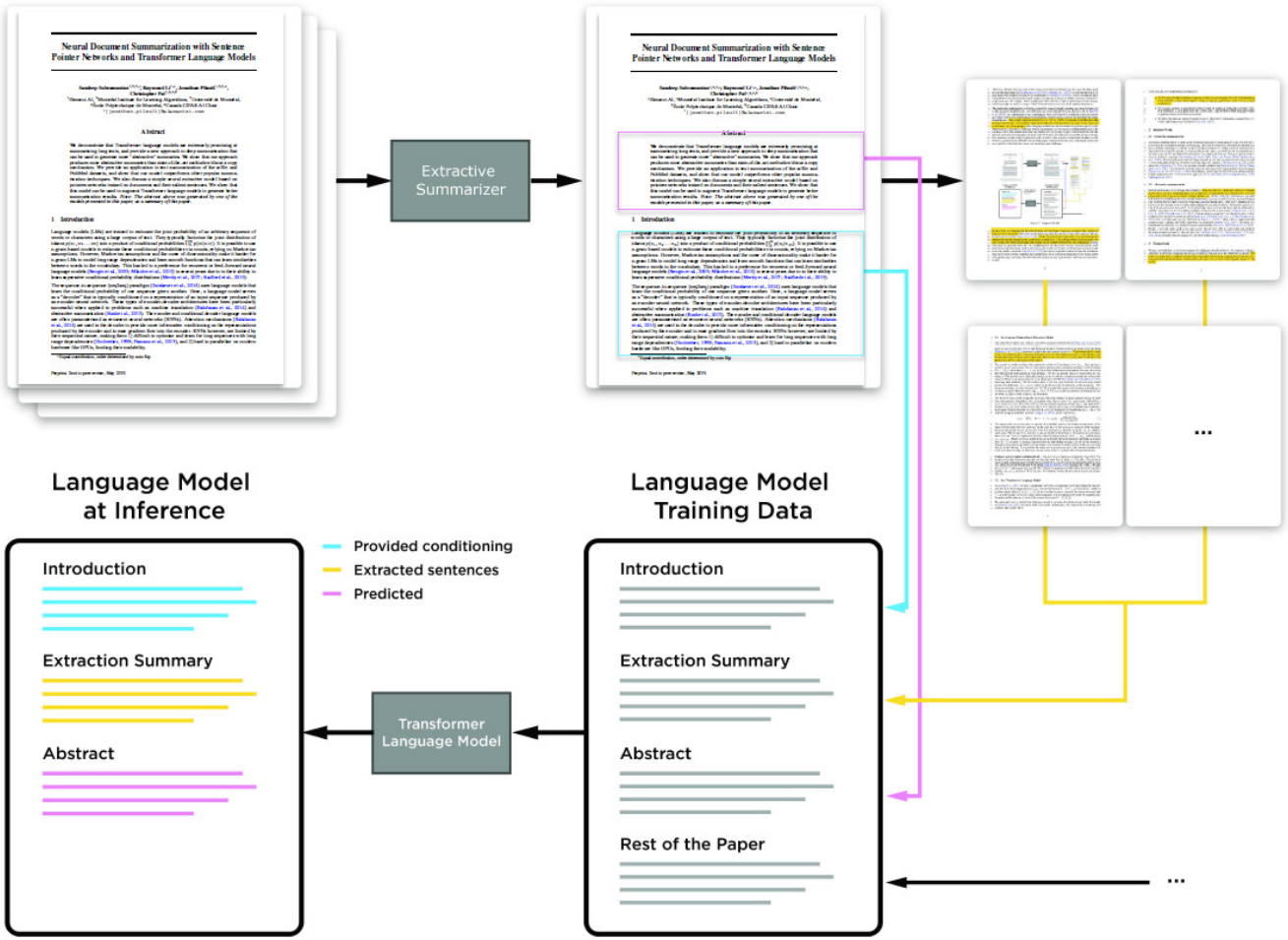
Суммаризация на основе трансформеров

Abstract

We present a method to produce abstractive summaries of long documents that exceed several thousand words via neural abstractive summarization. We perform a simple extractive step before generating a summary, which is then used to condition the transformer language model on relevant information before being tasked with generating a summary. We show that this extractive step significantly improves summarization results. We also show that this approach produces more abstractive summaries compared to prior work that employs a copy mechanism while still achieving higher rouge scores. *Note: The abstract above was not written by the authors, it was generated by one of the models presented in this paper.*

S.Subramanian, R.Li, J.Pilault, C.Pal. On Extractive and Abstractive Neural Document Summarization with Transformer Language Models. 2019.

Суммаризация на основе трансформеров



S.Subramanian, R.Li, J.Pilault, C.Pal. On Extractive and Abstractive Neural Document Summarization with Transformer Language Models. 2019.

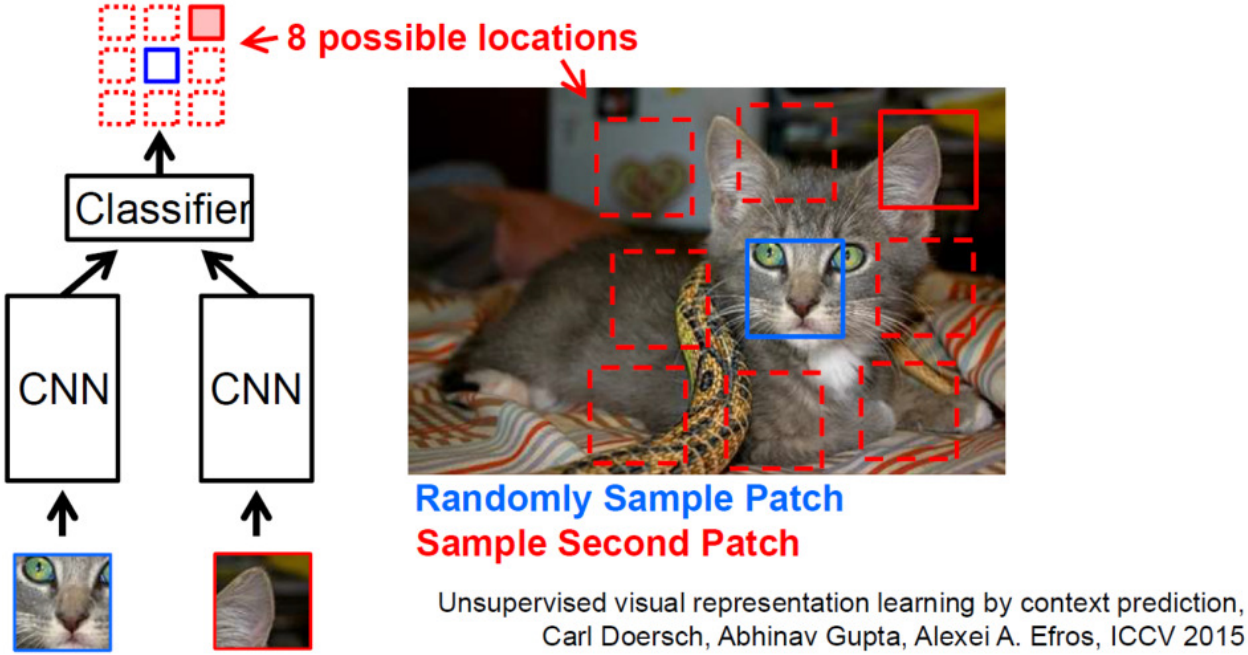
Сравнение с эталонными методами суммаризации

Model	Type	ROUGE 1	2	3	L
Previous Work					
SumBasic	Ext	29.47	6.95	2.36	26.3
LexRank	Ext	33.85	10.73	4.54	28.99
LSA	Ext	29.91	7.42	3.12	25.67
Seq2Seq	Abs	29.3	6.00	1.77	25.56
Pointer-gen	Mix	32.06	9.04	2.15	25.16
Discourse	Mix	35.80	11.05	3.62	**31.80**
Our Models					
Lead-10	Ext	35.52	10.33	3.74	31.44
Sent-CLF	Ext	34.01	8.71	2.99	30.41
Sent-PTR	Ext	42.32	15.63	7.49	38.06
TLM-I	Abs	39.80	12.20	4.42	22.36
TLM-I+E (M,M)	Mix	41.59	14.26	5.94	23.55
TLM-I+E (G,M)	Mix	**42.43**	**15.24**	**6.68**	24.08
Oracle					
Gold Ext	Orac	44.25	18.17	9.14	35.33
TLM-I+E (G,G)	Orac	46.52	18.19	8.73	26.88

S.Subramanian, R.Li, J.Pilault, C.Pal. On Extractive and Abstractive Neural Document Summarization with Transformer Language Models. 2019.

Концепция самообучения (self-supervised)

Сеть обучается предсказывать взаимное расположение двух фрагментов на одном изображении



Unsupervised visual representation learning by context prediction,
Carl Doersch, Abhinav Gupta, Alexei A. Efros, ICCV 2015

Преимущество: не нужна размеченная обучающая выборка, при этом сеть способна выучить векторные представления.

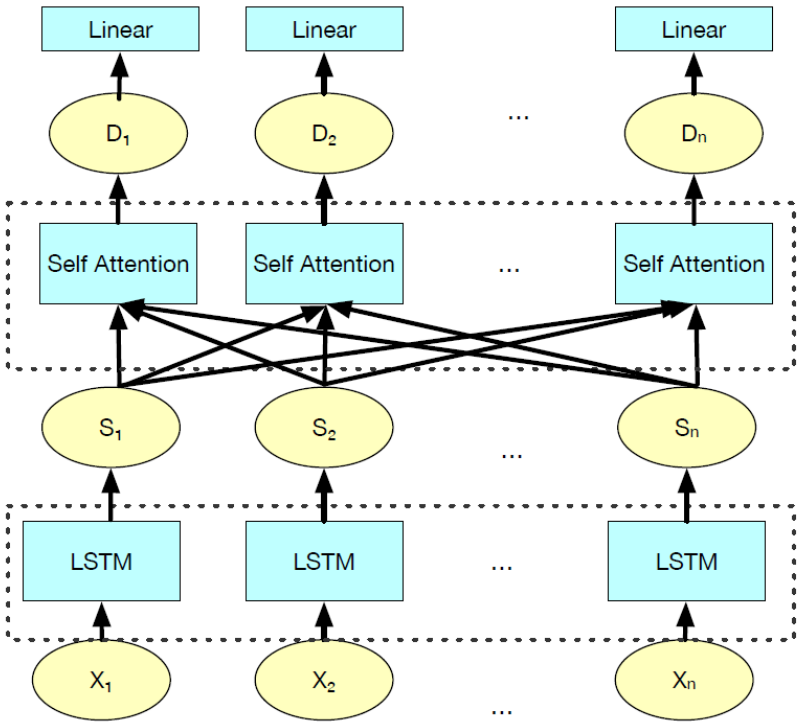
Базовая модель бинарной классификации предложений

Классы: 1 — включить в реферат, 0 — не включать

D_i — контекстные эмбеддинги предложений

S_i — вектор признаков предложения для классификации

X_i — эмбеддинг i-го предложения



Hong Wang, Xin Wang, Wenhan Xiong, Mo Yu, Xiaoxiao Guoz et al. Self-Supervised Learning for Contextualized Extractive Summarization. 2019.

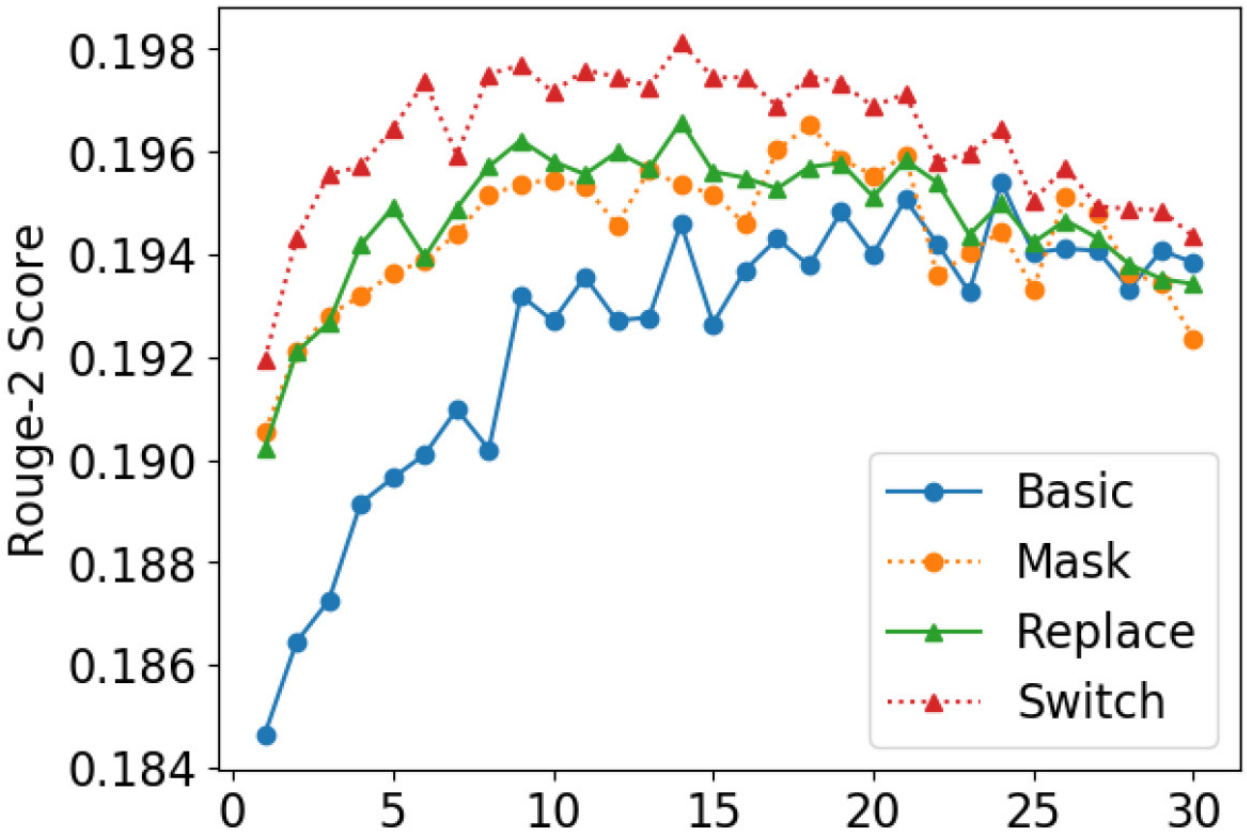
Три способа сгенерировать данные для self-supervised

- **Mask**
  — с вероятностью $P_m = 0.25$ пропускать предложение
  — предсказывать предложение из пула пропущенных T_m
- **Replace**
  — с вероятностью $P_r = 0.25$ заменять предложение случайным предложением из *другого документа*
  — предсказывать, было ли предложение заменено
- **Switch**
  — с вероятностью $P_s = 0.25$ заменять предложение случайным предложением из *данного документа*
  — предсказывать, было ли предложение заменено

Hong Wang, Xin Wang, Wenhan Xiong, Mo Yu, Xiaoxiao Guoz et al. Self-Supervised Learning for Contextualized Extractive Summarization. 2019.

Сравнение моделей суммаризации по метрике ROUGE

Зависимость ROUGE от числа итераций



Hong Wang, Xin Wang, Wenhan Xiong, Mo Yu, Xiaoxiao Guoz et al. Self-Supervised Learning for Contextualized Extractive Summarization. 2019.

Сравнение моделей суммаризации по метрике ROUGE

Method	Rouge-1	Rouge-2	Rouge-L
Basic	41.07	18.95	37.56
LEAD3	39.93	17.62	36.21
NEUSUM	41.18*	18.84	37.61
Mask	41.15*	19.06*	37.65*
Replace	41.21*	19.08*	37.73*
Switch	**41.36**	**19.20**	**37.86**
SentEnc	41.17*	19.04*	37.69*
Switch 0.15	41.35*	19.18*	37.85*
Switch 0.35	41.27*	19.12*	37.77*

Basic, Lead3, NeuSum — эталонные модели

SentEnc — случайная инициализация уровня self-attention

Hong Wang, Xin Wang, Wenhan Xiong, Mo Yu, Xiaoxiao Guoz et al.
Self-Supervised Learning for Contextualized Extractive Summarization. 2019.

Резюме

- Сегментация похожа на задачу разладки временного ряда
- Для сегментации можно взять любой критерий неоднородности текстов в двух последовательных окнах
- Суммаризация — некорректно поставленная задача, может иметь очень много разнообразных хороших решений
- Не существует идеального критерия качества суммаризации
- Абстрактивная (abstractive) суммаризация является открытой проблемой, сложной даже для нейронных сетей