Computational Linguistics and Intellectual Technologies:
Proceedings of the International Conference “Dialogue 2018”

Moscow, May 30—June 2, 2018

INTRA-TEXT COHERENCE AS A MEASURE OF TOPIC MODELS’ INTERPRETABILITY

Alekseev V. A. (wasya.alekseev@gmail.com),
Bulatov V. G. (bt.uytya@gmail.com),
Vorontsov K. V. (vokov@forecsys.ru)

Moscow Institute of Physics and Technology (State University)

The article is devoted to the problem of how to automatically measure the interpretability of topic models. Some new, intra-text, approaches to estimate the interpretability of the topics are proposed. Computational experiments are conducted with the use of text files from “PostNauka”, which is a collection of popular science content.

Keywords: topic modeling, topic coherence, topic interpretability, text segmentation, topic model, PLSA, LDA, BigARTM, text analysis, machine learning

ВНУТРИТЕКСТОВАЯ КОГЕРЕНТНОСТЬ КАК МЕРА ИНТЕРПРЕТИРУЕМОСТИ ТЕМАТИЧЕСКИХ МОДЕЛЕЙ ТЕКСТОВЫХ КОЛЛЕКЦИЙ

Алексеев В. А. (wasya.alekseev@gmail.com),
Булатов В. Г. (bt.uytya@gmail.com),
Воронцов К. В. (vokov@forecsys.ru)

Московский Физико-Технический Институт

Статья посвящена задаче измерения интерпретируемости и когерентности тематических моделей. Предлагается новый, внутритекстовый, подход к оценке меры согласованности темы. Вычислительные эксперименты проводятся на коллекции научно-популярного контента «ПостНаука».

Ключевые слова: тематическое моделирование, интерпретируемость, когерентность, сегментация, тематическая модель, PLSA, LDA, BigARTM, анализ текстов, машинное обучение

1. Introduction

Topic modeling is a text analysis method which aims to discover hidden thematic structure in large collections of texts. Topic models are used in information retrieval [10], documents' categorization [12], social networks' data analysis, [16, 15], recommendation systems [10, 7], exploratory search [5] and other areas. After the processing of documents' collection, a topic model gives a set of topics covered in the documents, the distribution of these topics in the documents, and words that characterize each topic [11].

The interpretability is a desirable property of a good topic model [19]. A topic is said to be well interpreted, if it corresponds to real-world concept of interest. However, the topics derived by topic models may not be clear and understandable, they may include words from different weakly related areas. [8]

Recently, an automated procedure estimating the interpretability was introduced. This method evaluates the list of the most frequent topic words and favorably compares to the human experts' judgements of the same list.

However, we believe that this approach suffers from several fundamental limitations. We argue that these limitations bring into question the common practice of treating coherence and interpretability as equivalent.

The aim of this paper is twofold. The first is to outline a class of issues inherent in a traditional notions of coherence. A key problem with this approach is that reducing the topic model to a short list of words loses too much precision. Previous studies linking coherence and interpretability failed to take this into account.

However, the proportion of text covered by these top frequent words is not controlled in any way. We show that in practice this proportion is too small to justify treating coherence and interpretability as equivalent.

The second purpose is to demonstrate the feasibility of the alternative approach which we call the *intra-text coherence*, defined as an average thematic similarity of terms, closely located in the text. To justify this new measure, we will adapt the procedure used in [3], [8] and [14].

2. Related work

For the topic modeling purposes, the *topic* is defined as a probability distribution over words. For example, the topic named "theatre" could be a probability distribution concen-trated on a words such as "actor", "play", "premiere", "parterre" and "spectator" (on the contrary, the probability of words such as "loan" and "vertebrates" would be extremely low or even zero).

The topic model can be described by two distributions: $\phi_{wt}=(w|t)$, the probability to draw a word w from the topic t and $\theta_{td}=p(t|d)$, the probability to find a topic t within the document d.

Early work on topic modeling conceptualized it as an intermediate stage of information retrieval pipeline. The possibility of meaningful interpretation was an afterthought. For measuring the quality of topics when evaluated against human judgments, several metrics were proposed.

Currently, there is a consensus among researchers that the evaluation of human interpretability should conform to the following framework:

1) Picking some small set of words for each topic (typically, a list of ten most frequent words, but the more sophisticated approaches are possible [2]). The term *top tokens* has come to be used to refer to this set.

2) a. Presenting this set to a human expert to obtain a human judgment of a set quality.

or

b. Gathering an array of co-occurrence statistics associated with members of this set and performing a series of calculations involving these numbers.

This framework was introduced in seminal works of Blei [14, 3] and Mimno [8] and then greatly developed by the topic modeling community. We will call this extensive category of metrics *top-tokens based*.

The main attraction of top-token based measures is their simplicity. Instead of evaluating the whole probability distribution, the researcher only has to look at the short list of the most "representative" words.

However, their inherent limitation is deeply rooted in the same thing. The list of top five to ten words reflects only part of the whole probability distribution, and poorly (if at all) characterizes how good topic model does represent the particular corpus.

We argue that the list of the most frequent words is inadequate in justifying the quality of topic model regardless of the method of its analysis. This applies equally to the human experts' ratings and the automated procedures based on the word co-occurrence counts.

3. Towards a better interpretability metric

As was previously noted, traditional coherence metrics consist of two steps: first, they use information from $(w|t)$ distribution; secondly, they retrieve the co-occurrence statistics.

The idea behind automated coherence measures is to find out how often do certain words appear together within the sliding context window and compare that number to the frequency predicted by pure coincidence. The topic is said to be coherent if the positions of its words tend to cluster, do not appear to be random.

This is reminiscent of the linguistic phenomenon of textual cohesion [1]: the sentences of natural language texts are connected to each other via syntactic and lexical devices such as word repetition, synonyms/near-synonyms, hyponyms and so on.

We conjecture that the natural language texts are divided into coherent spans which contain only small number of latent topics. According to this assumption, the purpose of topic modeling should be understood as an adequate segmentation of the initial text into thematically homogeneous fragments consisting of a handful of topics.

Note that frequent top-words co-occurrences is an indirect sign that the topic is represented in the text collection as a coherent text fragment.

Therefore, we argue that interpretability of a topic should be evaluated not only by the consistency of top-words use, but also by the consistency of all topic words use within text segments. We could obtain an automated measure of the model interpretability by examining the degree the topic model violates this consistency.

Instead of drawing inferences about the whole topic based on behaviour of the short list of ten most frequent words, one should start by examining words appearing together in a text and then proceed by comparing their $(t|w,d)$.

This procedure will be dealt with in more detail in the following section.

4. Coherences

In this paper, we present several automatic measures distinct from traditional *top-token based* approaches.

The first method—SemantiC (Semantic Closeness)—estimates semantic proximity of closely located in the text words as vectors with components $(t|w)$. To estimate the proximity between words one can calculate l2 distance between the corresponding vectors

$$SemantiC_{L2}|_t = -\langle [\rho(\boldsymbol{w}_i, \boldsymbol{w}_j) \le window] \|\boldsymbol{w}_i - \boldsymbol{w}_j\|_2 \rangle$$

where $\rho(\boldsymbol{w}_i, \boldsymbol{w}_j)$—text-distance between words (number of other words between them), window—window of words, in which $\boldsymbol{w}_i$ and $\boldsymbol{w}_j$ are considered to be close in text-distance. Minus sign makes coherence higher if words' vectors are close. In addition to the Euclidean distance, Cosine Similarity measure can be used:

$$SemantiC_{Cos}|_t = +\langle [\rho(\boldsymbol{w}_i, \boldsymbol{w}_j) \le window] cos(\boldsymbol{w}_i, \boldsymbol{w}_j) \rangle$$

The third proposed way to estimate semantic closeness by topic is to calculate variance between components corresponding to this topic:

$$SemantiC_{Var}|_t = Varianc(\boldsymbol{w}_i(t), \boldsymbol{w}_{i+1}(t), \ldots, \boldsymbol{w}_{i+window}+(t))$$

Before computing, all vectors were multiplied by 1000, so as to increase the result value for the coherence.

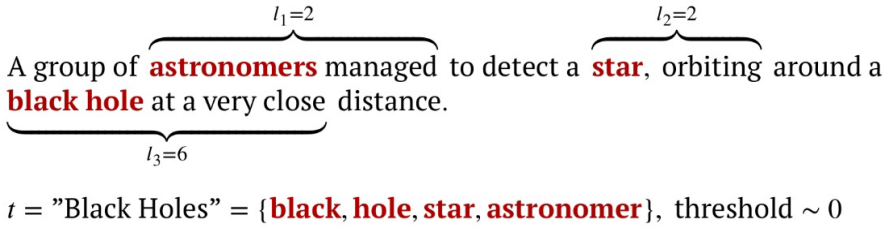

Figure 1. An example illustrating the idea of TopLen coherence

As long as words of a topic under interest are observed, they are counted. If some unrelated word is encountered it is also counted but gives a negative penalty. When the absolute value of total penalty appears to be quite big, the process stops, and the number of counted words gives one value of topic length.

Another method—TopLen (Topic Length)—calculates the average duration of the topic in text. The auxiliary function score (w_j,t) returns the difference between the component of the vector corresponding to the topic and the maximal component among the other topics. Non-negative parameter threshold smooths the effect when TopLen encounter words not from the topic while counting topic length, the process of counting continues as long as threshold (chosen to be 0.01) plus sum of scores is non-negative (see Figure 1 for an example).

The last proposed method—FoCon (Focus Consistency)—evaluates how much differ adjacent words throughout the whole text, summing the pairs of differences between corresponding components of $(t|w)$ vectors (components, by means of which the differences are calculated, are the maximal components of the adjacent words vectors). Minus sign serves the same role as in case of SemantiC—coherence rises when words differ less.

$$FoCon|_t = -\sum_{d\in D}\sum_{\substack{w_i,w_j\in W_d\\ j-i=1}} |w_i[t_1]-w_j[t_1]| + |w_i[t_2]-w_j[t_2]|$$

5. Experiments

5.1. Interpretation and representation

Automated coherence measures rest on the word co-occurrence counts. If top tokens often appear together within the context window, this set of words is said to be *coherent*, i.e. these words fit together in a natural or reasonable way.

It is implicitly assumed that if set of top tokens is coherent, then the whole topic is coherent as well. Such arguments were criticized before [21], but we wish to understand the issue quantitatively. What fraction of collection is represented in the co-occurrence counts related to the given top token set?

Let Q be a set of words. We will call the position of word $w \in Q$ *represented* if it has a non-zero contribution to the Q co-occurrence counts (see Figure 2). We will measure the *representational frequency* of two topic models.

Our primary dataset is a corpus consisting of articles published in "PostNauka", a popular Russian online magazine about science. We investigate a topic model consisting of 19 subject-related topics and a single background topic (see Figure 3).

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 ньютоновской 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

Figure 2: Words used to calculate coherence. We see a single top token ("частиц") and a wide range of weakly topical words, which are ignored while calculating coherence by the traditional methods.

Topic	First Top-Word	Second Top-Word	Third Top-Word
1: математика	математика (0.016)	задача (0.008)	декарт (0.008)
2: технологии	технология (0.015)	робот (0.012)	сеть (0.010)
3: физика	частица (0.027)	электрон (0.015)	кварк (0.015)
4: химия	химия (0.021)	молекула (0.019)	материал (0.016)
5: земля	земля (0.029)	планета (0.028)	атмосфера (0.012)
6: астрономия	звезда (0.039)	галактика (0.031)	вселенная (0.019)
7: биология	клетка (0.027)	организм (0.011)	мозг (0.010)
8: медицина	пациент (0.016)	препарат (0.012)	заболевание (0.012)
9: психология	психология (0.009)	мозг (0.009)	психолог (0.008)
10: экономика	экономика (0.016)	страна (0.010)	цена (0.008)
11: история	история (0.010)	историк (0.007)	власть (0.006)
12: политика	государство (0.014)	политика (0.012)	политический (0.011)
13: социология	социология (0.013)	социолог (0.009)	социальный (0.008)
14: культура	культура (0.015)	фильм (0.007)	искусство (0.006)
15: образование	университет (0.021)	образование (0.014)	школа (0.013)
16: язык	язык (0.077)	слово (0.037)	словарь (0.011)
17: философия	философия (0.018)	философ (0.013)	философский (0.008)
18: религия	святилище (0.010)	религия (0.007)	царь (0.006)
19: россия	россия (0.028)	страна (0.009)	русский (0.009)

Figure 3: PostNauka's topics, each represented by its 3 top-words

Next, we will focus on the topic model presented in [3], which uses a sample of Wikipedia articles. This model was identified as best based on assessment of top 10 tokens by human experts. This model consists of 50 topics.

As can be seen from the table 1, top-tokens cover a vanishing fraction of corpus. Informally speaking, top token-based measures ignore more than 98% of the collection!

Table 1: The proportion of corpus contributing to the co-occurrence counts of top 10 most frequent words for each topic

	PostNauka	Wikipedia
Minimum	0.000159	0.000065
Median	0.000483	0.000293
Mean	0.000619	0.000356
Maximum	0.002764	0.001149
Total	0.012027	0.016585

5.2. Ground truth

The evaluation of interpretability is extremely labor-intensive. The strength of top token-based measures is their ability to reduce topics of the topic model to the accessible list of words. Even then, gathering human judgments about a large number of topics is a daunting task.

This leaves us with a difficult problem. On one hand, we try to construct a measure taking into account the whole Φ and Θ matrices and the whole corpus. On the other hand, validating such measure requires comparing them to the human judgment. Therefore, one needs to somehow obtain human ratings about the whole corpus and the whole probability distribution.

We propose a way to circumvent this infeasible procedure: instead of asking human experts to produce a number of labels, we generate a semi-synthetic dataset with known labels. In this enterprise, the structure of PostNauka dataset is of a tremendous help. The topics of articles are general and diverse enough to make the majority of documents *monotopical*: i.e. every word of such document could be attributed either to a single specific topic or to background topic.

We use these monotopical documents to produce a semi-synthetic dataset. The idea is to "cut" the monotopical documents into smaller monotopical segments and then "sew" them together in random order. The intent of this semi-synthetic dataset is to serve as a ground truth by which topic models can be evaluated

The generation procedure ensures that we know true topic labels for every word. Given this information, it is possible to define segm to be the segmentation quality of any topic model. There are two natural ways to do this:

- soft: for each topic t the sum of $p(t \mid d, w)$ on all pairs (d, w), $d \in D, w \in W_d$ is calculated, with total result equals to the sum of these sums for all topics
- strict: for each topic t for all segments of topic t the number of coincidences of topic, predicted by the model for a word in a document, with the topic t of the segment to which this word belongs $[\operatorname{argmax}_\tau p(\tau \mid d, w) = t]$.

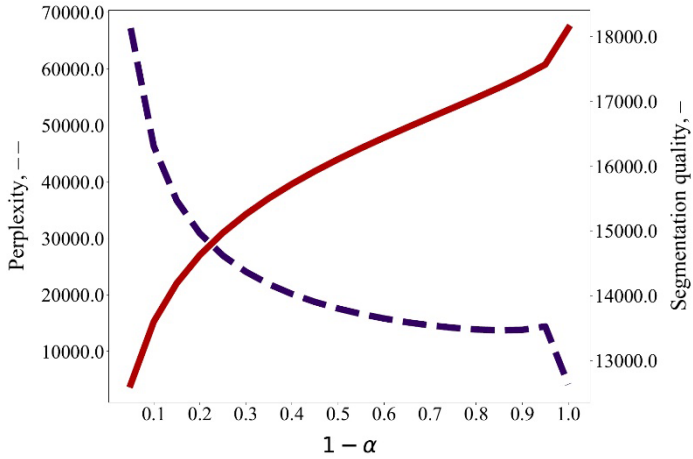

Figure 4. The relationship between segmentation quality and perplexity of topic model

On the X axis is the proportion of good Phi matrix: one minus α (degree of Phi degradation). The fact that segmentation quality monotonically increases when perplexity decreases implies that the proposed segmentation quality may be used as a measure of quality of topic models.

Having established the ground truth, we are able to evaluate different coherence measures. The quality of each candidate measure coh is defined to be a Spearman correlation coefficient between the function value and the segmentation quality.

For this purpose, we generated a number of different Φ matrices as a weighted combination of Φ_{good} (the topic model of PostNauka dataset, discussed above) and Φ_{bad} (a set of random columns taken from Dirichlet $(0.01^{|W|})$ distribution):

$$(\alpha)=\alpha\cdot\Phi_{bad}+(1-\alpha)\Phi_{good}$$

For each α, the segmentation quality and all the investigated coherence metrics were calculated. Thus, a sample $\{\langle \text{soft}(m), \text{strict}(m), c_1(m), c_2(m), \ldots c_n(m)\rangle \mid m \in M, c_i \in \text{Coh}, 1 \leq i \leq |\text{Coh}|\}$ was obtained. Four series of experiments were conducted, with different Φ_{bad} matrices.

Good Topic Model

topic 16: язык
Категория будущего времени в большинстве языков Африки отсутствует. Есть много способов говорить о будущем, но это более сложные способы, касающиеся предположения, желания. Нормальный африканский грамматический приём — не говорить "я это сделаю"или "это будет а сказать "это возможно"или "я хочу это сделать"они говорят о будущем, но "попадают"в будущее непрямым путём.

topic 12: политика
И я посылаю деньги борцам за независимость Курдистана, участвую в акциях поддержки курдских повстанцев и так далее. Вот такое наложение друг на друга разных членств, разных "гражданств". В литературе последних десяти лет бытуют такие выражения, как "гендерное гражданство "экономическое гражданство". Первое указывает на членство в воображаемом сообществе женщин, приверженных идеям феминизма.

SQ (S)	SQ (H)	N	M	SC L2	SC Cos	SC Var	TL	FC
16.0e3	**3.76e4**	**-3.65**	**-2.69**	**-3.70**	0.700	**-8.12e3**	**3.45**	**-5.44e4**

Bad Topic Model

topic 16: язык
Категория будущего времени в большинстве языков Африки отсутствует. Есть много способов говорить о будущем, но это более сложные способы, касающиеся предположения, желания. Нормальный африканский грамматический приём — не говорить "я это сделаю"или "это будет а сказать "это возможно"или "я хочу это сделать"они говорят о будущем, но "попадают"в будущее непрямым путём.

topic 12: политика
И я посылаю деньги борцам за независимость Курдистана, участвую в акциях поддержки курдских повстанцев и так далее. Вот такое наложение друг на друга разных членств, разных "гражданств". В литературе последних десяти лет бытуют такие выражения, как "гендерное гражданство "экономическое гражданство". Первое указывает на членство в воображаемом сообществе женщин, приверженных идеям феминизма.

SQ (S)	SQ (H)	N	M	SC L2	SC Cos	SC Var	TL	FC
5.54e3	1.10e4	-4.83	-3.12	-12.9	0.947	-37.0e3	2.87	-13.9e4

Figure 5. Illustration of a model segmentating semisynthetic text

The figure shows two segments of size 50 words from different topics after being processed by *Bad Topic Model* or *Good Topic Model* (discussed above). These segments were extracted from one of the generated documents, in which they were adjacent. Words that are not labeled were assigned topics different from the two of represented segments. Below the segments are coherence values. SQ (S)—stands for soft segmentation quality, SQ (H)—strict segmentation quality, N—Newman, M—Mimno, SC—SemantiC, TL—TopLen, FC—FoCon. Values in bold indicate that coherence function rises as model's quality increases.

Table 2: Spearman correlations between coherences and segmentation qualities (soft) for datasets with different sizes of segments: 50, 100, 200 and 400 words and with 5 topics in each document

Coh	Corr	Coh	Corr	Coh	Corr	Coh	Corr
Newman	0.75	Newman	0.94	Newman	0.80	Newman	0.85
Mimno	0.96	Mimno	0.96	Mimno	0.94	Mimno	0.97
SC L2	0.92	SC L2	0.91	SC L2	0.70	SC L2	0.59
SC Cos	−0.97	SC Cos	−0.96	SC Cos	−0.97	SC Cos	−0.96
SC Var	**1.00**	SC Var	**1.00**	SC Var	**1.00**	SC Var	**1.00**
TopLen	**1.00**	TopLen	**1.00**	TopLen	**1.00**	TopLen	**1.00**
FoCon	**1.00**	FoCon	**1.00**	FoCon	**1.00**	FoCon	**1.00**

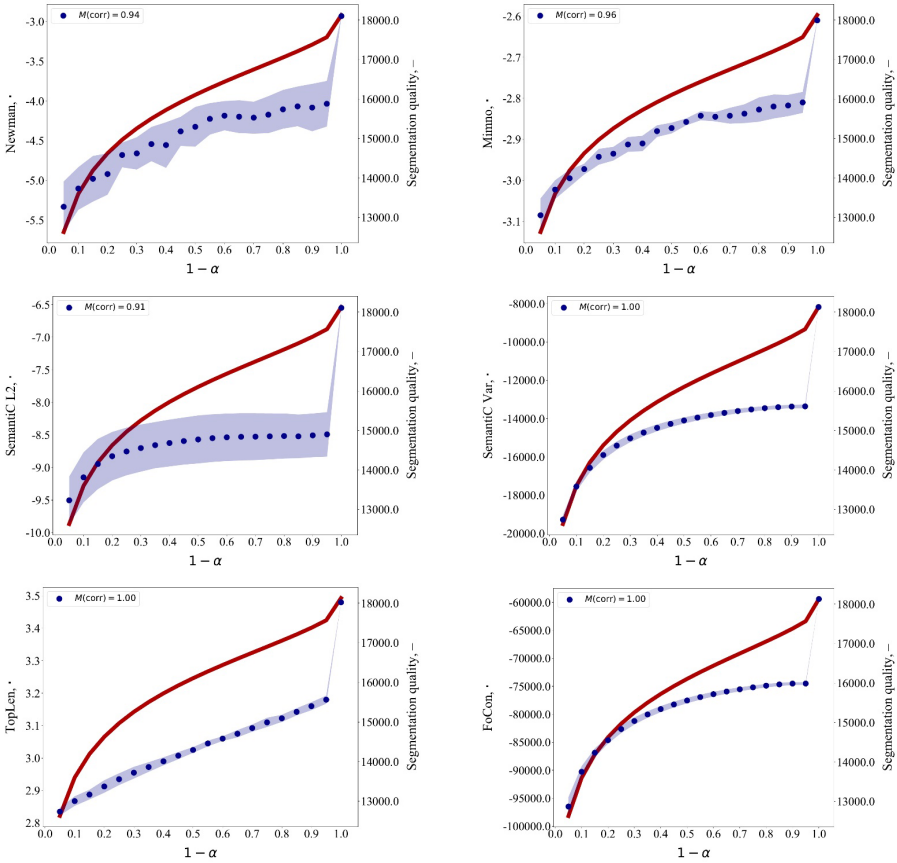

Figure 6. The comparison of different coherence measures with segmentation quality as a function of α, the topic model degradation parameter. Coherence values drawn on the plots are averaged values from 4 series (α) which differ in Φ_{bad} matrix

6. Results

Three new methods for estimating topic model's interpretability are presented: SemantiC, TopLen and FoCon,—which try to take into account all words of the text when evaluating coherence. The new methods show that this is possible to develop an indicator of interpretability able to overcome the shortfalls of top token-based measures.

Experiments on semisynthetic dataset, consisting of segments of different topics, were conducted in order to analyze some properties of new coherences and existing ones.

Proposed methods demonstrate high correlations with the quality of semisynthetic dataset segmentation. SemantiCVar and TopLen appear to perform best.

Acknowledgments

The work was supported by Government of the Russian Federation (agreement 05.Y09.21.0018) and the Russian Foundation for Basic Research grant 17-07-01536. We thank Alexander Romanenko and Irina Efimova for their assistance in data collection.

All experiments with the data were carried out with the use of the BigARTM library [9, 18].

References

1. *Harold W. Kuhn.* "The Hungarian method for the assignment problem". In: Naval Research Logistics (NRL) 2.1–2 (1955), pp. 83–97.
2. *David M. Blei and John D. Lafferty.* "Topic models". In: Text mining: classification, clustering, and applications 10.71 (2009), p. 34.
3. *Jonathan Chang et al.* "Reading Tea Leaves: How Humans Interpret Topic Models". In: Advances in Neural Information Processing Systems 22: 23rd Annual Conference on Neural Information Processing Systems 2009. Proceedings of a meeting held 7–10 December 2009, Vancouver, British Columbia, Canada. Ed. by Yoshua Bengio et al. Curran Associates, Inc, 2009, pp. 288–296. isbn: 9781615679119.
4. *David Newman, Sarvnaz Karimi, and Lawrence Cavedon.* "External evaluation of topic models". In: Australasian Document Computing Symposium, December 2009. 2009, pp. 11–18.
5. *Ianina A., Golitsyn L., Vorontsov K.* Multi-objective topic modeling for exploratory search in tech news. In: Artificial Intelligence and Natural Language. AINL 2017, St. Petersburg, Russia, September 20–23, 2017. Ed. by Filchenkov A., Pivovarova L., Žižka J. Communications in Computer and Information Science, vol 789. Springer, Cham, 2017.—pp 181–193.
6. *Robert K Nelson.* "Mining the dispatch". In: Mining the dispatch (2010). url: http://dsl. richmond.edu/dispatch/pages/intro.
7. *Sang Su Lee, Tagyoung Chung, and Dennis McLeod.* "Dynamic Item Recommendation by Topic Modeling for Social Networks". In: Eighth International Conference on Information Technology: New Generations, ITNG 2011, Las Vegas, Nevada, USA, 11–13 April 2011. Ed. by Shahram Latifi. IEEE Computer Society, 2011, pp. 884–889. isbn: 978-0-7695-4367-3.

8. *David Mimno et al.* “Optimizing Semantic Coherence in Topic Models”. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. EMNLP ’11. Edinburgh, United Kingdom: Association for Computational Linguistics, 2011, pp. 262–272. isbn: 978-1-937284-11-4.
9. *Kochedykov D., Apishev M., Golitsyn L., Vorontsov K.* Fast and Modular Regularized Topic Modelling. In: Proceeding Of The 21St Conference Of FRUCT (Finnish-Russian University Cooperation in Telecommunications) Association. The seminar on Intelligence, Social Media and Web (ISMW). Helsinki, Finland, November 6–10, 2017. Pp.182–193.
10. *Chong Wang and David M. Blei.* “Collaborative topic modeling for recommending scientific articles”. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, August 21–24, 2011. Ed. by Chid Apté, Joydeep Ghosh, and Padhraic Smyth. ACM, 2011, pp. 448–456. isbn: 978-1-4503-0813-7.
11. *David M. Blei.* “Probabilistic topic models”. In: Commun. ACM 55.4 (2012), pp. 77–84.
12. *Timothy N. Rubin et al.* “Statistical topic models for multi-label document classification”. In: Machine Learning 88.1–2 (2012), pp. 157–208.
13. *Nikolaos Aletras and Mark Stevenson.* “Evaluating topic coherence using distributional semantics”. In: Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013)—Long Papers. 2013, pp. 13–22.
14. *Jey Han Lau, David Newman, and Timothy Baldwin.* “Machine Reading Tea Leaves: Automatically Evaluating Topic Coherence and Topic Model Quality.” In: EACL. 2014, pp. 530–539.
15. *Julio Cesar Louzada Pinto and Tijani Chahed.* “Modeling Multi-topic Information Diffusion in Social Networks Using Latent Dirichlet Allocation and Hawkes Processes”. In: Tenth International Conference on Signal-Image Technology and Internet-Based Systems, SITIS 2014, Marrakech, Morocco, November 23–27, 2014. IEEE Computer Society, 2014, pp. 339–346. isbn: 978-1-4799-7978-3.
16. *Devesh Varshney, Sandeep Kumar, and Vineet Gupta.* “Modeling Information Diffusion in Social Networks Using Latent Topic Information”. In: Intelligent Computing Theory—10th International Conference, ICIC 2014, Taiyuan, China, August 3–6, 2014. Proceedings. Ed. by De-Shuang Huang, Vitoantonio Bevilacqua, and Prashan Premaratne. Vol. 8588. Lecture Notes in Computer Science. Springer, 2014, pp. 137–148. isbn: 978-3-319-09332-1.
17. *Michael Röder, Andreas Both, and Alexander Hinneburg.* “Exploring the space of topic coherence measures”. In: Proceedings of the eighth ACM international conference on Web search and data mining. ACM. 2015, pp. 399–408.
18. *Konstantin Vorontsov et al.* “BigARTM: Open Source Library for Regularized Multimodal Topic Modeling of Large Collections”. In: Analysis of Images, Social Networks and Texts—4th International Conference, AIST 2015, Yekaterinburg, Russia, April 9–11, 2015, Revised Selected Papers. Ed. by Mikhail Yu. Khachay et al. Vol. 542. Communications in Computer and Information Science. Springer, 2015, pp. 370–381. isbn: 978-3-319-26122-5.

19. *Potapenko A. A., Popov A. S., Vorontsov K. V.* Interpretable probabilistic embeddings: bridging the gap between topic models and neural networks. In: Artificial Intelligence and Natural Language. AINL 2017, St. Petersburg, Russia, September 20–23, 2017. Ed. by Filchenkov A., Pivovarova L., Žižka J. Communications in Computer and Information Science, vol 789. Springer, Cham, 2017.— pp. 167–180.
20. *Halliday M. A. K., Hasan R.* Cohesion in English.—Routledge, 2014.
21. *Benjamin M. Schmidt.* "Words alone: Dismantling topic models in the humanities". In: Journal of Digital Humanities 2.1 (2012), pp. 49–65