Записки научных
семинаров ПОМИ
Том 499, 2021 г.

E. Tutubalina, S. I. Nikolenko

TOPIC MODELS WITH SENTIMENT PRIORS BASED ON DISTRIBUTED REPRESENTATIONS

Abstract. In recent works, topic models for aspect-based opinion mining have been extended to automatically train sentiment priors for topic-word distributions, leading to automated discovery of sentiment words and improved sentiment classification. In this work, we propose an approach where sentiment priors are trained in the space of word embeddings; this allows us to both discover more aspect-related sentiment words and further improve classification. We also present an experimental study that validates our results.

§1. Introduction

Topic modeling has become the model of choice for a number of applications dealing with unsupervised analysis of large text collections. The basic Latent Dirichlet Allocation model [6, 11] has been subject to many extensions with different goals, including modeling interrelated topics [8, 16], topic evolution in time [4, 31, 34], supervised approaches with a response variable [5], and so on.

One important application of topic models has been in the field of sentiment analysis. Recently, topic models have been successfully applied to aspect-based opinion mining: topic models are able to identify latent topical aspects with sentiments towards them in reviews and other sentiment-related datasets in an unsupervised way [23, 29]. Recent studies usually define an aspect as an attribute or feature of the product that has been commented upon in a review and can be clustered into coherent topics, or *aspects* [17, 23, 38]; e.g., *cupcake* and *steak* are part of the topic *food* for restaurants.

Key words and phrases: topic modeling, natural language processing, sentiment analysis, social media.

The work of Elena Tutubalina was supported by a grant from the President of the Russian Federation for young scientists-candidates of science (MK-3193.2021.1.6) and he framework of the HSE University Basic Research Program and Russian Academic Excellence Project "5-100". The work of Sergey Nikolenko was supported by the St. Petersburg State University, research project "Artificial Intelligence and Data Science: Theory, Technology, Industrial and Interdisciplinary Research and Applications".

Topic models dealing with sentiment analysis usually incorporate sentiment labels for individual words. Such topic models as JST and Reverse-JST [17], ASUM [39], and USTM [38] use existing dictionaries of sentiment words to set β sentiment priors for individual words in certain topics. Tutubalina and Nikolenko [30] proposed a new approach, starting with a seed dictionary of sentiment words but then training new β priors with an expectation-maximization approach. This provides a possibility to discover new sentiment words, especially aspect-related sentiment words that could not be listed in a dictionary, have different sentiment priors for the same words in different aspects, and it has been shown to generally improve sentiment classification.

On the other hand, recent advances in distributed word representations have made it into a method of choice for modern natural language processing [10]. In this approach, words are embedded into Euclidean space, trying to capture semantic relations with the geometry of this semantic space. Starting from the works of Mikolov et al. [20, 21], distributed word representations have been applied for numerous natural language processing problems, including text classification, extraction of sentiment lexicons, part-of-speech tagging, syntactic parsing and so on. In particular, long short-term memory networks (LSTM) over word embeddings have been successfully applied to sentiment analysis by Wang et al. [33], while convolutional networks, which aim to discern local features (e.g., sentiment words in this case), have been used for sentiment analysis by Kalchbrenner et al. [13]. Several approaches that combine topic models and word vectors have already been proposed, e.g., the neural topic models of Cao et al. [7] and Gaussian mixture topic models of Yang et al. [37], but have not yet been extended to sentiment-based topic models.

In this work, we attempt to combine the strengths of both topic models and distributed representations, training a sentiment topic model with priors based on word embeddings. The underlying idea is that instead of training separate β priors independently for every word, we train a sentiment prior in the semantic space that automatically extends to highly similar, interchangeable words; this lets us significantly extend the trained sentiment dictionaries and improve sentiment classification. Note also that instead of a single unified sentiment prediction provided by, e.g., an LSTM this approach yields specific positive and negative words for individual aspects in a review, getting a more detailed and easily interpretable perspective on sentiment evaluation.

This paper is organized as follows. In Section 2, we describe related work, briefly surveying some sentiment-specific LDA extensions, optimization methods for model parameters, and word embeddings. In Section 3, we introduce our approach to optimize sentiment-specific hyperparameters. In Section 4, we present experimental results on the SentiRuEval-2015 dataset that show improvement in sentiment classification and qualitative results for the topic models. Section 5 concludes the paper.

§2. Related work

Traditional aspect-based approaches to sentiment analysis extract phrases that contain words from predefined and usually manually constructed lexicons or words that have been shown by trained classifiers to predict a sentiment polarity. These works usually distinguish *affective* words that express feelings (*happy*, *disappointed*) and *evaluative* words that express sentiment about a specific thing or aspect (*perfect*, *awful*); these words come from a known dictionary, and the model is supposed to combine the sentiments of individual words into a total estimate of the entire text and individual evaluations of specific aspects. For a recent overview of opinion mining see [18]; a sentiment lexicon plays a central role in most methods.

Recently, several topic models have been proposed and successfully used for sentiment analysis. Probabilistic topic models, usually based on Latent Dirichlet Allocation (LDA) and its extensions [17, 19, 38, 39], assume that there is a document-specific distribution over sentiments since sentiment depends on a document, and the models' priors are based on the lexicon.

Lin et al. [17] proposed sentiment modifications of LDA called Joint Sentiment-Topic (JST) and Reverse Joint Sentiment-Topic (Reverse-JST) models with the basic assumption that in the JST model, topics depend on sentiments from a document's sentiment distribution π_d and words are generated conditional on sentiment-topic pairs, while in the Reverse-JST model sentiments are generated conditional on the document's topic distribution θ_d (Fig. 1). Lin et al. [17] derive Gibbs sampling distributions for both models.

Similar to the JST, Jo and Oh [39] proposed the Aspect and Sentiment Unification Model (ASUM), where all words in a sentence are generated from one topic with the same sentiment. Topics (aspects from reviews) are generated from a sentence distribution over sentiments. ASUM achieved an improvement over supervised classifiers and other generative models, including JST.

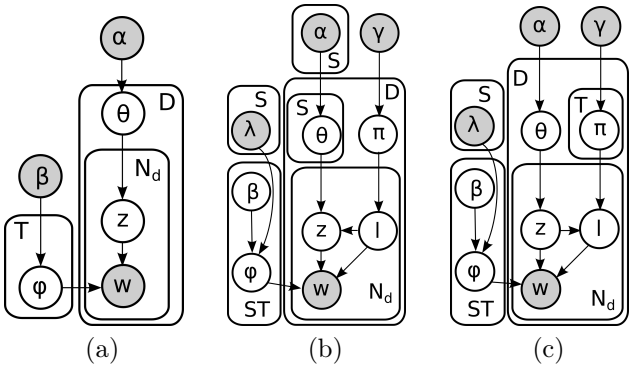

Figure 1. Sentiment LDA extensions: (a)LDA; (b) JST; (c) Reverse-JST.

Yang et al. [38] proposed User-Aware Sentiment Topic Models (USTM) which incorporate user meta-data with topics and sentiments. In this model, topics depend on the document's tags, and words are conditioned on the latent topics, sentiments and tags. USTM gave a substantial improvement over JST and ASUM in the prediction of reviews' sentiment; however, the authors did not analyze the minimal number of tags needed for high-performance training without possible overfitting.

We also note a nonparametric hierarchical extension of ASUM called HASM [14] and nonparametric extensions of USTM models, USTM-DP(W) and USTM-DP(S) [38].

Existing topic models for aspect-based sentiment analysis almost invariably assume a predefined dictionary of sentiment words, usually incorporating this information into the β priors for the word-topic distributions in the LDA model. It has been found that an asymmetric Dirichlet prior over the per document topic-based sentiment proportions yields an improvement in classification over models with symmetric priors [38]. Tutubalina and Nikolenko [30] propose a novel approach for automatic updates of sentiment labels for individual words in a semi-supervised fashion, starting from a small seed dictionary with optimization based on Expectation-Maximization. On each E-step, the sentiment priors β_{kw} are updated proportionally to the number of words w generated with sentiment label k in the corpus, with a coefficient that decreases with the number of iterations

to avoid overfitting. However, this approach treats every word as an independent dimension, and overall seems to suffer from too many independent variables.

Training sentiment β priors can be regarded as part of an effort for optimizing priors in topic models. In related work, topic hyperparameters α have been optimized with fixed-point iteration to maximize the log-evidence [22]. Seaghdha and Teufel [28] use the hyperparameters of a Bayesian latent variable model to investigate rhetorical and topical language, sampling them with Hamiltonian Monte Carlo. Hong et al. [12] utilize a mixture between EM and a Monte Carlo sampler to effectively learn all parameters in variational inference. Another group of optimization methods for sentiment model parameters minimizes the errors between observed and predicted ratings. Diao et al. [9] use gradient descent to minimize an objective function that consists of prediction error on user ratings and probability of observing the text conditioned on priors. Li et al. [15] construct a supervised user-item based topic model that optimizes priors and other parameters using textual topic distribution and employs user and item latent factors to predict ratings. We do not consider supervised topics models with observed labels in this paper but note them as a possible direction for future work.

Finally, in this work we use distributed word representations, i.e., models that map each word occurring in the dictionary to a Euclidean space, attempting to capture semantic relationships between the words as geometric relationships in the Euclidean space. Usually, one first constructs a vocabulary with one-hot representations of individual words, where each word corresponds to its own dimension, and then trains representations for individual words starting from there, basically as a dimensionality reduction problem [21]. For this purpose, researchers have usually employed a model with one hidden layer that attempts to predict the next word based on a window of several preceding words. Then representations learned at the hidden layer are taken to be the word's features; other variations include Glove [25] and other methods [1].

There have been several attempts to use distributed word representations to construct topic models. The Neural Topic Model developed by Cao et al. [7] models both topic-word and document-topic distributions with neural networks, training n-gram embeddings together with document-topic embeddings; this model has also been extended to the supervised

setting. Yang et al. [37] model a topic as a Gaussian cluster in the semantic space, thus making the topic model into a Gaussian mixture. Another way to adopt topic modeling into a neural model was proposed in [36], where a modification of TweetLDA, LDA for short texts [26], was employed to generate topics and topical keywords as extra information for an input message for a convolutional neural network.

Unlike neural topic models, in this work we develop a method that matches existing approaches to sentiment-based topic modeling more closely; we use already existing pretrained word embeddings and utilize them to improve sentiment classification. First, this lets us use word embeddings trained on very large corpora and encompassing many different language samples, much more than any sentiment-related dataset might provide. Second, this approach is easier to apply and extend in practical situations: for the English language one can download high-quality word embeddings trained on huge corpora such as Wikipedia, and for other languages one can train word embeddings with existing libraries such as *word2vec* [20] and its reimplementations [27].

§3. Sentiment priors as distributions in the semantic space

Previously, sentiment priors were introduced in the model either as predefined prior values drawn from a dictionary or as a set of independent values β_{kw} that have to be trained separately on the E-step. In this work, we change the underlying model of sentiment priors: instead of completely independent prior values β_{kw} for every sentiment value k and every word w, we assume that β_{kw} should be similar for words that are similar in the semantic Euclidean space of word embeddings. Suppose that we have found a set of nearest neighbors $\mathrm{Nei}(w)$ for every word w. This set can, for instance, result from a clustering model or simply from thresholding nearest neighbors with a distance tuned to provide good semantic matches.

We use the EM approach for training sentiment priors. On the E-step, we estimate p_{kw}, the probabilities that word w occurs with sentiment k in the corpus, with counters n_{kw} from the Gibbs sampling process. We add a new regularizer on the values of p_{kw} that captures that $p_{kw} \approx p_{kw'}$ for $w' \in \mathrm{Nei}(w)$, i.e., words with highly similar vectors should in all probability have the same sentiment. In the resulting optimization problem, on the E-step we augment log-likelihood of the model, which in this case is a multinomial distribution

$$\log L = \sum_{k,w} n_{kw} \log p_{kw},$$

with a regularizer $R(p)$ that accounts for this assumption; for convenience of optimization, we represent this regularizer in logarithmic form, as

$$R(w) = -\sum_{w' \in \mathrm{Nei}(w)} \frac{1}{d(w,w')} \sum_{k} \left(\log p_{kw} - \log p_{kw'}\right)^2.$$

In total, on the E-step we maximize

$$\log L + \sum_{w} R(w)$$
$$= \sum_{k,w} n_{kw} \log p_{kw} - \frac{\alpha}{2} \sum_{w} \sum_{w' \in \mathrm{Nei}(w)} \frac{1}{d(w,w')} \sum_{k} \left(\log p_{kw} - \log p_{kw'}\right)^2,$$

where α is a regularization coefficient and $d(w,w')$ is the distance between word vectors for w and w' in the semantic space (in the experiments, we tried Euclidean and cosine distances) under the constraints that $\sum_k p_{kw} = 1$ for every w. This is a quadratic optimization problem on $\log p_{kw}$, so it can be solved with off-the-shelf quadratic optimizers. We leave other possible forms of the word vector regularizer for further study.

Once we have found p_{kw}, we can set $\beta_{kw} \propto p_{kw}$. It is beneficial for the topic model to use a sparsity-inducing prior distribution with small parameters β, so in the experiments below we normalized $\sum_k p_{kw}$ to the maximum sum of fixed priors β based on $\mathrm{Nei}(w)$.

§4. Experimental evaluation

4.1. Datasets and settings. To demonstrate the effectiveness of the proposed optimization step, we have conducted an extensive experimental study using six datasets[1]. The **Hotel** dataset consists of reviews of hotels along with author names from TripAdvisor.com. Yang et. al. [38] used only 808 reviews with top 5 location tags, so we adopted the dataset from [32]. In order to apply USTM, we crawled the meta-data of review authors from more than 300,000 reviews, filtering about half out by requiring that user

[1]All datasets are available at `https://yadi.sk/d/82jgiXddsEtCG`.

meta-data has location, gender, and age, and the authors belong to top-50 most common locations. To avoid the sparsity issue, we considered the top 15 location tags along with 5 age tags and 2 gender tags. The **Amazon** dataset contains product reviews from Amazon.com[2] about computer, automotives, and home tools (further called **AmazonComp**, **AmazonAuto**, and **AmazonTools** respectively). In order to apply USTM, for each dataset we crawled the meta-data of review authors such as location, filtering to top-25 most common locations. The **Restaurant** and **Cars** datasets consist of Russian reviews crawled from on-line review websites Otzovik.com and Restoclub.ru, respectively; there is no information about review authors. In preprocessing, we removed punctuation, converted word tokens to lowercase, removed stopwords[3] except negations *нет* (*not*) and *не* (*no*), filtered out rare words that occur less than 5 times in the dataset and high-frequency words that occur in more than 30% of the reviews, and applied lemmatization for Russian texts using the *Mystem* library[4]. Table 1 provides detailed information about each dataset.

Dataset	# reviews	voc. size	# tokens	avg. len	lang
Hotel	26,683	18,253	2,044,215	84.53	EN
AmComp	76,192	22,078	4,594,010	67.01	EN
AmAuto	22,362	9,719	843,519	41.81	EN
AmTools	29,793	12,861	1,351,649	50.52	EN
Restaurant	45,777	26,589	3,202,689	75.86	RU
Cars	8,270	10,783	696,349	94.88	RU

Table 1. Summary statistics for the review datasets.

For word embeddings, we used continuous bag-of-words (CBOW) and skip n-gram *word2vec* models trained on a large Russian-language corpus with about 14G tokens in 2.5M documents [2, 24].

We integrate sentiment information in the described models by using asymmetric priors β. Our manually created lexicon for Russian consisted of 1079 positive words and 1474 negative words, and for English we adopted the MPQA Lexicon [35] with 2718 positive and 4911 negative words. We use symmetric priors for other words (possibly neutral) that are not found in the seed dictionary. Thus, we divided sentiment priors into three different values: neutral, positive, and negative. We first set the β priors for all words in the corpus to $\beta_{kw} = 0.01$; then, if a word belongs to the seed

[2] `https://snap.stanford.edu/data/web-Amazon.html`

[3] The stopword list is adopted from `https://pypi.python.org/pypi/stop-words`

[4] `https://tech.yandex.ru/mystem/`

sentiment dictionary, we set the sentiment priors for a positive word to $\beta_{*w} = (1, 0.01, 0.001)$ (1 for positive, 0.1 for neutral, and 0.001 for negative); for a negative word, to $\beta_{*w} = (0.001, 0.01, 1)$. Posterior inference for all models was done with 1000 Gibbs iterations with $K = 10$, $\alpha = \frac{50}{K}$, and $\gamma = 0.1$.

4.2. Results. For each dataset, we trained four topic models (see Section 2): JST, Reverse-JST, ASUM, and USTM. We compared them in three variations: (i) with fixed βs (without any optimization); (ii) with EM optimization as shown in [30] (marked with "+EM"); (iii) with our proposed optimization step (marked with "+W2V"). The priors are updated after every 50 iterations.

We held out 20% of the training set about hotels and restaurants as a validation set to set the regularization coefficient α. In order to to learn priors, we perform gradient descent with learning rate of 10^{-6} using the Theano library [3]. For both corpora, we set $\alpha = 1.0$ for all datasets.

For evaluation, we held out 10% of the reviews for testing purposes and used the remaining 90% to train topic models. We manually analyzed cosine similarities between pairs of words with opposite polarities like *хороший* [*good*] – *плохой* [*bad*], *трудно* [*hard*] – *нетрудно* [*not hard*], *ругать* [*abuse*] – *расхваливать* [*praise*] and chose similarity thresholds e for the distances $d(w, w')$ as 0.77 for Russian and 0.72 for English.

In the **Restaurant** dataset, each review is associated with a set of ratings providing scores between 0 (lowest) and 10 (highest) about food, interior, service. We mark these reviews with 'positive' sentiment if the average rating score is equal or greater than 7. We mark these reviews with negative sentiment if the rating score is equal to or less than 4. In other datasets, each review is associated with an overall rating between 0 (lowest) and 5 (highest). We mark these reviews from 5 datasets with "positive" or "negative" sentiment if the rating score is equal or greater than 4 or the rating score is equal or less than 2. The unmarked reviews are treated to be "neutral". Statistics of our corpora are presented in Table 2.

Following [38], the probabilities $p(l \mid d)$ were calculated based on the topic-sentiment-word distribution ϕ. In our experiments, a review d is classified as positive if its probability of positive label $p(l_{\text{pos}} \mid d)$ is higher than its probabilities of negative and neutral classes $p(l_{\text{neg}} \mid d)$ and $p(l_{\text{neu}} \mid d)$, and vice versa. Since ASUM, JST and RJS only consider positive or negative sentiments, we evaluate the performance of all models based only on reviews with either positive or negative ground truth labels. Table

Dataset	Labels			# tokens	
	pos.	neg.	neutr.	# pos.	neg.
Hotel	3,136	5,884	17,663	299,017	128,885
AmComp	13,098	7,056	56,038	576,994	309,347
AmAuto	3,225	1,554	17,583	129,185	55,771
AmTools	4,871	2,556	22,366	198,919	96,199
Restaurant	8,728	10,791	26,258	244,596	63,987
Cars	199	671	7,400	64,183	19,157

Table 2. Summary statistics of positive, negative, and neutral labels for the review datasets.

Model	Hotel				AmazonComputer				AmazonAuto			
	P	R	F1	Acc	P	R	F1	Acc	P	R	F1	Acc
JST	.987	.305	.465	.351	.870	.735	.797	.698	.913	.499	.645	.543
RSJT	.983	.771	.864	.775	.863	.256	.395	.368	.880	.308	.456	.378
ASUM	.975	.794	.875	.790	.903	.470	.619	.533	.869	.777	.821	.712
USTM	.947	.834	.887	.809	.839	.829	.834	.734	.876	.864	.870	.781
JST-EM	.973	.465	.629	.493*	.889	.597	.714	.614	.851	.874	.862	.765*
RSJT-EM	.968	.554	.705	.569	.771	.4378	.558	.439*	.883	.514	.650	.533*
ASUM-EM	.968	.749	.845	.745	.893	.504	.644	.547*	.845	.715	.775	.649*
USTM-EM	.937	.746	.831	.718	.824	.837	.831	.724	.862	.885	.874	.784
JST-W2V	.953	.888	.919	.846$^{*}_{\dagger}$	.852	.778	.813	.712	.856	.862	.859	.760*
RSJT-W2V	.986	.816	.893	.818$^{*}_{\dagger}$	.851	.428	.569	.479$^{*}_{\dagger}$	.915	.311	.465	.391*
ASUM-W2V	.977	.803	.882	.801$^{*}_{\dagger}$	.884	.531	.664	.566$^{*}_{\dagger}$	.896	.664	.763	.650
USTM-W2V	.947	.961	.954	.914$^{*}_{\dagger}$	.813	.970	.885	.796$^{*}_{\dagger}$	.863	.950	.905	.831$^{*}_{\dagger}$

Model	AmazonTools				Restaurant				Cars			
	P	R	F1	Acc	P	R	F1	Acc	P	R	F1	Acc
JST	.883	.404	.554	.462	.984	.731	.839	.758	.996	.438	.609	.453
RSJT	.832	.276	.415	.355	.960	.330	.492	.410	.997	.593	.743	.603
ASUM	.881	.592	.709	.597	.952	.836	.890	.822	.989	.756	.857	.755
USTM	.855	.918	.885	.803	N/A				N/A			
JST-EM	.844	.597	.699	.582*	.975	.792	.874	.804*	.981	.907	.943	.893*
RSJT-EM	.747	.481	.585	.495*	.879	.417	.566	.450*	.976	.722	.830	.714*
ASUM-EM	.851	.612	.712	.596	3.963	.510	.667	.562	.983	.658	.789	.658
USTM-EM	.821	.847	.834	.725	N/A				N/A			
JST-W2V	.838	.918	.876	.785$^{*}_{\dagger}$	.982	.729	.837	.754	.981	.974	.977	.956$^{*}_{\dagger}$
RSJT-W2V	.856	.345	.492	.410*	.943	.498	.652	.541$^{*}_{\dagger}$	.996	.636	.776	.644*
ASUM-W2V	.899	.647	.753	.647$^{*}_{\dagger}$	.984	.773	.866	.793	.996	.710	.829	.715$_{\dagger}$
USTM-W2V	.846	.977	.907	.833$^{*}_{\dagger}$	N/A				N/A			

Table 3. Comparison of topic models on several real-world datasets; ∗and †over the accuracy indicate statistically significant improvements over the corresponding model with static βs and βs optimized by the EM-algorithm, respectively, as measured by the Wilcoxon signed ranks test.

3 presents classification results. The reported results are macro-averaged based on 5-fold cross validation.

Several important observations can be derived from the results in Table 3. First, results of four models on the **Hotel** dataset are highly correlated with results in [38]. USTM as the state-of-the-art model achieved better

n.	type	unigrams
1	neu	немецкий [german], volkswagen, дизель [diesel], дизельный [diesel powered], мотор [engine], немец [german], ауди [audi], audi, пассат [passat], мерседес [mercedes], микроавтобус [minubus]
1	pos	toyota, японский [japanese], bmw, nissan, mitsubishi, тойота [toyota], honda, японец [japanese], высокий [high], внимание [attention], привод [drivegear], модель [model], полный [full]
1	neg	минус [minus], небольшой [little], достаточно [sufficient], маленький [small], трасса [road], вполне [just], низкий [low], бензин [gasoline], недостаток [weak point], приходиться [have to]
2	neu	выбор [choice], покупка [purchase], выбирать [to choose], вариант [option], решать [to decide], деньги [money], приобретать [to get], данный [current], находить [to find], хотеться [to wish]
2	pos	коробка [box], передача [transmission], трасса [road], автомат [automatic], быстро [fast], ехать [to drive], поворот [turn], педаль [push bar], мягкий [soft], управление [running], езда [ride]
2	neg	иномарка [foreign car], ваз [vas], калина [], приора [priora], отечественный [home made], лада [lada], автопром [car industry], лад [okay], российский [russian], приор [prior], грант [grant]
3	neu	дверь [door], передний [front], пассажир [passenger], кнопка [button], сзади [behind], кресло [chair], стекло [glass], водитель [driver], панель [panel], зеркало [mirror], нога [foot]
3	pos	форд [ford], ford, мазда [mazda], фокус [focus], комфортный [comfortable], класс [class], советовать [to recommend], довольный [satisfied], экономичный [efficient], skoda, друг [friend]
3	neg	ваз [VAZ], ремонт [repair], запчасть [the spare part], иномарка [foreign car], волга [Volga], состояние [condition], приходиться [have to], отечественный [home made], отец [father]

Table 4. Topics discovered in the Auto dataset by RJST+W2V.

results than RJST, JST, and ASUM on four English datasets. Second, for USTM, the results clearly show that USTM+W2V yields an improvement

over the original models with sentiment priors based on a predefined sentiment lexicon and USTM+EM. For JST and RJST, results are mixed: JST+W2V and RJST+W2V achieved better accuracy and F1-measure over JST+EM and RJST+EM, respectively, on half of the experiments. Results of ASUM+EM and ASUM+W2V are only slightly better or worse than original ASUM, which makes sense since ASUM presupposes that all words in a sentence are generated from the same sentiment, and we are trying to train sentiment priors for individual words.

4.3. User attribute prediction. We conducted experiments to predict the author's attributes of a review based on its lexical content, similar to [38]. For this experiment, we used the **Hotel** dataset with three-dimensional user attributes: location, gender, and age. Mean Average Precision (MAP) was used as an evaluation measure. Table 5 presents the results.

Model	PLDA	USTM	USTM-EM	USTM-W2V
MAP	.338	.446	.453	.475

Table 5. Topic models performance on the task of predicting the attributes of review authors.

USTM-W2V	**P**	**R**	**F1**	**Acc**
$e = .55$	.955	.927	.941	.892
$e = .60$	.960	.909	.933	.881
$e = .65$	.961	.943	.952	.912
$e = .72$	.947	**.961**	**.954**	**.914**
$e = .80$	**.969**	.811	.884	.802

Table 6. USTM-W2V performance with varying similarity threshold e (Hotel dataset).

Word2vec params				**P**	**R**	**F1**	**Acc**
s	w	n	v				
100	11	10	30	.889	.406	.558	.446
200	11	10	20	.916	.413	.569	.463
300	11	1	30	.943	.498	.652	.541

Table 7. Reverse-JST performance with different word embeddings (Restaurant dataset, $e = .77$).

Similar to the sentiment prediction task, the topic model with the proposed optimization USTM-W2V achieved better results than the baseline models PLDA and USTM.

4.4. Effect of Similarity Threshold and Regularization Coefficient α. In this paper, we propose an method of optimizing sentiment-based priors β based on distributed representations. In order to demonstrate the effects of the threshold distance between word vectors in the semantic space and the regularization coefficient in the function $R(w)$, we used USTM which obtained best results in the classification task.

First, we validate the effectiveness of the cosine similarity threshold from 0.55 to 0.80 on the **Hotel** dataset; evaluation results are presented in Table 6. Obviously, the smaller the threshold value chosen, the greater the number of words with at least one nearest neighbors produced. This threshold controls the density of clustering nearest words'priors. The numbers of unique words with $|nei(w)| \geqslant 1$ are 13496, 11493, 8801, 4789, and 1177 for e equal to 0.55, 0.60, 0.65, 0.72, and 0.80 respectively. Several observations can be made based on the results. First, USTM-W2V with the lowest thresholds $e = .55$ and $e = .60$ outperformed USTM (see Table 3). Second, USTM-W2V with $e = .80$ used only 6.45% of vocabulary to maximize the function on the E-step and achieved the lowest results in Table 6, while best results were obtained by USTM-W2V that used 26.23%.

We further investigate the impact of regularization coefficient on the **AmazonTools** dataset. Figure 2 presents the results of this experiment for four models; it shows that for all models the sentiment prediction accuracy reaches maximum value when the coefficient α is set from 0.5 to 1.5.

4.5. Comparison of Word Embeddings. Since we train our own word vectors using *word2vec* models for Russian texts, we conducted a set of experiments to compare different word embeddings. We have trained several word embedding with a high-performance GPU implementation of the CBOW model[5] with different parameters s (vector size), w (length of local context), n (negative sampling), and v (vocabulary cutoff: minimal frequency of a word to be included in the vocabulary). Table 7 shows classification results for some characteristic examples for the Reverse-JST model. In general, increasing word embedding dimension up to about 300 improved the results, while the n and v parameters had very little effect.

[5] `https://github.com/ChenglongChen/word2vec_cbow`.

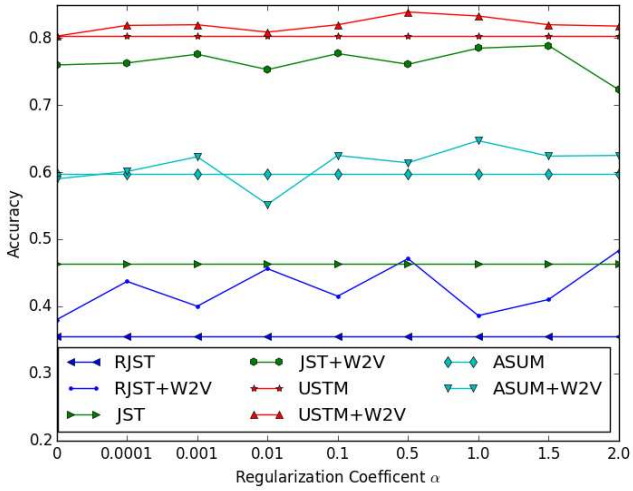

Figure 2. Accuracy of sentiment prediction by varying the regularization coefficient α (AmazonTools dataset).

Word Embeddings	P	R	F1	Acc
GloVe 100d	.959	.934	.946	.901
GloVe 200d	.951	.970	.961	.927
GloVe 300d	.955	.941	.948	.904
word2vec 300d	.947	.961	.954	.914

Table 8. USTM performance with different word embeddings (Hotel dataset, $e = .72$).

For USTM, we also examined publicly available GloVe word vectors trained on 6 billion words from newswire text data and Wikipedia [25]. As shown in Table 8, 200-dimensional GloVe embeddings slightly improved over word2vec embeddings on Hotel dataset. As shown in Table 9, manual probes of different words' sentiment priors confirmed that priors' values are more accurate for 200-dimensional vectors over 100-dimensional vectors.

4.6. Qualitative analysis. In this section, we present qualitative analysis on the topics discovered by RJST with w2v-based optimization step.

Word	Glove 200d			Glove 100d		
	neut	*pos*	*neg*	*neut*	*pos*	*neg*
bad	.182	.001	.817	.271	.001	.728
badly	.041	.001	.958	.020	.001	.979
crowded	.152	.001	.847	.115	.024	.861
benefit	.038	.902	.059	.034	.881	.085
fix	.465	.158	.376	.569	.209	.221
problem	.247	.161	.592	.274	.001	.725
wait	.725	.172	.102	.693	.207	.099
work	.704	.189	.106	.635	.097	.268
wow	.099	.827	.028	.068	.796	.136
incredibly	.102	.819	.078	.116	.784	.099
beautifully	.089	.785	.126	.083	.773	.143

Table 9. Sentiment priors of USTM after training with optimization based on Glove 200d and Glove 100d vectors (Hotel dataset).

The primary goal of modifying sentiment-specific priors based on distributed word representations is to compute similar priors for semantically-related words so that they have higher probabilities to represent related aspects and similar sentiment. To analyze the results according to this goal, we report samples of discovered sentiment topics in Table 4. Top ranked terms are illustrated for specific sentiment-related topics.

Table 4 indicates that RJST+W2V mostly extracts semantically-related aspects from reviews representing nouns like car brands in English and Russian (e.g., *volkswagen*, *toyota*, *ford*, *форд* [*ford*]). Second, negative topics show that people suffer with Russian car industry, old cars, and car repair (negative subtopics #2 and #3). Finally, the positive sample extracted by RJST+W2V contains certain aspects like *driveability transmission* (*transmission*, *fast*, *drivegear*), while neutral subtopics describe *car configuration* (e.g, *mirror*, *behind*, *panel*, *glass*) or *purchase process* (e.g, *money*, *option*, *to find*).

§5. Conclusion

In this work, we have presented a method to automatically update sentiment priors for interchangeable words based on distributed representations. Our experimental evaluation has shown that this idea leads to improvements in sentiment classification and prediction of user attributes

over topic models based on predefined priors and models that update sentiment priors for individual words. Qualitative analysis of the discovered topics also shows that our model with modified priors can find coherent topics in an accurate way. In further work, we plan to elaborate upon the interplay between sentiment priors in LDA extensions and distributed word representations; we hope it will be possible to incorporate distributed word representations directly into other priors.

References

1. R. Al-Rfou, B. Perozzi, and S. Skiena, *Polyglot: Distributed word representations for multilingual nlp*, Proc. 17th Conference on Computational Natural Language Learning (Sofia, Bulgaria), Association for Computational Linguistics, August 2013, pp. 183–192.
2. N. Arefyev, A. Panchenko, A. Lukanin, O. Lesota, P. Romanov, *Evaluating three corpus-based semantic similarity systems for russian*, Proceedings of International Conference on Computational Linguistics Dialogue, 2015.
3. J. Bergstra, O. Breuleux, F. Bastien, P. Lamblin, R. Pascanu, G. Desjardins, J. Turian, D. Warde-Farley, Y. Bengio, *Theano: a CPU and GPU math expression compiler*, Proc. Python for scientific computing conference (SciPy), vol. 4, Austin, TX, 2010, p. 3.
4. D. M. Blei, J. D. Lafferty, *Dynamic topic models*, Proc. 23rd International Conference on Machine Learning (New York, NY, USA), ACM, 2006, pp. 113–120.
5. D. M. Blei, J. D. McAuliffe, *Supervised topic models*, Advances in Neural Information Processing Systems **22** (2007).
6. D. M. Blei, A. Y. Ng, N. I. Jordan, *Latent Dirichlet allocation*, Journal of Machine Learning Research **3** (2003), no. 4–5, 993–1022.
7. Z. Cao, S. Li, Y. Liu, W. Li, H. Ji, *A novel neural topic model and its supervised extension*, Proc. 29th AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA., 2015, pp. 2210–2216.
8. J. Chang, D. M. Blei, *Hierarchical relational models for document networks*, Annals of Applied Statistics **4** (2010), no. 1, 124–150.
9. Q. Diao, M. Qiu, C.-Y. Wu, A. J. Smola, J. Jiang, C. Wang, *Jointly modeling aspects, ratings and sentiments for movie recommendation (jmars)*, Proc. 20th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2014, pp. 193–202.
10. Y. Goldberg, *A primer on neural network models for natural language processing*, CoRR **abs/1510.00726** (2015).
11. T. Griffiths, M. Steyvers, *Finding scientific topics*, Proceedings of the National Academy of Sciences **101 (Suppl. 1)** (2004), 5228–5335.
12. L. Hong, A. Ahmed, S. Gurumurthy, A. J. Smola, K. Tsioutsiouliklis, *Discovering geographical topics in the twitter stream*, Proc. 21st international conference on World Wide Web, ACM, 2012, pp. 769–778.

13. N. Kalchbrenner, E. Grefenstette, P. Blunsom, *A convolutional neural network for modelling sentences*, Proc. 52nd Annual Meeting of the Association for Computational Linguistics (Vol. 1: Long Papers) (Baltimore, Maryland), Association for Computational Linguistics, June 2014, pp. 655–665.
14. S. Kim, J. Zhang, Z. Chen, A. H. Oh, S. Liu, *A hierarchical aspect-sentiment model for online reviews*, Proc. Twenty-Seventh AAAI Conference on Artificial Intelligence, July 14-18, 2013, Bellevue, Washington, USA., 2013.
15. F. Li, S. Wang, S. Liu, M. Zhang, *Suit: A supervised user-item based topic model for sentiment analysis*, Proc. 28th AAAI Conference on Artificial Intelligence, 2014.
16. S. Z. Li, *Markov random field modeling in image analysis*, Advances in Pattern Recognition, Springer, Berlin Heidelberg, 2009.
17. C. Lin, Y. He, R. Everson, S. Ruger, *Weakly supervised joint sentiment-topic detection from text*, IEEE Transactions on Knowledge and Data Engineering **24** (2012), no. 6, 1134 – 1145 (und).
18. B. Liu, *Sentiment analysis: Mining opinions, sentiments, and emotions*, Cambridge University Press, 2015.
19. B. Lu, M. Ott, C. Cardie, B. K. Tsou, *Multi-aspect sentiment analysis with topic models*, Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference (2011), 81–88.
20. T. Mikolov, K. Chen, G. Corrado, J. Dean, *Efficient estimation of word representations in vector space*, CoRR **abs/1301.3781** (2013).
21. T. Mikolov, I. Sutskever, K. Chen, G. Corrado, J. Dean, *Distributed representations of words and phrases and their compositionality*, CoRR **abs/1310.4546** (2013).
22. T. Minka, *Estimating a dirichlet distribution*, 2000.
23. S. Moghaddam, M. Ester, *On the design of LDA models for aspect-based opinion mining*, Proc. 21st ACM international conference on Information and knowledge management, ACM, 2012, pp. 803–812.
24. A. Panchenko, N. V. Loukachevitch, D. Ustalov, D. Paperno, C. M. Meyer, N. Konstantinova, *Russe: The first workshop on Russian semantic similarity*, Proc. International Conference on Computational Linguistics and Intellectual Technologies (Dialogue), May 2015, pp. 89–105.
25. J. Pennington, R. Socher, C. Manning, *GloVe: Global vectors for word representation*, Proc. 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) (Doha, Qatar), Association for Computational Linguistics, October 2014, pp. 1532–1543.
26. D. Quercia, H. Askham, J. Crowcroft, *TweetLDA: supervised topic classification and link prediction in twitter.*, WebSci (Noshir S. Contractor, Brian Uzzi, N. W. Macy, and Wolfgang Nejdl, eds.), ACM, 2012, pp. 247–250.
27. R. **v**Rehů**v**rek and P. Sojka, *Software Framework for Topic Modelling with Large Corpora*, Proc. LREC 2010 Workshop on New Challenges for NLP Frameworks (Valletta, Malta), ELRA, May 2010, `http://is.muni.cz/publication/884893/en`, pp. 45–50 (English).
28. D. O. Séaghdha and S. Teufel, *Unsupervised learning of rhetorical structure with un-topic models.*, COLING, 2014, pp. 2–13.
29. I. Titov, R. McDonald, *Modeling online reviews with multi-grain topic models*, Proc. 17th International conference on World Wide Web, ACM, 2008, pp. 111–120.

30. E. Tutubalina, S. I. Nikolenko, *Inferring sentiment-based priors in topic models*, Proc. 14th Mexican International Conference on Artificial Intelligence, LNCS vol. 9414, Springer, 2015, pp. 92–104.
31. C. Wang, D. M. Blei, D. Heckerman, *Continuous time dynamic topic models*, Proc. 24th Conference on Uncertainty in Artificial Intelligence, 2008.
32. H. Wang, Y. Lu, C. Zhai, *Latent aspect rating analysis without aspect keyword supervision*, Proc. 17th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2011, pp. 618–626.
33. X. Wang, Y. Liu, C. Sun, B. Wang, X. Wang, *Predicting polarities of tweets by composing word embeddings with long short-term memory*, Proc. 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Vol. 1: Long Papers) (Beijing, China), Association for Computational Linguistics, July 2015, pp. 1343–1353.
34. X. Wang, A. McCallum, *Topics over time: a non-Markov continuous-time model of topical trends*, Proc. 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (New York, NY, USA), ACM, 2006, pp. 424–433.
35. T. Wilson, J. Wiebe, P. Hoffmann, *Recognizing contextual polarity in phrase-level sentiment analysis*, Proc. conference on human language technology and empirical methods in natural language processing, Association for Computational Linguistics, 2005, pp. 347–354.
36. Y. Wu, W. Wu, Z. Li, M. Zhou, *Topic augmented neural network for short text conversation*, arXiv preprint arXiv:1605.00090 (2016).
37. M. Yang, T. Cui, W. Tu, *Ordering-sensitive and semantic-aware topic modeling*, CoRR **abs/1502.0363** (2015).
38. Z. Yang, A. Kotov, A. Mohan, S. Lu, *Parametric and non-parametric user-aware sentiment topic models*, Proc. 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, 2015, pp. 413–422.
39. J. Yohan, A. H. Oh, *Aspect and sentiment unification model for online review analysis*, Proc. 4th ACM International Conference on Web Search and Data Mining (New York, NY, USA), WSDM '11, ACM, 2011, pp. 815–824.

Поступило 2 октября 2020 г.

Kazan Federal University,
Kazan, Russia;
National Research University
Higher School of Economics
Myasnitskaya ul., 20, Moscow 101000, Russia
E-mail: elvtutubalina@kpfu.ru

St. Petersburg State University
7/9 Universitetskaya nab.,
St. Petersburg, 199034 Russia;
St. Petersburg Department of
Steklov Institute of Mathematics,
St. Petersburg, Russia
E-mail: s.nikolenko@spbu.ru, sergey@logic.pdmi.ras.ru