

ИЗВЛЕЧЕНИЕ АСПЕКТОВ, ТОНАЛЬНОСТИ И КАТЕГОРИЙ АСПЕКТОВ НА ОСНОВАНИИ ОТЗЫВОВ ПОЛЬЗОВАТЕЛЕЙ О РЕСТОРАНАХ И АВТОМОБИЛЯХ

Иванов В. В. (nomemm@gmail.com),
Тутубалина Е. В. (tutubalinaev@gmail.com),
Мингазов Н. Р. (nicrotek547@gmail.com),
Алимова И. С. (alimovallseyar@gmail.com)

Казанский Федеральный Университет, Казань, Россия

Ключевые слова: анализ тональности текстов, SentiRuEval, отзывы пользователей, извлечение аспектов, категории аспектов

EXTRACTING ASPECTS, SENTIMENT AND CATEGORIES OF ASPECTS IN USER REVIEWS ABOUT RESTAURANTS AND CARS

Ivanov V. V. (nomemm@gmail.com),
Tutubalina E. V. (tutubalinaev@gmail.com),
Mingazov N. R. (nicrotek547@gmail.com),
Alimova I. S. (alimovallseyar@gmail.com)

Kazan Federal University, Kazan, Russia

This paper describes a method for solving aspect-based sentiment analysis tasks in restaurant and car reviews subject domains. These tasks were articulated in the SentiRu Evaluation for Russian (SentiRuEval-2015) initiative. During the SentiRuEval-2015 we focused on three subtasks: extracting explicit aspect terms from user reviews (tasks A), aspect-based sentiment classification (task C) as well as automatic categorization of aspects (task D).

In aspect-based sentiment classification (tasks C and D) we propose two supervised methods based on a Maximum Entropy model and Support Vector Machines (SVM), respectively, that use a set of term frequency features in a context of the aspect term and lexicon-based features. We achieved 40% of macro-averaged F-measure for cars and 40,05% for reviews about restaurants in task C. We achieved 65.2% of macro-averaged F-measure for cars and 86.5% for reviews about restaurants in task D. This method ranked first among 4 teams in both subject domains. The SVM classifier is based on unigram features and pointwise mutual information to calculate category-specific score and associate each aspect with a proper category in a subject domain.

In task A we carefully evaluated performance of a method based on syntactic and statistical features incorporated in a Conditional Random Fields model. Unfortunately, the method did not show any significant improvement over a baseline. However, its results are also presented in the paper.

Key words: aspect-based sentiment analysis, sentiment evaluation, user reviews, aspect extraction, aspect categories

1. Introduction

Over the past decade, opinion mining (also called sentiment analysis) has been an important concern for Natural Language Processing (NLP). Since online reviews significantly influence people's decisions about purchases, sentiment identification has a number of applications, including tracking people's opinions about movies, books, and products, etc.

In this study we describe our approaches for solving a task on sentiment analysis, which was formulated as a separate track in the Sentiment Evaluation for Russian (SentiRuEval-2015) initiative. The SentiRuEval task concerns aspect-based sentiment analysis of user reviews about restaurants and cars. The task consists of several subtasks: aspect extraction (tasks A and B), sentiment classification of explicit aspects (task C), and detection of aspects categories and sentiment summarization of a review (tasks D and E). The primary goal of the SentiRuEval task is to find words and expressions indicating important aspects of a restaurant or a car based on user opinions and to classify them into polarity classes and aspect categories (Loukachevitch et al., 2015).

There have been a large number of research studies in the area of aspect-based sentiment analysis, which are well described in Liu (2012) and Pand and Lee (2008). Traditional approaches in opinion mining are based on extracting high-frequency phrases containing adjectives from manually created lexicons (Turney, 2002; Popescu and Etzioni, 2007). State-of-the-art papers have implemented probabilistic topic models, such as Latent Dirichlet Allocation (LDA), and Conditional Random Field (CRF) for multi-aspect analysis tasks (Moghaddam and Ester, 2012; Choi and Cardie, 2010). Sentiment analysis in English has been explored in depth and there are many well-established methods and general-purpose sentiment lexicons that contain a few thousand terms. However, research studies of sentiment analysis in Russian have been less successful. In 2011–2013 studies have focused on solving a task on sentiment analysis during ROMIP sentiment analysis tracks (Chetviorkin and Loukachevitch, 2013; Kotelnikov and Klekovkina, 2012; Blinov et al., 2013; Frolov et al., 2013).

We use the Conditional Random Fields model applied to the aspect extraction task. In task C for aspect-based sentiment classification we propose a method based on a Maximum Entropy model that uses a set of term frequency features in a context of the aspect term and lexicon-based features. The classifier for aspect category detection is based on a SVM model with a set of category-specific features. We achieved 40% of macro-averaged F-measure for cars and 40,05% for reviews about restaurants in task C. We achieved 65.2% of F-measure for cars and 86.5% for reviews about restaurants in task D.

The rest of the paper is organized as follows. In Section 2 we introduce related work on sentiment analysis. In Section 3 we describe proposed approaches. Section 4 presents results of experiments. Finally, in Section 5 we discuss the results.

2. Related Work

In this paper, we focus on the detection of the three major cores in a review: aspect terms, sentiment about these aspects, and aspects' categories. During the last decade, a large number of methods were proposed to identify these elements.

Aspect term extraction. There are several widely used methods that treat the task as a classification problem (Popescu et al., 2005), as a sequence labeling problem (Jakob and Gurevych, 2010; Kiritchenko et al., 2014; Chernyshevich, 2014), as a topic modeling or a traditional clustering task (Moghaddam and Ester, 2012; Zhao et al., 2014). The classification problem is to determine whether nouns and noun phrases are target of an opinion or not. Popescu et al. (2005) used syntactic patterns in relation with sentiment from general-purpose lexicons to identify high-frequency noun phrases. Poria et al. (2014) proposed a rule-based approach, based on knowledge and sentence dependency trees. These approaches are limited due to lower results on extracting low-frequency aspects or hand-crafted dependency rules for complex extraction. In (Kiritchenko et al., 2014; Chernyshevich, 2014) the authors proposed two modifications of a standard scheme for sequence labeling models.

Aspect term polarity. Most of the early approaches for classifying aspects rely on seed words or a manually generated lexicon that contains strongly positive or strongly negative words. Turney (2002) proposed an unsupervised method, based on a sentiment score of each phrase that is calculated as the mutual information between the phrase and two seed words. Recent papers have widely applied machine learning methods to solve the tasks of sentiment classification (Pang et al., 2002; Pang and Lee, 2008; Blinov et al., 2013; Kiritchenko et al., 2014). Moghaddam and Ester (2012) proposed extensions of the LDA model to extract aspects and their sentiment ratings by considering the dependency between aspects and their sentiment polarities. However, topic models achieve lower performance on multi-aspect sentence classification than the SVM classifier in three different domains (Lu et al., 2011).

Aspect category detection. Automatic categorization of explicit aspects into aspect categories has been studied as the task of sentiment summarization. Moghaddam and Ester (2012) investigated it as a part of a latent aspect mining problem. There have been some works on grouping aspect terms from review texts for the sentiment analysis in the task 4 of the international workshop on Semantic Evaluation (SemEval-2014). The task was evaluated with the F-measure and the best results were achieved by SVM classifiers with bag-of-words features and information from unlabeled reviews (Pontiki et al., 2014; Kiritchenko et al., 2014).

Several studies about sentiment analysis have been done in Russian, related to evaluation events of Russian sentiment analysis systems (Chetviorkin and Loukachevitch, 2013). Frolov et al. (2013) proposed a dictionary-based approach with fact semantic filters for sentiment analysis of user reviews about books. Blinov et al.

(2013) showed benefits of machine learning method over lexical approach for user reviews in Russian and used manual emotional dictionaries.

3. System description

In this section we describe our approaches for three tasks of aspect-based sentiment analysis of user reviews about restaurants and cars. The CRF model was used for automatic extraction of explicit aspects (task A). We applied machine-learning approaches for the tasks C and D, based on bag-of-words model and a set of lexicon-based features that are described in Section 3.2 and 3.3, respectively. The morpho-syntactic analyzer Mystem was used for text normalization at the preprocessing step.

3.1. Aspect Extraction

The goal of aspect extraction is to detect extract major explicit aspects of a product (task A). Since the task can be seen as a particular instance of the sequence-labeling problem, we employ Conditional Random Fields (Lafferty et al., 2001).

Explicit aspects denote some part or characteristics of a described object such as *передний привод* (*front-wheel drive*), *руль* (*steering wheel*), *динамика* (*dynamics*) in cars reviews; *столук* (*table*), *официант* (*waiter*), *блюдо* (*dish*) in reviews about restaurants. In the following examples we consider user phrases about explicit aspects.

We use Inside-Outside-Begin scheme and Passive Aggressive algorithm for training CRF; brief description of the features used to represent the current token w_i are presented below: the current token w_i , the current token w_i within a window (w_{i-2}, \dots, w_{i+2}); the part of speech tag of the current token; the part of speech tag of the token within a tag window ($tag_{i-2}, \dots, tag_{i+2}$); the number of occurrences of the tokens in the training set; the presence of the token in manually created domain-dependent dictionaries.

3.2. Aspect-based sentiment classification

The task of sentiment classification aims to predict polarity (positive, negative, neutral, or both) of each aspect from the product reviews. We applied the Maximum Entropy classifier with default parameters, based on a bag-of-words model and a set of lexicon-based features that are described in Section 3.2.2.

The following examples illustrate the aspects (marked in *italic*) with different polarities from the reviews. Some phrases like “*персонал улыбчивый, приветливый.*” (“smiling, friendly staff”), “общее впечатление: отличная *машина*” (“overall impression: great car”) or “просторный *салон*, удобно сидеть пассажиру сзади” (“*spacious interior*, a passenger could sit *comfortably* behind the driver’s seat”) contain strong positive or negative context near the aspect term. Therefore, such cases could

be correctly classified extracting bigrams in the phrases. Complex analysis of sentiment phrases such as “заказывал *бифштекс*, нет слов как *вкусно*” (“I ordered a beefsteak, there are no words to describe just how *tasty* this was”) and “в городском цикле *компьютер* будет показывать очень неприятные цифры” (“in the city the *computer* will show very unpleasant figures”) shows that there is a distance between the polarity words *вкусно* (*tasty*), *неприятные* (*unpleasant*) and the aspect terms. We use combinations of the aspect term and a context term to classify these cases. Difficult phrases with both sentiments such as “отмечу некоторую *жесткость сидений*, но привыкаешь, главное сидеть удобно” (“I note some *rigidity of the seats*, but you get used to it, the main thing is sit conveniently”) or “горячее неплохое, но на гриль было непохоже” (“*hot dishes* are quite good, but not similar to a grill”) could be recognized by presence of the conjunction word *но* (*but*).

Given a context of the aspect term, two types of word bigrams are generated for feature extraction: (i) context bigrams, using a text within a context window of the aspect term; (ii) aspect-based bigrams as a combination of the aspect term itself and a context word within the context window. The context window of the aspect term w_i denotes a sequence $(w_{i-4}, \dots, w_{i+4})$.

3.2.1. Manually created sentiment lexicon

We collected user rated reviews from *otzovik.com*: 7,526 reviews about restaurants and 4,952 reviews about cars. To make corpus more accurate, we included only *Pros* reviews with an overall rating 5 into positive corpus and *Cons* reviews with an overall rating 1 or 2 into negative corpus. *Pros* (*Преимущества*) and *Cons* (*Недостатки*) are parts of a review that describe strong reasons why an author of the review likes or dislikes the product, respectively. For each domain we selected the top K adverbs, adjectives, verbs, reducing noun words that express aspects, action verbs and most common adjectives. The manually created dictionary consists of about 741 positive and 362 negative words in restaurants domain and includes 1,576 positive and 741 negative words in cars domain. We combine two dictionaries to achieve better evaluation results.

For lexicon-based features we use the following scores: each word in the sentence is weighted by its distance from the given aspect:

$$score(w) = \frac{sc(w)}{e^{|i-j|}}$$

where i, j is the positions of the aspect term and the word, $sc(w)$ is the sentiment word's score, that equals 1 for positive words and -1 for negative words, extracted from the sentiment dictionary.

3.2.2. Classification Features for Aspect Term Polarity

Each review is represented as a feature vector, for each aspect features are extracted from the aspect and its context in a sentence. A brief description of the features that we use is presented below:

- **character n-grams**: lowercased characters n-grams for $n = 2, \dots, 4$ with document frequency greater than two were considered for feature selection.

- **lexicon-based unigrams:** unigrams from the sentiment lexicon are extracted for feature selection.
- **context n-grams:** unigrams (single words) and bigrams are extracted from the context window. We extract these n-grams for several combinations: (i) replacement of the aspect term with the word *aspect*; (ii) replacement of sentiment words with the polarity word *pos* or *neg*; (iii) replacement of sentiment words with a part of speech tag.
- **aspect-based bigrams:** bigrams generated as a combination of the aspect term itself and a word within the context window. We extract these bigrams for several combinations that described above.
- **lexicon-based features:** the features are calculated as follows: the maximal sentiment score; the minimum sentiment score; the sum of the words' sentiment scores; the sum of positive words' scores; the sum of negative words' scores. Sentiment words with negations shift the sentiment score towards the opposite polarity.

Due to limited size of the context window and difficulty in classifying the aspect with both negative and positive sentiment towards its term, we create hand-crafted rule for such cases: if the sentence (*s*) contains the aspect term, a conjunction word *но*, *а* (*but*) and the classifier predicts the neutral label for the aspect, we mark the aspect by the both label.

3.3. Automatic categorization of explicit aspects into aspect categories

The goal of task D is to classify each aspect to one of predefined categories. In restaurant reviews there are the following aspect categories: *food*, *service*, *interior*, *price*, *general*. For automobiles aspect categories are: *drivability*, *reliability*, *safety*, *appearance*, *comfort*, *costs*, *general*.

We describe the task of automatic categorization of explicit aspects in the following examples. Some aspects such as food products (e.g., *бифитекс* (*beefsteak*), *утка по-пекински* (*Peking duck*)) or car components (e.g., *гидроусилитель* (*power steering*), *двигатель* (*engine*)) are classified by a human annotator's explicit knowledge. The categories of food products and car components are *food* and *drivability*, respectively. The category label of some explicit aspects depends on a context of a user review. In the examples “*машина свои деньги отработала полностью*” (“the car is worth its price”), “*пробовал отпустить руль машина едет ровно*” (“have experimented with the driving wheel and the car running smoothly”), “*машина предназначена для фанатов*” (“the *car* is intended for fans”) and “*довольно красивая машина*” (“quite beautiful *car*”) the categories of the aspect term *машина* (*car*) are *costs*, *drivability*, *whole*, *appearance*, respectively.

We addressed the task as a text classification problem and trained the SVM classifier with the sequential minimal optimization (SMO). For each aspect term w_i we extracted the aspect term itself and the features from the context window (w_{i-2}, \dots, w_{i+2}). Category-specific lexicons are based on a score for each term w in the training test:

$$score(w) = PMI(w, cat) - PMI(w, oth)$$

where PMI is pointwise mutual information, cat denotes all aspects' contexts in the particular category, oth denotes aspects' contexts in other categories.

The SVM classifier is based on bag-of-words model and other features described below:

- word n-grams: the aspect term and unigrams from the context of the aspect term are extracted for feature selection.
- category-specific features: the following features are calculated separately for each category: the maximal score in the context; the minimum score in the context; the sum of the words' scores in the context; the average of the words' scores in the context;

4. Experimental Results

For experimental purposes we used the training set of 200 annotated reviews and the testing set of 200 reviews for each domain provided by the organizers of the SentiRuEval task.

4.1. Performance results

The official results obtained by our approaches on the testing set are presented in Tables 1, 2a, 2b and 3. The tables show the official baseline results and the results of other participants according to macro-average F-measure as the main quality measure in the task (Loukachevitch et al., 2015).

For task A exact matching and partial matching were used to calculate F1-measure. Table 1a and 1b show that our method based on the CRF model did not have any significant improvement over a baseline.

For task C macro-averaged F-measure is calculated as the average value between F-measure of the positive class, negative class and F-measure of the both class. Tables 2a show that according to macro-averaged F1-measure, our classifier does not pay off when compared with the approach with run_id 4_1, that is based on a Gradient Boosting Classifier model. Our approach has 0.13% and 0.06% improvements in macro-averaged F1-measure over the approach with run_id 3_1, ranked second in restaurants and banks domain, respectively. Our runs could not be evaluated due to technical problems with the submission.

Table 3 shows the official baseline results and the results of the method, ranked second according to macro-averaged F-measure in task D. This method ranked first among 4 teams in both subject domains. The best approach has 0.06% and 0.09% improvements in macro F1-measure over the baseline in restaurants and cars domains, respectively.

Table 1a. Performance metrics in extraction of explicit aspects in restaurants domain (task A)

	Exact matching			Partial matching		
	Macro P	Macro R	Macro F	Macro P	Macro R	Macro F
Our method	0.3515	0.5331	0.5331	0.6507	0.4399	0.5109
An approach, ranked first	0.5506	0.6901	0.6070	0.6886	0.7916	0.7284
Official baseline	0.5570	0.6903	0.6084	0.6580	0.6960	0.6651

Table 1b. Performance metrics in extraction of explicit aspects in cars domain (task A)

	Exact matching			Partial matching		
	Macro P	Macro R	Macro F	Macro P	Macro R	Macro F
Our method	0.6411	0.5363	0.5749	0.7264	0.6117	0.6498
An approach, ranked first	0.6619	0.6560	0.6513	0.7917	0.7272	0.7482
Official baseline	0.5747	0.6287	0.5941	0.7449	0.6720	0.6966

Table 2a. Performance metrics in the classification task in restaurants domain (task C)

Run_id	Micro P	Micro R	Micro F	Macro P	Macro R	Macro F
Official baseline	0.7104	0.7104	0.7104	0.3209	0.2506	0.2671
1_1	0.6194	0.6194	0.6194	0.2517	0.2454	0.2379
1_2	0.6194	0.6194	0.6194	0.2517	0.2454	0.2379
3_1	0.6696	0.6696	0.6696	0.3223	0.2430	0.2696
4_1	0.8249	0.8249	0.8249	0.5872	0.5569	0.5545
Our approach	0.7671	0.7671	0.7671	0.4582	0.3729	0.4081

Table 2b. Performance metrics in the classification task in cars domain (task C)

Run_id	Micro P	Micro R	Micro F	Macro P	Macro R	Macro F
Official baseline	0.6192	0.6192	0.6192	0.2949	0.2685	0.2648
1_1	0.6471	0.6471	0.6471	0.3399	0.3194	0.3293
1_2	0.6531	0.6531	0.6531	0.3563	0.3297	0.3422
3_1	0.5589	0.5589	0.5589	0.3016	0.2621	0.2794
4_1	0.7428	0.7428	0.7428	0.5725	0.5667	0.5684
1_3	0.6252	0.6252	0.6252	0.3507	0.3262	0.3345
Our approach	0.7110	0.7111	0.7111	0.4481	0.3761	0.4001

Table 3. Performance metrics in categorization of aspects in both subject domains (task D)

	Restaurants			Cars		
	Macro P	Macro R	Macro F	Macro P	Macro R	Macro F
Our approach	0.8960	0.8414	0.8653	0.6854	0.6355	0.6521
Second result	0.8627	0.7963	0.8110	0.7146	0.5750	0.6077
Official baseline	0.8742	0.7737	0.7996	0.6672	0.5190	0.5636

4.2. Ablation Experiments

We performed ablation experiments to study the benefits of features, which are used for the CRF model and machine learning methods. Tables 4a, 4b and 5 show ablation experiments for tasks A and C on the testing set, removing one each individual feature category from the full set. Error analysis and Tables 4a and 4b show that the features on the set of two previous and two next tokens decrease our results in task A in restaurants domain. The most effective features for task C are based on aspect-based bigrams that include combinations of the aspect term and other words from the context window.

Table 4a. Results for the ablation experiments in aspect extraction about restaurants (task A)

	Exact matching			Partial matching		
	P	R	F1	P	R	F1
all features	0.3515	0.5331	0.5331	0.6507	0.4399	0.5109
w/o dictionaries	0.3382	0.4971	0.3961	0.3850	0.6921	0.4821
w/o frequencies	0.6503	0.4322	0.5068	0.7313	0.4755	0.5612
w/o all tokens within (w_{i-2}, \dots, w_i)	0.6105	0.4065	0.4751	0.7118	0.4667	0.5471
w/o all tokens within (w_i, \dots, w_{i+2})	0.6471	0.4375	0.5104	0.7272	0.4865	0.5681
w/o tokens that contained all features within (w_{i-1}, \dots, w_{i+1})	0.7311	0.4801	0.5644	0.6476	0.4416	0.5120

Table 4b. Results for the ablation experiments in aspect extraction about cars (task A)

	Exact matching			Partial matching		
	P	R	F1	P	R	F1
all features	0.6411	0.5363	0.5749	0.7264	0.6117	0.6498
w/o dictionaries	0.6451	0.5421	0.5798	0.7303	0.6191	0.6556
w/o frequencies	0.6380	0.5364	0.5742	0.7148	0.6121	0.6455
w/o all tokens within (w_{i-2}, \dots, w_i)	0.6281	0.5217	0.5609	0.7341	0.6077	0.6498
w/o all tokens within (w_i, \dots, w_{i+2})	0.6144	0.5328	0.5624	0.7022	0.6197	0.6453
w/o tokens that contained all features within (w_{i-1}, \dots, w_{i+1})	0.6414	0.5356	0.5742	0.7264	0.6091	0.6472

Table 5. Results for the ablation experiments in sentiment classification towards aspects (task C)

	Restaurants			Cars		
	macro P	macro R	macro F	macro P	macro R	macro F
All features	0.4582	0.3729	0.4081	0.4481	0.3761	0.4001
w/o character n-grams	0.4479	0.3659	0.4000	0.4480	0.3750	0.3994
w/o lexicon-based unigrams	0.4259	0.3651	0.3921	0.4213	0.3669	0.3869
w/o aspect-based bigrams	0.4261	0.3396	0.3728	0.4380	0.3746	0.3951
w/o context n-grams	0.4355	0.3586	0.3906	0.4370	0.3717	0.3941
w/o lexicon- based scores	0.4629	0.3681	0.4050	0.4374	0.3747	0.3959

Table 6. Results for feature ablation experiments in categorization of aspects (task D)

Combinations of features	Restaurants			Cars		
	P	R	F	P	R	F
word n-grams	0.7650	0.7193	0.7388	0.6554	0.6060	0.6219
word n-grams + single cumulative score	0.8185	0.7705	0.7914	0.6800	0.6296	0.6461
word n-grams + domain-specific scores	0.8960	0.8414	0.8653	0.6854	0.6355	0.6521

The experiments for task D are presented in Table 6. Through these feature ablation experiments we show that most important features are the domain-specific features, that are based on pointwise mutual information for the category and include four different calculations of scores in the context of the aspect term.

5. Conclusion

In this paper we described supervised methods for sentiment analysis of user reviews about restaurants and cars. In extraction of explicit aspects (task A) we proposed the method based on syntactic and statistical features incorporated in the Conditional Random Fields model. The method did not show any significant improvement over the official baseline. In extraction of sentiments towards explicit aspects (task C) our method was based on the Maximum Entropy model on a set of lexicon-based features and two types of term frequency features: context n-grams and aspect-based bigrams. We demonstrated that by using these features, classification performance increases from baseline macro-averaged F-measures of 0.267 to 0.408 for restaurants and of 0.265 to 0.4 for cars. In categorization of explicit aspects into aspect categories (task D) we proposed the SVM classifier, based on unigram features and pointwise mutual information to calculate category-specific score. We achieved 65.2% of macro-averaged F-measure for cars and 86.5% for reviews about restaurants in task D. This method ranked first among 4 teams in both subject domains. For future work we plan to provide error analysis of the described methods.

Acknowledgments

This work was funded by the subsidy of the Russian Government to support the Program of competitive growth of Kazan Federal University and supported by Russian Foundation for Basic Research (RFBR Project 13-07-00773).

References

1. *Blinov P., Klekovkina M., Kotelnikov E., Pestov O.* (2013), Research of lexical approach and learning methods for sentiment analysis, *Computational Linguistics and Intellectual Technologies*, Vol. 2(12), pp. 48–58.
2. *Chernyshevich M.* (2014), IHS R&D Belarus: Cross-domain Extraction of Product Features using Conditional Random Fields, *SemEval 2014*, pp. 309–313.
3. *Chetviorkin I., Loukachevitch N.* (2013), Evaluating Sentiment Analysis Systems in Russian, *ACL 2013*, p. 14.
4. *Choi Y., Cardie C.* (2010), Hierarchical sequential learning for extracting opinions and their attributes, *Proceedings of the ACL 2010 conference short papers*, pp. 269–274.
5. *Jakob N., Gurevych I.* (2010), Extracting opinion targets in a single-and cross-domain setting with conditional random fields, *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pp. 1035–1045.

6. Frolov A. V., Polyakov P. Yu., Pleshko V. V. (2013), Using semantic filters in application to book reviews sentiment analysis, available at: www.dialog-21.ru/digests/dialog2013/materials/pdf/FrolovAV.pdf
7. Kiritchenko S., Zhu X., Cherry C., Mohammad S. M. (2014), NRC-Canada-2014: Detecting aspects and sentiment in customer reviews, *SemEval 2014*, pp. 437–442.
8. Lafferty J., McCallum A., Pereira F. C. (2001), Conditional random fields: Probabilistic models for segmenting and labeling sequence data, *Proceedings of the 18th International Conference on Machine Learning 2001 (ICML 2001)*, pp. 282–289.
9. Liu B. (2012), Sentiment analysis and opinion mining, *Synthesis Lectures on Human Language Technologies*, vol. 5(1), pp. 1–167.
10. Loukachevitch N., Blinov P., Kotelnikov E., Rubtsova Y., Ivanov V., Tutubalina E. (2015), SentiRuEval: testing object-oriented sentiment analysis systems in Russian, *Proceedings of International Conference Dialog-2015*, pp. 3–9.
11. Lu B., Ott M., Cardie C., Tsou B. K. (2011), Multi-aspect sentiment analysis with topic models, *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference*, pp. 81–88.
12. Moghaddam S., Ester M. (2012), On the design of LDA models for aspect-based opinion mining, *Proceedings of the 21st ACM international conference on Information and knowledge management*, pp. 803–812.
13. Pang B., Lee L., Vaithyanathan S. (2002), Thumbs up?: sentiment classification using machine learning techniques, *Proceedings of the ACL-02 conference on Empirical methods in natural language processing*, vol. 10, pp. 79–86.
14. Pang B., Lee L. (2008), Opinion mining and sentiment analysis, *Foundations and trends in information retrieval*, vol. 2(1–2), pp. 1–135.
15. Pontiki M., Papageorgiou H., Galanis D., Androutsopoulos I., Pavlopoulos J., Manandhar S. (2014), Semeval-2014 task 4: Aspect based sentiment analysis, *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pp. 27–35.
16. Popescu A. M., Etzioni O. (2007), Extracting product features and opinions from reviews, *Natural language processing and text mining*, pp. 9–28.
17. Poria S., Cambria E., Ku L. W., Gui C., Gelbukh A. (2014), A rule-based approach to aspect extraction from product reviews, *SocialNLP 2014*, pp. 28–37.
18. Turney P. D. (2002), Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews, *Proceedings of the 40th annual meeting on association for computational linguistics*, pp. 417–424.
19. Zhao Y., Qin B., Liu T. (2014), Clustering Product Aspects Using Two Effective Aspect Relations for Opinion Mining, *Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data*, pp. 120–130.