ВЫДЕЛЕНИЕ АСПЕКТНЫХ ТЕРМИНОВ В ОТЗЫВАХ С ИСПОЛЬЗОВАНИЕМ МОДЕЛИ УСЛОВНЫХ СЛУЧАЙНЫХ ПОЛЕЙ

Рубцова Ю. В. (yu.rubtsova@gmail.com),
Кошельников С. А. (koshelnikovsa@gmail.com)

Ключевые слова: извлечение аспектов, CRF, извлечение мнений, отзывы пользователей

ASPECT EXTRACTION USING CONDITIONAL RANDOM FIELDS

Rubtsova Y. V. (yu.rubtsova@gmail.com),
Koshelnikov S. A. (koshelnikovsa@gmail.com)

This paper describes the aspect extraction system that was presented at SentiRuEval-2015: aspect-based sentiment analysis of users' reviews in Russian. The proposed system uses a conditional random field algorithm for extracting aspects mentioned in the text. We used a set of morphological and syntactic features for machine learning and demonstrated that using lemmas as a feature can improve aspect extraction results. The system was used to perform two subtasks, Task A—automatic extraction of explicit aspects and Task B—automatic extraction of all aspects (explicit, implicit and sentiment facts), and tested on two domains—restaurants and cars. Both subtasks, A and B, in both domains have been completed with quite a high level of precision which meant that the system was capable of rather accurate recognition of aspect terms. But lower recall results implied that the system found enough aspect terms that could not be treated as aspects according to the gold standard. Our systems performed competitively and showed the results comparable to those of the other 10 participants.

Key words: aspect detection, aspect extraction, CRF, opinion mining, reviews

1. Introduction

With the popularity of blogs, social networks, and sites with user reviews of products and services growing every year web users post more and more reviews. As a result an enormous pool of reviews, evaluations, and recommendations in various domains has been accumulated that data attracts attention of both the researchers dealing with opinion mining, sentiment analysis and trend recognition and businessmen who are more interested in the practical application of reputation marketing.

Automatic sentiment analysis is mostly used at the following levels:

- Document level (Turney, 2002; Pang et.al, 2002; Rubtsova, 2014),
- Sentence or phrase level (Wilson et.al, 2009),
- Aspect level (Liu 2012; Zhang, Liu, 2014; Marrese-Taylor et.al, 2014).

As a rule people express their opinions not on the product or service as a whole but on some part, feature or characteristic thereof and that is the aspect that shall be extracted from the text and subjected to sentiment analysis. The aspect-level sentiment analysis can give us much more useful information on the author's opinion on various features of the product or service under analysis than sentiment analysis of the whole text.

Dialogue conference included Dialogue Evaluation section: evaluation of sentiment analysis systems for the Russian SentiRuEval (Loukachevitch et.al, 2015). The participants of the evaluation were to perform the following 5 subtasks:

A. Extract explicit aspects from the offered review,
B. Extract all the aspects from the offered review,
C. Perform sentiment analysis of the explicit aspects,
D. Categorize the aspects terms by predefined categories,
E. Evaluate the aspects categories as related to the offered review in general.

This paper describes the system that was used to perform Tasks A and B during SentiRuEval competition.

The rest of the paper is structured as follows. In Section 2 we discuss the current state of the art and different mechanisms of aspects extraction from product reviews. In Section 3 we describe our system. Section 4 demonstrates the performance of our system as compared to the results of systems of other SentiRuEval participants. Section 5 presents details conclusions and prospects of the future development.

2. Related work

There are four major approaches to extract aspects from texts. The first one is based on the frequency of nouns and/or noun phrases. Commonly people use similar terms to describe the features and their attitude to the products and differing terms used to describe other details (situation, required accompanying information) in their comments. Thus, counting frequency of the most common nouns and/or phrases in the texts of the same domain helps to extract explicit aspect terms from a large number of reviews (Hu and Liu, 2004). The precision level of that algorithm later has been improved by 22% (Popescu and Etzioni, 2005). As common words appear frequently in texts and are often defined as aspects, a filtering mechanism was invented to exclude most common non-aspect nouns and/or phrases from the analysis results (Moghaddam and Ester, 2011).

The second approach is based on simultaneous extraction of both sentiment words (user opinions) and aspects. As any opinion is expressed in relation to an object, by looking for sentiment words we can find aspects they relate to. Hu and Liu used this approach to find low-frequency aspects (Hu and Liu, 2004). Another approach is supervised machine learning. Generally, for the purposes of aspect extraction supervised machine learning is focused on sequence labeling tasks because aspects and opinions on the products are often interrelated and constitute a sequence of words. The most wide-spread methods of supervised machine learning are hidden Markov modeling (HMM) (Jin et al., 2009) and conditional random fields (CRF) (Lafferty et al., 2001; Sutton and McCallum, 2006; Jakob and Gurevych, 2010). The fourth approach is unsupervised machine learning or topic modeling. Topic modeling assumes that each document consists of a mixture of topics and each topic is a probability distribution (Titov and McDonald, 2008; Brody and Elhadad, 2010). The most works on aspect extraction with the use of topic modeling approach are based on the methods of extended probabilistic latent semantic analysis (pLSA) model (Hofmann, 2001) and latent Dirichlet allocation (LDA) model (Blei et al., 2003).

To perform complex tasks such as simultaneous aspect extraction and sentiment analysis or simultaneous aspect extraction and categorization, one can use combination of different approaches such as max entropy и latent Dirichle allocation (Zhao W. X. et al, 2010) or semi supervised model with the topic modeling approach when user provides some seed words for a few aspect categories (Mukherjee and Liu, 2012).

3. System description

We participated into two evals:

- Extract the explicit aspects, i.e. extract a part of the object under analysis or one of its characteristics such as *engine* for the domain of cars or *service* for the domain of restaurants,
- Extract all the aspects of the object under analysis that includes extraction of explicit aspects, implicit aspects (an aspect + the author's unambiguous opinion on the aspect) and sentiment facts (when the author uses no opinion expressions but specifies a fact that unambiguously reveals his or her attitude to the object).

To extract opinion targets or aspects from sentences containing opinion expressions, we utilized CRF. CRF shows comparatively good results for the task of aspect extraction from reviews. For instance, for SemEval-2014 shared task related to aspect-based Sentiment Analysis, two best results have been obtained by systems that were based on CRF (Pontiki rt al., 2014).

Conditional Random fields is proposed as an undirected sequence model, which models a conditional probability $p(Y|X)$ over hidden sequence Y given observation sequence X. That is, the conditional model is trained to label an unknown observation sequence X by selecting the hidden sequence Y which maximizes $p(Y|X)$. As a software implementation of CRF, we utilized the Mallet tool (McCallum, 2002).

3.1. Pre-processing

Jakob and Gurevych (Jakob and Gurevych, 2010) represented the possible labels following the Inside-Outside-Begin (IOB) labelling schema: B-Target, identifying the beginning of an opinion target; I-Target, identifying the continuation of a target, and O for other (non-target) tokens. Therefore as we used sequential labeling, we assigned a label to each word in the sentence where *s-e* indicated the start of an explicit aspect term, *c-e* indicated the continuation of an explicit aspect term, *s-i* indicated the start of an implicit aspect term, *c-i* indicated the continuation of an implicit aspect term (just as for facts-terms: *s-f* for start fact, *c-f* for continuation fact) and *O* indicated a non-aspect term.

To extract syntactic features (e.g. POS and lemma) described in the next section, we used TreeTagger for the Russian language (Sharoff et al., 2008).

We also noticed that car brands are often written in the Latin alphabet and/or contain numbers such as Nissan Micra or VAZ 2109. So for the collection of cars we added the rules that made it possible to recognize a full car name (or brand) as a single explicit term. As you can see in Table 3, this had some positive results—the System was ranked 3rd by the exact matching variant of F-measure.

We also converted all the capital letters into lowercase as the software tools may take *Engine* and *engine* as two different aspects, which is not true.

3.2. Features

Word

Strings of the current token were used as features. We extracted one previous and one subsequent word and used them as additional word features to get more information on the context the word is used in.

POS

The part-of-speech (POS) tag of the current token was used as a feature. Aspect terms are often expressed by nouns. POS tagging adds useful information on the part of speech the word belong to. To determine the part of speech we used TreeTagger—a tool that performs complete syntactic analysis. We reduce complete morphologic analysis up to the parts of speech such as *N* for *engine* and *V* for *driving*.

Lemma

The lemma of the current token was used as a feature. Due to the enormous number of word-forms in Russian language we added the normal form of word as a feature. To extract lemmas we also use a TreeTagger.

3.3. Architecture

We built two systems:

- System 1: CRF with all the above-mentioned labels. We used *s-e, c-e and O* labels for explicit aspect extraction to perform Task A and *s-e, c-e, s-i, c-i, s-f, c-f, O* to extract all the aspects for Task B.

- System 2: Combination of the results of two CRFs—CRF for extraction of explicit aspect terms and CRF for extraction of implicit aspect terms + sentiment facts terms (not explicit).
  Task A was performed using System 1 and Task B—using both systems.

4. Results

The results of Tasks A and B were evaluated by F-measure. Two cases of F-measure were calculated: exact matching and partial matching. Macro F1-measure means in this case calculating F1-measure for every review and averaging the obtained values. To measure partial matching, the intersection between gold standard and extracted term was calculated. Tables 1 to 4 demonstrate how the System performance of Task A and Tables 5 to 8 refer to performance of Task B. The results of the System were compared to the baseline and the two best results of SentiRuEval participants.

As you can see from Table 1 to 4, the System demonstrated high precision level in both domains (2nd position in Task A for both cars and restaurants by Precision metrics). It shall be noted that in the domain of cars the results were better when lemma feature was not in use—it may be concerned to pre-processing rules to the car collection. In Task B both systems built also showed a rather high precision level (see Table 5–8). In the domain of restaurants system 1 with word+pos+lemma features ranked 3rd amount all the participants by the partial matching case of F-measure.

Table 1 Task A results, Restaurant domain, exact matching

System	Precision	Recall	F-measure
baseline	0.557	0.6903	0.6084
№1	0.7237	0.5738	0.6319
№2	0.6358	0.6327	0.6266
Word+POS	0.661	0.515	0.5704
+lemma	0.6674	0.5417	0.5899

Table 2 Task A results, Restaurant domain, partial matching

System	Precision	Recall	F-measure
baseline	0.658	0.696	0.6651
№1	0.8078	0.6165	0.728
№2	0.7458	0.7114	0.7191
Word+POS	0.738	0.563	0.6277
+lemma	0.7485	0.5937	0.652

Table 3 Task A results, Car domain, exact matching

System	Precision	Recall	F-measure
baseline	0.5747	0.6287	0.5941
№1	0.76	0.6218	0.6761
№2	0.6619	0.656	0.6513
Word+POS	0.7109	0.5454	0.6075
+lemma	0.704	0.5785	0.6256

Table 4 Task A results, Car domain, partial matching

System	Precision	Recall	F-measure
baseline	0.7449	0.6724	0.6966
№1	0.7917	0.7272	0.7482
№2	0.8561	0.6551	0.7304
Word+POS	0.797	0.6047	0.6747
+lemma	0.7908	0.6485	0.6991

Table 5 Task B results, Restaurant domain, exact matching

System	Precision	Recall	F-measure
baseline	0.546577	0.647729	0.587201
№1	0.609432	0.600621	0.600128
№2	0.733599	0.513197	0.596179
System 1 Word+POS	0.639256	0.456334	0.52577
+lemma	0.639798	0.487202	0.546905
System 2 Word+POS	0.652145	0.458471	0.531644
+lemma	0.67152	0.491622	0.56153

Table 6 Task B results, Restaurant domain, partial matching

System	Precision	Recall	F-measure
baseline	0.671626	0.593093	0.619285
№1	0.756213	0.610754	0.667928
№2	0.668677	0.637097	0.645234
System 1 Word+POS	0.710428	0.493393	0.5692
+lemma	0.709915	0.529354	0.595303
System 2 Word+POS	0.724649	0.457863	0.547813
+lemma	0.752364	0.493553	0.585126

Table 7 Task B results, Car domain, exact matching

System	Precision	Recall	F-measure
baseline	0.597886	0.589612	0.588623
№1	0.7701	0.553546	0.636623
№2	0.656321	0.616423	0.630149
System 1 Word+POS	0.690826	0.476309	0.556107
+lemma	0.670594	0.518742	0.578086
System 2 Word+POS	0.718995	0.482064	0.568331
+lemma	0.701193	0.520375	0.589311

Table 8 Task B results, Car domain, partial matching

System	Precision	Recall	F-measure
baseline	0.783254	0.605976	0.674288
№1	0.814283	0.650998	0.714762
№2	0.795431	0.646999	0.704189
System 1 Word+POS	0.793637	0.53216	0.625502
+lemma	0.777257	0.584768	0.656113
System 2 Word+POS	0.808562	0.509979	0.61308
+lemma	0.782394	0.558153	0.638947

4.1. Error analysis

An analysis of the errors indicated some common mistakes: not recognized and excessively recognized. In general there is one more type of error for the task of aspect extraction—partially recognized aspect terms. Due to provided evaluation scripts we won't be able to observe third type of mistake. From Table 9, we can find that a major bunch of errors is related to not recognized aspect terms.

Table 9. Error type distribution for the task A (exact matching)

	Restaurants	Cars
Word+POS		
not recognized	65%	68%
excessively recognized	35%	32%
Word+POS+lemma		
not recognized	63%	65%
excessively recognized	37%	35%

We can also observe that adding Lemmas as a CRF feature leads to increasing excessively recognized terms. We compared two our systems and find out that second one can better deal with collocation. For instance it extracted "duck soup" («суп из утки») instead of just "soup" («суп») extracted by system 1. However collocations extraction is also a drawback of system 2 because occasionally it extracts to much irrelevant terms. For example "sea food pasta to husband" («пасту с морепродуктами, мужу»).

In the future, we would like to experiment with additional statistical and lexical features of CRF. Using additional text collections and topic modeling preprocessing can also make further improvements.

5. Conclusions

We presented two aspect extraction systems built on the basis of conditional random field algorithm. Realization of these systems demonstrated that preprocessing and use of lemmas for the Russian language as a CRF feature shows comparatively good the overall F-measure. The performance of our systems was comparable to the best results of SentiRuEval participants. Subsequently we are going to add statistical methods as a CRF feature. We are also planning to make a research and find a way to improve the recall results without reduce a precision.

References

1. *Blei D. M., Ng A. Y., Jordan M. I.* (2003). Latent dirichlet allocation. the Journal of machine Learning research, 3, 993–1022.
2. *Brody S., Elhadad N.* (2010). An unsupervised aspect-sentiment model for online reviews. In Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pp. 804–812.
3. *Hofmann T.* (2001). Unsupervised learning by probabilistic latent semantic analysis. Machine learning, 42(1–2), pp. 177–196.
4. *Hu M., Liu B.* (2004). Mining and summarizing customer reviews. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 168–177.
5. *Jakob N., Gurevych I.* (2010). Extracting opinion targets in a single-and cross-domain setting with conditional random fields. In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, ACL, pp. 1035–1045.
6. *Jin W., Ho H. H., Srihari R. K.* (2009, June). OpinionMiner: a novel machine learning system for web opinion mining and extraction. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 1195–1204.
7. *Lafferty J., McCallum A., Pereira F.* (2001). Conditional random fields: probabilistic models for segmenting and labeling sequence data. In Proceedings of International Conference on Machine Learning (ICML-2001).

8. *Liu B.* (2012). Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5(1), pp. 1–167.
9. *Loukachevitch N. V., Blinov P. D., Kotelnikov E. V., Rubtsova Yu. V., Ivanov V. V., Tutubalina E.* (2015), SentiRuEval: Testing Object-oriented Sentiment Analysis Systems in Russian, Proceedings of International Conference Dialog.
10. *Marrese-Taylor E., Velásquez J. D., Bravo-Marquez F.* (2014). A novel deterministic approach for aspect-based opinion mining in tourism products reviews. Expert Systems with Applications, 41(17), pp. 7764–7775.
11. *McCallum A. K.* (2002). MALLET: A Machine Learning for Language Toolkit.
12. *Moghaddam S., Ester M.* (2011). ILDA: interdependent LDA model for learning latent aspects and their ratings from online product reviews. In Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval, pp. 665–674
13. *Mukherjee A., Liu B.* (2012). Aspect extraction through semi-supervised modeling. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1, pp. 339–348
14. *Pang B., Lee L., Vaithyanathan S.* (2002). Thumbs up?: sentiment classification using machine learning techniques. In Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10, pp. 79–86.
15. *Pontiki M., Papageorgiou H., Galanis D., Androutsopoulos I., Pavlopoulos J., Manandhar S.* (2014). Semeval-2014 task 4: Aspect based sentiment analysis. In 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. In Natural language processing and text mining, pp. 9–28.
17. *Rubtsova Yu. V.* (2014). Development and research domain independent sentiment classifier, In SPIIRAS Proceedings, 5(36), pp. 59–77.
18. *Sharoff S, Kopotev M, Erjavec T, Feldman A, Divjak D.* (2008) Designing and evaluating Russian tagsets. In LREC.
19. *Sutton C., McCallum A.* (2006). An introduction to conditional random fields for relational learning. Introduction to Statistical Relational Learning. MIT Press.
20. *Titov I., McDonald R.* (2008). Modeling online reviews with multi-grain topic models. In Proceedings of the 17th international conference on World Wide Web, ACM, pp. 111–120.
21. *Turney P. D.* (2002). Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In Proceedings of the 40th annual meeting on association for computational linguistics, ACL, pp. 417–424.
22. *Wilson T., Wiebe J., Hoffmann P.* (2009). Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. Computational linguistics, 35(3), 399–433.
23. *Zhang L., Liu B.* (2014). Aspect and entity extraction for opinion mining. In Data mining and knowledge discovery for big data, pp. 1–40.
24. *Zhao W. X., Jiang J., Yan H., Li X.* (2010). Jointly modeling aspects and opinions with a MaxEnt-LDA hybrid. In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, ACL, pp. 56–65.