Computational Linguistics and Intellectual Technologies:
Proceedings of the International Conference "Dialogue 2016"

Moscow, June 1–4, 2016

EXTRACTING ASPECTS, CATEGORY AND SENTIMENT OF ASPECTS IN RUSSIAN USER REVIEWS IN RESTAURANTS DOMAIN

Mashkin D. O. (daniil.mashkin@gmail.com)

Vyatka State University, Kirov, Russia

In this paper we present our approach to the SemEval 2016 Task on the aspect-based sentiment analysis of user reviews and datasets with the rules of evaluation. We describe the possibility of the subtasks solution with the help of the classification pipeline. Our research examines the possibility of the sequence labeling and the usage of the supervised machine learning CRF model for the aspect term extraction and for the category detection. For the aspect term polarity detection we used the search of neighboring words in the sentiment dictionary.

Key words: SemEval, aspect-based sentiment analysis, sequence labeling, aspect term extraction, aspect term category detection

1. Introduction

Many users willingly leave reviews about goods and services on different websites. The opinion mining and the sentiment analysis has become important tasks of Natural Language Processing. There exist detailed overviews in these fields (Pang and Lee, 2008; Liu, 2012; Pavlopoulos, 2014)

The sentiment analysis makes it possible to determine the aspects to which the opinion is expressed. For example, in the restaurant domain these aspect categories may be: service, food and restaurant in general. The aspect extraction was described for the first time in (Hu and Liu, 2004). The authors also determined the difference between *explicit* and *implicit* aspects. This task is called the aspect-based sentiment analysis and includes the following subtasks (Liu, 2012):

- the aspect term extraction;
- the aspect category detection;
- the aspect term polarity detection.

This paper describes the possibility of the subtasks solution with the help of the classification pipeline, which consists of Condition Random Fields (CRF) (Sutton and McCallum, 2011) models and Naive Bayes (NB). This paper examines the sequence labeling application for the task of the aspect term extraction and for the aspect category detection.

Section 2 gives a brief overview of previous works on the aspect-based sentiment analysis. Section 3 describes the datasets provided by the organizers of the international evaluation SemEval 2016 Task 5: Aspect-based sentiment analysis[1]. Section 4 describes Subtask 1 (sentence level ABSA) solution algorithm. Section 5 presents the results of the algorithm work. Section 6 describes the error analysis. Section 7 outlines the conclusions and directions for future work.

2. Related work

Researches in the field of the aspect-based sentiment analysis has seen mush interest thanks to the SemEval evaluations (Pontiki et al., 2015) and Dialogue (Loukachevitch et al. 2015). The solutions which the researchers suggest are state-of-the-arts.

There exist some approaches for the solution of the subtask of the aspect term extraction (Liu, 2012):

- supervised learning;
- frequent nouns and noun phrases;
- relation-based methods;
- topic modeling.

Hu and Liu (2004) was first to consider frequent nouns and noun phrases as possible object aspects, towards which the opinion is expressed. They used the relation-based methods for the rare aspects extraction. This work (Hu and Liu, 2004) served as the base for further researches. Popescu et. al. (2005) could achieve the improvement of exactness, removing noun phrases, which cannot be the object aspects. Moghaddam and Ester (2012) solved this subtask with the help of the LDA topic models.

The solution of the subtask of the aspect term extraction as sequence labeling task with the help of CRF is one of the most popular (Jakob and Gurevych, 2010; Kiritchenko et al., 2014; Chernyshevich, 2014; Hamdan et. al., 2015; Toh and Su, 2015; Tarasov, 2015). For the subtask of the aspect category detection many authors use the machine learning Support Vector Machines (SVM) method with bag-of-words features (Kiritchenko et al., 2014; Ivanov et al., 2015). For the aspect term polarity detection SVM is also used, as well as manual emotional dictionaries (Pontiki et al., 2015).

---

1 http://alt.qcri.org/semeval2016/task5 — SemEval 2016. Task 5 ABSA.

3. Text data

The SemEval-2016 organizers provided the labeled reviews for the restaurant domain. The train and test reviews are stored in xml format documents. Each review is divided into sentences, where the aspect terms are extracted specifying the following fields:

- *target*—an aspect term, a word or word combination;
- *from*—a symbol number, which an aspect term begins;
- *to*—a symbol number, which an aspect term finishes;
- *category*—the category and the subcategory of an aspect term;

- *polarity*—the polarity of an aspect term.

If the *target* is *NULL*, the aspect term is implicit. Fields *to* and *from* for the implicit aspect terms are equal to zero. Field *category* has a pair of CATEGORY#SUBCATEGORY and can be of 12 types. The category RESTAURANT consists of three subcategories—PRICE, MISCELLANEOUS and GENERAL. Categories FOOD and DRINKS include three subcategories PRICE, QUALITY, STYLE&OPTIONS. The rest three categories AMBIENCE, SERVICE and LOCATION contain only one subcategory GENERAL.

Table 1 presents detailed quantitative characteristics of the review corpus.

Table 1. Detailed quantitative characteristics of the reviews SemEval 2016

	Train	**Test**
Reviews	312	103
Sentences	3,548	1,168
Words	40,094	13,181
Aspect terms	4,089	1,300
Explicit aspect terms	3,159	972
Implicit aspect terms	930	328
Aspect terms of word combination	579	187
Aspect terms of one word	2,580	785

4. Algorithm

This section explains the solution of the SemEval 2016 Task 5 ABSA Subtask 1 with the help of the aspect-based sentiment analysis.

Subtasks Slot 2: Opinion Target Expression (OTE) and Slot 1: Aspect category detection mean the extraction both the explicit and implicit aspect terms. To find out the explicit aspect terms and their categories we used the sequence labeling and the CRF classification pipeline. The implicit aspects terms extraction and the category detection were solved with the help of NB estimators with bag-of-words features. For the aspect polarity detection we used the manual sentiment dictionary and NB estimator.

4.1. Explicit aspect term extraction and category detection

To solve the task of the extraction and the category detection of the explicit aspect terms it is necessary to create a token sequence from the review corpus and to define a set of features.

We used the classification pipeline for the extraction and the category detection of the explicit aspect terms. During the research we tested the possibility of the sequence labeling and the supervised machine learning CRF model for the aspect term extraction and for the category detection. That is why the pipeline consisted of three following each other CRF models (see Fig. 1) and the Adaptive Regularization of Weight Vector (AROW) (Mejer, 2010) algorithm. Each CRF model solved its own subtask: the estimation of the BIO label, the estimation of the category label and the estimation of the subcategory label. Each model received the same set of features and additional features of predictions of the previous model. As the additional feature for the CRF models when defining the category and the subcategory on detection stages we used appropriate to NB models subtask predictions.

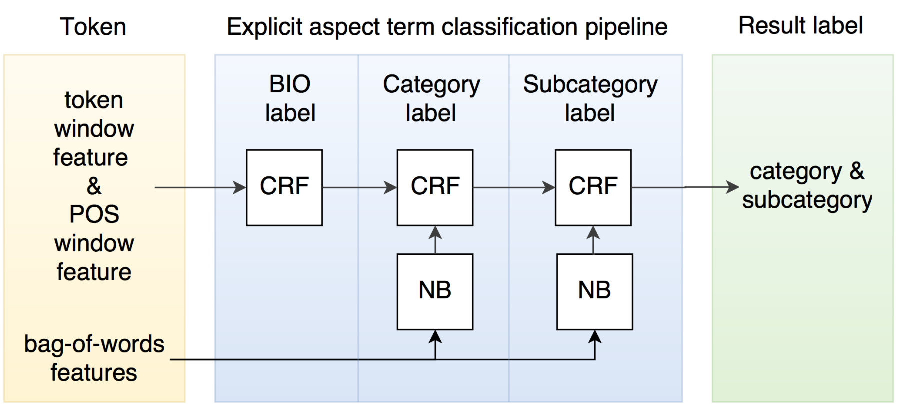

Fig. 1. Classification pipeline for explicit aspect term extraction and category detection

Each CRF model input sequence element has a set of features, consisted of tokens within window (w_{i-1}, …, w_{i+5}) and POS window (p_{i-2}, …, p_{i+5}). The bag-of-word features for NB estimator were formed of the context window (w_{i-2}, …, w_{i+2}) for the categories and the context window (w_{i-4}, …, w_{i+4}) for the subcategories. The left and right context window borders were estimated experimentally.

4.2. Implicit aspect term extraction and category detection

The author can express his opinion towards the object implicitly. For example, the following sentences are found in the restaurant reviews: *«Обязательно придем* *к вам еще!»* ("*We will surely come again!*") or «*И я, и все гости остались очень довольны»* ("*My guests and me liked it lot*"), where the positive opinion towards the restaurant in general is expressed. Such opinions are not attached to a definite word in the sentence and become implicit aspects.

The implicit extraction and the category detection tasks cannot be decided with the help of the sequence labeling methods. The algorithm does not receives the sequence of words, but the sequence of sentences with the bag-of-word features. These subtasks were solved with two NB estimators. The first estimator is used to classify the element of the sequence to one of the two classes, if there is an implicit aspect term or not. The second estimator is used to classify the category and subcategory of the implicit aspect term. The estimator selects the most possible class out of 12 category types.

4.3. Aspect term polarity detection

In Slot 3: Sentiment Polarity each predefined aspect term has to be assigned to one of the following polarity labels: *positive, negative or neutral.*

Kotelnikov and Pletneva (2016) used the search of neighboring words in the sentiment dictionary for the aspect term polarity detection. Positive words had weight +1, negative −1. If the sum of the weights of the neighboring sentiment words was positive, then the positive polarity was assigned to the current token. If the sum of the weights was negative, the negative polarity was assigned to the current token.

In the case of the polarity detection of the explicit aspect terms, context window (w_{i-5}, ..., w_{i+5}) was checked. For the implicit aspect terms the whole sentence was the context window.

If the sum of the weights of neighboring words was equal to zero, then the polarity detection was rated with the help of the NB estimator with the bag-of-word features.

5. Results

Tables 2–4 present the comparison of the classification pipeline described in this article and the baseline provided by the organizers. We got the results of the classification pipeline with different sets of features.

The organizers provided the baseline for each subtask. The SVM estimator was used when preparing the baseline. The estimator received *n* bag-of-words features, extracted from the train data for each tuple (*category, target, polarity*). The frequency-based approach was used when preparing the baseline for Slot 2.

The field *With NB* shows, whether the additional feature, obtained from the NB estimator (Fig.1), was used or not. We used the fields *Token window* (w_{i-n}, ..., w_{i+m}) and *POS window* (p_{i-n}, ..., p_{i+m}) window to rate the borders for the token interval features. We carried out the algorithm on the train and the test datasets.

Table 2. Slot 1: Aspect Category Detection

Features			**Train data**			**Test data**		
With NB	**Token Window**	**POS Window**	**Precision (%)**	**Recall (%)**	**F1 (%)**	**Precision (%)**	**Recall (%)**	**F1 (%)**
Baseline								52.78
no	**2;2**	**2;2**	**73.84**	**62.74**	**67.84**	**54.32**	**44.24**	**48.77**
no	3;3	3;3	77.53	63.03	69.53	51.66	44.52	47.83
no	5;5	5;5	82.86	62.23	71.08	49.55	41.46	45.15
no	1;1	1;1	77.28	51.96	62.14	60.70	33.67	43.31
no	2;5	1;5	77.14	62.74	69.20	48.42	38.58	42.95
yes	2;5	1;5	66.03	53.80	59.29	38.53	31.63	34.74

Table 3. Slot 2: Opinion Target Expression

Features			**Train data**			**Test data**		
With NB	**Token Window**	**POS Window**	**Precision (%)**	**Recall (%)**	**F1 (%)**	**Precision (%)**	**Recall (%)**	**F1 (%)**
Baseline								46.78
no	**2;2**	**2;2**	**77.13**	**61.94**	**68.70**	**46.78**	**42.85**	**44.73**
no	3;3	3;3	82.72	62.91	71.47	38.13	38.13	38.13
no	1;1	1;1	73.10	39.68	51.44	52.48	28.88	37.26
yes	1;5	2;5	80.71	57.59	67.22	36.52	32.03	34.13
no	1;5	2;5	80.06	60.31	68.80	35.41	32.14	33.70
no	5;5	5;5	87.45	59.70	70.96	30.07	29.41	29.73

Table 4. Slot 1&2

Features			**Train data**			**Test data**		
With NB	**Token window**	**POS window**	**Precision (%)**	**Recall (%)**	**F1 (%)**	**Precision (%)**	**Recall (%)**	**F1 (%)**
Baseline								36.88
no	**2;2**	**2;2**	**65.98**	**53.55**	**59.12**	**38.98**	**30.23**	**34.05**
no	3;3	3;3	69.88	54.48	61.23	33.45	28.00	30.48
no	1;1	1;1	68.09	41.91	51.89	42.42	21.53	28.57
no	1;5	2;5	68.98	53.41	60.20	29.10	22.38	25.30
no	5;5	5;5	75.84	54.68	63.54	77.13	61.94	23.57
yes	1;5	2;5	58.88	43.82	50.25	25.43	19.00	21.75

6. Discussion

Tables 2–4 show, that:

- the additional feature containing predicted category from the NB estimator does not improve the results;
- the interval features with left and right borders equal to 2 tokens showed the best results in all subtasks;
- the results of the algorithm on the train data exceeded approximately twice the results on the test data.

All the runs showed the results according to F1-measure worse than baseline. There is a great difference between the runs on the train and the test data. It can indicate that the algorithm is overfitted.

For the comparison of the results, the algorithm using for category detection SVM with bag-of-words features was employed. The comparison of the best result of the CRF classification pipeline, the SVM and the baseline is presented in Table 5.

Table 5. CRF and CRF + SVM F1-measure comparison

	Slot 1	**Slot 2**	**Slot 1&2**
Pipeline CRF	48.77	44.73	34.05
Baseline	52.78	46.78	36.88
CRF + SVM	52.93	45.48	31.41

7. Conclusion

This paper presents the results obtained while using the classification pipeline of the same type estimators and the features which proved to be overfitting. The CRF for the category detection showed the worse results than baseline and SVM. However, we plan to do a deep analysis to find a more accurate set of features for the CRF on the stage of the explicit aspect term extraction.

References

1. *Chernyshevich M.* (2014), IHS R&D Belarus: Cross-domain Extraction of Product Features using Conditional Random Fields, Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), pp. 309–313.
2. *Hu M., Liu B.* (2004), Mining and summarizing customer reviews, Proceedings of the tenth ACM SIGKDD international conference on knowledge discovery and data mining, pp. 168–177.

3. *Ivanov V. V., Tutubalina E. V., Mingazov N. R., Alimova I. S.* (2015), Extracting Aspects, Sentiment and Categories of Aspects in User Reviews about Restaurants and Car, Proceedings of International Conference Dialog, Vol. 2, pp. 22–34.
4. *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.
5. *Kiritchenko S., Zhu X., Cherry C., Mohammad S.* (2014), NRC-Canada-2014: Detecting Aspects and Sentiment in Customer Reviews, Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), Dublin, pp. 437–442.
6. *Kotelnikov E. V., Pletneva M. V.* (2016), Text Sentiment Classification based on Genetic Algorithm and Word and Document Co-clustering, Journal of Computer and Systems Sciences International, Vol. 55 (1), pp. 106–114.
7. *Liu B.* (2012), Sentiment Analysis and Opinion Mining, Synthesis Lectures on Human Language Technologies, Vol. 5 (1).
8. *Mejer A., Crammer K.* (2010), Confidence in Structured-Prediction using Confidence-Weighted Models, Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 971–981.
9. *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.
10. *Pang B., Lee L.* (2008), Opinion mining and sentiment analysis, Foundations and trends in information retrieval, Vol. 2(1–2), pp. 1–135.
11. *Pavlopoulos J.* (2014), Aspect based sentiment analysis.
12. *Pontiki M., Galanis D., Papageorgiou H., Manandhar S., Androutsopoulos I.* (2015), SemEval-2015 Task 12: Aspect Based Sentiment Analysis. Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), pp. 486–495
13. *Popescu A. M., Nguyen B., Etzioni O.* (2005), OPINE: Extracting product features and opinions from reviews, Proceedings of HLT/EMNLP on interactive demonstrations, pp. 32–33.
14. *Sutton, C., McCallum, A.* (2011), An introduction to conditional random fields, Machine Learning, Vol. 4(4), pp. 267–373.
15. *Tarasov D. S.* (2015), Deep Recurrent Neural Networks for Multiple Language Aspect-based Sentiment Analysis of User Reviews, Proceedings of International Conference Dialog, Vol. 2, pp. 53–65.
16. *Toh, Z., & Su, J.* (2015), NLANGP: Supervised Machine Learning System for Aspect Category Classification and Opinion Target Extraction, Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), pp. 496–501.