ТЕСТИРОВАНИЕ ПРАВИЛ ДЛЯ СИСТЕМЫ АНАЛИЗА ТОНАЛЬНОСТИ

Кузнецова Е. С. (knnika@yandex.ru)

ГК «Геострим», Москва, Россия

Лукашевич Н. В. (louk_nat@mail.ru),
Четверкин И. И. (ilia2010@yandex.ru)

МГУ им. М. В. Ломоносова, Москва, Россия

Ключевые слова: анализ тональности, общественно-политическая область, РОМИП, правила

TESTING RULES FOR A SENTIMENT ANALYSIS SYSTEM

Kuznetsova E. S. (knnika@yandex.ru)

GK «Geostream», Moscow, Russia

Loukachevitch N. V. (louk_nat@mail.ru),
Chetviorkin I. I. (ilia2010@yandex.ru)

Lomonosov Moscow State University, Moscow, Russia

The paper is devoted to testing rules useful for sentiment analysis of Russian. First, we describe the working principles of the POLYARNIK sentiment analysis system, which has an extensive sentiment dictionary but a minimal set of rules to combine sentiment scores of opinion words and expressions. Then we present the results achieved by this system in ROMIP-2012 evaluation where it was applied in the sentiment analysis task of news quotes. The analysis of detected problems became a basis for implementation of several new rules, which were then tested on the ROMIP-2012 data.

Key words: sentiment analysis, socio-political domain, ROMIP, rules

Introduction

In recent years sentiment analysis is one of the most rapidly developing branches of computational linguistics. A lot of studies in this area have been conducted for English language, also there already exist working systems (for example, TwitterSentiment), different resources (WordNet, SentiWordNet and others) and natural language processing tools, which simplify the task (Liu, 2010; Pang, Lee, 2008).

For Russian language the sentiment analysis task is complicated by the lack of analogues of the above mentioned resources and tools in Russia and a significantly smaller number of related research papers. Recently, there is some growth of interest in the sentiment analysis task for Russian, both from organizations and researchers. In connection with this interest, the evaluation of sentiment analysis systems for Russian language was organized in 2011 and 2012 as a part of the Russian Information Retrieval Seminar — ROMIP (Chetviorkin et al., 2012; Chetviorkin, Loukachevitch, 2013).

There are two main approaches to the sentiment analysis task (Liu, 2010; Pang, Lee, 2008):

- Machine learning methods, when a system is trained using a labeled text collection,
- Dictionary-based methods, which are based on usage of opinion lexicons, linguistic rules and taking into account contexts of the words (Taboada et al., 2011; Pazelskaya, Solovyev, 2011).

In this paper we investigate the influence of linguistic rules on the quality of sentiment analysis in Russian on the example of the POLYARNIK system. At first, we describe the basic principles of the system, which has an extensive dictionary of opinion words and expressions, but a minimal set of rules to combine them (section 2). Then we provide the results of the POLYARNIK at ROMIP-2012 news-based sentiment analysis task and analyze revealed problems (section 3). Finally, we add a set of rules taking into account multiword sentiment expressions, multiword operators, irrealis markers and evaluate the impact of these added rules to the performance of the system (section 4).

1. Linguistic rules in sentiment analysis systems

The most wide-spread linguistic rules used in the sentiment classification are as follows:

- use of operator words, which increases the base score of a sentiment word (*очень, значительно*) or revert the score to the opposite (*не, нет*…) (Taboada et al., 2011; Pazelskaya, Solovyev, 2011; Chetviorkin, Loukachevitch, 2010);
- aggregation of the scores of sentiment words (Taboada et al., 2011; Liu, 2010)

The rule-based system for sentiment analysis of texts in English is described and tested in the detailed study (Taboada et al., 2011). Additionally, the study describes several rules that take into account the appearance of irrealis markers (words indicating that a certain situation or action is not known to have happened) — the score of sentiment words that occur in the same fragment with them is nullified. The list of these markers includes modals, conditional markers (*if*), some negative sentiment

words like *any*, *anything*, certain verbs (for example, *expect* and *doubt*), questions, and words enclosed in quotes.

The analysis of rules for sentiment classification systems from various Russian and English-language studies on the basis of ROMIP-2012 training collection of news quotations is conducted in the study (Kuznetsova, 2012).

2. POLYARNIK system for sentiment classification of socio-political texts

Sentiment classification of socio-political texts is distinguished from the other domains by the fact that this domain includes a wide variety of topics. This fact hampers the creation of a training collection for machine learning algorithms within realistic time limits. For this reason it is necessary to use engineering technologies for creating sentiment dictionaries, searching dictionary items in texts and combining them with linguistic rules.

The four main dictionaries for socio-political sentiment classification are presented in the POLYARNIK system:

- **Positive words and expressions dictionary** (*новаторский, нравственный, огромный потенциал, очень убедительный* — about 7 thousand words and expressions*)*. The +1 value was assigned to the most part of entries in this dictionary;
- **Negative words and expression dictionary** (*осквернить (desecrate), отсутствие диалога (absence of the dialog)* — about 15 thousand words and expressions*)*. The −1 value was assigned to the most part of these entries;
- **Dictionary of operators**, which can revert or intensify the value of the sentiment expressions. Operators can intensify the base value of expressions (*очень, значительно (very much, considerably),* etc.) or revert the value to the opposite (*не, отменить (not, abolish)* etc.). There are about 140 operators in the dictionary;
- **Dictionary of stop-expressions** — the list of multiword expressions containing sentiment words, but not expressing any overall sentiment, for example, *фонд эффективной политики (foundation of effective politics) etc.* This dictionary contains about 250 items.

To create these dictionaries the following procedure was used:

At the first stage the list of sentiment word candidates was made on the basis of news text collection *Coll_news* (2 million documents). For this purpose the sentiment words and expressions extracted for the movie domain (see Chetviorkin, Loukachevich, 2010) were taken. Documents that contain more than 3 different sentiment words from this list were chosen from the news collection *Coll_news*. It was supposed that if there were at least three sentiment expressions in a text, then this text was likely to contain more sentiment words. In this way the sub-collection *Coll_Sent* with presumably high share of sentiment words was constructed. Lemmas (words in a dictionary form) that appeared at least in 100 documents were extracted from this *Coll_Sent* sub-collection. As a result, thirty thousand lemmas were obtained.

So-called *weirdness* formula (Ahmad et al, 1999) was applied to the sub-collection lemmas:

$$Weirdness = \frac{P_s(w)}{P_g(w)}$$

where $P_s(w)$ — probability of the word appearance in documents of *Coll_sent* sub-collection, $P_g(w)$ — probability of the word appearance in documents of *Coll_news* collection.

The first ten thousand lemmas ordered by *weirdness* were manually refined and it appeared that these lemmas contained more than 30% of sentiment words. These sentiment words formed the first version of the sentiment dictionary.

At the second stage derivational variants (adverbs formed from adjectives; participles formed from verbs, etc.) of the described sentiment words were added to the obtained dictionary.

At the third stage POLYARNIK system was used in the sentiment analysis of news articles, and during the analysis of the results the dictionaries were specified and supplemented. For this purpose we selected big articles with a large share of sentiment words, and thus a single article could become a source of various additional sentiment expressions.

The algorithm of assigning the sentiment value to a document or a text fragment in POLYARNIK system was as follows:

- The words in the processed text are matched with the dictionaries. If an ambiguous match between the text fragment and dictionary entries is found, the longest entry is chosen;
- When an operator word is found at a distance of 5 words (this parameter of the algorithm can be changed), the system is looking to the right of it for a sentiment word, to which this operator can be applied. The search is performed until a punctuation mark is found.

3. The POLYARNIK system in ROMIP-2012 news-based opinion classification task

One of the ROMIP-2012 evaluation tasks was the task of sentiment classification of news-based opinions (direct and indirect speech, further *quotations*) extracted from news articles. The task was to classify quotations as neutral, positive or negative speaker comment about the topic of the quotation. The example of a negative quote (opinionated expressions are underlined): *По мнению эксперта, глава белорусского государства больше всего **боится (afraid of)**, что страну все-таки **лишат права** (deprive the right) провести чемпионат мира по хоккею в 2014 году.*

The results of POLYARNIK system in the news-based sentiment classification are shown in Table 1.

Table 1. The best results in ROMIP-2012
news-based sentiment classification task

Run_ID	Macro_P, %	Macro_R,%	Macro_F1,%	Accuracy,%
POLYARNIK	**62.6**	**61.6**	**62.1**	**61.6**
xxx-11	60.6	57.9	59.2	57.1
xxx-15	56.3	56.0	56.2	58.2

So, POLYARNIK sentiment analysis system obtained the best results in the news-based sentiment classification task, and to our opinion this fact can be explained from the above-described technique of dictionary creation, which allowed us to extract an actual sentiment lexicon for the socio-political domain.

To evaluate additional rule types, which could improve our existing sentiment classification system, we analyzed the reasons of incorrect classification on the basis of 140 news quotations from the training collection. Altogether we have analyzed 40 quotations, which were incorrectly classified by the base version of POLYARNIK system.

Most errors in the news-quotation sentiment classification occurred due to the lack of sentiment words and expressions in the system dictionary — 16 quotations (40 %), including the lack of sentiment expressions or stop-expressions — 13 quotations (32.5%). Examples of these expressions are shown in Table 2.

At the same time, the dictionary can also contain wrong (or not always relevant) data about sentiment of a word or expression. Thus, in the system dictionary it was indicated that expression *не думать (not to think)* has negative sentiment, while this expression did not have this sentiment in one of the analyzed examples: *Сам игрок заявил, что пока не думает о переходе в другой клуб (The player said that he does not think of moving to another club…).*

Table 2. The examples of expressions that are
useful to include in a sentiment dictionary

Expression	Automatically assigned sentiment	Real sentiment
индустрия гостеприимства (hospitality industry)	Positive	neutral
падение личности (fall of personality)	Neutral	negative
заработать удаление (remove from the field)	Neutral	negative
чувствовать себя как дома (feel at home)	Neutral	positive
не видеть в этом смысл (do not see the point)	Neutral	negative

Table 3. Exploring the use cases of the linguistic rules

Quotations	Automatically assigned /real sentiment	Rule needed to improve analysis
Он заявил, что речь идет о **прискорбном недоразумении**, ведь он всегда считал, что литература и **искусство** должны служить **морали**	0/−	Irrealis factor: the sentiment score of the fragment going after expression like «*думал, что*» (*thought that*) should be reduced
Секретарь президиума генсовета «Единой России», запредседателя Госдумы Сергей Неверов в субботу заявил, что партия **не боится раскола в** связи с появлением в ней разных идеологических платформ	0/+	The negation operator should be applied to a group of sentiment words rather than to a single word (*do not afraid of a split*)
Алла Джиоева заявила, что не пойдет на выборы и в качестве избирателя, потому что «не видит в этом смысла и **не верит** в их **объективность**»	0/−	The negation operator should be applied to a group of sentiment words rather that to a single word (*do not believe in their objectivity*)
Один из руководящих сотрудников Hermitage Capital, жизни которого уже **не раз угрожали** из России, утверждает, что явное сотрудничество между британской полицией и российским МВД подвергает его семью **опасности**	+/−	The negation operator is a part of the other operator «*не раз*» (*not a single time*), which should be applied as an intensifier

Additionally, authors of the paper did not agree with assessor classification of 5 quotations (12.5%). Therefore, the misclassification of 25 quotations from 40 examples came from either the dictionary or from the complexity of short utterance sentiment classification.

At the same time it was found that the classification quality of 4 quotations could be improved by developing existing rules and implementing new rules. Table 3 shows the examples of such quotations and indicates rules that could be applied here. The sentiment words and expressions found by system are underlined. The rule definitions are partially based on rules described in the work (Kuznetsova, 2012).

Therefore, as we can see from the analysis, using the additional rules can improve the performance for 4 quotations, what equals to more than 3% of classification accuracy growth, and quotations appear to be the appropriate material for testing different types of rules proposed in the literature.

4. Testing various rules for news-based sentiment classification

To evaluate the impact of linguistic rules to the sentiment classification quality, the following scheme was used. We implement a certain set of rules and test its performance on the ROMIP-2012 quotation training set. After all rules are tested, we evaluate the sentiment analysis system quality on the separate ROMIP-2012 test set.

The set of linguistic rules, which was tested on available quotation sets, can be divided in two groups. The first group is the consideration of various combinations of sentiment words and operators. The second is a group of rules considering the irrealis factor of a text fragment and reducing the sentiment score in such fragment. Further, a fragment (=clause) is a part of a sentence between two punctuation marks.

The first group contains the following set of rules (referred to below as *algo*)
1.1. If an operator word is a part of a longer stop-word or sentiment expression, it does not act as an operator;
1.2. If a group of operators appears together, their scores are multiplied;
1.3. If there is unknown hyphenated word appeared in a text fragment, it is divided in two words and their scores are considered separately;
1.4. If there is a sentiment word sequence, and a negative word appears among them then the score of the whole sequence becomes negative, otherwise positive;
1.5. An operator is applied to the resulting score of a group of sentiment words.

The second group contains the following set of rules (referred below as *rules*). The rules were modified from (Kuznetsova, 2012):
2.1. If there is a question mark in a sentence, and the sentence does not begin with the words *почему/зачем (why, for what),* its sentiment score should be reduced;
2.2. If there is *если (if)* in a clause, the sentiment scores of the words in this fragment that go after *если* should be reduced;
2.3. If there is *ли* particle in a clause, and there is no such words as *чуть/то/вряд/видишь/видите/мало/едва/что* just before *ли,* the sentiment score of the clause should be reduced;
2.4. If there is *бы* particle in a clause then the sentiment score of the words in this clause, which go after *бы,* should be reduced.

Note that it was supposed in (Kuznetsova, 2012) that all the above mentioned rules of the second group, result in nullifying the corresponding fragment sentiment score, but our experiments demonstrated that the reduction of the sentiment score is more efficient. The sentiment score of a fragment containing irrealis is reduced by current algorithms with a certain specified coefficient (in this version 0.4).

Table 4. The results of both groups of rules on
the ROMIP-2012 training collection

	Macro_P, %	Macro_R, %	Macro_F1, %	Accuracy, %
Baseline	60.9	61.0	60.9	60.5
Baseline + rules	61.1	61.3	61.2	60.9
Baseline + algo	61.4	61.5	61.5	61.4
Full composition	**61.5**	**61.6**	**61.6**	**61.5**

Table 4 shows the results of the aforementioned rule group implementation in POLYARNIK system. The evaluation metrics used in ROMIP-2012 (Chetviorkin, Loukachevitch, 2013) are applied here. Table 5 shows how the number of correctly and incorrectly classified quotations changes depending on the rule set.

There were 3893 quotations in the training collection, and the scores of 333 of them changed in case of the full set of rules. Therefore we can see that we managed to improve the system performance without any changes in the sentiment dictionaries.

Table 5. The quality of quotation sentiment classification with various rule sets

	Number of quotations changed to the correct class	Number of quotations changed to the incorrect class	Growth of correctly classified quotations compared to the baseline
Baseline	—	—	—
Baseline + rules	20	7	13
Baseline + algo	53	21	32
Full composition	60	22	38

The new version of POLYARNIK system was applied to the test collection of ROMIP-2012 news sentiment classification task for the final evaluation. Table 6 shows the quality metrics of the system with various groups of rules on the test collection. The resulting quality of the full rule set is less than on the training set, but in general we can see performance improvements for all groups of rules.

Table 6. The results of both groups of rules on
the ROMIP-2012 test collection

	Macro_P, %	Macro_R, %	Macro_F1, %	Accuracy, %
Baseline	62.6	61.6	62.1	61.60
Baseline + rules	62.8	61.9	62.3	61.90
Baseline + algo	63.0	62.2	62.6	62.25
Full composition	62.9	62.2	62.6	**62.32**

Conclusion

In this paper POLYARNIK sentiment analysis system was presented. The system performance yielded the best results in the ROMIP-2012 news-based sentiment classification task, what in our opinion is due to the extensive system dictionaries, which were created beforehand.

Then without any changes to the sentiment lexicon we implemented the set of rules to take into account groups of opinion words and operators and irrealis markers. Using these new rules, the system performed better both on the train and test collections. In prospect we suppose to continue incorporation of different kinds of rules into POLYARNIK system and testing them on the available quotation collections. Furthermore, we plan to examine rules performance in sentiment analysis in specific domains such as movies, books, etc.

Acknowledgements

References

1. *Ahmad K., Gillam L., Tostevin L.* University of Surrey participation in Trec8: Weirdness indexing for logical documents extrapolation and retrieval. In the Proceedings of Eigth Text Retrieval Conference (Trec-8), 1999.
2. *Chetviorkin I., Loukachevich N.* Automatic extraction of domain specific opinion words. Computational Linguistics and Intellectual Technologies. Proc. of International Conference Dialog, 2010, pp. 565–571.
3. *Liu B.* Sentiment analysis and Subjectivity. Handbook of Natural Language Processing, CRC Press, Taylor and Francis Group, Boca Raton, 2010, pp. 1–38.
4. *Chetviorkin I., Braslavski P., Loukachevitch N.* Sentiment analysis track at ROMIP 2011. Computational Linguistics and Intellectual Technologies. Proc. of International Conference Dialog, 2012, pp. 739–746.
5. *Chetviorkin I., Loukachevitch N.* Sentiment analysis track at ROMIP 2012. Computational Linguistics and Intellectual Technologies. Proc. of International Conference Dialog, 2013.
6. *Kuznetsova E. S.* Linguistic support for sentiment analysis of opinionated quotes in Russian. Dipoma thesis of Lomonosov Moscow State University, 2012, Moscow.
7. *Pang B., Lee L.* Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval, 2008, v. 2, n. 1–2, pp. 1–135.
8. *Pang B., Lee L., Vaithyanathan S.* Thumbs up? Sentiment classification using machine learning techniques. In Proceeding of the conference on empirical methods in natural language processing, 2002, Philadelphia, PA, USA, pp. 79–86.
9. *Pazelskaya A., Solovyev A.* method of sentiment analysis in Russian texts. Computational Linguistics and Intellectual Technologies. Proc. of International Conference Dialog, 2010, pp. 510–522.
10. *Taboada M., Brooke J., Tofiloski M., Voll K., Stede M.* Lexicon-based methods for Sentiment Analysis. Computational linguistics, 37(2), 2011.