АВТОМАТИЧЕСКОЕ ОПРЕДЕЛЕНИЕ ТОНАЛЬНОСТИ ОБЪЕКТОВ С ИСПОЛЬЗОВАНИЕМ СЕМАНТИЧЕСКИХ ШАБЛОНОВ И СЛОВАРЕЙ ТОНАЛЬНОЙ ЛЕКСИКИ

Поляков П. Ю. (pavel@rco.ru),
Калинина М. В. (kalinina_m@rco.ru),
Плешко В. В. (vp@rco.ru)

ООО «ЭР СИ О», Москва, Россия

Ключевые слова: определение тональности, анализ мнений, тональность объектов, тональность атрибутов, синтактико-семантический анализ, семантические шаблоны

AUTOMATIC OBJECT-ORIENTED SENTIMENT ANALYSIS BY MEANS OF SEMANTIC TEMPLATES AND SENTIMENT LEXICON DICTIONARIES

Polyakov P. Yu. (pavel@rco.ru),
Kalinina M. V. (kalinina_m@rco.ru),
Pleshko V. V. (vp@rco.ru)

RCO LLC, Moscow, Russia

This paper studies use of a linguistics-based approach to automatic object-oriented sentiment analyses. The original task was to extract users' opinions (positive, negative, neutral) about telecom companies, expressed in tweets and news. We excluded news from the dataset because we believe that formal texts significantly differ from informal ones in structure and vocabulary and therefore demand a different approach. We confined ourselves to the linguistic approach based on syntactic and semantic analysis. In this approach a sentiment-bearing word or expression is linked to its target object at either of two stages, which perform successively. The first stage includes usage of semantic templates matching the dependence tree, and the second stage involves heuristics for linking sentiment expressions and their target objects when syntactic relations between them do not exist. No machine learning was used. The method showed a very high quality, which roughly coincides with the best results of machine learning methods and hybrid approaches (which combine machine learning with elements of syntactic analysis).

Key words: sentiment analysis, object-oriented sentiment analysis, aspect-based sentiment analysis, opinion mining, syntactic and semantic analysis, semantic templates

1. Introduction

The task of automatic sentiment analysis of natural language texts has become extremely in demand. Many commercial companies producing goods and services are interested in monitoring social networking websites and blogs for users' opinions about their products and services. However, until recently there were no tagged text corpora in Russian on which developers could test and compare quality of their methods. This gap was filled by ROMIP and later SentiRuEval sentiment analysis evaluation conferences with their sentiment analysis tracks. However, the task of the previous conferences was to detect general sentiment of a text (for example, see Chetviorkin I., Braslavski P. I., Loukachevitch N. [2]), while at the present conference the task was brand new—object-oriented sentiment analysis, which is more difficult and requires more sophisticated algorithms; for, in case of general sentiment detection, selection of positive and negative terms and defining of their weights are important, while, in case of object-oriented sentiment detection, syntactic relations between a target object and a word expressing sentiment are also of great importance.

Such object-oriented method is not new for us; we have already used similar approach in our previous research. For instance, we evaluated sentiment-oriented opinions in regard to car makes on the material of the LiveJournal blog AUTO_RU (see description of the method in Ermakov A. E. [4]). It should be mentioned, however, that in all the previous cases results had only been evaluated by ourselves. Participation in SentiRuEval gave us a chance to have an independent evaluation of our method and compare our results with other participants'.

In this paper we present results of applying a linguistics-based approach involving syntactic and semantic analysis to the task of automatic object-oriented sentiment analysis. We confined ourselves to a linguistic method only, having excluded machine learning, because it was interesting to see what results a pure linguistic approach without machine learning methods would provide.

The task was to find sentiment-oriented opinions (positive and negative) about telecom companies in tweets.

2. Related Work

Usually object-oriented or aspect-oriented approaches either rely only on statistics-based algorithms, word distance count, machine learning, etc. to find opinion targets (starting with the first work on opinion target extraction by Hu and Liu [5]); or they may use shallow parsing to segment a sentence, find significant conjunctions, negations, and modifiers (ex., Kan D. [7]). Other approaches are looking for syntactic dependency between a sentiment term and its target (ex., Popescu A., Etzioni O. [9]), ignoring sentiment-bearing words which are not syntactically related to any target object. The distinctive feature of our approach is that using a deep linguistic method we take into account not only syntactically related sentiment terms (which provides high precision) but also independent sentiment-bearing words and phrases (which provides high recall).

Some researchers try combine statistical and linguistic methods in order to achieve the best results; for example, in Jakob N., Gurevych I. [6] authors use, among other, the dependency parse tree to link opinion expressions and the corresponding targets; and the experiments show that adding the dependency path based feature yields significant improvement to their method. However, their algorithm is searching for short and direct dependency relations only; therefore, their approach has difficulties with more complex sentences. Furthermore, they do not distinguish between a target object (ex., *camera*), its attributes or parts (ex., *lens cap, strap*), and its qualities (ex., *usability*); and, hence, they label the closest noun phrase as a target of the opinion. In contrast, we use a very basic ontology to distinguish between a target object, attributes, and qualities; and having found a sentiment related to an attribute or quality our algorithm goes down the dependency parse tree searching for a target object. If not found syntactically, the target object is being searched for by a heuristic, based on the clause distance. When the target object is found, the sentiment labeled to its attribute is assigned to the object.

3. Methods

To perform the task we based on our previous researches and solutions. Detailed description of these methods can be found in Ermakov A. E., Pleshko V. V. [3] and Ermakov A. E. [4]. New to the approaches described in [3] and [4] was adding so-called 'Free Sentiment Detection', which will be described in Section 3.2.

The text analysis algorithm has the following stages in regard to the sentiment detection task:

1) Tokenization;
2) Morphological analysis;
3) Object extraction;
4) Syntactic analysis;
5) Fact extraction (use of semantic templates);
6) Free sentiment detection.

Stages 1, 2, and 4 were implemented by standard RCO tools for general text analysis. At stage 3 we paid more attention to the objects concerning the given subject (names of mobile companies, telecom terminology, etc.). Stages 5 and 6 were core to the sentiment detection task and, therefore, will be described in detail.

3.1. Semantic Templates

The main method of sentiment analysis involved usage of semantic templates.

Semantic template is a directed graph representing a fragment of a syntactic tree with certain restrictions applied to its nodes. The syntactic tree of a sentence contains semantic and syntactic relations between words, which are defined by the syntactic parser. The restrictions in the templates can be applied to a part of speech, name, semantic type, syntactic relations, morphological forms, etc. Fact extraction is performed by finding a subgraph in the syntactic tree of a sentence which is isomorphic to the template (with all restrictions applied).

RCO syntactic analyzer, based on the dependency tree approach, has been used. The semantic network built by the syntactic parser is invariant to the word order and voice; for example, sentences (1) *Оператор украл деньги со счета* and (2) *Деньги украдены оператором со счета* will have the same semantic net. Such semantic network constitutes an intermediate representation level between the semantic scheme of a situation and its verbal expression, that is, a deep-syntactic representation, abstracted from the surface syntax.

Settings of the semantic interpreter allow filtering negative and 'unreal' (imperative, conditional, etc.) statements, which don't correspond to real events and should not be analyzed. As a result, examples like (3) *если Билайн будет плохо работать; сеть якобы падает; связь бы обрывалась; не Билайн плохо работает* can be excluded from the sentiment detection.

To decrease the number of templates describing semantic frames, we have so-called auxiliary templates, which add new nodes and relations into the semantic network. In the process of semantic analysis and fact extraction auxiliary templates work before all other templates, so that semantic templates can base on the net built by both the syntactic analyzer and the auxiliary templates. For example, if we interpret phrases like (4) *X does Y, X begins to do Y*, and (5) *X decides to do Y* as equal for a particular semantic frame, instead of creating a semantic template for each example we can have one auxiliary template, which will mark the subject of the main verb as the subject of the subordinate verb, and one simple semantic template—(4) *X does Y.*

Semantic templates can have so-called 'forbidding nodes' which impose restrictions on the context, defining in which context the template should not match. For example, (6) *У Билайна надежная связь* is a positive statement, while adding the adverb *наименее* changes its sentiment to opposite: (7) *У Билайна наименее надежная связь*. By the means of forbidding nodes we can distinguish between these two sentences, stating that the adjective should not be modified by the adverb *наименее*. Usage of forbidding nodes significantly increases the precision of sentiment analysis.

Fig. 1 demonstrates a semantic template used to detect sentiment expressed by a verb or adverb in sentences like: (8) *Билайн ловит хорошо; Интернет летает*.

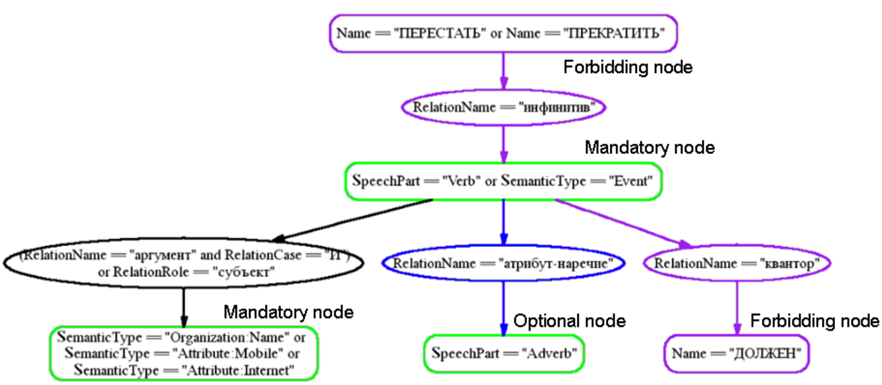

Fig. 1. Example of a semantic template

Nodes contain restrictions on parts of speech (*SpeechPart == "Verb"; SpeechPart == "Adverb"*), lexical items (*Name == "ПЕРЕСТАТЬ" or Name == "ПРЕКРАТИТЬ"*), semantic categories (*SemanticType == "Organization:Name" or SemanticType == "Attribute:Mobile"*). Restrictions on semantic and syntactic relations between words include: relation name (*RelationName == "аргумент»; RelationName == «квантор»*), semantic role (*RelationRole == "субъект»*), case (*RelationCase == "И"*). Forbidding nodes state that the verb expressing sentiment should not be controlled by the verbs *перестать* or *прекратить* or modified by the predicative *должен*. Thus, this template will match the sentence (8) *Билайн хорошо ловит* (which is positive), but not (9) *Билайн перестал хорошо ловить* (which is negative) or (10) *Билайн должен хорошо ловить* (which we consider neutral).

Restrictions of the semantic templates were enriched by the use of special dictionaries (so-called filters), containing vocabulary for positive and negative appraisals. This vocabulary includes nouns, adjectives, verbs, adverbs, and collocations. A word from a filter must be syntactically related to the target of evaluation. Selection of terms for the filters was manual, performed by a linguistic expert. Examples of positive terms: *супербыстрый, шустро, красота, крутяк, блистать, радовать, обеспечивать уверенный прием*. Examples of negative terms: *завышенный, препротивнейший, позорище, тормознутость, обдирать, терять соединение, фигово.*

For example, a set of particular words from the semantic filters are applied to the template in Fig.1 as restrictions: verbs or verbal nouns parameterize the node with the restriction *SpeechPart == "Verb" or SemanticType == "Event"*; adverbs parameterize the node with the restriction *SpeechPart == "Adverb"*, both these nodes have the semantic role 'Appraisal'.

Ultimate targets of evaluation were main Russian mobile phone providers (Beeline, Megafon, MTS, Rostelecom, Tele2), but also users' appraisals of providers' attributes were taken into account (communication quality, mobile Internet, customer service, etc.).

Analyzing users' comments and opinions on social networking sites and forums experts defined a set of attributes which were most frequently mentioned by mobile phone users. Thus, a list of most important things for users was made. Given attributes were divided into three classes: 1) Mobile Attributes—terms strictly connected to the mobile telephony: *SMS, MMS, 3G, LTE, SIM-card, roaming, etc.*; 2) Internet Attributes—terms strictly connected to the Internet: *Internet, ping, etc.*; 3) General Attributes—terms often used related to the mobile telephony but which can also refer to other domains: *call center, signal, network, customer support, balance, etc*. Each list was extended by synonyms and spelling variants (*интернет=инет=и-нет; lte=лте =ltешечка =лте-шечка; баланс счета=состояние счета=средства на счету=деньги на счету, etc*.). When a sentiment related to a certain attribute was detected, given sentiment was also ascribed to the corresponding mobile provider.

In Fig.1 the node with the restriction *SemanticType == "Organization:Name" or SemanticType == "Attribute:Mobile" or SemanticType == "Attribute:Internet"* is parameterized by names of mobile operators, mobile attributes or Internet attributes; the semantic role of the node is 'Target Of Evaluation'.

This method provides a very high precision, though not so high recall.

3.2. 'Free' Sentiment

Although usage of semantic templates provides very good accuracy, this method has its disadvantage—a word expressing sentiment must be in the same sentence as the target of evaluation and must be syntactically related to it. As it is not always so in natural texts, some cases of clearly expressed sentiment will be omitted by this method, and the recall will suffer. This problem becomes extremely significant when we analyze informal texts—forums, social networking websites, blogs, etc. Writing an informal text message, users often disregard punctuation and spelling rules, mistype, because of which the syntactic parser may fail to correctly analyze the structure of a sentence and build a semantic network. Users often express their sentiment through interjections, which are not a part of the syntactic tree; hence the semantic templates are of no use in this case. We call words that express sentiment but have no syntactic relation to the target of evaluation (or such relation has not been built by the parser) 'free sentiment'.

To solve this problem another method has been applied. We used an algorithm which is looking for free sentiment in the text using dictionaries (or profiles) of positive and negative lexicon, and if such sentiment has been found tries to relate it to the target object.

These two methods complement each other, with the semantic template method working first. In this regard, the classifier 'ignores' terms already found and related to the target object by templates, because we assume that the accuracy provided by the semantic templates is close to 100%.

As profiles for positive and negative classes we used corresponding filters, having removed context-dependent sentiment words and leaving only explicit emotional or evaluative vocabulary. For example, we removed verbs *УМЕРЕТЬ, ПРОИГРЫВАТЬ*, because although they are obviously negative in the context like: (11) *интернет умер;* (12) *оператор X проигрывает оператору Y*; but in another context, not related to the mobile telephony, they may be neutral and just state a fact. At the same time we enriched our profiles with interjections and other emotional expressions which cannot be syntactically related to the object of evaluation, for example: (13) *не надо так! что за нах; ни фига себе; ну как так можно, etc*.

Having found a sentiment, our algorithm was looking for an object of evaluation—a name of a mobile company—in the given text and ascribed this sentiment to the target. If several mobile operators were mentioned in the text, the appraisal was ascribed to the nearest operator. If both positive and negative sentiment was detected related to the same mobile provider mentioned, we gave preference to the negative sentiment, regarding positive expressions as sarcasm.

No machine learning had been used. The methods applied were based on linguistic analysis only.

4. Dataset

The training and test collection granted by organizers consisted of 5,000 labeled and 5,000 not labeled tweets containing sentiment-oriented opinions or positive and negative facts about telecom companies.

As the main goal of social networks sentiment analysis is to find sentiment-oriented opinions, we labeled texts containing reprints of news and additionally measured sentiment detection quality for the training collection with news reprints excluded. We excluded news texts from the final dataset because we believe that the difference in structure and vocabulary between formal (news) and informal (posts, blogs, tweets) texts is crucial. As a rule, in news texts authors don't express their attitude openly; news is more likely to contain coverage of events and facts, which can be interpreted as positive or negative for the newsmaker, rather than explicit sentiment; and therefore analyzing news demand a different approach. Furthermore, vocabulary of informal texts is quite different from vocabulary of formal texts.

That is why we additionally estimated the method performance on the collection with news reprints and companies' press releases excluded from the dataset. Since our method is based on linguistic analysis only, we did not use training collection.

5. Results

Initially, for the purpose of estimation of coincidence between assessors we asked our expert to evaluate the test collection manually and marked each reference to mobile phone companies as being positive, negative or neutral. Results of our expert's evaluation are presented in Table 1. F1-measure macro- and micro-averaged was used as a primary evaluation metric [1]. Additionally, for convenience, recall and precision are also present in the tables. As shown in Table 1, the estimation of tweets by our expert differed from one granted by the organizers. We consider the score given by our expert as the highest possible for an automatic sentiment detection system for the given collection. The agreement between our expert and organizers' labeling was higher when we excluded news from the dataset, which confirms our assumption that a different approach should be used for sentiment analysis of news.

Table 1. The estimation of coincidence between expert and assessors

	Macro-average			Micro-average		
	Recall	**Precision**	**F1**	**Recall**	**Precision**	**F1**
With news	0.722	0.686	0.703	0.771	0.728	0.749
Without news	0.785	0.694	0.737	0.831	0.735	0.780

The results of all participants are shown in Fig. 1, our results are highlighted by bold lines and are labeled as "RCO". It is interesting that several methods probably based on different approaches demonstrate very similar high scores of F1 (about 0.5), nevertheless, these scores are sufficiently less than theoretical maximum that corresponds to coincidence between assessors (see bars "Expert" on Fig. 1). It could prove that automatic sentiment detection task is still a challenging problem.

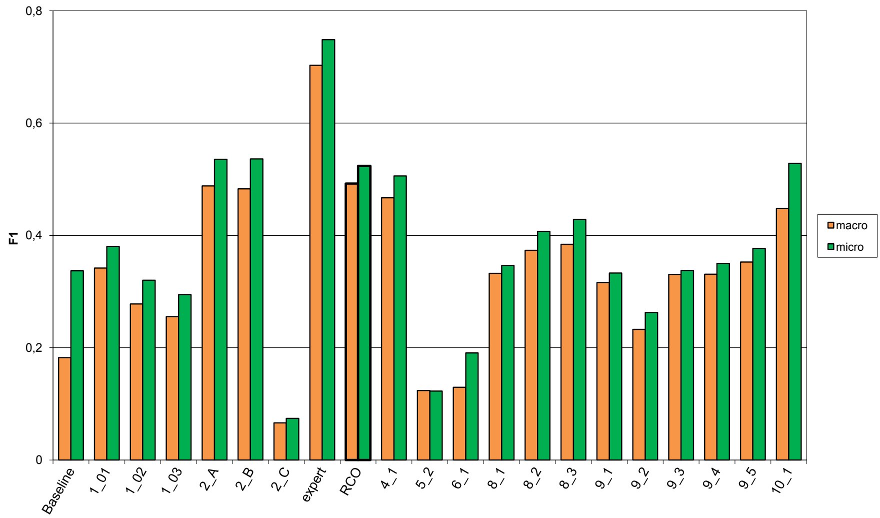

Fig. 2. Macro- and micro-averaged F1 measure calculated on test collection for all participants. The scores for our method are labeled as "RCO". The scores of expert's evaluation are labeled as "expert"

The detailed results of our method are presented in Table 2. We calculated recall, precision and F1 for original collection (labeled as "With news") and for collection with exclusion of messages contained news and press releases (labeled as "Without news"). For comparison, the best scores among the methods of all participants are presented.

Table 2. The performance of our method and best F1 measure among the methods of all participants

	Macro-average			Micro-average		
	Recall	**Precision**	**F1**	**Recall**	**Precision**	**F1**
With news	0.436	0.566	0.480	0.451	0.585	0.509
Without news	0.465	0.562	0.492	0.475	0.583	0.524
Best result			0.492			0.536

6. Conclusion

Our combined linguistic method showed a very high quality, which roughly coincides with the best results of machine learning methods and hybrid approaches (combining machine learning with elements of syntactic analysis). In the future we are planning to add machine learning to our linguistic approach.

References

1. *Blinov P. D., Kotelnikov E. V.* (2014), Using distributed representations for aspect-based sentiment analysis, Dialog '14, Bekasovo.
2. *Chetviorkin I., Braslavski P. I., Loukachevitch N.* (2012), Sentiment analysis track at ROMIP 2011, Bekasovo.
3. *Ermakov A. E., Pleshko V. V.* (2009), Abstract Semantic Interpretation in Computer Text Analysis Systems [Semanticheskaya interpretatsiya v sistemakh kompyuternogo analiza teksta], Information Technologies [Informacionnye tehnologii], Vol. 6, pp. 2–7.
4. *Ermakov A. E.* (2009), Knowledge Extraction from Text and its Processing: Current State and Prospects [Izvlecheniye znaniy iz teksta i ikh obrabotka: sostoyaniye i perspektivy], Information Technologies [Informacionnye tehnologii], Vol. 7, pp. 50–55.
5. *Hu M., Liu B.* (2004), Mining and summarizing customer reviews, International Conference on Knowledge Discovery and Data Mining (ICDM).
6. *Jakob N., Gurevych I.* (2010), Extracting Opinion Targets in a Single-and Cross-Domain Setting with Conditional Random Fields, Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2010).
7. *Kan D.* (2012), Rule-based approach to sentiment analysis at ROMIP '11 , Bekasovo.
8. *Loukachevitch N., Blinov P., Kotelnikov E., Rubtsova Yu., Ivanov V., Tutubalina E.* (2015), SentiRuEval Testing Object-Oriented Sentiment Analysis Systems in Russian.
9. *Popescu A., Etzioni O.* (2005), Extracting product features and opinions from reviews, Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP) .
10. *Polyakov P. Yu., Kalinina M. V., Pleshko V. V.* (2012), Research of applicability of thematic classification to the problem of book review classification. Dialog '12. Naro-Fominsk.
11. *Polyakov P. Yu., Frolov A. V., Pleshko V. V.* (2013), Using semantic categories in application to book reviews sentiment analysis, Dialog'13, Bekasovo.