Sentiment Attitudes and Their Extraction from Analytical Texts

N.L. Rusnachenko [1] N.V. Loukachevitch [2]

Bauman Moscow State Technical University, Moscow, Russia
kolyarus@yandex.ru

Lomonosov Moscow State University, Moscow, Russia
louk_nat@mail.ru

11-14 September, 2018

Sentiment Analysis: genres of documents [1]

Users' reviews or short posts in social networks (Twitter)

- SentiRuEval competition in Russian (related: SemEval, task A)
- Posts are limited and short in length;
- Mostly user reviews $\Rightarrow$ considered a single object for analysis.

Sentiment Analysis: genres of documents [2]

News or analytical reports

- Large amount of named entities (*NE*):

  Ukraine, Russia, Russian Federation, ...
- Large amount of attitudes between *NE*;
- Has complicated structure.

Example

As is apparent in Washington, there is no place for objectivity on the subject of Russia, irrespective of facts and events[1]

- Washington is negative to Russia
- Author is negative to Washington
- Autor attitude towards Russia?

[1] https://www.counterpunch.org/2017/05/26/ukraine-and-the-nato-military-alliance/

Outline

- Corpus of analytical articles **RuSentRel** annotated with sentiment attitudes;
- Experiments on extracting sentiments with machine-learning methods
  - Baselines
  - Features
  - Human performance in the same task

New sentiment-annotated collection RuSentRel [1]

- **RuSentRel**[2][LR18] consisted of analytical articles from Internet-portal `inosmi.ru` (foreign mass media);
- Text attitudes – manual annotation, as triplets:

$$\langle Object, Subject, Label \rangle$$

  - Object – named entity or "author"
  - Subject – named entity
  - Label $\in$ {pos, neg}

[2] `https://github.com/nicolay-r/RuSentRel/tree/v1.0`

The relations between Finland and Sweden can be considered as good. These were the countiries of the same kingdom until 1809. Countries united by the fact that they are not a part of NATO, but alliance partners. Besides, Finland and Sweden increase bidirectional partnership in defence domain.[3]

[3] https://inosmi.ru/politic/20160623/236948867.html

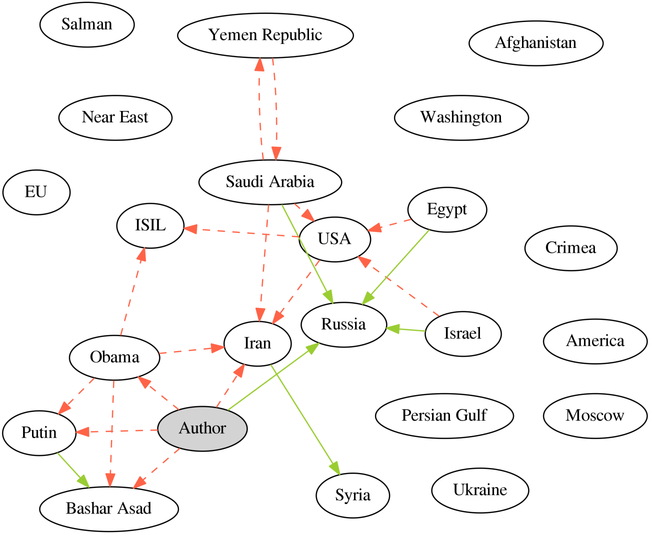
Despite all this, the discussions in Kultarante led to disagreements. Former Swedish Minister of Defense Karin Enström resented by fact that in next month, the Finland president meets the president of Russia in Helsinki. ... By opinion of the second participant of Swedish discussion, Vladimir Putin was not welcome in Sweden.[4]

[4] https://inosmi.ru/politic/20160623/236948867.html

Whole Text Labeling

Obama, Asad, neg
USA, ISIL, neg
Iran, Asad, pos
USA, IRAK, neg
USA, Afganistan, neg
Japan, USA, pos
South Korea, USA, pos
Australia, USA, pos
Author, Obama, pos

Picture of whole text



Named Entities labeling

- Automatic, recognizer based on CRF [ML16];
- List of synonymous *NE* – manually implemented [5].

Russia, Russian Federation, ...
EU, Europe, Eurounion, ...

[5] `https://github.com/nicolay-r/RuSentRel/blob/v1.0/synonyms.txt`

Task

- Classification of attitudes between named entities into three classes: positive, negative, neutral
- Measure: averaged sum of F-measure of positive class and negative class
- The first attempt
  - Summer School "Natural Language Processing and Data mining" (2017)
  - Higher School of Economy

Dataset Statistics

- 73 large analytical articles divided into **Training** and **Test** collections (44 in train, 29 in test);

Average per doc.	Training collection	Test collection
sentences	74.5	137
text attitudes (pos.)	6.23	14.7
text attitudes (neg.)	9.33	15.6
NE	**194**	**300**
NE (unique)	33.3	59.9

Table 1: Statistics of RuSentRel 1.0 corpus

Entities' features [1]

- word2vec similarity between entities
  - vectors of multiword expressions are calculated as the averaged sum of the component vectors;
- the named entity type according to NER recognizer:
  - person, organization, location, or geopolitical entity;
- the presence in the lists of countries or their capitals;
- the relative frequency of a NE or the whole synonym group in the document;
- the order of two named entities;
- Concrete lemmas of named entities **are not used**

Context Features [2]

- the number of sentiment words from **RuSentiLex** vocabulary:
  - the number of pos. words, number of neg. words;
  - *avg* sentiment score of the sentence;
  - *avg* sentiment score **before** the first NE, **between** named entities, and **after** the second NE according to RuSentiLex
- the distance between named entities in lemmas
- the number of other named entities between the target pair
- number of commas between the named entities
- *max*, *min* and *avg* for all features

Experiments

Table 2: Results for sentiment attitudes extraction from RuSentRel corpus

method	*precision*	*recall*	$F_1(P, N)$
Baseline neg	0.03	0.39	0.05
Baseline pos	0.02	0.40	0.04
Baseline distr	0.05	0.23	0.08
Naïve Bayes Gauss	0.06	0.15	0.11
Naïve Bayes Bernoulli	0.13	0.21	0.16
KNN	0.18	0.06	0.09
SVM (GRID)	0.09	**0.36**	0.15
Random forest (GRID)	0.41	0.21	**0.27**
CNN	0.41	0.23	0.31
PCNN	0.42	0.23	0.31
Expert agreement	0.62	0.49	0.55

References

- **Corpus:** https://github.com/nicolay-r/RuSentRel/tree/v1.0
- **Research:** https://github.com/nicolay-r/sentiment-relation-classifiers/tree/tsd_2018

Piecewise CNN

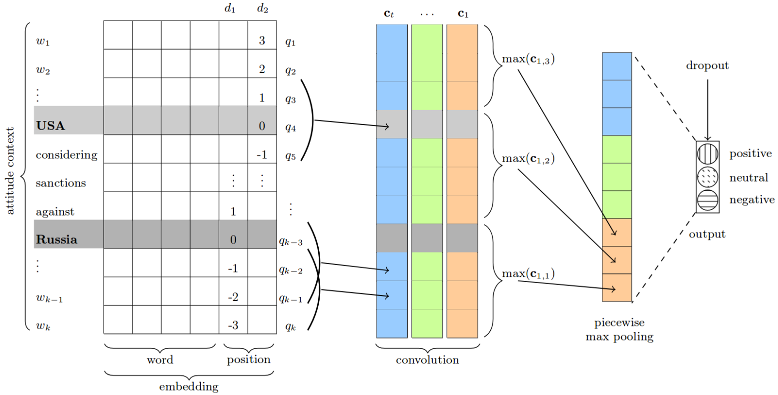


Figure 1: Piecewise Convolutional Neural Network

References I

- N. Loukachevitch and N. Rusnachenko, *Extracting sentiment attitudes from analytical texts*, Proceedings of International Conference of Computational Linguistics and Intellectual Technologies Dialog-2018 (2018).
- A. Mozharova, V. and V. Loukachevitch, N., *Combining knowledge and crf-based approach to named entity recognition in russian*, International Conference on Analysis of Images, Social Networks and Texts (2016), 185–195.