Attention-Based Neural Networks for Sentiment Attitude Extraction using Distant Supervision

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

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

Research Computing Center of Moscow State University, Moscow, Russia
louk_nat@mail.ru

The 10th International Conference on Web Intelligence, Mining and Semantics
June 30 – July 3, 2020
Biarritz, France

Introduction

Microbloging posts (Twitter)

- Mostly user reviews $\Rightarrow$ considered a single object for analysis.

Analytical articles:

- Large amount of named entities (*NE*): Ukraine$_e$, Russia$_e$, Russian Federation$_e$;
- Has complicated structure:

  Trump$_e$ accused China$_e$ and Russia$_e$ of "playing devaluation of currencies"

Related:

- Text Analysis Conference (TAC), Knowledge Base Population (KBP) track[1];
- MPQA 3.0: sentiment attitudes towards entities and events;

1 https://tac.nist.gov/2014/KBP/Sentiment/index.html

Sentiment Attitude Extraction Task

1. Input (Context):

   «As is apparent in Washington_{subj}, there is no place for objectivity on the subject of Russia_{obj}, irrespective of facts and events»

2. Output (Extract):

   Washington $\rightarrow$ Russia, negative

Resources

RuSentRel: Contents

- 73 large analytical articles;
- **Text attitudes** – manual annotation, sentiment towards *named entities* (*NE*) as triplets $\langle Object, Subject, Label \rangle$, where:
  - Subject – *NE* or "author"
  - Object – *NE*
  - Label $\in$ {pos, neg}
- **Named Entities** – automatic (CRF based recognizer);
- List S of synonymous *NE* – manually implemented.

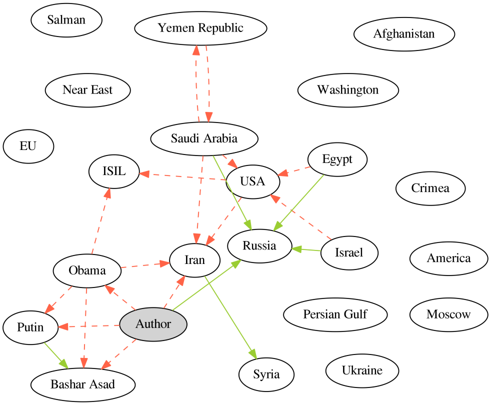


Figure 1: Document opinions

RuSentiFrames Lexicon Structure

Describes sentiments and connotations conveyed with a predicate in a verbal or nominal form.

1. Role Designation:
   - A0 is an argument exhibiting features of a Prototypical Agent;
   - A1 is a Theme.
2. Dimentions:
   - the attitude of the author of the text towards mentioned participants;
   - **polarity** – sentiment between participants;
   - **effects** to participants;
   - mental **states** of participants related to the described situation.

Frame	"Одобрить" (Approve)
roles	A0: who approves A1: what is approved
polarity	A0 → A1, pos, 1.0 A1 → A0, pos, 0.7
effect	A1, pos, 1.0
state	A0, pos, 1.0 A1, pos, 1.0

Table 1: Example description of frame "Одобрить" (Approve) in RuSentiLex lexicon.

RuAttitudes: Collection of automatically labeled news

1. 13.4 K news texts (gathered);
2. News structure – title and list of sentences;
3. Text attitudes – annotated with attitudes between participants (positive or negative); sentiment attitude annotation methods, applied to the news title:
   - **Pair-Based** – utilizing the pre-assigned attitudes (list of pairs);
   - **Frame-Based** – utilizing frame entries from the RuSentiFrames lexicon; matching the following pattern:

$$\ldots \quad \text{Subject}_e \quad \ldots \quad \{\boldsymbol{frame}_{A0 \rightarrow A1}\}_k \quad \ldots \quad \text{Object}_e \quad \ldots$$

RuAttitudes: News Example

Title
McCain: USA_e continue_{pos} supporting_{pos} Georgia_e

$\downarrow$ $\texttt{USA}\rightarrow\texttt{Georgia}_{pos}$ (sentence: 5)

« USA_e and in further continue_{pos} support_{pos} freedom, sovereignty and territorial integrity Georgia_e within the internationally recognized borders of the country», – he said.

$\downarrow$ $\texttt{USA}\rightarrow\texttt{Georgia}_{pos}$ (sentence: 11)

29'th december prime-minister Kvirikashvili_e reported, that the government of Georgia_e has established first contacts with the new USA_e administration.

Attention Models

Sentiment Attitude Extraction

1. Introducing **context attitude** – a pair with its named entities (source: Subject, target: Object) in a context

   «Talking about the separation of the Caucasus region$_e$ due to the confrontation between Russia$_{subj}$ and Turkey$_{obj}$ is not necessary, although there is a danger»

2. We predict a sentiment label of a pair (Subject→Object) in following formats:
   - **Two-scale** – positive or negative (classification);
   - **Three-scale** – positive, negative or *neutral* (extraction).

Context as a sequence of «terms»

Terms:

- Words;
- Entities (Masked);
- Frames;
- Tokens;

context
Talking about the separation of the Caucasus region$_e$ due to the confrontation between Russia$_{subj}$ and Turkey$_{obj}$ is not necessary, although there is a danger.

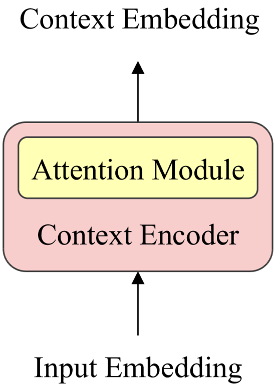
↓

terms
Talking about the separation of the E$_e$ due to the confrontation$_{neg}$ between E$_{subj}$ and E$_{obj}$ is not-necessary$_{neg}$ <COMMA> al- though there is a danger <DOT>

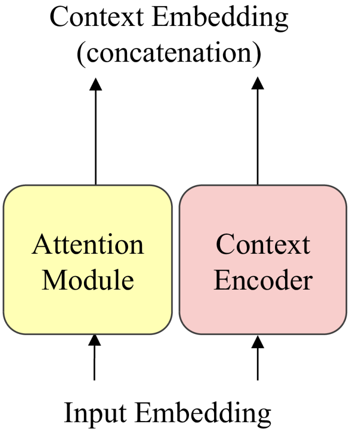
Input embedding

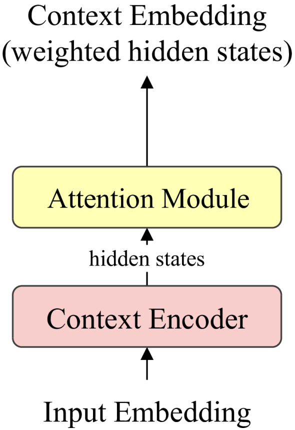
1. Pretrained Word2Vec model (size of 1000);
2. Additional parameters (size of 5 each):
    - **Distance embedding** – is vectorized distance in terms from attitude participants of entry pair (E_{subj} and E_{obj} respectively) to a given term;
    - **Closest to synonym** distance embedding – is a vectorized abs. distance in terms from a given term towards the nearest entity, synonymous to E_{subj} and E_{obj} ;
    - **Part-of-speech** embedding;
    - `A0→A1` **polarity** embedding – is a vectorized «positive» or «negative» value for frame entries in RuSentiFrames;

What is Attention?



Attention embed options





Attention Modules (Types and References)

- **Based on input embedding**: AttCNN[1] (PCNN[2]);
- **Based on hidden states**: Att-BLSTM[3] (BiLSTM as an encoder);

[1] Yatian Shen and Xuanjing Huang. "Attention-Based Convolutional Neural Network for Semantic Relation Extraction". In: *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers* (Dec. 2016), pp. 2526–2536.

[2] Nicolay Rusnachenko and Natalia Loukachevitch. "Using convolutional neural networks for sentiment attitude extraction from analytical texts". In: *EPiC Series in Language and Linguistics* 4 (2019), pp. 1–10.

[3] Peng Zhou et al. "Attention-based bidirectional long short-term memory networks for relation classification". In: *Proceedings of the 54th Annual Meeting of the ACL (Volume 2: Short Papers)* (2016), pp. 207–212.

Feature-Attentive context encoder

1. Feature f: attitude ends, frames;

$$h_i = [x_i, f]$$

2. Weight calculation:

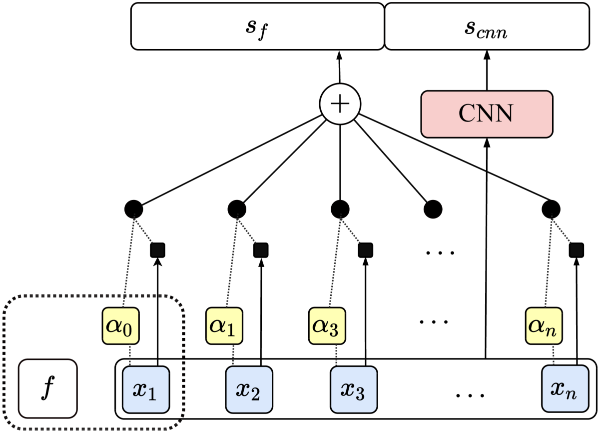
$$u_i = W_a\,(\tanh(W_{we} \cdot h_i + b_{we})) + b_a$$

3. Attention Embedding:

$$\alpha = softmax(u)$$

$$\hat{s} = \sum_{i=1}^{n} x_i \cdot \alpha_i$$

$$s_f = avg_{j=1..k}(\hat{s})$$



Self-Attentive context encoder

1. Hidden state **w**.

$$h_i = \overrightarrow{h_i} + \overleftarrow{h_i}, \ i \in \overline{1..n}.$$

2. Weight calculation:

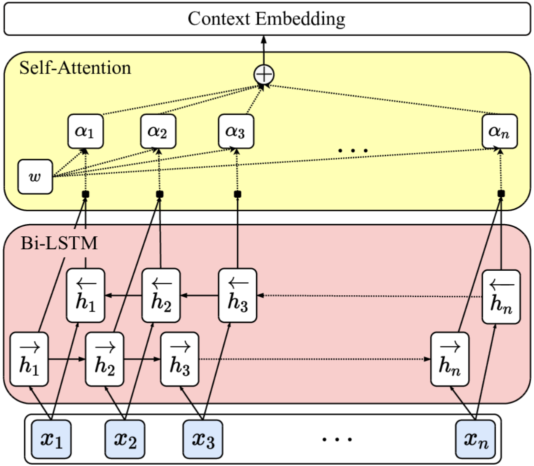
$$m_i = \tanh(h_i)$$

$$u_i = m_i^T \cdot \mathbf{w}$$

3. Attention Embedding:

$$\alpha = softmax(u)$$

$$s = \tanh(H \cdot \alpha)$$



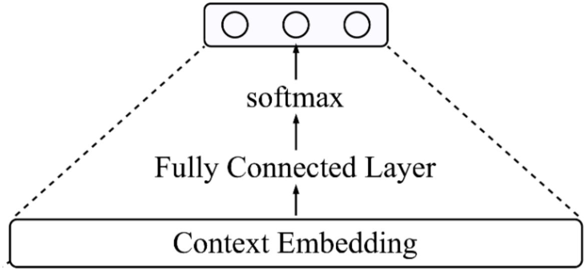
Output Class Probabilities

s – is a context embedding;

$$r = W_r \cdot \tanh(s) + b_r$$

$$o = softmax(r)$$

$$label = argmax(o)$$



Experiments

Experiment Details

1. Experiment formats:
   - Two-scale – pos/neg (classification);
   - Three-scale – pos/neg + *neutral* (extraction).
2. Output: document level opinions;
3. Evaluation:
   - Using F1-score measure;
   - F_{avg} – 3-Fold CV average results;
   - F_{test} – Fixed Test sep. of RuSentRel.

1. Training Formats:
   - **SL** – RuSentRel;
   - **DS** – RuSentRel + RuAttitudes for fine-tuning.
2. Model groups:
   - BiLSTM;
   - PCNN;
   - CNN.

Results

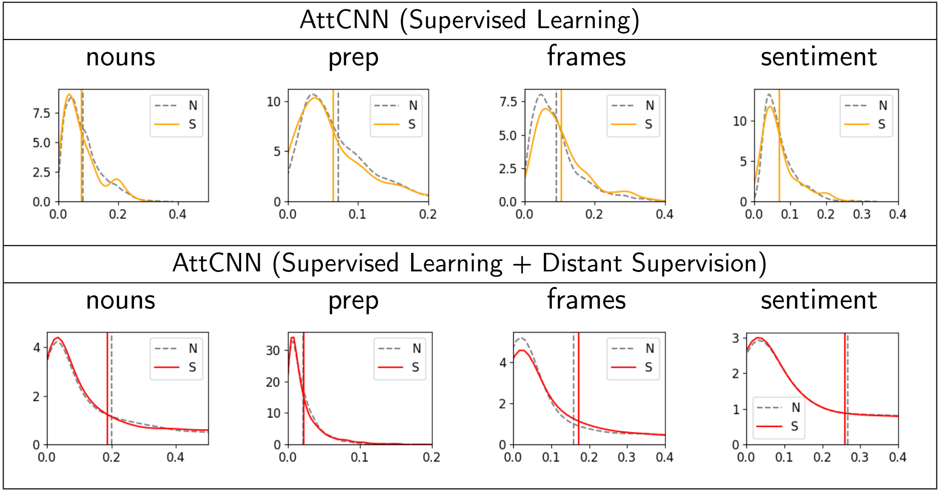
Model	DS	2-scale $F1_{avg}$	2-scale $F1_{test}$	3-scale $F1_{avg}$	3-scale $F1_{test}$
Att-BLSTM	✓	**0.67**	0.68	**0.33**	0.38
BiLSTM	✓	0.65	**0.70**	0.31	0.39
Att-BLSTM		0.64	0.68	0.31	0.32
BiLSTM		0.63	0.67	0.29	0.34
$AttPCNN_e$	✓	0.65	0.66	0.31	**0.41**
PCNN	✓	0.60	0.63	0.32	0.40
$AttPCNN_e$		0.62	0.67	0.30	0.35
PCNN		0.61	0.66	0.29	0.32
$AttCNN_e$	✓	0.63	0.66	0.32	**0.41**
CNN	✓	0.63	0.68	0.31	0.40
$AttCNN_e$		0.64	0.62	0.27	0.30
CNN		0.55	0.59	0.27	0.31

Weight Distribution Analysis

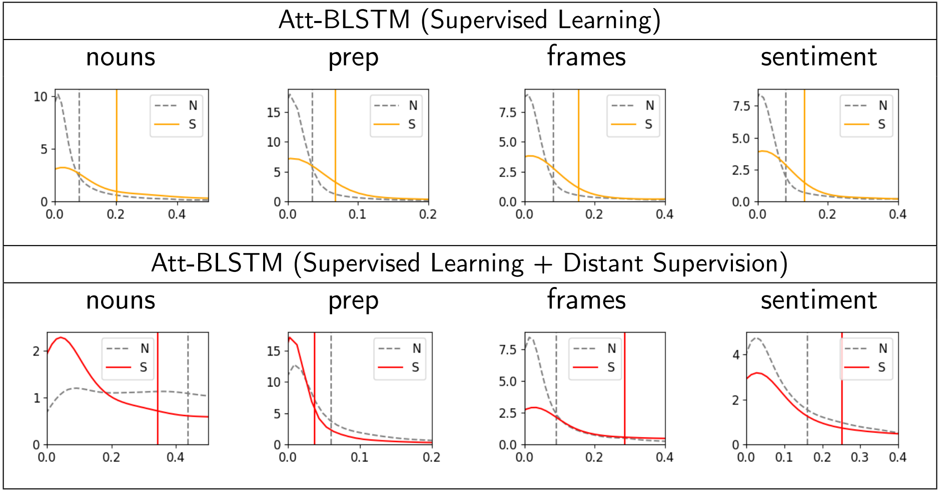
1. Considering models of 3-scale experiment;
2. **Term groups** presented in analysis:
   - Frames;
   - Words → Nouns, Verbs, Prepositions, Sentiment[2];
3. **Context-level weight** of a particular term group is a weighted sum of terms which both appear in the context and belong to the corresponding term group.
4. We utilize distributions (ρ_S, ρ_N) of context-levels weights across:
   - Sentiment contexts (S) – contexts labeled with positive or negative labels;
   - Neutral contexts (N) – contexts labeled as neutral;

2 Contains words and expressions of the Russian language with sentiment labels

Weight Distribution Analysis (AttCNN)



Weight Distribution Analysis (Att-BLSTM)



Weight distribution visualization on sentiment contexts

Att-BLSTM (Supervised Learining)

leading such a game , E_{subj} will finally ***lose***$_{pos}$ ***trust-in***$_{pos}$ E_{obj} and country E

...

But E_{subj} consequently emphasizes its ***interest***$_{pos}$ in ***normalizing***$_{pos}$ relationships with E_{obj} (<NUM> february <NUM> year <DOT> took place the visit E at E and its ***conversation***$_{pos}$ with the spiritual leader E and with president E)

Att-BLSTM (Supervised Learining + Distant Supervision)

leading such a game , E_{subj} will finally ***lose***$_{pos}$ ***trust-in***$_{pos}$ E_{obj} and country E

...

But E_{subj} consequently emphasizes its ***interest***$_{pos}$ in ***normalizing***$_{pos}$ relationships with E_{obj} (<NUM> february <NUM> year <DOT> took place the visit E at E and its ***conversation***$_{pos}$ with the spiritual leader E and with president E)

$$\ldots \quad \text{Subject}_e \quad \ldots \quad \{\textit{frame}_{A0 \rightarrow A1}\}_k \quad \ldots \quad \text{Object}_e \quad \ldots$$

Conclusion

- We consider sentiment attitude extraction task as two-scale and three-scale classification tasks.
- We study the attention-based neural networks. Application of Distant Supervision results in 10% increase by F1 for models that with non-attentive encoders; replacing the latter with attentive encoders results in 3% increase by F1.
- Model with self-attentive encoders (Att-BLSTM) illustrates the greatest discrepancy in weight distributions on sentiment and neutral contexts across all the term groups presented in the analysis;

Links

1. Resources:
   - **RuSentRel**: `https://github.com/nicolay-r/RuSentRel/tree/v1.1`
   - **RuSentiFrames**: `https://github.com/nicolay-r/RuSentiFrames/tree/v1.0`
   - **RuAttitudes**: `https://github.com/nicolay-r/RuAttitudes/tree/v1.0`
2. Word2Vec News embedding[3];
3. Experiments[4] (Based on AREkit[5] framework)

3 http://rusvectores.org/static/models/rusvectores2/news_mystem_skipgram_1000_20_2015.bin.gz

4 https://github.com/nicolay-r/attitude-extraction-with-attention-and-ds

5 https://github.com/nicolay-r/AREkit/blob/0.20.3-wims-rc