Sentence Level Representation and Language Models in the task of Coreference Resolution for Russian

Le T.A., Petrov M.A., Kuratov Y.M., Burtsev M.S.

Neural Networks and Deep Learning Lab -
Moscow Institute of Physics and Technology

May 31, 2019

This work was supported by National Technology Initiative and PAO Sberbank project ID 0000000007417F630002.

Overview

The task of Coreference Resolution

Goal: Finding all expressions that refer to the same entity in a text

My sister has a friend called John. She thinks he is so funny.

Challenges:

- Long inputs: Paragraphs, even documents with hundreds of sentences.
- Document-level prediction
- Nested mentions

Available datasets

English:

- CoNLL-2012 shared task

Russian:

- Dialogue 2014 Shared Task (RuCor)
- Dialogue 2019 Shared Task (AnCor)

Datasets	Language	Mentions	Chains
CoNLL 2012 Shared Task	En	194,480	44,221
RuCor	Ru	16,558	3,638
AnCor	Ru	28,961	5,678

Table: Coreference resolution datasets. Mentions and chains number computed for train + dev + test sets.

End-to-end coreference resolution model[1] as Baseline

Two-stages approach: Mention detection, mention clustering

- Generate vector representation of spans: $\boldsymbol{g}_i$
- Calculate mention scores for every possible spans:

$$s_{\mathrm{m}}(i) = \boldsymbol{w}_{\mathrm{m}}^{\top}\mathrm{FFNN}_{\mathrm{m}}\left(\boldsymbol{g}_i\right)$$

- Compute antecedent scores:

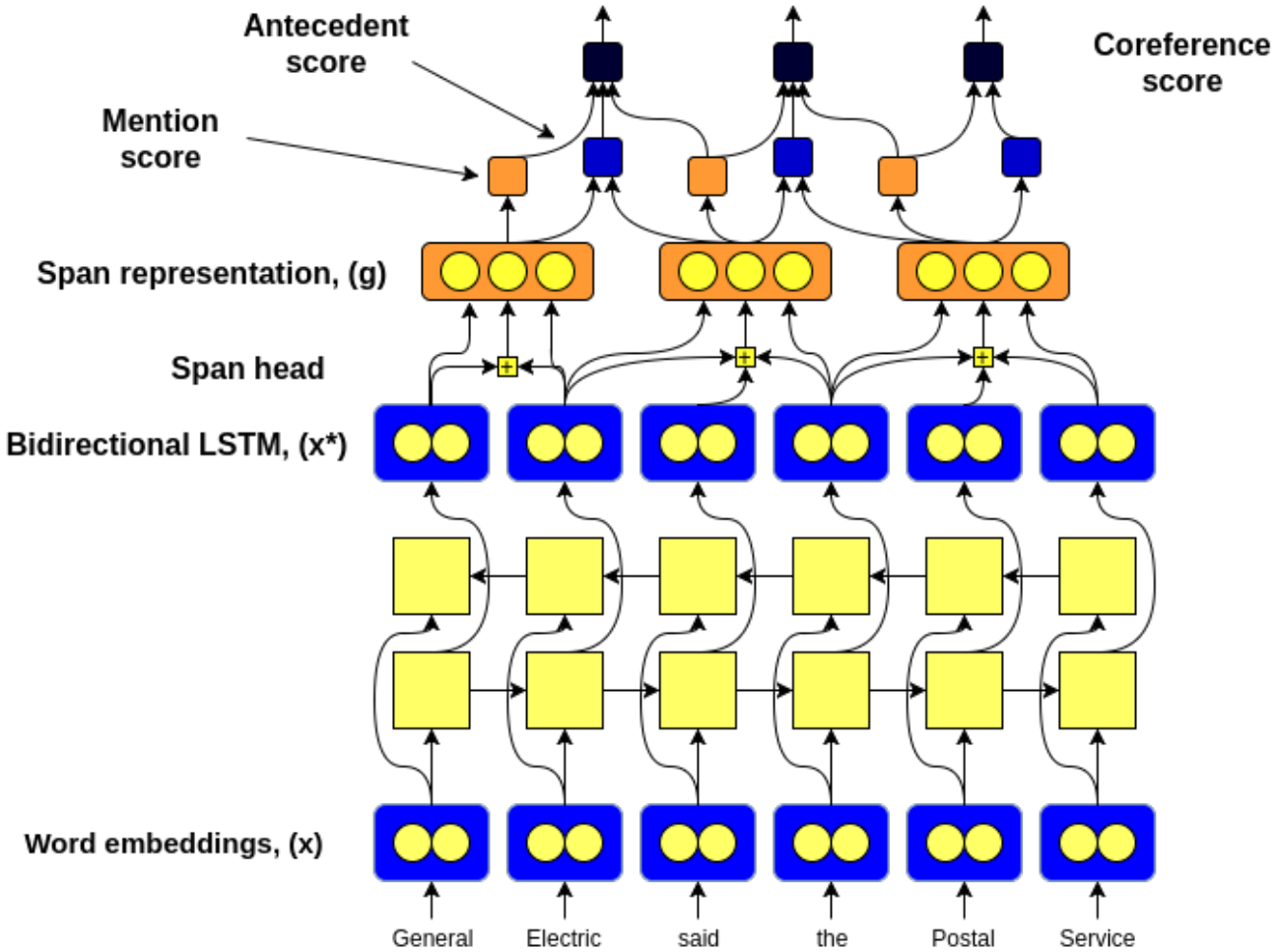
$$s_{\mathrm{a}}(i,j) = \boldsymbol{w}_{\mathrm{a}}^{\top}\mathrm{FFNN}_{\mathrm{a}}\left(\left[\boldsymbol{g}_i, \boldsymbol{g}_j, \boldsymbol{g}_i \circ \boldsymbol{g}_j, \phi(i,j)\right]\right)$$

- Coreference score:

$$s(i,j) = s_{\mathrm{m}}(i) + s_{\mathrm{m}}(j) + s_{\mathrm{a}}(i,j)$$

[1] Lee, Kenton, et al. "End-to-end Neural Coreference Resolution." Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017.

End-to-end coreference resolution model[1] as Baseline



[1] Lee, Kenton, et al. "End-to-end Neural Coreference Resolution." Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017.

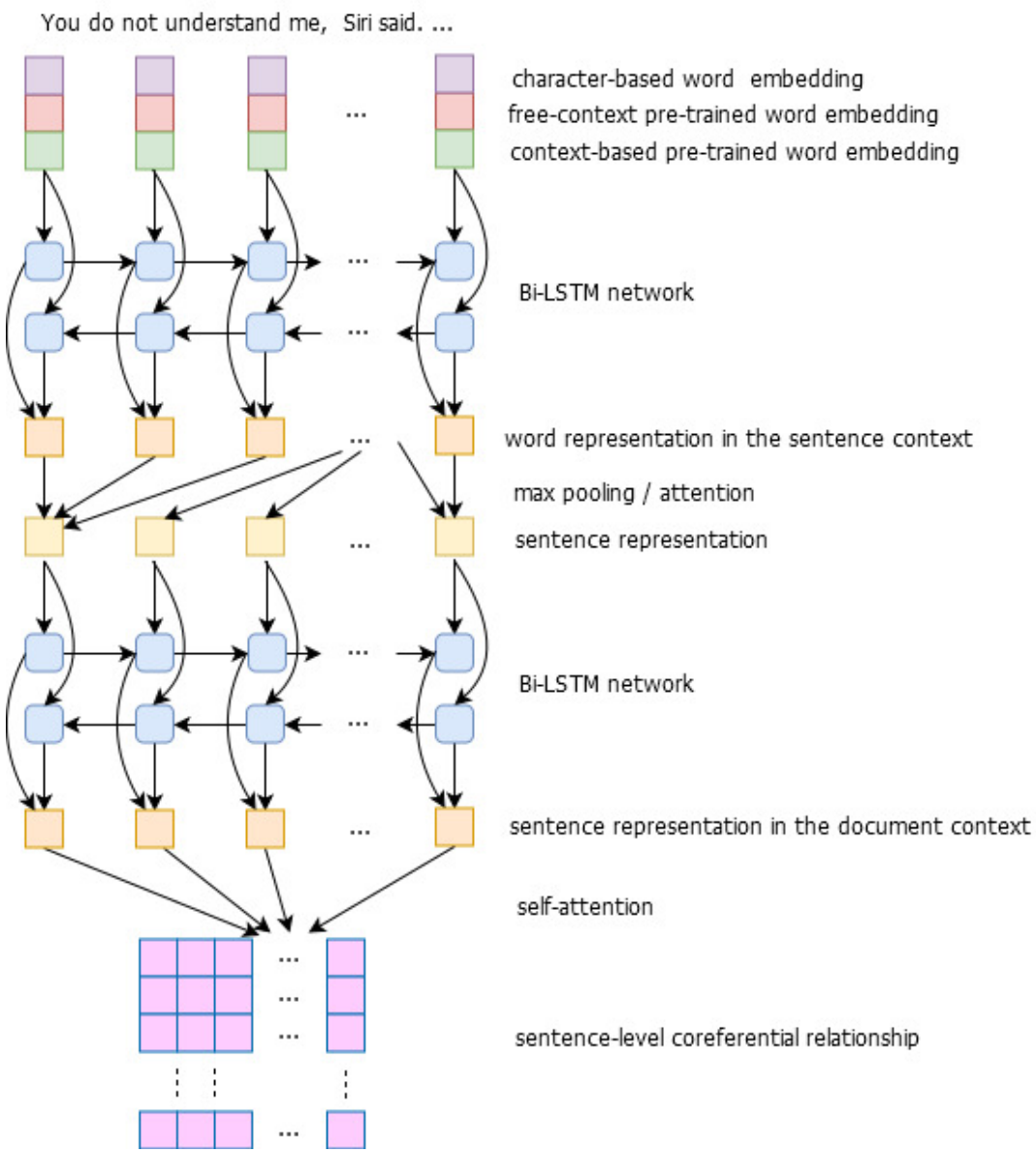
Sentence-level Coreferential Relationship Model

- Idea: Capture sentence relationship in the coreference context.
- Input: document with n sentences $D = \{s_1, s_2, .., s_n\}$
- Output: square matrix $M[n, n]$ capturing the coreferential relationship between sentences.

For example:
A: My sister knows you.
B: Oh, really?
A: Yes, she thinks you are so kind.

0.3	0.2	0.9
0.2	0.1	0.3
0.9	0.3	0.1

Sentence-level Coreferential Relationship Model



Effect of sentence-level coreferential relation

Dataset	F1 (test set)
Original OntoNotes 5.0	73.00
OntoNotes 5.0 + sent.-level coref. relation with 20% of noise	74.13
OntoNotes 5.0 + sent.-level coref. relation with 16% of noise	74.73
OntoNotes 5.0 + sent.-level coref. relation with 9% of noise	75.56
OntoNotes 5.0 + sent.-level coref. relation with 6% of noise	76.36
OntoNotes 5.0 + sent.-level coref. relation with 3.5% of noise	77.01
OntoNotes 5.0 + sent.-level coref. relation with 1.5% of noise	77.92
OntoNotes 5.0 + ground-truth sent.-level coref. relation	78.84

Table: Effect of sentence-level coreferential relation on the baseline model performance

Sentence-level Coreferential Relationship-based Model

- Bi-LSTM + CNN + Self attention + log scale distance + weighted classes
- Bi-LSTM + CNN + Self attention + log scale distance + weighted classes + ELMo
- Bi-LSTM + CNN + Self attention + log scale distance + weighted classes + BERT

Model	On the dev. set				On the test set			
	Acc.	P	R	F	Acc.	P	R	F
Model 1	80.60	60.10	70.10	61.80	80.20	59.10	68.10	60.20
Model 2	82.60	62.50	75.50	66.10	81.60	61.40	72.30	63.80
Model 3	84.00	65.49	75.32	67.69	82.71	64.09	71.14	64.81

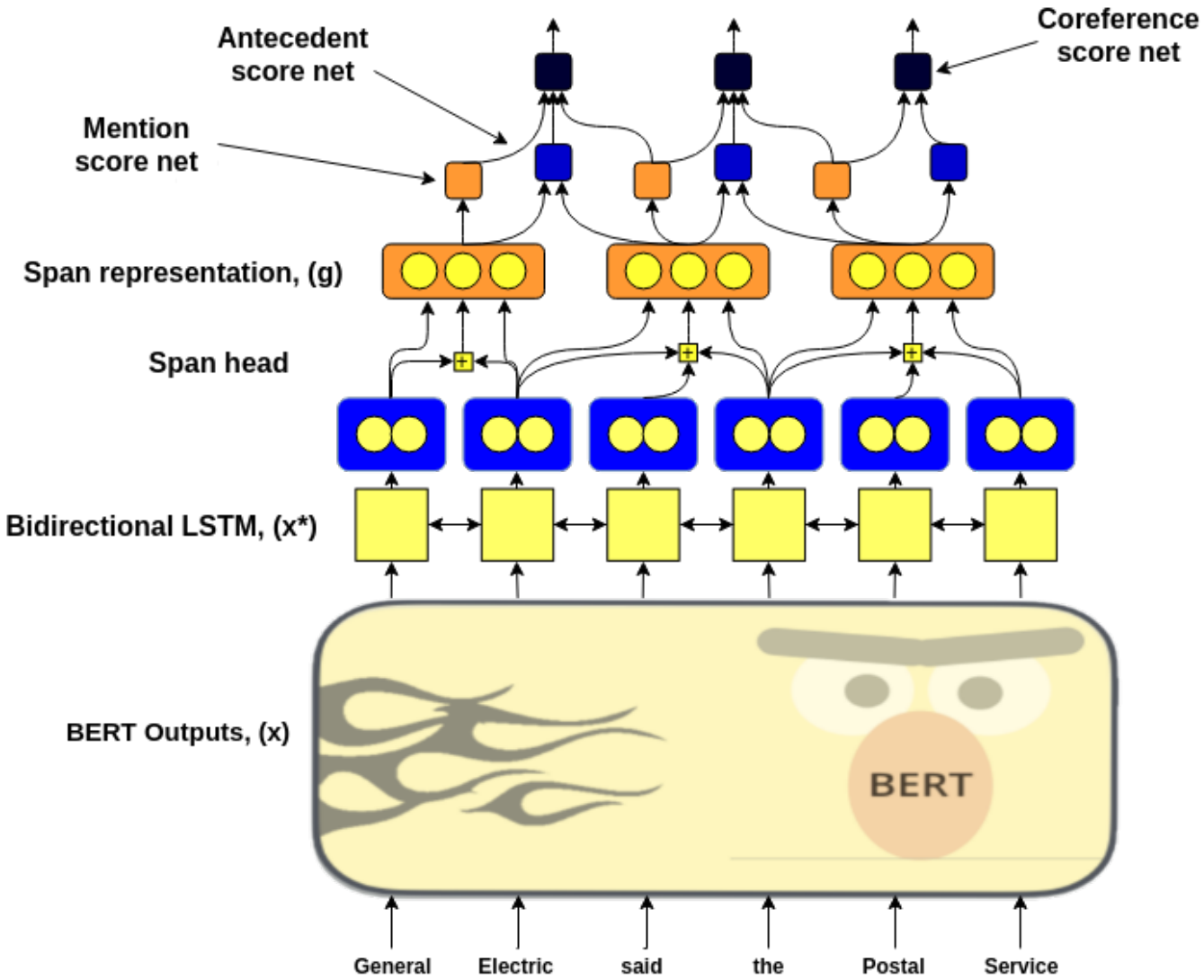
Table: Training Sentence-level Coreferential Relationship-based Model on OntoNotes 5.0

Models used for Dialogue 2019 Shared Task (AnCor)

- Baseline End-to-End Coreference model
- Baseline with features from pretrained Language Models (ELMo, BERT)

We tested these models on full coreference pipeline task and on setting with gold mention boundaries.

Baseline model with contextual embeddings



Experiments and results

We tested our models in two settings:

Model	*muc*	*bcube*	*ceafe*	*avg.*F_1
Baseline + ELMo	90.22	83.41	59.44	77.69
Baseline + RuBERT(1-6-12)	91.04	84.38	63.07	79.50
Baseline + ELMo + RuCor	91.51	84.16	61.33	79.01
Baseline + RuBERT(1-6-12) + RuCor	**91.47**	**84.49**	**63.81**	**79.92**

Table: Set. 1: Results on AnCor dataset, gold mentions.

Experiments and results

Model	*muc*	*bcube*	*ceafe*	*avg.*F_1
Baseline + ELMo	50.29	48.89	46.99	51.72
SCRb	60.00	48.89	50.39	53.61
Baseline + RuBERT(1-6-12)	60.95	51.08	49.24	53.76
Baseline + ELMo + RuCor	65.01	52.67	50.19	55.96
Baseline + RuBERT(1-6-12) + RuCor	**66.74**	**54.88**	**51.72**	**57.78**

Table: Set. 2: Results on AnCor dataset, full pipeline.

Experiment details

- Server properties: 10x NVIDIA GTX 1080 Ti
- Train data: AnCor + RuCor
- Tried to use different BERTs layers for constructions tokens embeddings
- 10 folds crossvalidation
- 10 models ensemble
- Tuning "top span ratio" for russian

Dialogue 2019 Shared Task Results (coreference)

Team/model	*muc*	*bcube*	*ceafe*	*avg.*F_1
SagTeam	**62.23**	52.79	**52.29**	**55.77**
Baseline + RuBERT(1-6-12) + RuCor	62.06	**53.54**	51.46	55.68

Table: Dialogue 2019 Shared Task Results. Coreference track. Using only text as training data.

Team/model	*muc*	*bcube*	*ceafe*	*avg.*F_1
Baseline + RuBERT(1-6-12) + RuCor	**82.62**	**73.95**	**72.14**	**76.24**
Legacy	75.83	66.16	64.84	68.94
MorphoBabushka	61.36	53.39	51.95	55.57

Table: Dialogue 2019 Shared Task Results. Coreference track. Gold mentions boundaries.

Dialogue 2019 Shared Task Results (anaphora)

Team/model	F_1(strong)	F_1(*soft*)
Baseline + RuBERT(1-6-12) + RuCor	**67.80**	**74.80**
Legacy	59.90	69.40
Etap	49.40	59.30
MorphoBabushka	37.10	54.90
Meanotek	36.90	45.20

Table: Dialogue 2019 Shared Task Results. Anaphora track. Using only text as training data.

Conlcusions

- Pretrained Language Models help a lot for the task of coreference resolution
- Sentence-level coreference resolution is a promising way of further research

Thank you for your attention!