

# Распознавание именованных сущностей с использованием синтактико-семантических признаков и нейросетей

МФТИ, кафедра компьютерной лингвистики  
Юсупов Идрис ([i.yusupov@phystech.edu](mailto:i.yusupov@phystech.edu))

# Нейронные сети

- Показывают state-of-the-art результаты.
- Признаки:
  - векторные представления слов,
  - минимальное количество вручную построенных признаков (газетеры и капитализация для NER).

# Высокоуровневые признаки

- Морфология (mystem),
- синтаксис (Malt),
- семантика (VabelNet),
- семантико-синтаксический анализатор (Compreno).

Нейронные сети + высокоуровневые  
признаки = ?

# Цель работы

## Цель:

- Исследовать возможность использования семантико-синтаксического анализатора Comreno в качестве источника высокоуровневых признаков для задачи распознавания именованных сущностей (NER) на англоязычном корпусе CoNLL 2003 в рамках нейросетевого подхода.

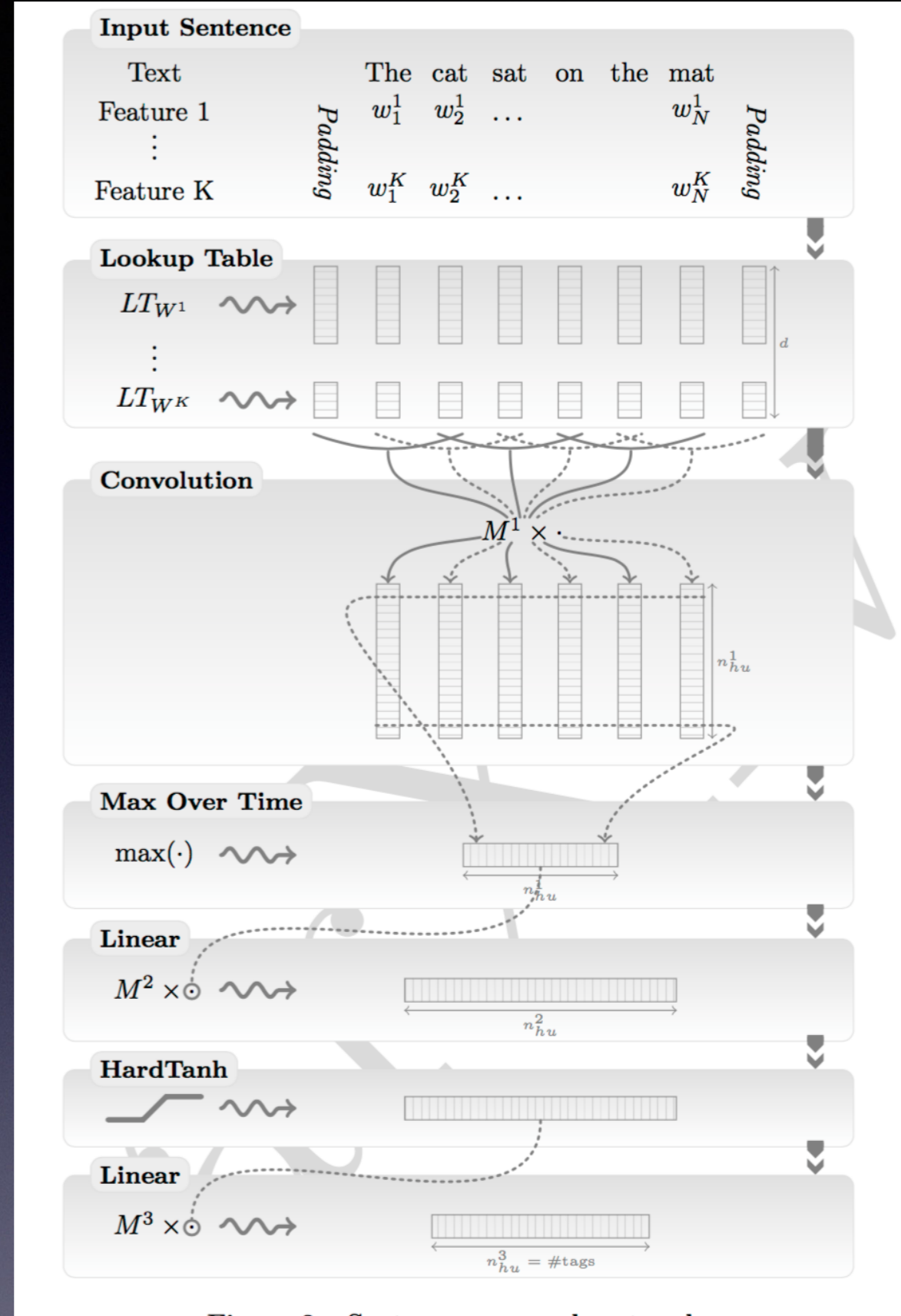


Figure 9. Sentence embedding network.

Сверточная нейронная сеть принимающая на вход предложение [Collobert R. et al 2011]

# Результаты без синтактико-семантических признаков

Модель	Используемые признаки	Датасет	Метод оптимизации	Полученная F1, %	F1 в статье [Collobert R. et al 2011], %
Win	Embeddings, Capitalization	train	Mini-batch gradient descent	86.27	-
Win	Embeddings, Capitalization	train	Stochastic gradient descent	-	86.96
ConvNet + CRF	Embeddings, Capitalization, Position	train	Stochastic gradient descent	-	88.67
ConvNet + CRF	Embeddings, Capitalization, Position, Gazetteer*	train	Stochastic gradient descent	-	89.59
ConvNet	Embeddings, Capitalization, Position	train	Stochastic gradient descent	86.77	-
ConvNet	<b>Embeddings, Capitalization, Position, Gazetteer*</b>	<b>train</b>	<b>Stochastic gradient descent</b>	<b>87.89</b>	-
ConvNet	Embeddings, Capitalization, Position, Gazetteer*	train+dev	Stochastic gradient descent	88.37	-
ConvNet	<b>Embeddings, Capitalization, Position, Gazetteer*</b>	<b>train</b>	<b>Mini-batch gradient descent</b>	<b>87.49</b>	-

\*Gazetteer - газетир предоставляемый соревнованием CoNLL 2003

# Синтактико-семантические признаки

- Вершины семантико-синтаксического дерева Comreno кодировались бинарным представлением.
- В результате для каждого токена были получены разрежённые вектора размерности 83950.
- Инженерия над признаками не проводилась.



# Опробованные способы внедрения синтактико-семантических признаков

- Сжать признаки Comreno с помощью SVD и добавить в сверточную нейросеть как еще один признак.
- Добавить еще одну нейросеть для Comreno признаков и оптимизировать её вместе со сверточной.

# Результаты с синтактико-семантическими признаками сжатыми SVD

Модель	Используемые признаки	Датасет	Метод оптимизации	F1, %
ConvNet	Position, Compreno SVD	train	Mini-batch gradient descent	75.89
ConvNet	Capitalization, Position, Gazetteer, Compreno SVD 1024	train	Mini-batch gradient descent	81.83
ConvNet	Embeddings, Capitalization, Position, Gazetteer, Compreno SVD 1024	train	Mini-batch gradient descent	86.85
ConvNet	Embeddings, Capitalization, Position, Gazetteer	train	Mini-batch gradient descent	87.49

# Совместная оптимизация Convolution и Compresso нейросетей

## Compresso нейросеть

input -> (1) -> (2) -> (3) -> (4) -> (5) -> (6) -> (7) -> output

(1): nn.SparseLinear(83951, 256)

(2): nn.Dropout(0.5)

(3): nn.HardTanh

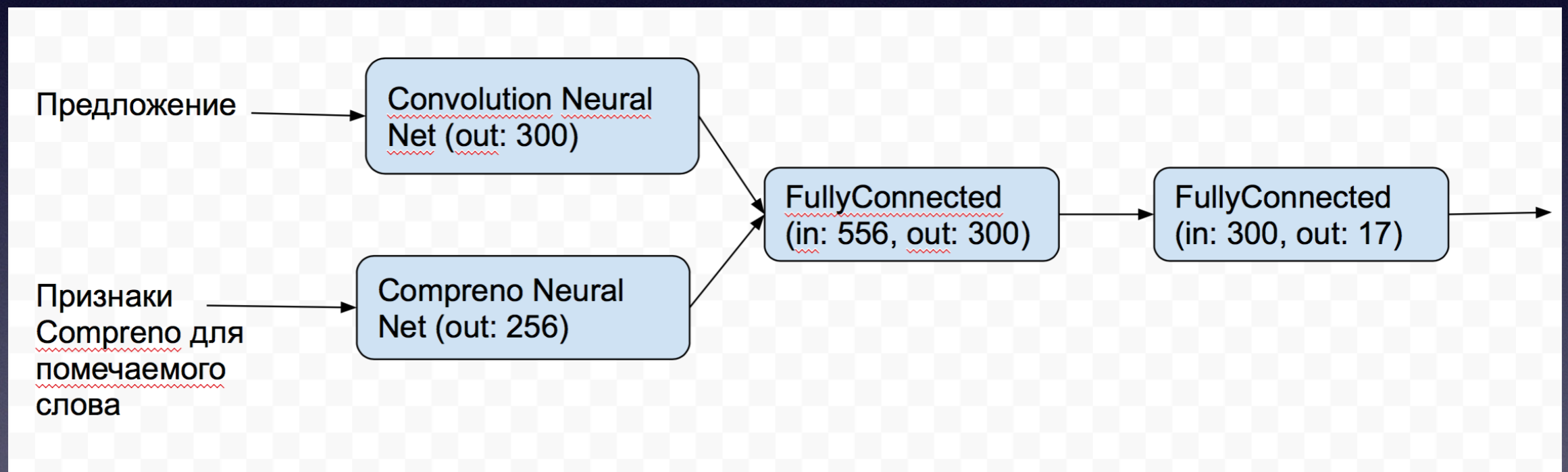
(4): nn.Linear(256 -> 256)

(5): nn.Dropout(0.5)

(6): nn.HardTanh

(7): nn.Linear(256 -> 17)

# Совместная оптимизация Sentence и Comreno нейросетей



# Результаты с синтактико-семантическими признаками для совместно-оптимизированной нейросети

Модель	Используемые признаки	Датасет	Метод оптимизации	F1, %
Compreno Net	Compreno Sparse Features	train	Mini-batch gradient descent	72.85
ConvNet	Embeddings, Capitalization, Position, Gazetteer	train	Mini-batch gradient descent	87.49
<b>ConvNet + Compreno Net</b>	<b>Embeddings, Capitalization, Position, Gazetteer, Compreno Sparse Features</b>	<b>train</b>	<b>Mini-batch gradient descent</b>	<b>88.47</b>
ConvNet + Compreno Net	Embeddings, Capitalization, Position, Gazetteer, Compreno Sparse Features	train+dev	Mini-batch gradient descent	88.81

# Немного о программной реализации

- Код для воспроизведения экспериментов выложен по адресу: <http://github.com/sld/torch-conv-ner>
- Реализовано с использованием deep learning фреймворка torch [13].

# Итог

- Был найден простой вариант подключения признаков Comreno к сверточной нейронной сети за счет которого F1-мера повысилась с 87.49% до 88.47%.

# Планы на будущее

- Добавить CRF в существующую архитектуру для повышения F1 меры.
- Исследовать работу предложенного решения на других корпусах.



# Литература

1. Collobert R. et al. Natural language processing (almost) from scratch //The Journal of Machine Learning Research. – 2011. – Т. 12. – С. 2493-2537.
2. Luo G. et al. Joint Named Entity Recognition and Disambiguation. In Proceedings of the 2015 Conference on empirical Methods in Natural Language Processing pages 879–888, Lisbon, Portugal, September. Association for Computational Linguistics. 2015.
3. Erik F. Tjong Kim Sang and Fien De Meulder, Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition. In: *Proceedings of CoNLL-2003*, Edmonton, Canada, 2003, pp. 142-147.
4. Passos A., Kumar V., McCallum A. Lexicon Infused Phrase Embeddings for Named Entity Resolution //CoNLL-2014. – 2014. – С. 78.
5. Chiu J. P. C., Nichols E. Named Entity Recognition with Bidirectional LSTM-CNNs //arXiv preprint arXiv:1511.08308. – 2015.
6. Bian J., Gao B., Liu T. Y. Knowledge-powered deep learning for word embedding //Machine Learning and Knowledge Discovery in Databases. – Springer Berlin Heidelberg, 2014. – С. 132-148.
7. Xu C. et al. Rc-net: A general framework for incorporating knowledge into word representations //Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management. – ACM, 2014. – С. 1219-1228.
8. Zhiheng Huang, Wei Xu, and Kai Yu. Bidirectional LSTM-CRF models for sequence tagging. CoRR , abs/1508.01991. 2015.
9. Lin D., Wu X. Phrase clustering for discriminative learning //Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2. – Association for Computational Linguistics, 2009. – С. 1030-1038.
10. Goldberg Y. A Primer on Neural Network Models for Natural Language Processing //arXiv preprint arXiv:1510.00726. – 2015
11. Multi-Task Cross-Lingual Sequence Tagging from Scratch, Zhilin Yang, Ruslan Salakhutdinov, William Cohen. -2016. - Preprint
12. <http://scikit-learn.org/stable/modules/decomposition.html#truncated-singular-value-decomposition-and-latent-semantic-analysis>
13. <http://torch.ch>