# **Learning representations** for text-level discourse parsing

Gregor Weiss

Faculty of Computer and Information Science
University of Ljubljana
Večna pot 113, Ljubljana, Slovenia (EU)

gregor.weiss@student.uni-lj.si

## Abstract

In the proposed doctoral work we will design an end-to-end approach for the challenging NLP task of text-level discourse parsing. Instead of depending on mostly hand-engineered sparse features and independent components for each subtask, we propose a unified approach **completely based on deep learning architectures**. To train better dense vector representations that capture communicative functions and semantic roles of discourse units and relations between them, we will **jointly learn all discourse parsing subtasks** at different layers of our stacked architecture and **share their intermediate representations**. By combining unsupervised training of word embeddings and related NLP tasks with our guided layer-wise multi-task learning of higher representations we hope to reach or even surpass performance of current state-of-the-art methods on annotated English corpora.

## Discourse parsing

- **discourse**: a piece of text meant to communicate specific information, function, or knowledge (clauses, sentences, or even paragraphs)
- understood only in relation to other discourses and their joint meaning is larger than individual unit's meaning alone
- information from related NLP tasks helps [2.4]

**Penn Discourse Treebank** [1] adopts the predicate-argument view and independence of discourse relations:
- 2159 articles from the Wall Street Journal
- 4 sense classes, 16 types, 23 subtypes

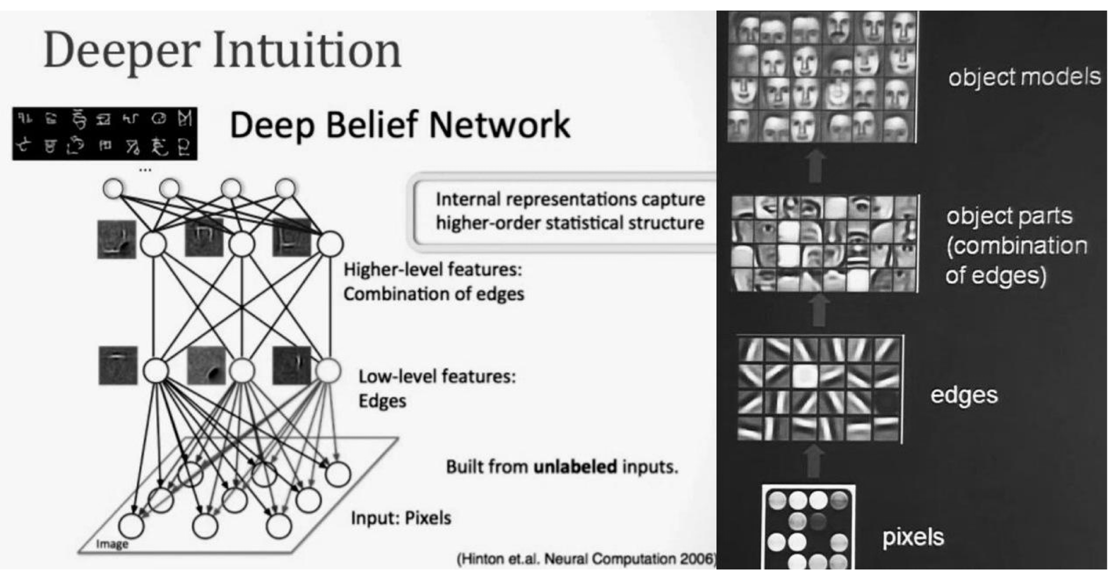
[Index arbitrage doesn't work]$_{arg1}$,
*and* [it scares natural buyers of stock]$_{arg2}$.
— PDTB-style, id: 14883, type: explicit, sense: Expansion.Conjunction

[But]$_{arg2}$
*if* [this prompts others to consider the same thing]$_{arg1}$,
*then* [it may become much more important]$_{arg2}$.
— PDTB-style, id: 14905, type: explicit, sense: Contingency.Condition

PDTB-style discourse parsing goals:
- locate explicit or implicit discourse *connectives*
- locate text spans for *argument 1 and 2*
- predict *sense* that characterizes the nature of the relation

## Deep learning

- multiple layers of learning blocks stacked on each other
- beginning with raw data, its representation is transformed into increasingly higher and more abstract forms in each layer, until finally features for a given task are reached

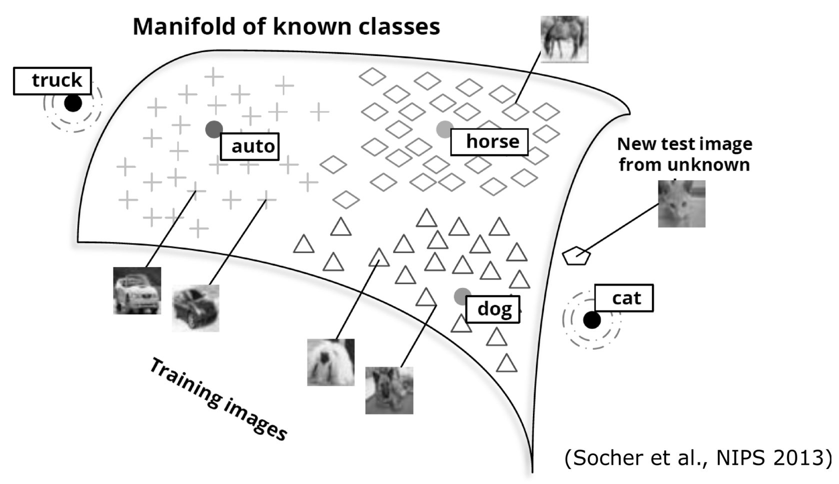



(Andrew Ng, 2013)

Text documents are usually treated as a sequence of words with different lengths:
- transition-based processing mechanisms
- **recurrent neural networks** (RNNs): apply the same set of weights over the sequence (temporal dimension) or structure (tree-based)

Represent text documents/words as numeric vectors of fixed size:
- **word embeddings** (word2vec) [3]
- character-level convolutional networks

**Pre-training** helps deep networks to develop natural abstractions and combined with multi-task learning [4] it can significantly improve their performance in the absence of hand-engineered features.




(Socher et al., NIPS 2013)

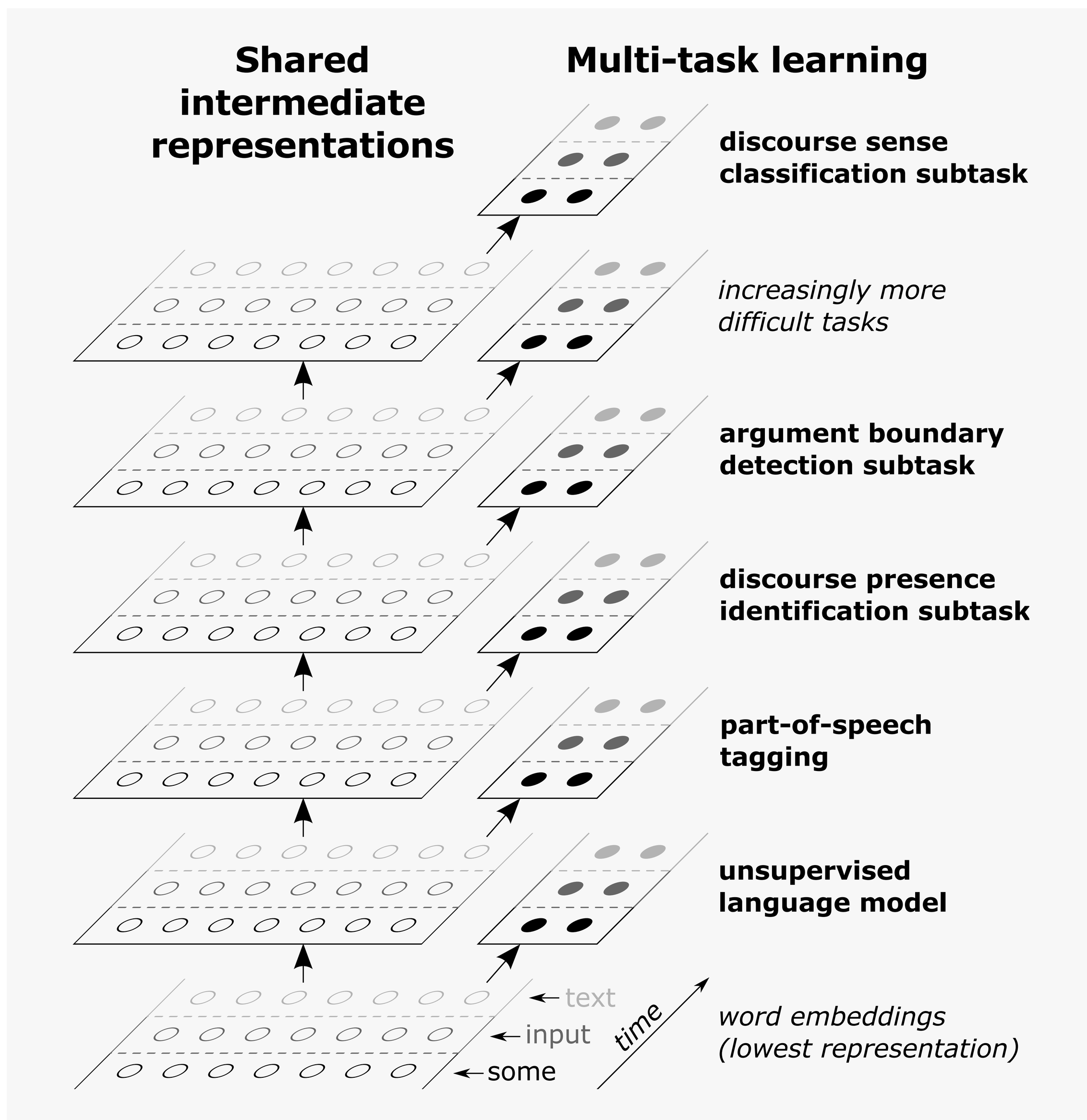
## Motivation

Natural language processing (NLP):
- large pipelines of **independently-constructed components**
- **hand-engineered sparse features** based on language/domain/task specific knowledge
- still room for improvement on more challenging NLP tasks

**Deep learning architectures**:
- one learning algorithm that could unify learning of all components
- latent features/representations are automatically learned as distributed dense vectors
- surprising results for a number of NLP tasks

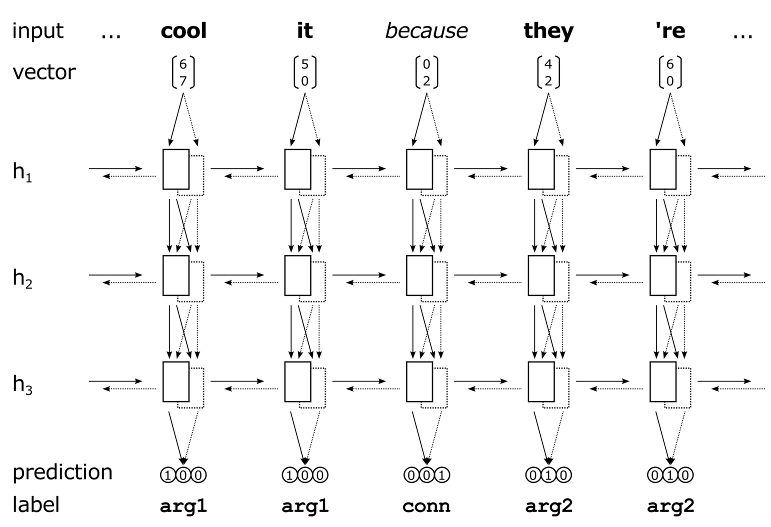
## Our approach

Lets design a PDTB-style end-to-end discourse parser almost without any hand-engineered NLP knowledge:
- **unified end-to-end architecture**
  - one learning algorithm for all discourse parsing subtasks and related NLP tasks
- **automatic learning of representations**
  - completely based on deep learning architectures (bidirectional deep RNN)
- **shared intermediate representations**
  - partially stacked on top of each other to benefit from each others representations
- **guided layer-wise multi-task learning**
  - jointly learning all discourse parsing subtasks and related NLP tasks
  - *lowest representation*: unsupervised training of word embeddings
  - *lower layers*: training on related NLP tasks
  - *higher layers*: training on increasingly more difficult discourse parsing subtasks




## Progress

To confirm that our approach would make sense for discourse parsing, we experimented with single-task learning with bidirectional deep RNN for discourse sense tagging:
- long training time for randomly initialized weights
- overfited training data




**Technology**:
- *Python*
- *Theano*: fast tensor manipulation library
- *Keras*: modular neural network library

**Resources**:
- pre-trained word2vec lookup table on Google News dataset to initialize word embeddings
- tokenized text documents as input
- POS tags of input tokens

**Evaluation** (from CoNLL 2015 shared task):
- performance in terms of precision/recall/F1-score
- explicit connectives, argument 1, 2 and combined extraction, sense classification, overall

**Future experiments**:
- different deep learning architectures
- different representation structures
- long short-term memory (LSTM)
- neural Turing machines (NTM)

## References

[1] R. Prasad, N. Dinesh, A. Lee, E. Miltsakaki, L. Robaldo, A. Joshi, and B. Webber, "The Penn Discourse TreeBank 2.0," Proc. Sixth Int. Conf. Lang. Resour. Eval., pp. 2961–2968, 2008.
[2] F. Kong, H. Tou, and N. Guodong, "A Constituent-Based Approach to Argument Labeling with Joint Inference in Discourse Parsing," in Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014, pp. 68–77.
[3] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural Language Processing (almost) from Scratch," J. Mach. Learn. Res., vol. 12, pp. 2493–2537, 2011.
[4] R. Collobert and J. Weston, "A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning," in Proceedings of the 25th international conference on Machine learning, 2008, vol. 20, no. 1, pp. 160–167.
[5] O. Irsoy and C. Cardie, "Deep Recursive Neural Networks for Compositionality in Language," in Advances in Neural Information Processing Systems (NIPS), 2014, pp. 2096–2104.