

# EMNLP 2015 Highlights

Yonatan Belinkov

October 2015

# Tutorials

- “Transparent Machine Learning for Information Extraction: State-of-the-Art and the Future”, Laura Chiticariu, Yunyao Li and Frederick Reiss
- Transparent ML
  - Easy to comprehend, debug, enhance

# Tutorials

- “Computational Analysis of Affect and Emotion in Language”, Saif Mohammad and Cecilia Ovesdotter Alm
  - Emotion/affect vs sentiment
  - Linguistic/psychological/cognitive theories
  - Techniques and applications

# Tutorials

- “Semantic Similarity Frontiers: From Concepts to Documents”, David Jurgens and Mohammad Taher Pilehvar
- “Learning Semantic Relations from Text”, Preslav Nakov, Vivi Nastase, Diarmuid Ó Séaghdha and Stan Szpakowicz

# Workshops

- “Vision and Language”
  - “Visually-Verifiable Textual Entailment”
  - “Linguistic Analysis of Multi-Modal RNNs”
  - “Image Representations and New Domains in Neural Image Captioning”
  - ...

# Workshops

- “Computational Approaches to Subjectivity, Sentiment and Social Media Analysis”
- “Health Text Mining and Information Analysis”
- “Linking Models of Lexical, Sentential and Discourse-level Semantics”
  - Keynote by Jacob Eisenstein
- “Cognitive Aspects of Computational Language Learning”

# Keynotes

- “Deep Learning of Semantic Representations”, Yoshua Bengio
- “Measuring How Elected Officials and Constituents Communicate”, Justin Grimmer

# Best Papers

- “Broad-coverage CCG Semantic Parsing with AMR”, Yoav Artzi, Kenton Lee and Luke Zettlemoyer
- “Semantically Conditioned LSTM-based Natural Language Generation for Spoken Dialogue Systems”, Tsung-Hsien Wen, Milica Gasic, Nikola Mrkšić, Pei-Hao Su, David Vandyke and Steve Young



# Datasets

- Stanford Natural Language Inference corpus
  - Given photo caption, Turkers write:
    - True, potential, and false captions
  - Photo not shown
  - 570,000 pairs, 10% manually validated
- WikiQA
  - Questions from Bing queries
  - Answers from associated Wikipedia pages
  - 3,000 questions, 2/3 w/o answers

# Datasets

- QA-SRL
  - Use question-answer pairs to specify argument-predicate structure
  - Given sentence+verb, annotators write QA pairs
  - 3,000 sentences, 20,000 QA pairs
  - Question generation, answer identification
- Arabic sentiment dataset (ASTD) and lexicon (SLSA)

# Neural nets for text representation

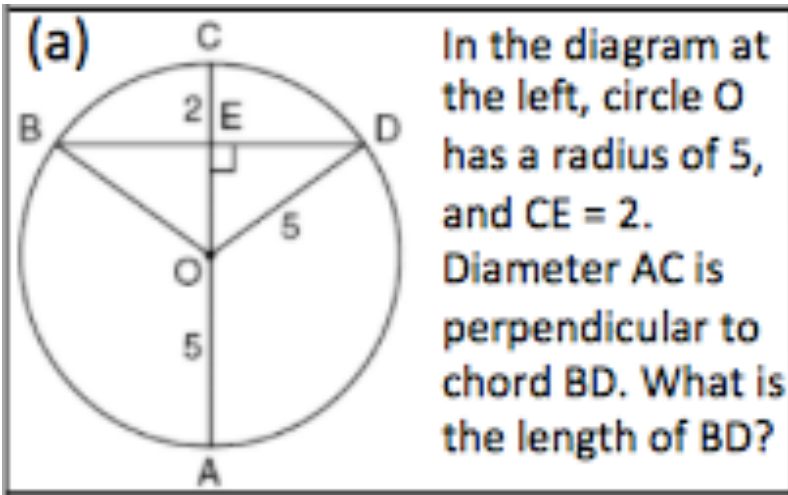
- Many Recurrent NN (mostly LSTM) papers
- Some Convolutional NN
- Few Recursive (using trees)
- Some combinations of the above

# Problem solving

A woman is paid 20 dollars for each day she works and **forfeits** a 5 dollars for each day she is idle. At the end of 25 days she nets 450 dollars. How many days did she work?

Waterford Spa had a full appointment calendar booked today. Help Janice figure out the schedule by matching each masseuse to her client, and determine the total price for each.

**Question 1** “When a baby shakes a rattle, it makes a noise. Which form of energy was changed to sound energy?” [Answer: mechanical energy]



1. Hannah paid more than Teri's client.
2. Freda paid 20 dollars more than Lynda's client.
3. Hannah paid 10 dollars less than Nancy's client.
4. Nancy's client, Hannah and Ginger were all different clients.
5. Hannah was either the person who paid \$180 or Lynda's client.

Clients: *Aimee, Ginger, Freda, Hannah.*

Prices: *\$150, \$160, \$170, \$180.*

Masseuses: *Lynda, Nancy, Tery, Whitney.*

# Problem solving

- “Solving Geometry Problems”
- “Learning to Automatically Solve Logic Grid Puzzles”
- “Solving General Arithmetic Word Problems”
- “Learn to Solve Algebra Word Problems Using Quadratic Programming”
- “Answering Elementary Science Questions by Constructing Coherent Scenes using Background Knowledge”

# Speech-related

- “A Model of Zero-Shot Learning of Spoken Language Understanding”
  - Dialogue act classification in restaurant domain
  - Multi-label classification with word and concept embeddings
- “Topic Identification and Discovery on Text and Speech”
  - Compare i-Vector, LDA, SAGE, LSA for topic ID
- “Hierarchical Latent Words Language Models for Robust Modeling to Out-Of Domain Tasks”
  - Out-of-domain ASR

# Speech-related

- “Turn-taking phenomena in incremental dialogue systems”
- “A Coarse-Grained Model for Optimal Coupling of ASR and SMT Systems for Speech Translation”

# Neural nets for text representation

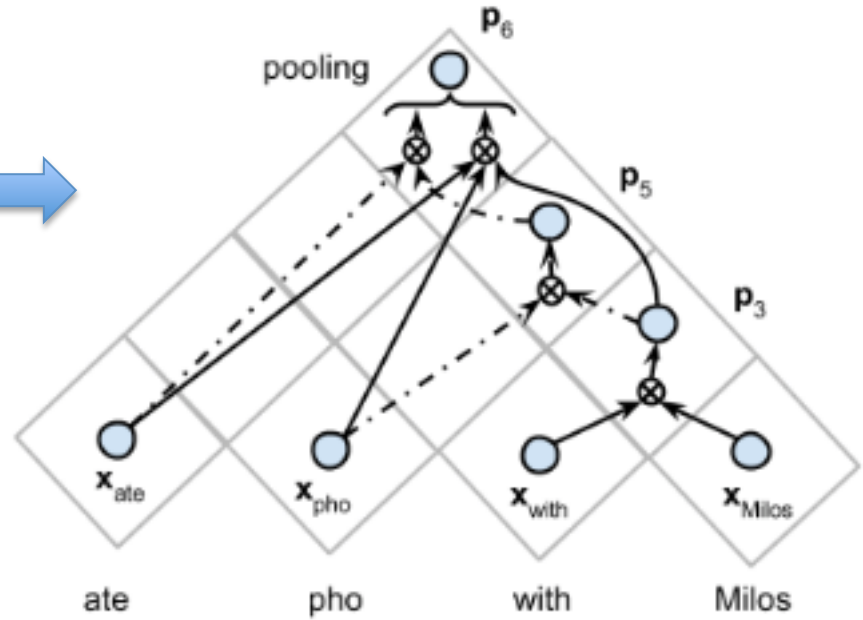
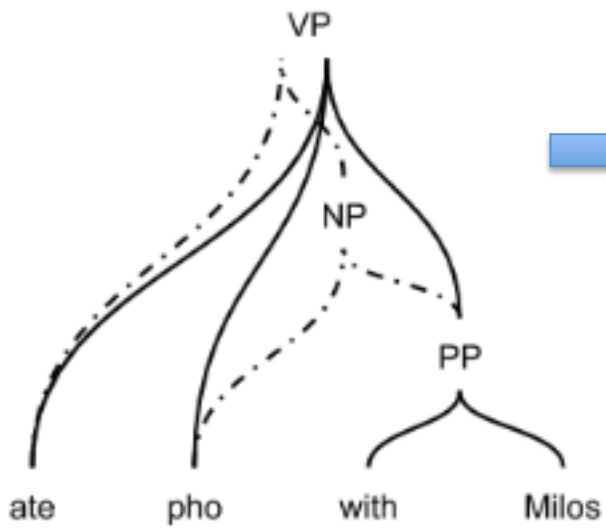
- Many Recurrent NN (mostly LSTM) papers
- Some Convolutional NN
- Few Recursive (using trees)
- Some combinations of the above



# CNN with trees

- “Semantic Relation Classification via CNNs with Simple Negative Sampling”
  - CNN on words along a dependency path
- “The Forest Convolutional Network”
  - Start with a forest of parse trees
  - Combine in a chart with convolutions
- “Discriminative Neural Sentence Modeling by Tree-Based Convolution”

# Forest CNN



# CNNs

- “Molding CNNs for text: non-linear, non-consecutive convolutions”
- “Multi-Perspective Sentence Similarity Modeling with Convolutional Neural Networks”
  - Multiple pooling types, granularities, similarity measurements, window sizes, etc.
  - Evaluate on paraphrasing and sentence similarity

# More CNNs

- “Distant Supervision for Relation Extraction via Piecewise Convolutional Neural Networks”
- “Deep Convolutional Neural Network Textual Features and Multiple Kernel Learning for Utterance-Level Multimodal Sentiment Analysis”
  - Sentiment in video segments: speech, image, text
  - CNN activations as features in SVM (or MKL)

# Trees or no trees?

- “When Are Tree Structures Necessary for Deep Learning of Representations?”
  - Compare Recursive to Recurrent NN models on several tasks
  - Recursive models mostly good for relation extraction, with long-distance dependencies
  - Otherwise, recurrent models are sufficient



# RNNs with trees

- “Classifying Relations via Long Short Term Memory Networks along Shortest Dependency Paths”
  - LSTM on words along a dependency path
  - Parallel LSTMs on POS tags, Wordnet, grammatical relations
  - Hidden layers combines everything

# RNNs with Attention

- “A Neural Attention Model for Sentence Summarization”
  - Encode sentence: BOW / BOW+Attention / CNN
  - Generate summary with a LM conditioned on input and current context
- “Effective Approaches to Attention-based Neural Machine Translation”
  - Global vs local attention → good for long sentences
  - MT evaluation



# Document models

- “Hierarchical Recurrent Neural Network for Document Modeling”
  - Word-level and sentence-level RNN LM
  - Tasks: sentence reordering, spoken Chinese-English translation
- “Document Modeling with Gated Recurrent Neural Network for Sentiment Classification”
  - Sentence representation w/ CNN or LSTM
  - Document representation w/ bidirectional gated units
  - Sentiment classification task

# Document models

- “Multi-Timescale Long Short-Term Memory Neural Network for Modelling Sentences and Documents”
  - Partition LSTM states, update at different time scales
  - Tasks: sentiment and question classification

# Embeddings

- “Distributional vectors encode referential attributes”
- “Building a shared world: mapping distributional to model-theoretic semantic spaces”
- “Deriving Boolean Structures from Distributional Vectors”
- “Learning Better Embeddings for Rare Words Using Distributional Representations”
- “Syntax-Aware Multi-Sense Word Embeddings for Deep Compositional Models of Meaning”
- “Trans-gram, Fast Cross-lingual Word-embeddings”
- “Do Multi-Sense Embeddings Improve Natural Language Understanding?”