Lecture 22: Representation Learning

Kai-Wei Chang
CS @ University of Virginia
kw@kwchang.net

Course webpage: http://kwchang.net/teaching/NLP16
Feature Representations

Feature Representation

- Color_red
- Shape_round
- Has_leaf
...

Learning Algorithm
Feature Representation

- E.g., Conditional Random Field
- \( P(t|w) \propto \prod_i \exp(\sum_k \lambda_k f_k(t_i, w) + \sum_l \eta_l g_l(t_i, t_{i-1}, w)) \)
Feature Representation

- High-order combinations – kernel trick
Tree Kernel

- How to measure the similarity between two parse trees?
Learning representations via NN

- Identify high-order combinations
- NN architecture for encoding language structures
- Learn hierarchical representations
  - Representations for token/phrases/sentences
How to represent words?

- Token, bi-gram, n-gram (one-hot features)
- Word embeddings
- Task-specific word embeddings
  - E.g., for sentiment analysis
How to represent phrases/sentences?

- Recursive NN [Socher, Manning, Ng 11]
- Many follow-up approaches
Unsupervised Feature learning & deep learning, Andrew Ng
Auto-encoder and auto-decoder
Sequence to sequence models
[Sutskever, Vinyals & Le 14]

- Have been shown effective in machine translation, image captioning and many structured tasks
NLP problems are structural
  - Output variables are inter-correlated
  - Need joint predictions

Traditional approaches
  - Graphical model approaches
    - E.g., Probabilistic graphical models, structured perceptron
  - Sequence of decisions
    - E.g., incremental perceptron, L2S, transition-based methods
Recent trends

- Landscape of methods in Deep∩Structure
  - Deep learning/hidden representation
    e.g., seq2seq, RNN
  - Deep features into factors, traditional factor graph inference
    e.g., LSTM+CRF, graph transformer networks
  - Globally optimized transitional-based approaches
    e.g., beam-search seq2seq, SyntaxNet
- …