COMP90042 Web Search & Text Analysis

Workshop Week 7

Zenan Zhai

April 16, 2019

University of Melbourne

Outline

Introduction

- Motivation
- Challenges
- Definition
- Tagsets

Methods

- N-gram Models
- Rule-based Methods
- Probabilistic Graphic Models (PGMs)
- Feed-Forward Neural Networks (FFNNs)
- Recurrent Neural Networks (RNNs)

Introduction

Motivation:

- Understanding syntactical structure of a sentence.
- Informative features for downstream tasks (e.g. dependency parsing, relation extraction, named entity recognition)

Challenges:

- Ambiguity of words allows multiple possible tags.

Introduction

Definition:

- Assigning Part-of-Speech tags to each token in the sequence.

Tagsets:

- Varies in different datasets.
- Penn Treebank Tagset
- `https://www.sketchengine.eu/penn-treebank-tagset/`

Exercise:

- Pierre Vinken, 61 years old, will join the board as a nonexecutive director Nov. 29.

Outline

Introduction

- Motivation
- Challenges
- Definition
- Tagsets

Methods

- N-gram Models
- Rule-based Methods
- Probabilistic Graphic Models (PGMs)
- Feed-Forward Neural Networks (FFNNs)
- Recurrent Neural Networks (RNNs)

Rule-based & N-gram Models

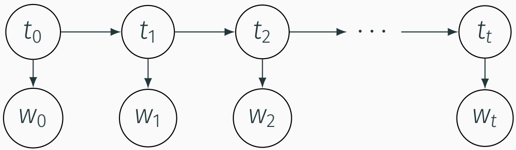
Rule-based

- Hand-crafted heuristics
- From possible POS narrow down to one tag

N-gram POS Tagging

- $P(t_i|w_i)$
- $P(t_i|w_i, t_{i-1}, t_{i-2})$

Probabilistic Graphic Model



Nodes denotes events, edges denotes dependencies.

Independent Assumption in HMM:

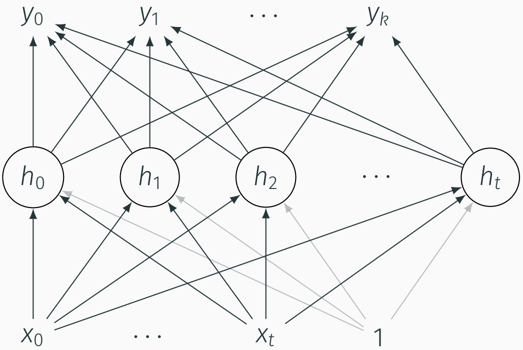
Transition Prob.	$P(t_i \vert t_{i-1})$	t_i only depends on t_{i-1}
Emission Prob.	$P(w_i \vert t_i)$	w_i only depends on t_i

Maximum Entropy Markov Model (MEMM)

- Discriminative Model: Use features as observations
- Model $P(t_i|w_i)$ instead of $P(w_i|t_i)$ in HMM

Conditional Random Fields (CRFs): Unidirectional PGM

Feed-Forward Neural Networks



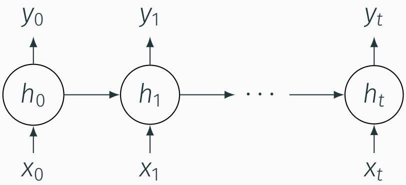
hidden states: $h_i = \sigma(\sum_j w_{ij} x_j + b_i)$

Inference: $\vec{P} = softmax(\vec{y})$

Loss in tagging: $\mathcal{L} = -\sum_i log(P(t_i | w_{i-2}, w_{i-1}, w_i, t_{i-2}, t_{i-1}))$

Loss in LM: $\mathcal{L} = -\sum_i log(P(w_i | w_{i-2}, w_{i-1}))$

Recurrent Neural Networks



$$h_t = \sigma_h(Wx_t + Vh_{t-1} + b_h)$$
$$y_t = \sigma_y(Uh_t + b_y)$$

t: time step
x: input vector
y: output vector
h: hidden states
W, U, V: Weight matrices
b: Bias
σ: activation function
(e.g. Sigmoid)

For sequence labeling

- $P(tag_i) = softmax(y_i)$
- $\mathcal{L} = -\sum_i log(P(tag_i))$

For language model

- $P(w_{i+1}) = softmax(y_i)$
- $\mathcal{L} = -\sum_i log(P(w_i))$