A UNIFIED FRAMEWORK FOR STRUCTURED PREDICTION

Wei Lu
StatNLP Research Group
Singapore University of Technology and Design
Outline

• Introduction
• Decoding, Learning
• Semi-Markov, Latent CRF, Latent SSVM
• Parsing with CRF, Hybrid Tree, and Predicting Overlapping Structures
• Pipeline, Mean Field and Neural CRF
1. Introduction
Structured Prediction

Fruit flies like a banana
Part-of-Speech Tagging

<table>
<thead>
<tr>
<th>A</th>
<th>N</th>
<th>V</th>
<th>D</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fruit</td>
<td>flies</td>
<td>like</td>
<td>a</td>
<td>banana</td>
</tr>
</tbody>
</table>
Fruit flies like a banana.
Constituency Parsing

S
  VP
    NP
      A
      Fruit
    NP
      N
      flies
    V
      like
    NP
      D
      a
      banana
Fruit flies like a banana.
Semantic Parsing

\text{LIKE}(F_{102}, B_{87})

Fruit flies like a banana
Semantic Parsing

\[
\begin{align*}
\text{LIKE} & \\
F_{102} & B_{87}
\end{align*}
\]

Fruit flies like a banana
Sentiment Analysis

( neutral )

Fruit flies like a banana

( positive )
Nested Chunking

Fruit flies like a banana
This Tutorial

- Shares a conceptually new way of thinking about building structured prediction models.
- Presents a unified structured prediction framework that encompasses classic models, and is able to model structures that standard Graphical Models cannot.
- Provides a way to rapidly prototype novel structured prediction models for new tasks.
A Unified Framework
Structured Prediction

One Assumption
Structures are constructed by following a collection of discrete actions.
States, Actions

S: shift
L: left-arc
R: right-arc

Fruit flies like a banana
States, Actions

S: shift
L: left-arc
R: right-arc

Fruit flies like a banana
States, Actions

\[ S: \text{shift} \]
\[ L: \text{left-arc} \]
\[ R: \text{right-arc} \]
States, Actions

S: shift
L: left-arc
R: right-arc

Example:
Fruit flies like a banana

Diagram:

S
S1 -> S2 -> L -> S3 -> S4

amod

Fruit flies like a banana
States, Actions

S: shift
L: left-arc
R: right-arc

Fruit flies like a banana

States: $S_1$, $S_2$, $S_3$, $S_4$, $S_5$

Actions: S, S, L, S, L

Roles: amod, nsubj

Fruit - amod
flies - nsubj
like -
a -
banana
States, Actions

S: shift
L: left-arc
R: right-arc
States, Actions

S: shift
L: left-arc
R: right-arc

Fruit flies like a banana.
States, Actions

S: shift
L: left-arc
R: right-arc

Fruit flies like a banana.
Fruit flies like a banana.
States, Actions

S: shift
L: left-arc
R: right-arc

Fruit
flies
like
a
banana

amod
nsubj
root

dobj
det
States, Actions

S: shift
L: left-arc
R: right-arc

Fruit flies like a banana.
States, Actions, Paths

A → AN → ANV → ANVD → ANVDN

Fruit flies like a banana
Fruit flies like a banana.
Score of a Path

Fruit flies like a banana
Score of a Path

Score of the path $p$

Score of each edge $e$

$S_w(p) = \sum_{e \in p} s_w(e)$

Parameters

Score of a Path

Fruit flies like a banana
Score of a Path

\[ S_w(p_1) = 7.2 + 3.5 + 0.9 + 2.0 + 3.2 - 10.2 = 6.6 \]

\[ S_w(p_2) = -0.2 - 4.2 - 1.7 - 3.0 + 2.2 + 12.8 = 6.9 \]
Search
Exhaustive Search
Beam Search
Heuristics Search

Fruit flies like a banana

D → DN → DNV → DNVDN
A → AN → ANV → ANVDN

7.2 3.5 0.9 2.0 3.2
-0.2 -4.2 -1.7 -3.0 2.2

−10.2 12.8
Score of an Edge

\[ s_w(e) = w \cdot f(e) = w \cdot f(x, [s, a]) \]
Score of an Edge

$s_w(e)$

D N V D

D N V

Fruit flies like a banana
Score of an Edge

\[ s_w(e) = w \cdot f(x, [s, a]) \]
Score of an Edge

\[ w = \begin{bmatrix} 1.2 \\ \vdots \\ -3.1 \end{bmatrix} \]

\[ f(x, [s, a]) = f(x, [\text{DNV}, D]) = \begin{bmatrix} 1 \\ \vdots \\ 0 \end{bmatrix} \]

\[ s_w(e) = w \cdot f(x, [s, a]) \]
Score of an Edge

\[ w = \begin{bmatrix} 1.2 \\ \vdots \\ -3.1 \end{bmatrix} \]

\[ f(x, [s, a]) = f(x, [\text{D\text{N}V}, \text{D}]) = \begin{bmatrix} 1 \\ \vdots \\ 0 \end{bmatrix} \]

\[ s_w(e) = w \cdot f(x, [s, a]) = 2.0 \]

Fruit flies like a banana
Score of an Edge

\[ w = \begin{bmatrix} 1.2 \\ \vdots \\ -3.1 \end{bmatrix} \]

\[ f(x, [s, a]) = f(x, [\text{DNV}, D]) = \begin{bmatrix} 1 \\ \vdots \\ 0 \end{bmatrix} \]

\[ s_w(e) = w \cdot f(x, [s, a]) = 2.0 \]
Score of an Edge

\[ \mathbf{w} = \begin{bmatrix} 1.2 \\ \vdots \\ -3.1 \end{bmatrix} \]

\[ \mathbf{f}(x, [s, a]) = \mathbf{f}(x, [\text{D N V}, \text{D}]) = \begin{bmatrix} 0 \\ \vdots \\ 1 \end{bmatrix} \]

\[ s_{\mathbf{w}}(e) = \mathbf{w} \cdot \mathbf{f}(x, [s, a]) = 1.8 \]
Score of an Edge

\[ s_w(e) = w \cdot f(e) = w \cdot f(x, [s, a]) \]

Fruit flies like a banana
Search Graph

\[ s_w(e) = \mathbf{w} \cdot \mathbf{f}(e) = \mathbf{w} \cdot \mathbf{f}(x, [s, a]) \]
\[ s_w(e) = w \cdot f(e) = w \cdot f(x, [s, a]) \]
Search Graph

\[ s_w(e) = \mathbf{w} \cdot \mathbf{f}(e) = \mathbf{w} \cdot \mathbf{f}(x, [s, a]) \]

Fruit flies like a banana

\[ s_w(e) = \mathbf{w} \cdot \mathbf{f}(e) = \mathbf{w} \cdot \mathbf{f}(x, [s, a]) \]
Fruit flies like a banana

\[ s_w(e) = w \cdot f(e) = w \cdot f(x, [s, a]) \]

Search Graph

Search Graph

\[ s_w(e) = w \cdot f(e) = w \cdot f(x, [s, a]) \]
Fruit flies like a banana.
Fruit flies like a banana.
Fruit flies like a banana.
Fruit flies like a banana.
Fruit flies like a banana.
A Compact Search Graph

Fruit flies like a banana
Fruit flies like a banana.
Decoding
Decoding

<table>
<thead>
<tr>
<th>Given</th>
<th>Find</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x, w$</td>
<td>$y$</td>
</tr>
</tbody>
</table>

Given a model $P(x|w)$, find $y$.
The Search Problem

\[
\max_y w \cdot f(x, y) = \max_y \sum_j w \cdot f(x, [y^j, y^{j+1}])
\]
The Search Problem

\[
\max_y w \cdot f(x, y) = \max_y \sum_j w \cdot f(x, [y^j, y^{j+1}]) \\
\arg \max_y w \cdot f(x, y)
\]
Viterbi

\[
\max_y w \cdot f(x, y) = \max_y \sum_j w \cdot f(x, [y^j, y^{j+1}])
\]

\[
\arg\max_y w \cdot f(x, y)
\]
Viterbi

\[ \max_y w \cdot f(x, y) = \max_y \sum_j w \cdot f(x, [y^j, y^{j+1}]) \]
Viterbi

\[ \max_y w \cdot f(x, y) = \max_y \sum_j w \cdot f(x, [y^j, y^{j+1}]) \]
Viterbi

$$\max_y w \cdot f(x, y) = \max_y \sum_j w \cdot f(x, [y^j, y^{j+1}])$$

Fruit flies like a banana
Viterbi

\[ \max_y w \cdot f(x, y) = \max_y \sum_j w \cdot f(x, [y^j, y^{j+1}]) \]
Viterbi

$$\max_y w \cdot f(x, y) = \max_y \sum_j w \cdot f(x, [y^j, y^{j+1}])$$
Viterbi

$$\max_y w \cdot f(x, y) = \max_y \sum_j w \cdot f(x, [y^j, y^{j+1}])$$
Viterbi

$$\max_y w \cdot f(x, y) = \max_y \sum_j w \cdot f(x, [y^j, y^{j+1}])$$
Viterbi

\[
\max_y w \cdot f(x, y) = \max_y \sum_j w \cdot f(x, [y^j, y^{j+1}]) = 9.7
\]
Viterbi

\[
\max_y w \cdot f(x, y) = \max_y \sum_j w \cdot f(x, [y^j, y^{j+1}]) = 9.7
\]

\[
\arg \max_y w \cdot f(x, y)
\]
Viterbi

\[
\max_y w \cdot f(x, y) = \max_y \sum_j w \cdot f(x, [y^j, y^{j+1}]) = 9.7
\]

\[
\arg \max_y w \cdot f(x, y)
\]
Viterbi

\[
\max_y w \cdot f(x, y) = \max_y \sum_j w \cdot f(x, [y^j, y^{j+1}]) = 9.7
\]

\[
\arg \max_y w \cdot f(x, y)
\]
Viterbi

\[
\max_y w \cdot f(x, y) = \max_y \sum_j w \cdot f(x, [y^j, y^{j+1}]) = 9.7
\]

\[
\text{arg} \max_y w \cdot f(x, y)
\]
Viterbi

\[
\max_y w \cdot f(x, y) = \max_y \sum_j w \cdot f(x, [y^j, y^{j+1}]) = 9.7
\]

arg \max_y w \cdot f(x, y)
Viterbi

\[ \max_y w \cdot f(x, y) = \max_y \sum_j w \cdot f(x, [y^j, y^{j+1}]) = 9.7 \]

arg \( \max_y w \cdot f(x, y) \)
Viterbi

\[
\max_y w \cdot f(x, y) = \max_y \sum_j w \cdot f(x, [y^j, y^{j+1}]) = 9.7
\]

\[
\text{arg max}_y w \cdot f(x, y)
\]
**MAP Inference**

\[
\max_y w \cdot f(x, y) = \max_y \sum_j w \cdot f(x, [y^j, y^{j+1}]) = 9.7
\]

\[
\arg \max_y w \cdot f(x, y)
\]
Inference

MAP

max

w
So Far

Given the parameters $w$, how to search for the optimal path (and its score)

Next...

How to learn the parameters $w$?