3. Semi-Markov CRF, Latent CRF, Parsing with CRF, Hybrid Tree and Predicting Overlapping Structures
Noun-Phrase Chunking

Fruit flies like a banana.
Chunking with CRF

\[
\min_w \sum_i \left( -w \cdot f(x_i, y_i) + \log \sum_y \exp (w \cdot f(x_i, y)) \right)
\]

\[
f(x, [y^j, y^{j+1}])
\]
Chunking with CRF

\[
\min_w \sum_i \left( -w \cdot f(x_i, y_i) + \log \sum_y \exp (w \cdot f(x_i, y)) \right)
\]

\[
f(x, [y^j, y^{j+1}])
\]
Chunking with CRF

\[
\min_w \sum_i (-w \cdot f(x_i, y_i) + \log \sum_y \exp (w \cdot f(x_i, y))) \\
\quad f(x, [y^j, y^{j+1}])
\]
Semi-Markov CRF

$$\min_w \sum_i \left( -w \cdot f(x_i, y_i) + \log \sum_y \exp \left( w \cdot f(x_i, y) \right) \right)$$

$$f_1(x, [y^{0-1} = N, y^2 = O])$$

Fruit flies like a banana

Semi-Markov CRF

\[
\min_w \sum_i \left( -w \cdot f(x_i, y_i) + \log \sum_y \exp (w \cdot f(x_i, y)) \right)
\]

\[
f_1(x, [y^{0-1} = N, y^2 = O])
\]

Semi-Markov CRF

\[
\min_w \sum_i \left( -w \cdot f(x_i, y_i) + \log \sum_y \exp (w \cdot f(x_i, y)) \right) \\
\]

\[
f_1(x, [y^{0-1} = N, y^2 = O])
\]
Semi-Markov CRF

\[
\min_w \sum_i \left( -w \cdot f(x_i, y_i) + \log \sum_y \exp (w \cdot f(x_i, y)) \right) \\
= f_1(x, [y^{0-1} = N, y^2 = O])
\]
Weak Semi-Markov CRF

$$\min_w \sum_i \left( -w \cdot f(x_i, y_i) + \log \sum_y \exp (w \cdot f(x_i, y)) \right)$$

$$f_1(x, [y^{0-1} = N]) \quad f_2(x, [y^1 = N, y^2 = O])$$

Fruit  flies  like  a  banana

Weak Semi-Markov CRF

\[
\min_w \sum_i \left( -w \cdot f(x_i, y_i) + \log \sum_y \exp \left( w \cdot f(x_i, y) \right) \right)
\]

\[
f_1(x, [y^{0-1} = N]) \quad f_2(x, [y^1 = N, y^2 = O])
\]
Weak Semi-Markov CRF

$$\min_w \sum_i \left( -w \cdot f(x_i, y_i) + \log \sum_y \exp (w \cdot f(x_i, y)) \right)$$

$$f_1(x, [y^{0-1} = N]) \quad f_2(x, [y^1 = N, y^2 = O])$$
Latent-Variable CRF

\[ \max_w \log p(y|x) \]
Latent-Variable CRF

\[ \max_w \log p(y|x) \]
Latent-Variable CRF

$$\max_w \log p(y|x) = \max_w \log \sum_h p(y, h|x)$$
Latent-Variable CRF

$$\max_w \log \left( \frac{\sum_h \exp (w \cdot f(x, h, y))}{\sum_{h', y'} \exp (w \cdot f(x, h', y'))} \right)$$

Fruit
flies
like
a
banana
Latent-Variable CRF

$$\min_w \sum_i \left( -\log \sum_h \exp(w \cdot f(x_i, h, y_i)) + \log \sum_{h', y} \exp(w \cdot f(x_i, h', y)) \right)$$

Fruit flies like a banana.
Latent-Variable CRF

$$\min_w \sum_i \left( - \log \sum_h \exp \left( w \cdot f(x_i, h, y_i) \right) + \log \sum_{h', y} \exp \left( w \cdot f(x_i, h', y) \right) \right)$$
Latent-Variable CRF

\[
\min_w \sum_i \left( -\log \sum_h \exp (w \cdot f(x_i, h, y_i)) + \log \sum_{h', y} \exp (w \cdot f(x_i, h', y)) \right)
\]
Latent-Variable CRF

\[
\min_w \sum_i \left( -\log \sum_h \exp (w \cdot f(x_i, h, y_i)) + \log \sum_{h', y} \exp (w \cdot f(x_i, h', y)) \right)
\]
Latent-Variable SSVM

\[
\min_w \sum_i \left( -\log \sum_h \exp (w \cdot f(x_i, h, y_i)) + \log \sum_{h', y} \exp (w \cdot f(x_i, h', y)) \right)
\]

\[
\min_w \sum_i \left( -\max_h (w \cdot f(x_i, h, y_i)) + \max_{h', y} (\Delta(y_i, y, h') + w \cdot f(x_i, h', y)) \right)
\]

Latent-Variable SSVM

$$\min_w \sum_i \left( -\max_h (w \cdot f(x_i, h, y_i)) + \max_{h', y} \left( \Delta(y_i, y, h') + w \cdot f(x_i, h', y) \right) \right)$$
Latent-Variable SSVM

$$\min_w \sum_i \left( -\max_h (w \cdot f(x_i, h, y_i)) + \max_{h', y} \left( \Delta(y_i, y, h') + w \cdot f(x_i, h', y) \right) \right)$$
Learning

Structured Perceptron, SSVM, Latent SSVM, ...

\[ \Delta(y, y') \]

\[ L \]

\[ w^{(k)} \]

\[ U \]

\[ \max \]

\[ w^{(k+1)} \]
Learning

Linear/Semi/Latent/Softmax-margin CRF, ...

\[ w^{(k)} \rightarrow \log \sum \exp \rightarrow w^{(k+1)} \]

\[ \Delta(y, y') \rightarrow \log \sum \exp \rightarrow \log \sum \exp \]
So Far

We focused on predicting structures in the form of linear chains.

Next

Structured prediction problems beyond linear structures.
Parsing with CRF, Hybrid Tree, and Predicting Overlapping Structures
Constituency Parsing

S
  /  \\      \\
 NP  VP
   /  \\   /
  NP  D  NP
  |   |  |
 A   V  N
 Fruit flies like a banana
Constituency Parsing

NP -> D N

VP -> V NP

NP -> A N

S

NP

A
Fruit

NP

N
flies

VP

V
like

NP

D
a

NP

N
banana
Constituency Parsing

NP
V

S

NP

VP

NP -> D N
NP -> A N

VP -> V NP

A
Fruit

N
flies

V
like

D
a

N
banana
Parsing with CRF

\[
\min_w \sum_i \left( -w \cdot f(x_i, y_i) + \log \sum_y \exp \left( w \cdot f(x_i, y) \right) \right)
\]
Fruit flies like a banana.

\[
\min_w \sum_i \left( -w \cdot f(x_i, y_i) + \log \sum_y \exp \left( w \cdot f(x_i, y) \right) \right)
\]
Hyperpath

$$\min_{w} \sum_{i} \left( -w \cdot f(x_i, y_i) + \log \sum_{y} \exp (w \cdot f(x_i, y)) \right)$$
Hyperpath, Hypergraph

Fruit flies like a banana.
Parsing with CRF

\[
\min_w \sum_i \left(-w \cdot f(x_i, y_i) + \log \sum_y \exp (w \cdot f(x_i, y))\right)
\]
Semantic Parsing

\[ \text{LIKE}(F102, B87) \]

Fruit flies like a banana
Semantic Parsing

\[
\text{LIKE}(F102, B87) = \text{LIKE}(A, B) = \text{LIKE}(A: F102, B: B87)
\]

\[
\text{Fruit flies like a banana}
\]
Hybrid Tree

\[
\min_w \sum_i \left( -\log \sum_h \exp (w \cdot f(x_i, h, y_i)) + \log \sum_{h', y} \exp (w \cdot f(x_i, h', y)) \right)
\]

Hybrid Tree

$$\min_w \sum_i \left( - \log \sum_h \exp (w \cdot f(x_i, h, y_i)) + \log \sum_{h', y} \exp (w \cdot f(x_i, h', y)) \right)$$
Hybrid Tree

$$\min_w \sum_i \left( -\log \sum_h \exp (w \cdot f(x_i, h, y_i)) + \log \sum_{h', y} \exp (w \cdot f(x_i, h', y)) \right)$$
Hybrid Tree

\[
\min_w \sum_i \left( - \log \sum_h \exp \left( w \cdot f(x_i, h, y_i) \right) + \log \sum_{h', y} \exp \left( w \cdot f(x_i, h', y) \right) \right)
\]
Hybrid Tree

$$\min_w \sum_i \left( -\log \sum_h \exp (w \cdot f(x_i, h, y_i)) + \log \sum_{h', y} \exp (w \cdot f(x_i, h', y)) \right)$$
Hybrid Tree

\[
\min_w \sum_i \left( -\log \sum_h \exp (w \cdot f(x_i, h, y_i)) + \log \sum_{h', y} \exp (w \cdot f(x_i, h', y)) \right)
\]

LIKE(F102, B87)

IF102
LIKE(A, B)
B:B87

Fruit flies like a banana
Hybrid Tree

\[ \text{LIKE}(F102, B87) \]

\[ \text{LIKE}(A, B) \]

\[ A: F102 \]

\[ B: B87 \]

\[ \text{Fruit} \rightarrow \text{flies} \rightarrow \text{like} \rightarrow a \rightarrow \text{banana} \]
Hybrid Tree

Fruit

Hybrid Tree

flies

like

a

banana

LIKE(A, B)

LIKE[AWB]

LIKE[A]

LIKE[WB]

LIKE[B]

LIKE[w]

B87

B87[w]

F102

F102[w]
Hypergraphs

Q1. How is this formalism related to Factor Graphs or Graphical Models?

A. Our Hypergraphs are able to capture context-specific independence (CSI) conveniently.
Q2. Other than CSI, can we do things with Hypergraphs that standard Graphical Models cannot do?

A. Yes. See next example.
Predicting Overlapping Structures
Nested Chunking

What can we do with conventional Graphical Models?

Fruit flies like a banana
Approach 1

Pipeline

Fruit flies like a banana
Approach 2
Joint

B I O B I
O B O O B

Fruit flies like a banana
Approach 3

Tree

Fruit like a banana
Overlapping Structures

What can we do with our hypergraphs?
Hyperpath
Separable Hyperpath

A hyperpath that visits each node in the hypergraph at most once.
Separable Hyperpath
Separable Hypergraphs

A hypergraph whose hyperpaths are all separable.

All the examples that we have seen so far are separable hypergraphs.

What about non-separable hypergraphs?
Non-separable Hyperpath
Non-separable Hypergraph

Fruit flies like a banana
Fruit flies like a banana
Fruit flies like a banana.
Fruit flies like a banana
Labeled Hypergraph

Fruit flies like a banana
Fruit flies like a banana
Fruit flies like a banana
Inference

Marginal

\[ \log \sum \exp W \]
Learning

Linear/Semi/Latent/Parsing/Softmax-Margin CRF, Hybrid Tree, Mention Hypergraphs, ...

$L \xrightarrow{} \log \sum \exp \xrightarrow{} w^{(k+1)}$

$w^{(k)} \xrightarrow{} \log \sum \exp \xrightarrow{} w^{(k+1)}$

$U \xrightarrow{} \log \sum \exp \xrightarrow{} w^{(k+1)}$

$\Delta(y, y') \xrightarrow{} \log \sum \exp \xrightarrow{} w^{(k+1)}$
Learning

Structured Perceptron, SSVM, Latent SSVM, ...

\[ L \]
\[ w^{(k)} \]
\[ U \]
\[ \Delta(y, y') \]
\[ \max \]
\[ w^{(k+1)} \]
So Far

Semi-Markov CRF, Latent CRF, Parsing with CRF, Hybrid Tree and Predicting Overlapping Structures.

Next...

Pipeline, Joint Models and Neural CRF