Neural Segmental Hypergraphs for Overlapping Mention Recognition

Bailin Wang [1] Wei Lu[2]

[1]University of Massachusetts Amherst

[2]Singapore University of Technology and Design

November 2, 2018

Overlapping Mention Recognition

A sentence from the GENIA corpus

In addition , we demonstrated that the $\langle$ $\langle$ EBNA - 1 $\rangle_{protein}$ gene $\rangle_{DNA}$ in infected $\langle$ thymocytes $\rangle_{cell\_type}$ was transcribed from the $\langle$ Fp promoter $\rangle_{DNA}$, rather than from the $\langle$ Cp / Wp promoter $\rangle_{DNA}$ which is used in $\langle$ latently infected $\langle$ B cells $\rangle_{cell\_type}$ $\rangle_{cell\_type}$.

Mention:

1) a reference to something

2) associated with a semantic type

Overlapping Mention Recognition

A sentence from the GENIA corpus

In addition , we demonstrated that the $\langle$ $\langle$ EBNA - 1 $\rangle_{protein}$ gene $\rangle_{DNA}$ in infected $\langle$ thymocytes $\rangle_{cell\_type}$ was transcribed from the $\langle$ Fp promoter $\rangle_{DNA}$, rather than from the $\langle$ Cp / Wp promoter $\rangle_{DNA}$ which is used in $\langle$ latently infected $\langle$ B cells $\rangle_{cell\_type}$ $\rangle_{cell\_type}$.

Overlapping mentions are frequent:

1) In GENIA, around 20% mentions overlap with one another.

2) In ACE datasets, the number is around 40%.

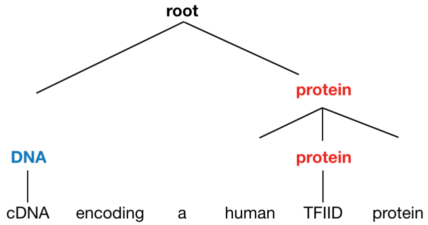
Challenge

$\langle$ cDNA $\rangle_{DNA}$ encoding a $\langle$ human $\langle$ TFIID $\rangle_{protein}$ protein $\rangle_{protein}$

- The search space of possible mention combinations increases to $2^{O(mn^2)}$, compared with non-overlapping mention recognition whose search space is $O((m+1)^n)$. [1]
- Traditional sequence models like linear-chain CRF are unable to model overlapping mentions.

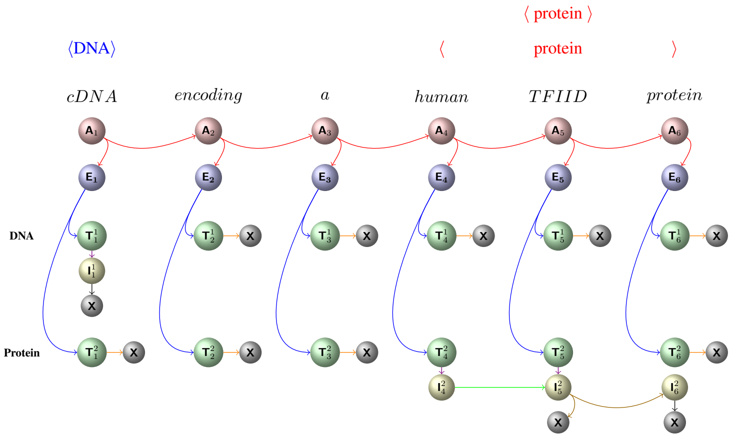
[1] m: number of semantic types, n: number of words.

Constituency Parsing (Finkel and Manning (2009))



Issue: chart-based parsing has the cubic time complexity in the number of words.

Mention Hypergraph (Lu and Roth (2015))



Mention Hypergraph (Lu and Roth (2015))

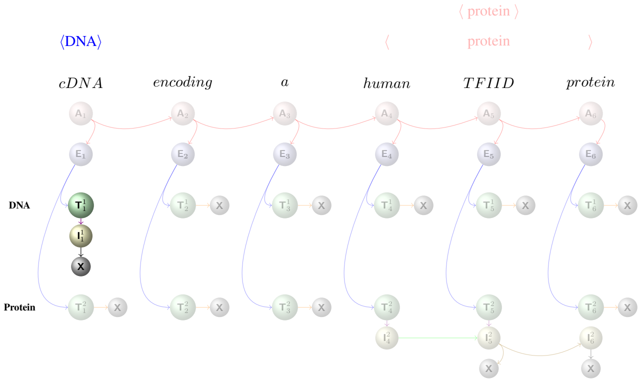
Nodes:

- $\mathbf{T}_i^k$ represents all mentions of type k **starting** with the i-th word
- $\mathbf{I}_i^k$ represents all mentions of type k **containing** the i-th word
- $\mathbf{X}$ marks the end of a mention.

Hyperedges (Production Rules):

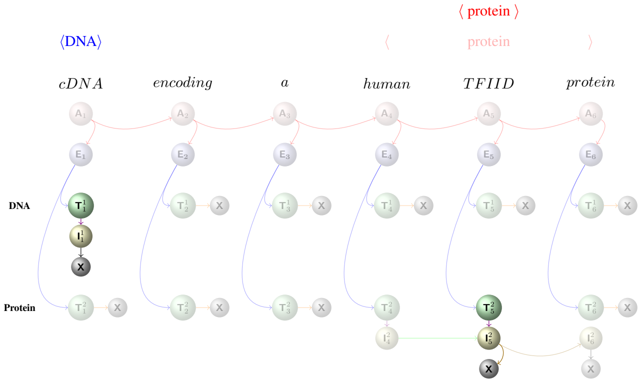
- $\{\ \mathbf{T}_i^k \rightarrow \mathbf{I}_i^k\ \}$, $\{\ \mathbf{T}_i^k \rightarrow \mathbf{X}\ \}$
- $\{\ \mathbf{I}_i^k \rightarrow \mathbf{I}_{i+1}^k\ \}$, $\{\ \mathbf{I}_i^k \rightarrow \mathbf{X}\ \}$, $\{\ \mathbf{I}_i^k \rightarrow \mathbf{I}_{i+1}^k, \mathbf{X}\ \}$

Mention Hypergraph (Lu and Roth (2015))



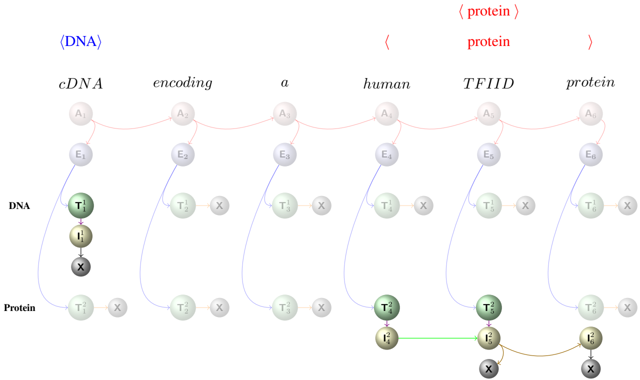
- $\mathbf{T}_i^k$ represents all mentions of type k **starting** with the i-th word
- $\mathbf{I}_i^k$ represents all mentions of type k **containing** the i-th word
- **X** marks the end of a mention.

Mention Hypergraph (Lu and Roth (2015))



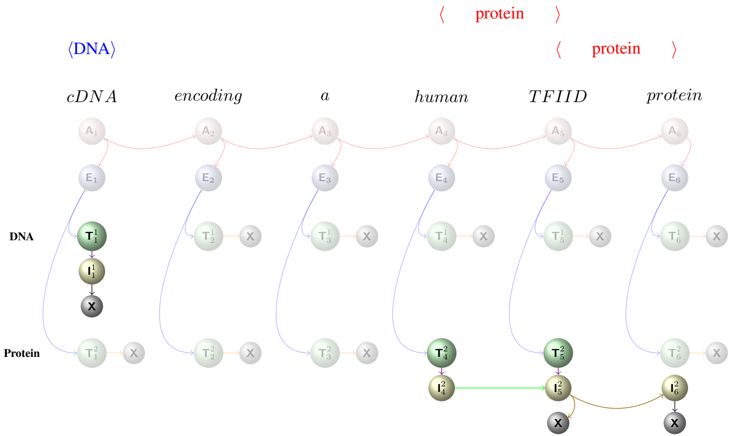
- $\mathbf{T}_i^k$ represents all mentions of type k **starting** with the i-th word
- $\mathbf{I}_i^k$ represents all mentions of type k **containing** the i-th word
- **X** marks the end of a mention.

Mention Hypergraph (Lu and Roth (2015))



- $\mathbf{T}_i^k$ represents all mentions of type k **starting** with the i-th word
- $\mathbf{I}_i^k$ represents all mentions of type k **containing** the i-th word
- **X** marks the end of a mention.

Structural Ambiguity of Mention Hypergraph



The hypergraph has multiple interpretations, such as the one shown above.

Segmental Hypergraph

Basic idea: Model the left and right boundaries of mention simultaneously.

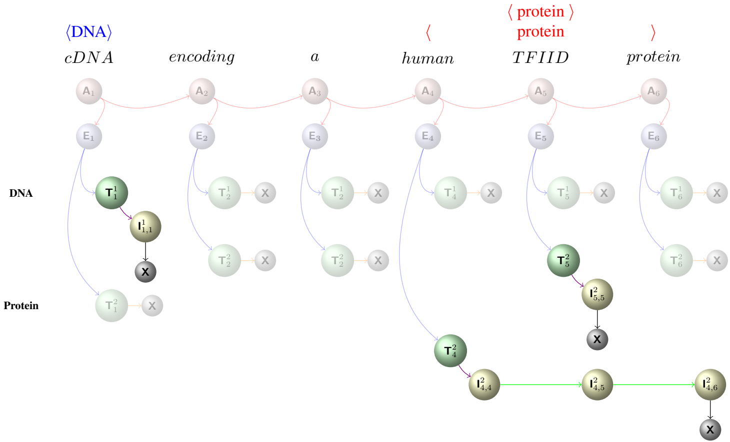
Nodes:

- $\mathbf{T}_i^k$ represents all mentions of type k **starting** with the i-th word
- $\mathbf{I}_{i,j}^k$: all mentions of type k that contain the j-th word and start with the i-th word
- $\mathbf{X}$ marks the end of a mention.

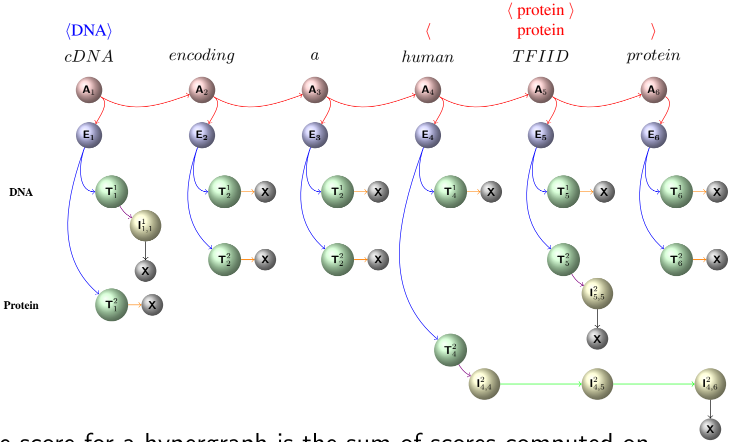
Hyperedges (Production Rules):

- $\{ \mathbf{T}_i^k \rightarrow \mathbf{I}_{i,i}^k \}$, $\{ \mathbf{T}_i^k \rightarrow \mathbf{X} \}$
- $\{ \mathbf{I}_{i,i}^k \rightarrow \mathbf{I}_{i,i+1}^k \}$, $\{ \mathbf{I}_{i,i}^k \rightarrow \mathbf{X} \}$, $\{ \mathbf{I}_{i,i}^k \rightarrow \mathbf{I}_{i+1}^k, \mathbf{X} \}$

Segmental Hypergraph



Scoring Segmental Hypergraph



- The score for a hypergraph is the sum of scores computed on each hyperedge e: $f(\boldsymbol{x}, \boldsymbol{y}) = \sum_{e \in \mathcal{G}_{\boldsymbol{y}}} \psi(e, \boldsymbol{x})$
- Hyperedges with parent node being $\mathbf{I}_{i,j}^{k}$ involve span-level features.
- In our neural settings, both span-level and word-level features could be learned using biLSTM efficiently .

Learning of Segmental Hypergraph

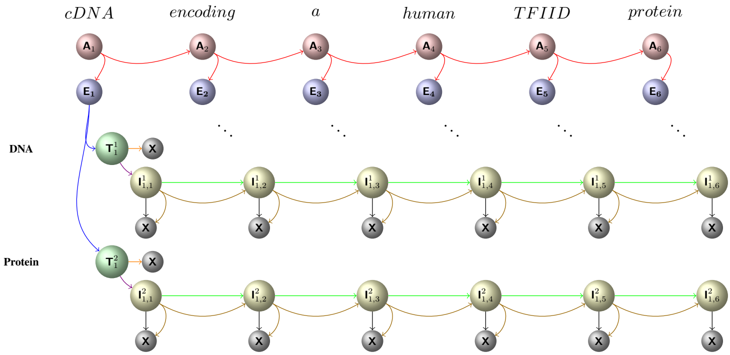


Figure: Complete segmental hypergraph

- Learning Objective: Maximize $p(\boldsymbol{y}|\boldsymbol{x}) = \frac{\exp f(\boldsymbol{x},\boldsymbol{y})}{\sum_{\boldsymbol{y}'} \exp f(\boldsymbol{x},\boldsymbol{y}')}$
- Computation in the complete segmental hypergraph: $Z(\boldsymbol{x}) = \sum_{\boldsymbol{y}'} \exp f(\boldsymbol{x},\boldsymbol{y}')$
- Time Complexity: corresponds with the number of nodes $O(mn^2)$

Length Restriction

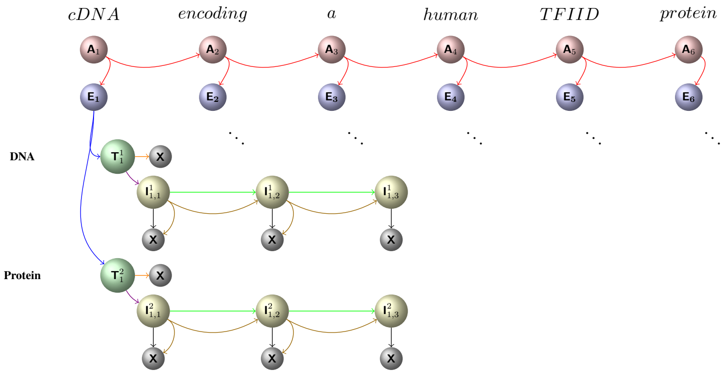


Figure: A segmental hypergraph with length restriction $c = 3$

- Restrict the maximal length of a mention: time complexity is then reduced to $O(cmn)$, analogous to semi-CRF.

Experiment Results (Non-neural version)

		ACE-2004			ACE-2005			GENIA		
		P	R	F_1	P	R	F_1	P	R	F_1
Non-Neural	CRF (LINEAR)	71.8	40.8	52.1	69.5	44.5	54.2	77.1	63.3	69.5
	CRF (CASCADED)	78.4	46.4	58.3	74.8	49.1	59.3	75.9	66.1	70.6
	Semi-CRF (c=6)	76.1	41.4	53.6	72.8	45.0	55.6	74.5	66.0	70.0
	Semi-CRF (c=n)	66.7	42.0	51.5	67.5	46.1	54.8	74.2	65.8	69.7
	Finkel and Manning (2009)	-	-	-	-	-	-	75.4	65.9	70.3
	Lu and Roth (2015)	70.0	56.9	62.8	66.3	59.2	62.5	74.2	66.7	70.3
	Muis and Lu (2017)	72.7	58.0	64.5	69.1	58.1	63.1	75.4	66.8	70.8
	SH (-NN, c=6)	69.4	57.0	62.0	70.3	55.8	62.2	77.0	66.1	71.1
	SH (-NN, c=n)	71.1	60.6	65.4	69.5	60.7	64.8	76.2	67.5	71.6

- SH (-NN): segmental hypergraphs with handcrafted features.
- c: maximal length of a mention, n: length of a given sentence

Experiment Results (Neural Version)

		ACE-2004			ACE-2005			GENIA		
		P	R	F_1	P	R	F_1	P	R	F_1
Neural	FOFE Xu et al. (2017) (c=6)	68.2	54.3	60.5	67.4	55.1	60.6	71.2	64.3	67.6
	FOFE Xu et al. (2017) (c=n)	57.3	46.8	51.5	56.3	44.6	49.8	63.2	59.3	61.2
	Katiyar and Cardie (2018)	73.6	71.8	72.7	70.6	70.4	70.5	**79.8**	68.2	73.6
	Ju et al. (2018) [2]	-	-	-	74.2	70.3	72.2	78.5	71.3	74.7
	Wang et al. (2018)	74.9	71.8	73.3	74.5	71.5	73.0	78.0	70.2	73.9
	SH (c=6)	79.1	67.3	72.7	75.7	69.6	72.5	76.6	71.0	73.7
	SH (c=n)	77.7	72.1	74.5	76.6	71.9	74.2	76.1	72.9	74.5
	SH (c=6) + *char*	**80.1**	67.5	73.3	75.9	70.0	72.8	76.8	71.8	74.2
	SH (c=n) + *char*	78.0	**72.4**	**75.1**	**76.8**	**72.3**	**74.5**	77.0	**73.3**	**75.1**

- SH: neural segmental hypergraphs
- +*char*: add character-level representations for each word (inspired by Lample et al. (2016))

Experiment Results (Full version)

		ACE-2004			ACE-2005			GENIA		
		P	R	F_1	P	R	F_1	P	R	F_1
Non-Neural	SH (-NN, c=6)	69.4	57.0	62.0	70.3	55.8	62.2	77.0	66.1	71.1
	SH (-NN, c=n)	71.1	60.6	65.4	69.5	60.7	64.8	76.2	67.5	71.6
Neural	SH (c=6)	79.1	67.3	72.7	75.7	69.6	72.5	76.6	71.0	73.7
	SH (c=n)	77.7	72.1	74.5	76.6	71.9	74.2	76.1	72.9	74.5
	SH (c=6) + *char*	**80.1**	67.5	73.3	75.9	70.0	72.8	76.8	71.8	74.2
	SH (c=n) + *char*	78.0	**72.4**	**75.1**	**76.8**	**72.3**	**74.5**	77.0	**73.3**	**75.1**

Neural models perform much better than non-neural models.

How well does it handle overlapping mentions?

	Overlapping			Non-Overlapping			w/s
	P	R	F_1	P	R	F_1	
Lu and Roth (2015)	68.1	52.6	59.4	64.1	65.1	64.6	503
Muis and Lu (2017)	70.4	55.0	61.8	67.2	63.4	65.2	253
Wang et al. (2018)	77.4	70.5	73.8	76.1	69.6	72.7	1445
SH (c=6)	80.2	68.3	73.8	74.8	70.0	72.3	248
SH (c=n)	80.6	73.6	76.9	75.5	71.5	73.4	157

Table: Results on different types of sentences (ACE05), w/s: # of words decoded per second.

What if the data has no overlapping mentions?

Model	F_1
SH (c=6)	89.6
SH (c=6) + *char*	90.5
SH (c=n)	89.2
SH (c=n) + *char*	90.2
Collobert et al. (2011)	88.7
Chiu and Nichols (2016)	90.9
Lample et al. (2016)	90.9
Ma and Hovy (2016)	91.2
Xu et al. (2017)	90.7
Strubell et al. (2017)	90.5

Table: Results on CoNLL-2003.

Summary

- A novel **segmental hypergraph** that is capable of modeling arbitrary combinations of mentions, capturing both span-level and word-level features, with no structural ambiguity.
- Our model features the time complexity of $O(mn^2)$, which can reduced to $O(cmn)$ if the length restriction is made.
- Our model achieves the state-of-the-art performance in three standard benchmark datasets.
- Code available: `http://statnlp.org/research/ie`

Thank you.

References I

Jason PC Chiu and Eric Nichols. 2016. Named entity recognition with bidirectional lstm-cnns. *Transactions of the Association for Computational Linguistics*, 4:357–370.

Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. 2011. Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, 12(Aug):2493–2537.

Jenny Rose Finkel and Christopher D Manning. 2009. Nested named entity recognition. In *Proc. of EMNLP*.

Meizhi Ju, Makoto Miwa, and Sophia Ananiadou. 2018. A neural layered model for nested named entity recognition. In *Proc. of NAACL-HLT*.

Arzoo Katiyar and Claire Cardie. 2018. Nested named entity recognition revisited. In *Proc. of NAACL-HLT*.

References II

Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. 2016. Neural architectures for named entity recognition. In *Proc. of NAACL-HLT*.

Wei Lu and Dan Roth. 2015. Joint mention extraction and classification with mention hypergraphs. In *Proc. of EMNLP*.

Xuezhe Ma and Eduard Hovy. 2016. End-to-end sequence labeling via bi-directional lstm-cnns-crf. In *Proc. of ACL*.

Aldrian Obaja Muis and Wei Lu. 2017. Labeling gaps between words: Recognizing overlapping mentions with mention separators. In *Proc. of EMNLP*.

Emma Strubell, Patrick Verga, David Belanger, and Andrew McCallum. 2017. Fast and accurate entity recognition with iterated dilated convolutions. In *Proc. of EMNLP*.

References III

Bailin Wang, Wei Lu, Yu Wang, and Hongxia Jin. 2018. A neural transition-based model for nested mention recognition. In *Proc. of EMNLP*.

Mingbin Xu, Hui Jiang, and Sedtawut Watcharawittayakul. 2017. A local detection approach for named entity recognition and mention detection. In *Proc. of ACL*.