FOFE-based Deep Neural Networks for Entity Discovery and Linking

Nargiza Nosirova

Mingbin Xu, **Nargiza Nosirova**, **Kelvin Jiang**, **Feng Wei** and **Hui Jiang**

Lassonade School of Engineering, York University, Canada

2017 TAC Workshop

FOFE-based Deep Neural Networks for Entity Discovery and Linking

Overview

- What is FOFE?
- FOFE-based model for Entity Discovery
- Ensemble modeling for Entity Discovery
- Multi-task model for Entity Discovery
- FOFE-based model for Entity Linking
- Experiments
- Conclusion

What is FOFE?

Definition (Fixed-size Ordinally Forgetting Encoding)

- $S = w_1, w_2, ..., w_n$ is a sequence of any discrete symbols;
- w_i is represented as $\boldsymbol{e_i}$ in 1-hot representation;
- the encoding of a partial sequence up to the t-th word is recursively defined as:

$$\boldsymbol{z_t} = \begin{cases} \boldsymbol{e_t}, & \text{if } t = 1 \\ \alpha \cdot \boldsymbol{z_{t-1}} + \boldsymbol{e_t}, & \text{otherwise} \end{cases}$$

- $\alpha \in (0, 1)$ and $t \in \{\mathbb{Z} | 1 \leq x \leq n\}$

What is FOFE? (continued)

Example

- $A = [1, 0, 0]$
- $B = [0, 1, 0]$
- $C = [0, 0, 1]$
- $ABC = [\alpha^2, \alpha, 1]$

What is FOFE? (continued)

Any **variable length** sequence is **losslessly** encoded into a **fixed-size** vector.

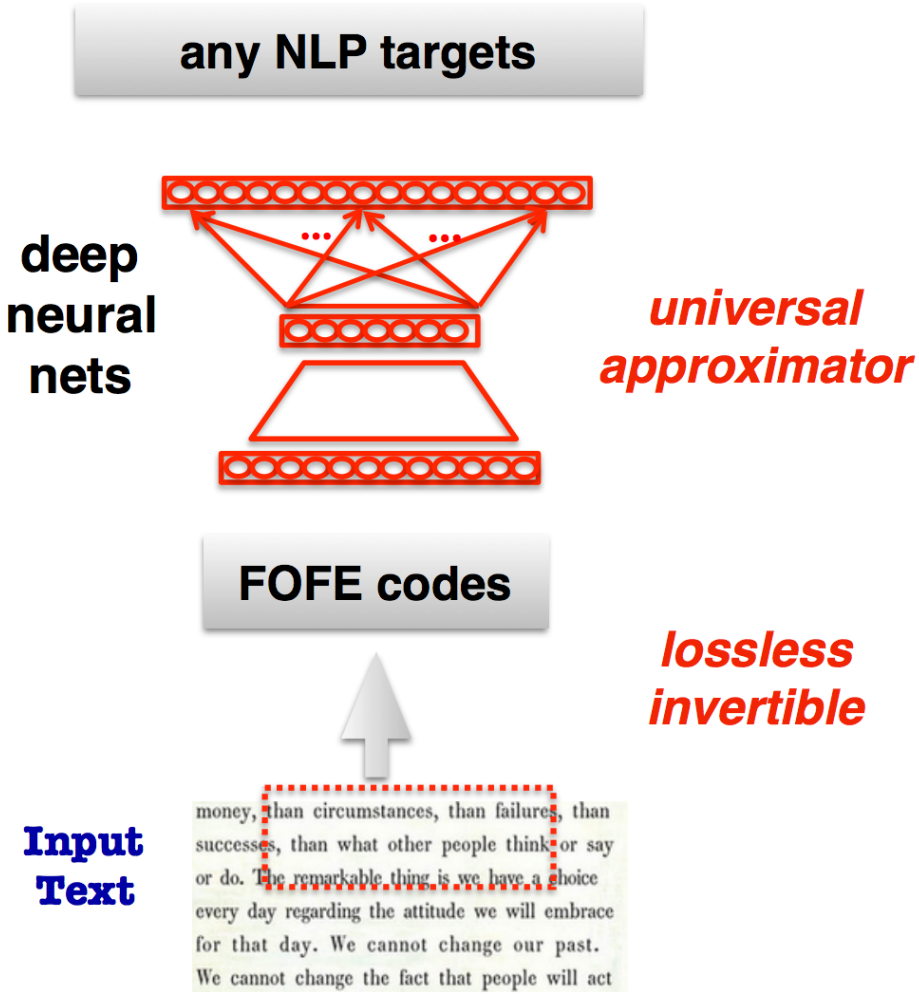
WORD	1-HOT
w_0	1000000
w_1	0100000
w_2	0010000
w_3	0001000
w_4	0000100
w_5	0000010
w_6	0000001

Table: Vocab of size 7

PARTIAL SEQUENCE	FOFE
w_6	$0, 0, 0, 0, 0, 0, 1$
w_6, w_4	$0, 0, 0, 0, 1, 0, \alpha$
w_6, w_4, w_5	$0, 0, 0, 0, \alpha, 1, \alpha^2$
w_6, w_4, w_5, w_0	$1, 0, 0, 0, \alpha^2, \alpha, \alpha^3$
w_6, w_4, w_5, w_0, w_5	$\alpha, 0, 0, 0, \alpha^3, 1+\alpha^2, \alpha^4$
$w_6, w_4, w_5, w_0, w_5, w_4$	$\alpha^2, 0, 0, 0, 1+\alpha^4, \alpha+\alpha^3, \alpha^5$

Table: Partial encoding of $w_6, w_4, w_5, w_0, w_5, w_4$

Universal Framework for NLP



FOFE-based model for Entity Discovery

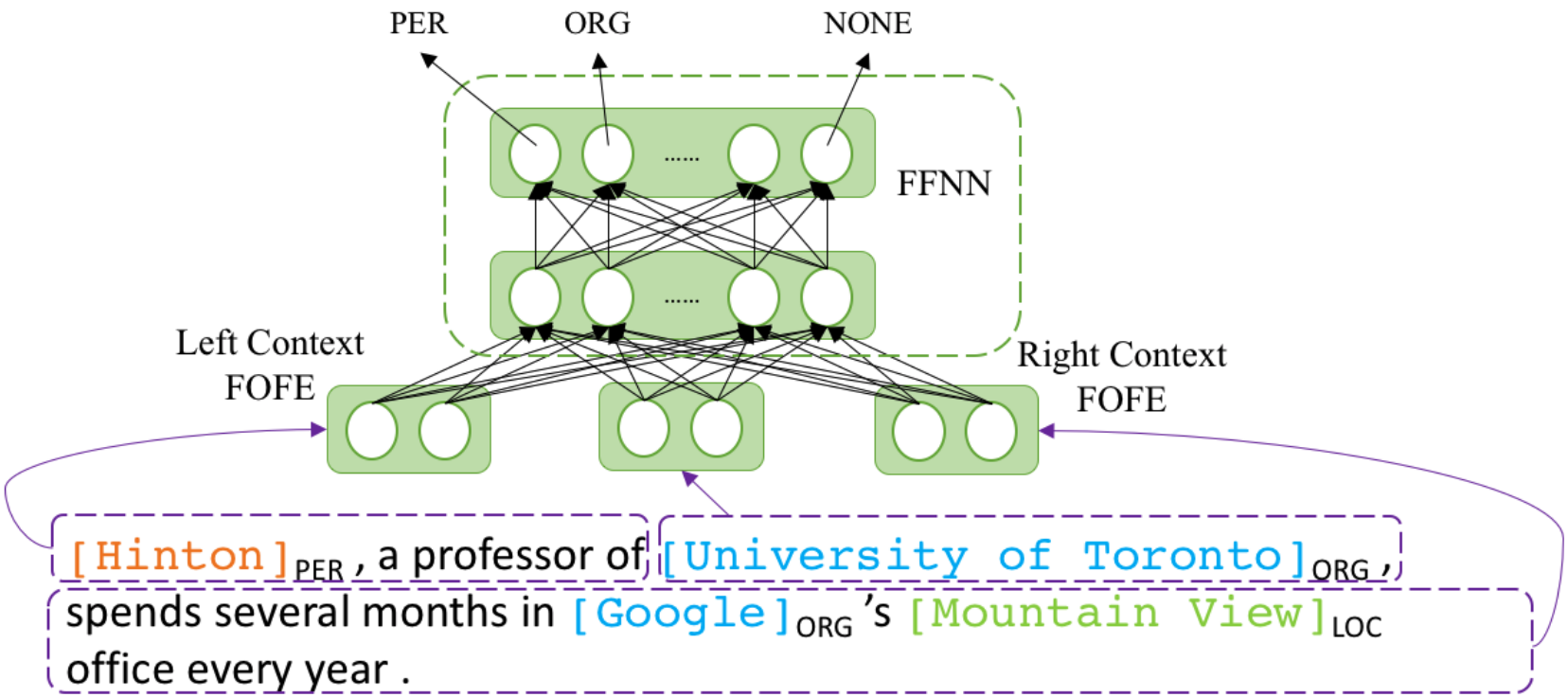


Figure: Illustration of the local detection approach for Entity Discovery using FOFE codes as input and FFNN as model.

FOFE-based Model for Entity Discovery (continued)

Features used:

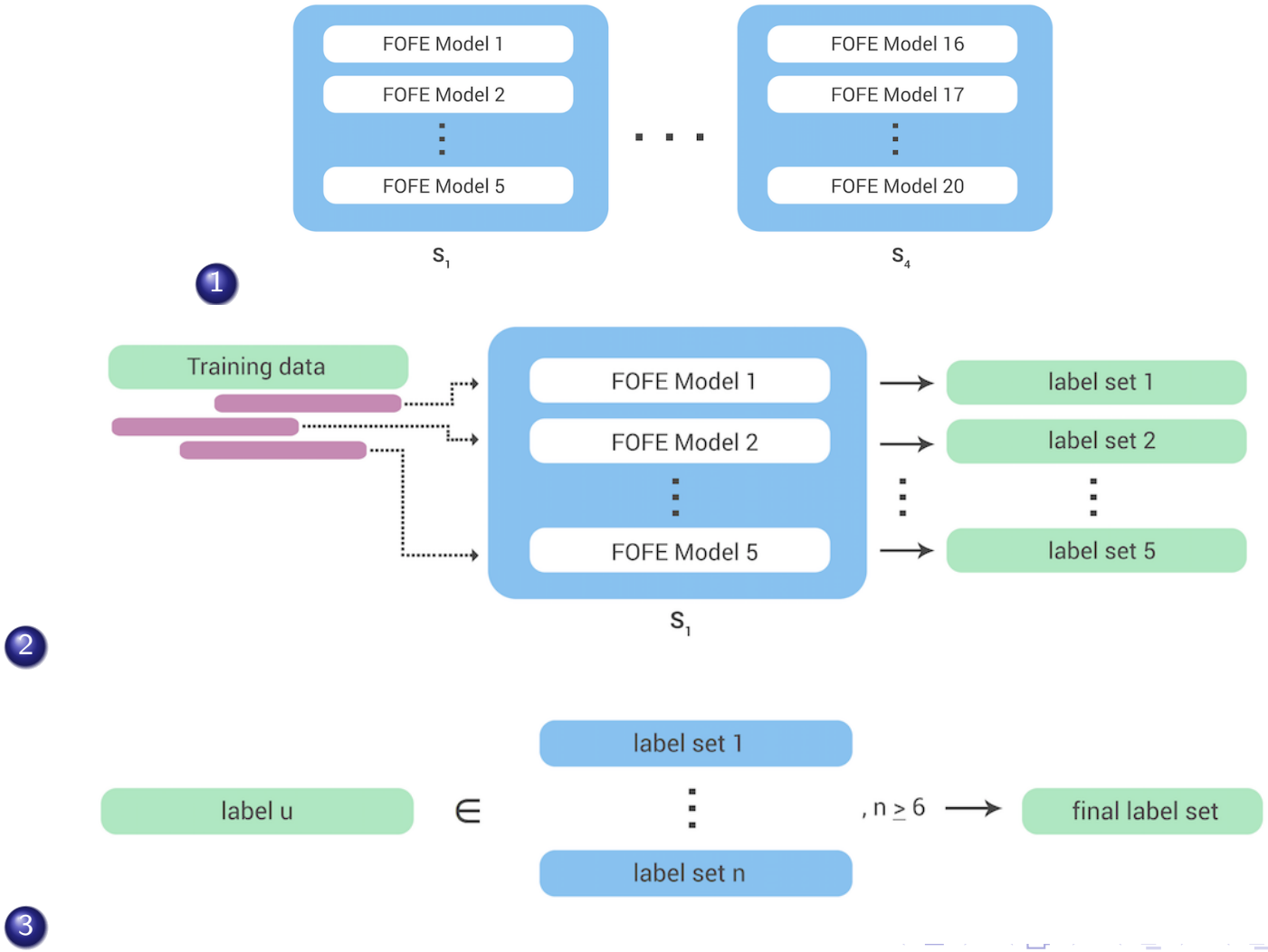
Word-level:

- BOW vector of the fragment
- FOFE codes of the left/right contexts

Character-level:

- FOFE code of the fragment
- Char CNN

Ensemble Modeling



Multi-task Model for Entity Discovery using FOFE

Multi-task Learning:

Concurrently learning a task alongside related (auxiliary) tasks by using a shared representation.

- Word and character level features are also FOFE based.
- Make use of different datasets, each treated as separate tasks.

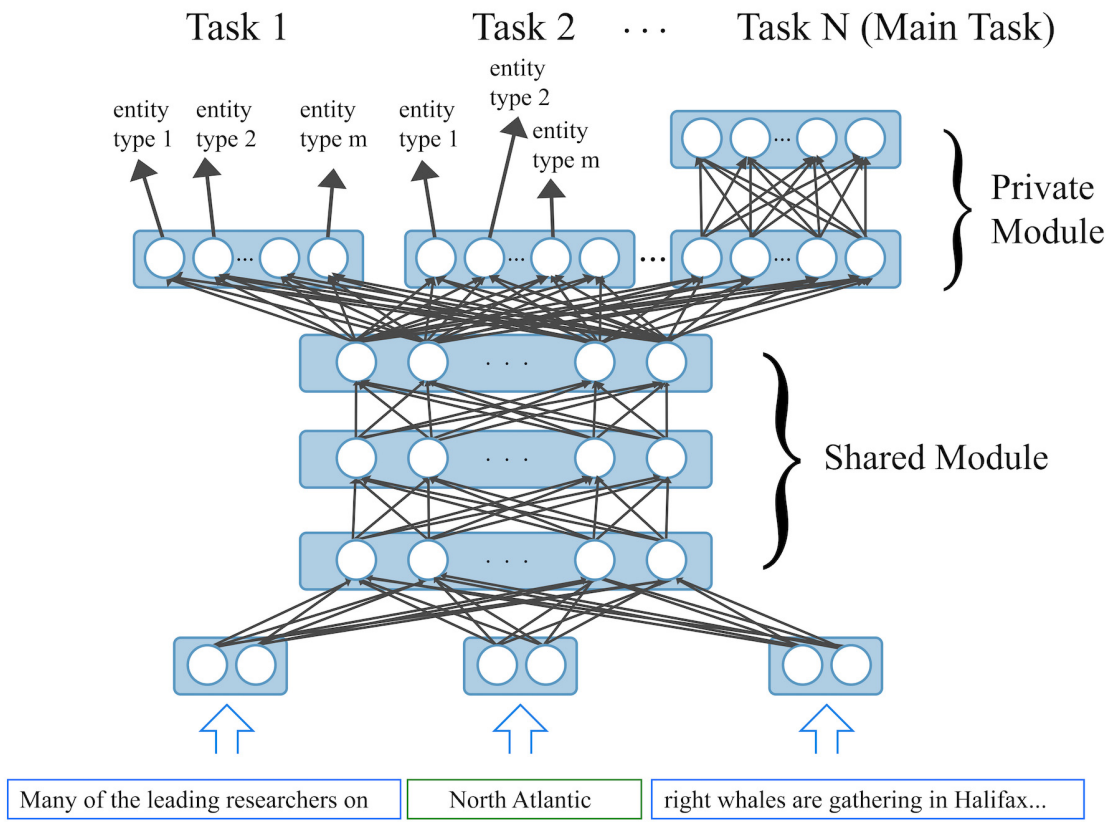
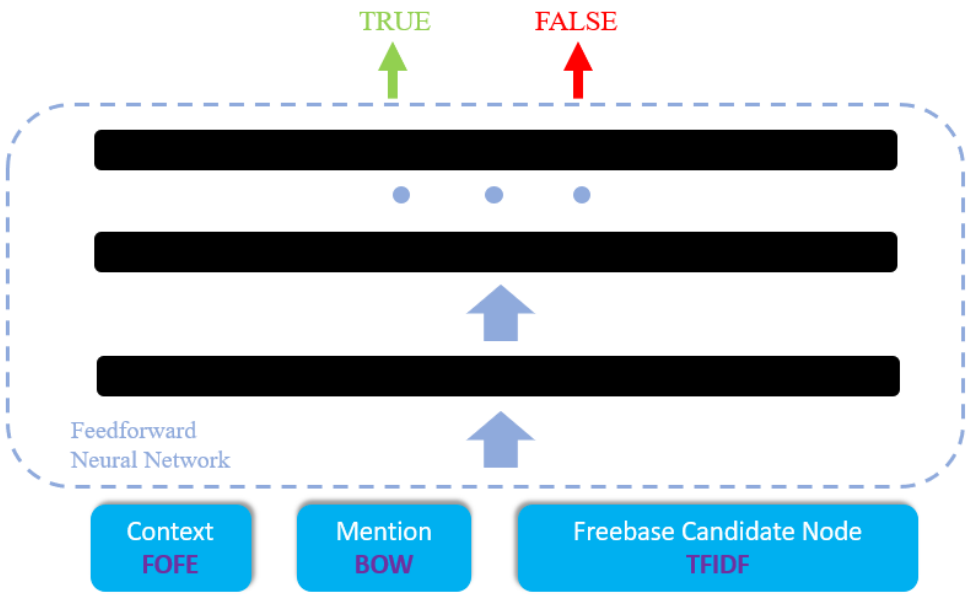


Figure: Illustration of the multi-task FFNN approach for Entity Discovery.

FOFE-based Entity Linking

1. Rule-based candidate generation
   - Generated based on knowledge bases (KB), such as Freebase, Wikipedia
   - Outputs a candidate list (Freebase nodes)
2. Neural Network based probability ranking
   - Candidate with the highest probability is chosen as the final linking result

FOFE-based Entity Linking (continued)



Features used:

- FOFE codes of left/right context.
- BOW of mention.
- Mention's candidates KB description as BOW normalized by tf-idf.

Figure: Illustration of our FOFE-based Entity Linking system using FFNNs.

Results: Datasets

- **FOFE-based Entity Discovery and Entity Linking models**
  - *Training data*: KBP 2015 (train & eval), KBP 2016 (eval), and iFLYTEK's in-house dataset.
- **Multitask FOFE-based model**
  - Main task: KBP 2017 EDL task
  - Auxiliary tasks:
    - *English:* CoNLL-2003, OntoNotes 5.0
    - *Spanish & Chinese:* DEFT Light ERE dataset

Ensemble Modeling Results

LANG	single model			model ensemble		
	P	R	F_1	P	R	F_1
ENG	0.801	0.745	0.772	0.808	0.774	**0.791**
CMN	0.775	0.660	0.713	0.793	0.726	**0.758**
SPA	0.856	0.715	0.779	0.839	0.773	**0.805**
ALL	-	-	-	0.817	0.747	**0.781**

Table: Entity Discovery (ED) performance of model ensemble in the KBP 2017 trilingual EDL evaluation.

Multi-task Learning Results

LANG	Single-task model			Multi-task model		
	P	R	F_1	P	R	F_1
ENG	0.866	0.706	0.778	0.878	0.705	**0.782**
CMN	0.795	0.635	0.707	0.789	0.665	**0.722**
SPA	0.919	0.631	0.748	0.844	0.738	**0.787**
ALL	-	-	-	0.830	0.698	**0.758**

Table: Entity Discovery (ED) performance for multi-task learning in the KBP 2017 trilingual EDL evaluation.

Entity Linking Results

LANG	baseline1		baseline2		FOFE-EL	
	NERLC	CEAFmC	NERLC	CEAFmC	NERLC	CEAFmC
ENG	**0.646**	0.630	0.572	0.615	0.648	**0.631**
CMN	**0.617**	**0.650**	0.579	0.615	0.641	**0.674**
SPA	**0.569**	0.568	0.538	0.547	0.577	**0.576**
ALL	**0.611**	**0.607**	0.565	0.586	**0.624**	**0.620**

Table: Performance on the KBP2017 EDL evaluation of our three entity linking systems. (*NERLC* denotes for *strong_typed_all_match* and *CEAFmC* for *typed_mention_ceaf*)

Conclusion

A local detection approach to Entity Discovery and MD by applying FFNN on top of FOFE

An extended multi-task approach to Entity Discovery using FOFE

No feature engineering and No external knowledge

Strong results on the KBP 2017 EDL track

Thank You!

(Q&A)