ICANN19
28th International Conference on Artificial Neural Networks

YORK UNIVERSITÉ UNIVERSITY

Dual-FOFE-net Neural Models for Entity Linking with PageRank

arXiv

Feng Wei, Uyen Trang Nguyen, Hui Jiang
Dept. of EECS, York University, Canada
{fwei,utn,hj}@cse.yorku.ca

Simple and Efficient Neural Linking Nets

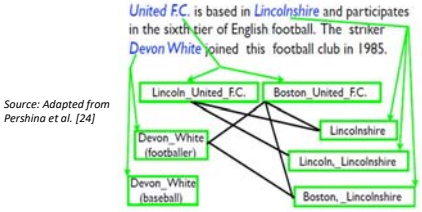
This paper presents a simple and computationally efficient approach for entity linking, which achieves state-of-the-art performance on both TAC2016 and TAC2017 task datasets.

PageRank based Distillation

PageRank based distillation not only enhances the candidate coverage, but also speeds up the whole model.

Limitations of conventional candidate generation module:

- Too many candidates
- Too much noise



Source: Adapted from Pershina et al. [24]

We rank the candidates of each mention based on their numbers of outbound links to all the recognized mentions in the same document, and keep the top τ candidates for each mention.

Simple and Efficient Dual-FOFE-net

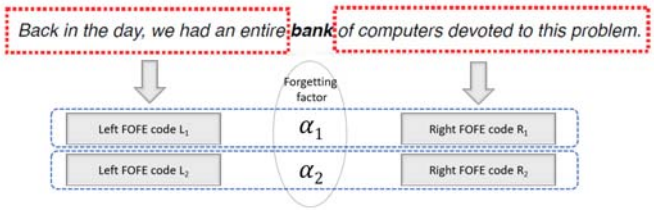
WORD	1-OF-K
w_0	1000000
w_1	0100000
w_2	0010000
w_3	0001000
w_4	0000100
w_5	0000010
w_6	0000001

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

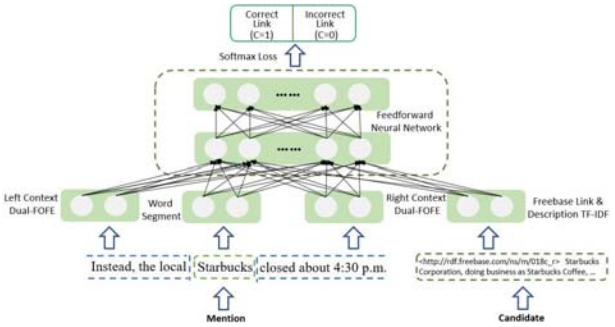
$$z_t = \alpha \cdot z_{t-1} + e_t \; (1 \le t \le n)$$

- Fixed-size ordinally-forgetting encoding (FOFE) codes, as an alternative sequence embedding representations, can almost uniquely encode any variable-length sequence of words into a fixed-size representation without losing any information.
- A smaller forgetting factor is helpful for representing the positional information of all words in the sequence; a larger forgetting factor is useful for maintaining the long-term dependency of past context.

Encode the context with **dual-FOFE**:



- Given a mention, its left and right contexts are encoded by dual-FOFE, and projected into a dense vector.
- The principle idea of dual-FOFE is to generate augmented FOFE encoding codes by concatenating two FOFE codes using two different forgetting factors.



We propose a simple and computationally efficient feed-forward neural ranking model, compared with recurrent neural networks or convolutional neural networks, by using dual-FOFE codes as input.

Results on TAC-KBP Task Datasets

	[17] TAC Rank 1		Our proposed models	
	NERLC	CEAFmC	NERLC	CEAFmC
Trilingual	64.7	66.0	**65.9**	**67.1**
English	66.6	67.6	**67.7**	**69.0**
Chinese	65.0	70.2	**66.4**	**70.7**
Spanish	61.6	63.5	**62.5**	**64.4**

On the TAC2016 dataset, our model outperforms the best system by 1.2% (NERLC F_1) and 1.1% (CEAFmC F_1), respectively, in terms of the overall trilingual entity linking performance.

	[33] TAC Rank 1		Our proposed models	
	NERLC	CEAFmC	NERLC	CEAFmC
Trilingual	67.8	70.5	**68.0**	**70.9**
English	66.8	68.8	**67.2**	**69.3**
Chinese	71.0	73.2	**71.6**	72.4
Spanish	65.0	68.9	64.8	**69.3**

On the TAC2017 dataset, our model achieves higher F_1 scores, by 0.2% (NERLC F_1) and 0.4% (CEAFmC F_1), respectively, compared to the best system in terms of the overall trilingual entity linking performance.

Deep Learning & Special Session: Machine Learning with Graphs: Algorithms and Applications

ID: 329