Exploring Neural Networks for Entity Discovery and Linking (EDL)

Dan Liu[1], Wei Lin[1], Shiliang Zhang[2], Si Wei[1], Hui Jiang[3]

[1]iFLYTEK Research, Hefei, Anhui, China

[2]University of Science and Technology of China, Hefei, Anhui,

Mingbin Xu [3], Feng Wei[3], Sed Watchara[3], Yuchen Kang[3], Hui Jiang[3]

[3]Dept. of Electrical Engineering and Computer Science

York University, Toronto, Canada

Outline

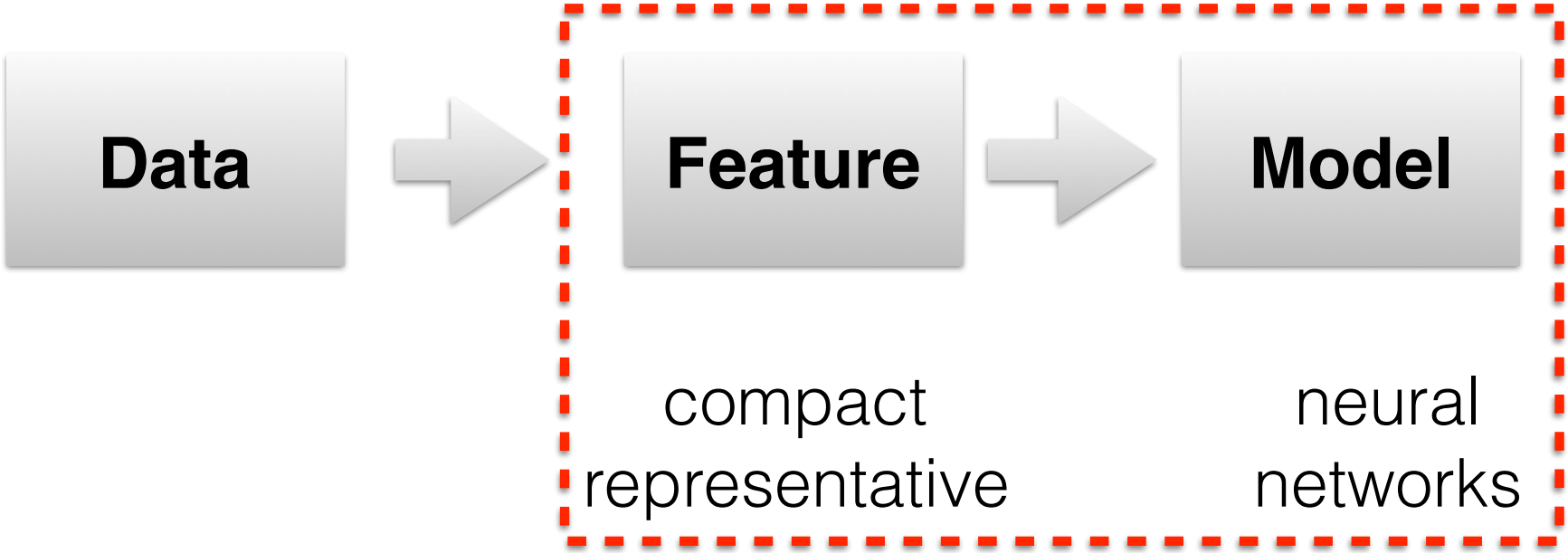
- Introduction
  - Deep Learning for NLP
- EDL Pipeline
- Two submitted systems
  - USTC_NELSLIP
  - YorkNRM
- Experiments and Discussions
- Conclusions

Deep Learning for NLP

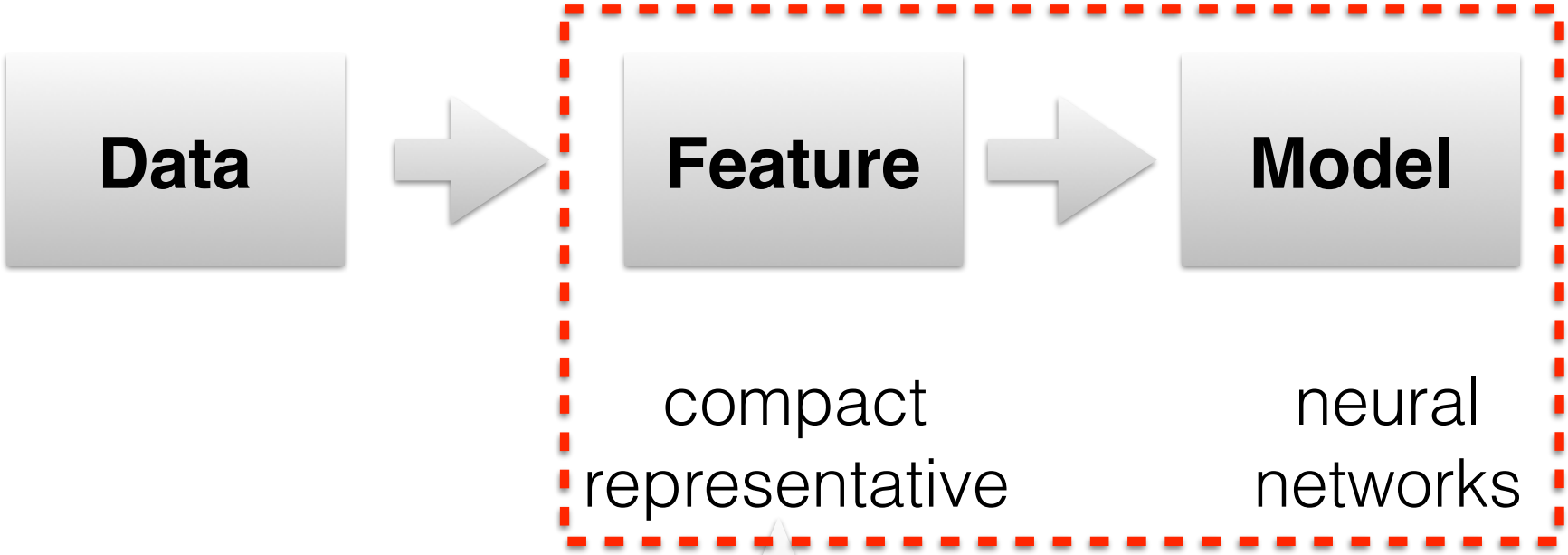
compact
representative

neural
networks

Deep Learning for NLP



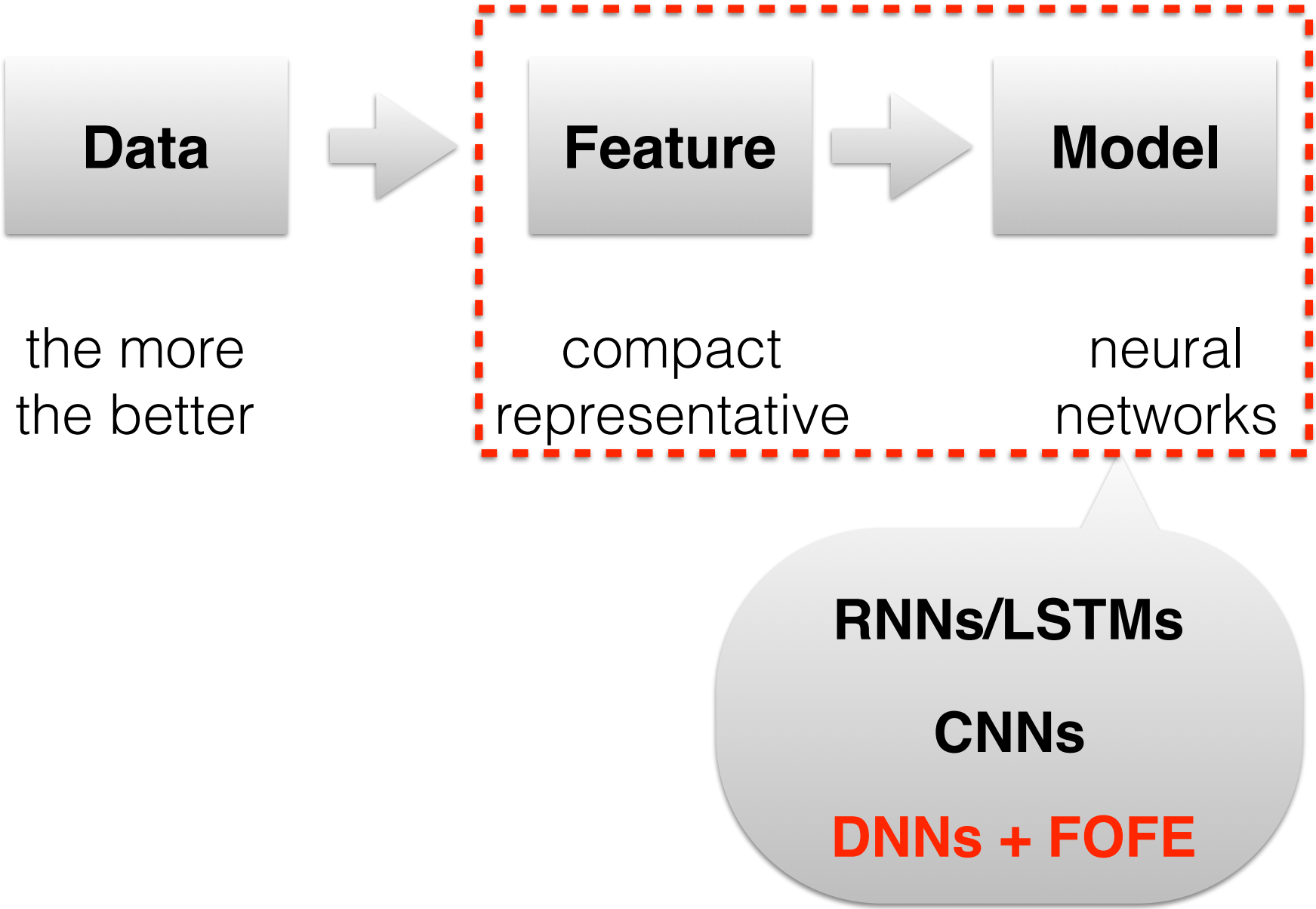
Deep Learning for NLP



Word: *word embedding*

sentence/paragraph/document:
variable-length *word sequences*

Deep Learning for NLP



Fixed-size Ordinally-Forgetting Encoding (FOFE)

- **FOFE: a fixed-size and unique encoding method for variable length sequences *[Zhang et. al., 2015]***
- **Excel in some NLP tasks: language modelling, …**

$$\mathbf{z}_t = \alpha \cdot \mathbf{z}_{t-1} + \mathbf{e}_t \quad (1 \leq t \leq T)$$

A: [1 0 0]
B: [0 1 0]
C: [0 0 1]

ABC: [a^2, a, 1]
ABCBC:
[a^4, a^3+a, 1+a^2]

Fixed-size Ordinally-Forgetting Encoding (FOFE)

- **FOFE: a fixed-size and unique encoding method for variable length sequences *[Zhang et. al., 2015]***
- **Excel in some NLP tasks: language modelling, …**

$$\mathbf{z}_t = \alpha \cdot \mathbf{z}_{t-1} + \mathbf{e}_t \quad (1 \leq t \leq T)$$

A: [1 0 0]
B: [0 1 0]
C: [0 0 1]

ABC: [a^2, a, 1]
ABCBC:
[a^4, a^3+a, 1+a^2]

Theorem 1 *If the forgetting factor α satisfies $0 < \alpha \leq 0.5$,* FOFE *is unique for any K and T.*

Theorem 2 *For $0.5 < \alpha < 1$, given any finite values of K and T,* FOFE *is almost unique everywhere for $\alpha \in (0.5, 1.0)$, except only a finite set of countable choices of α.*

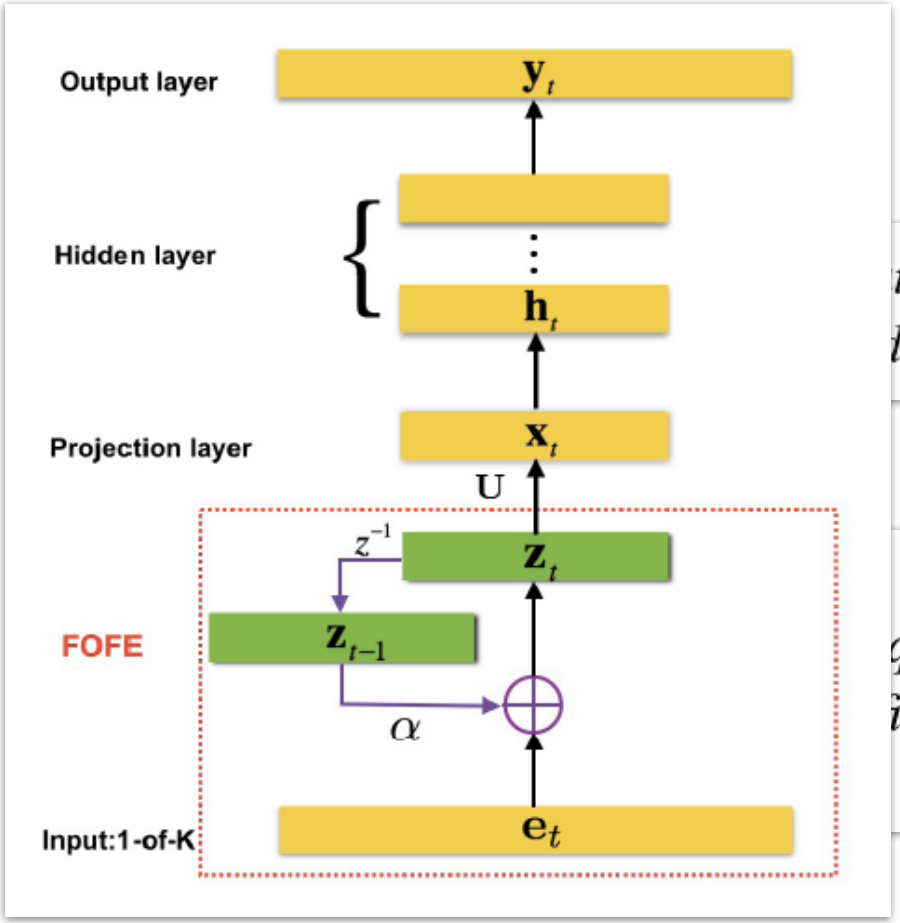
Fixed-size Ordinally-Forgetting Encoding (FOFE)

- **FOFE: a fixed-size and unique encoding method for variable length sequences *[Zhang et. al., 2015]***
- **Excel in some NLP tasks: language modelling, …**

$$\mathbf{z}_t = \alpha \cdot \mathbf{z}_{t-1} + \mathbf{e}_t \quad (1 \leq t \leq T)$$

A: [1 0 0]
B: [0 1 0]
C: [0 0 1]

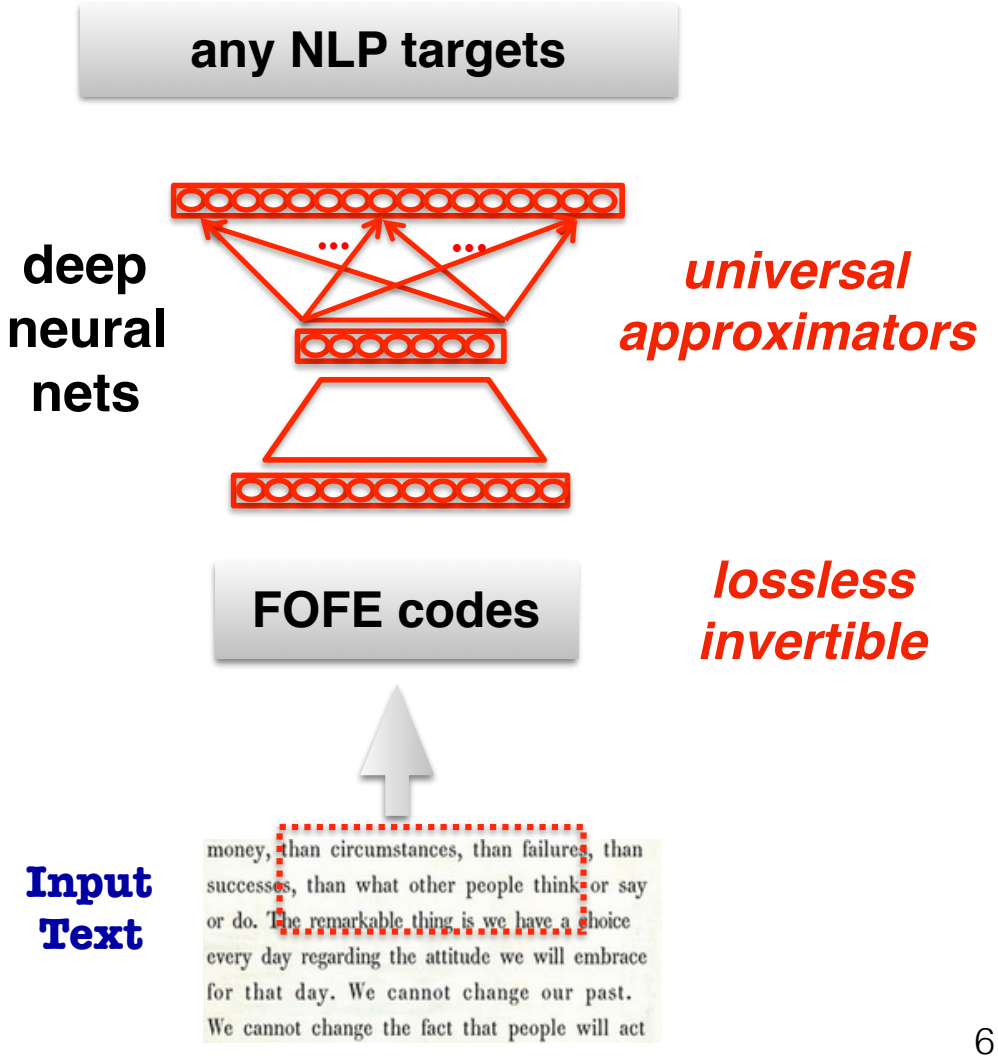
ABC: [a^2, a, 1]
ABCBC:
[a^4, a^3+a, 1+a^2]



tisfies 0 <
l T.

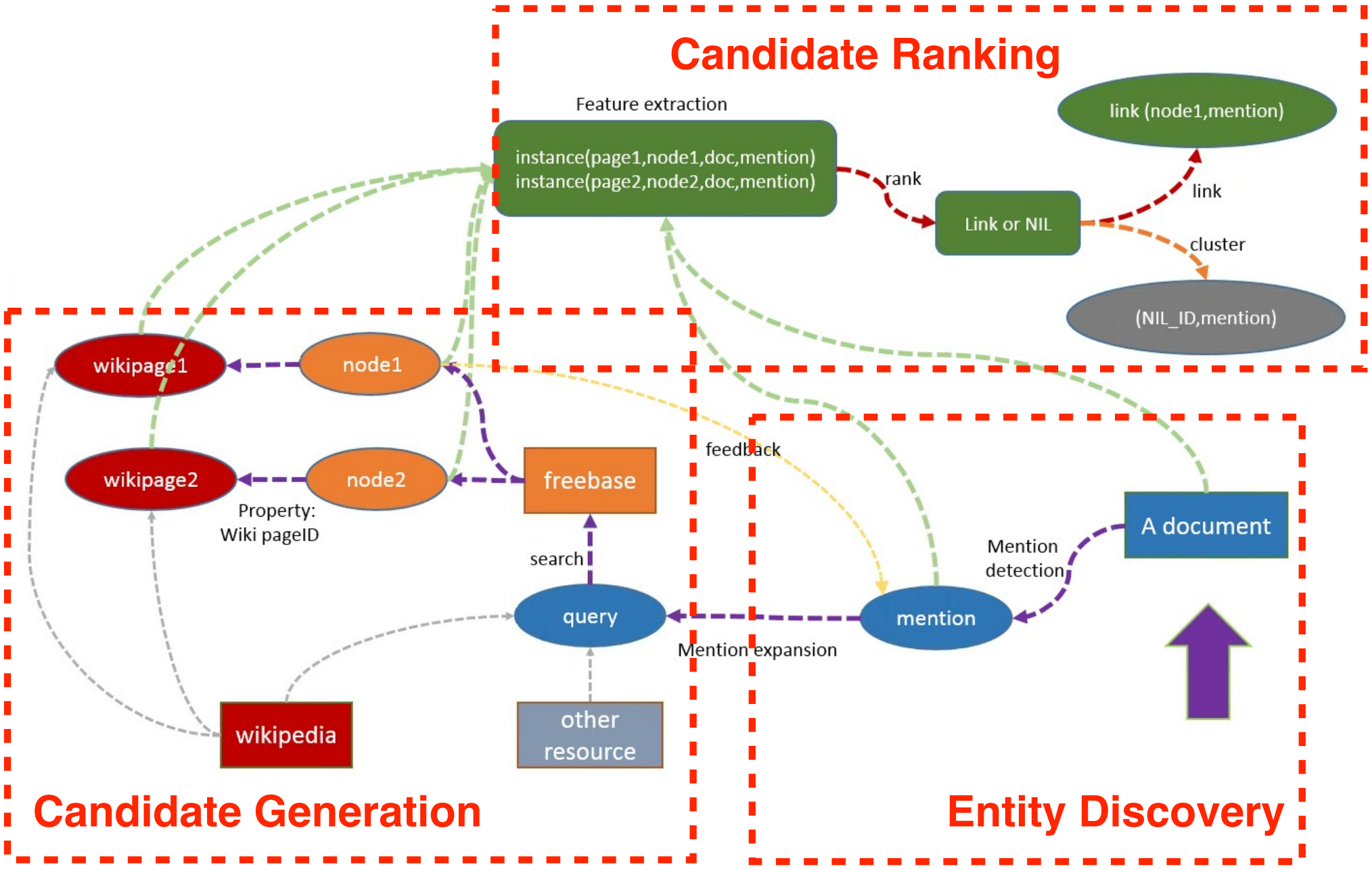
any finite
que every-
inite set of

FOFE+DNN for all NLP tasks

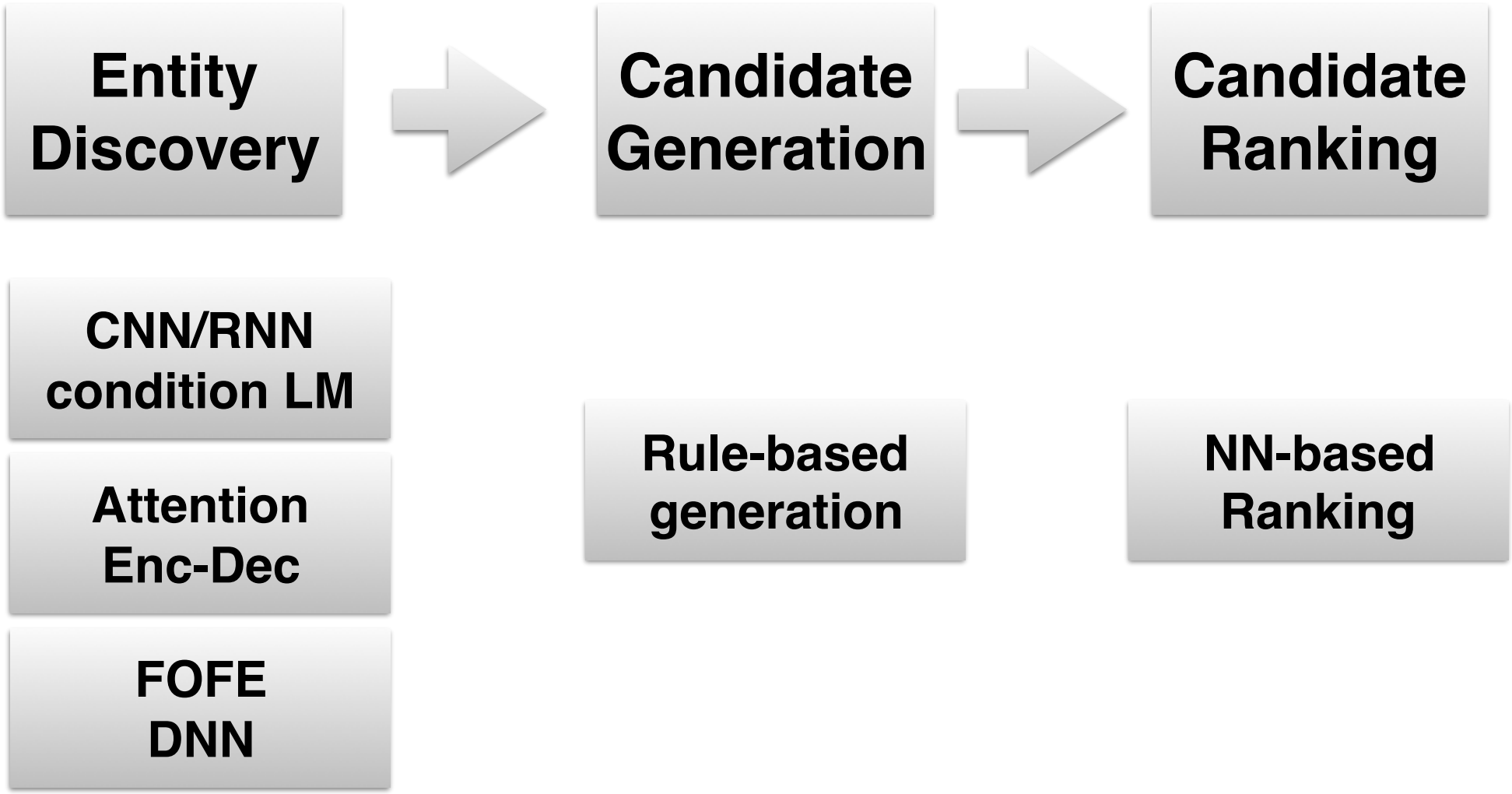


- **Theoretically sound**
- **No feature engineering**
- **Simple models**
- **General methodology**
  - ***not only*** **sequence labeling problems**
  - ***but also*** **(almost) all NLP tasks**

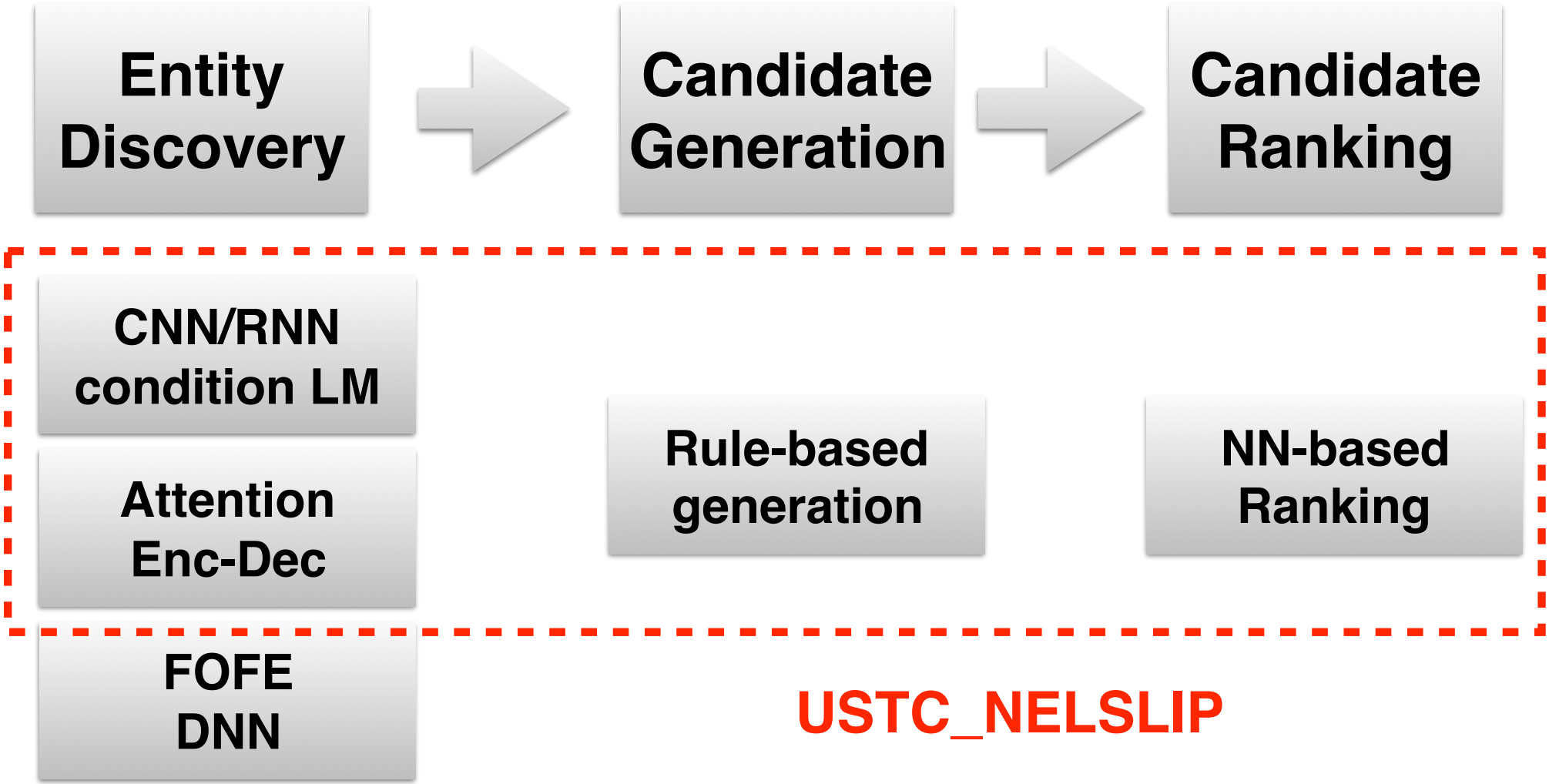
EDL Pipeline



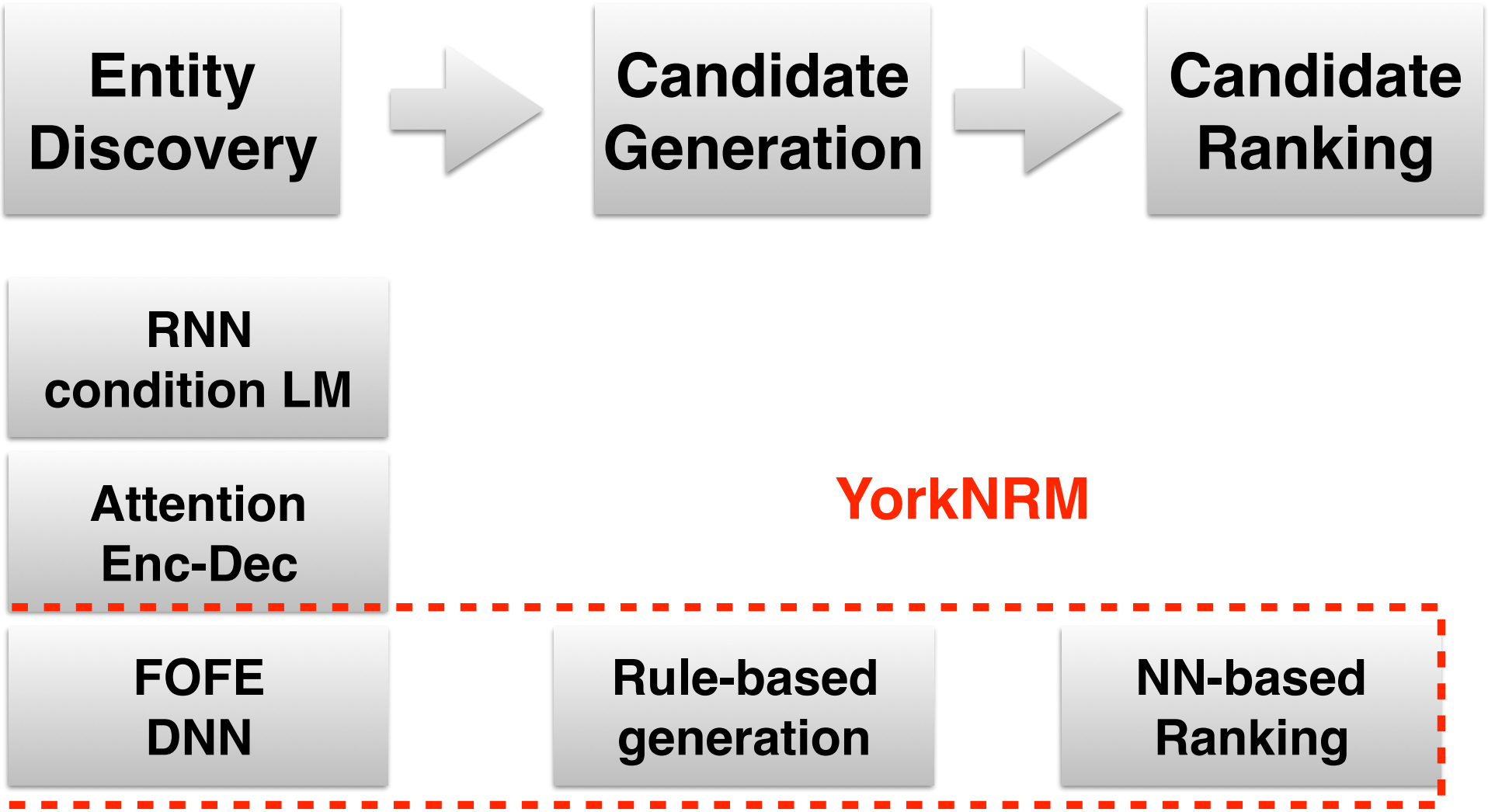
EDL System 1: USTC



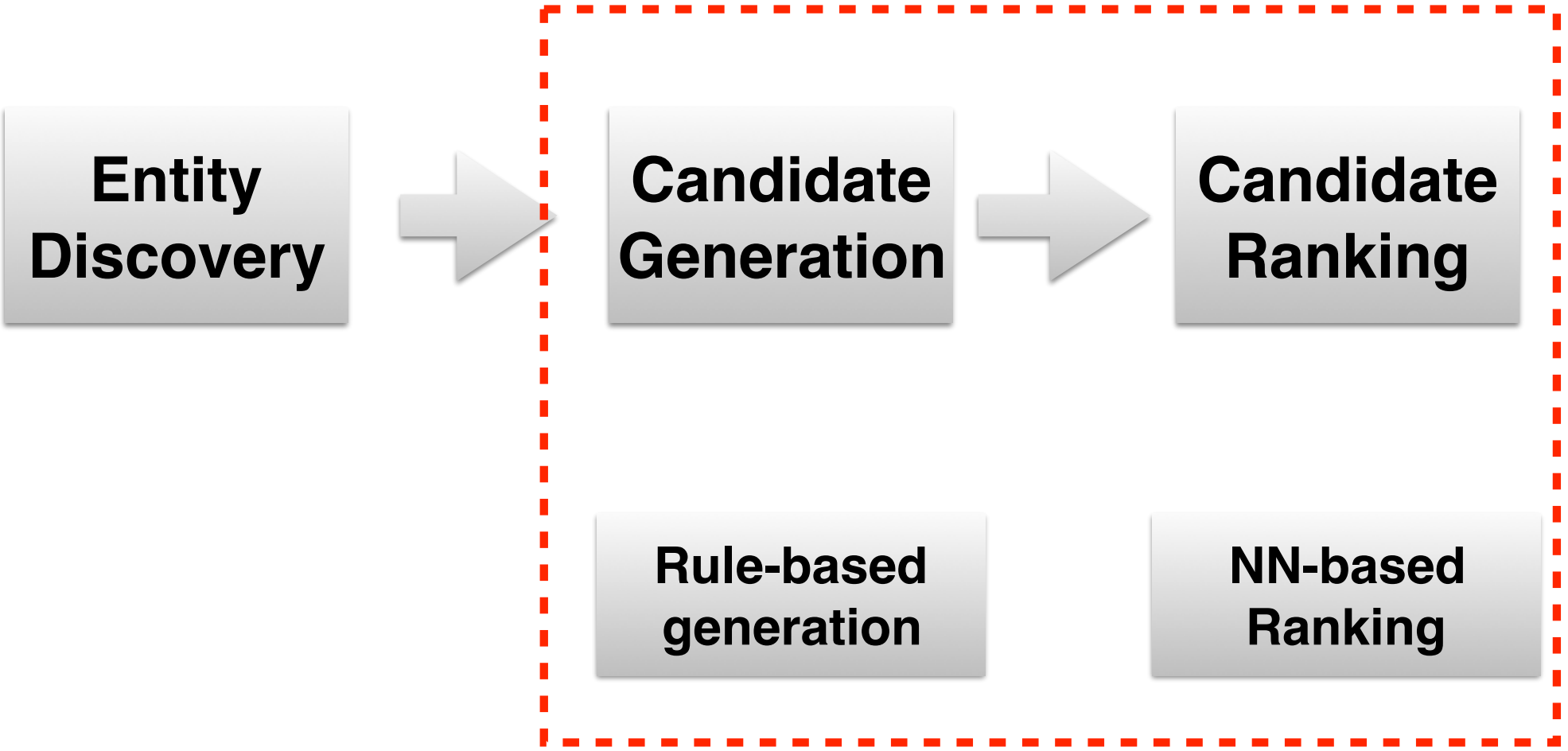
EDL System 1: USTC



EDL Sytem 2: York

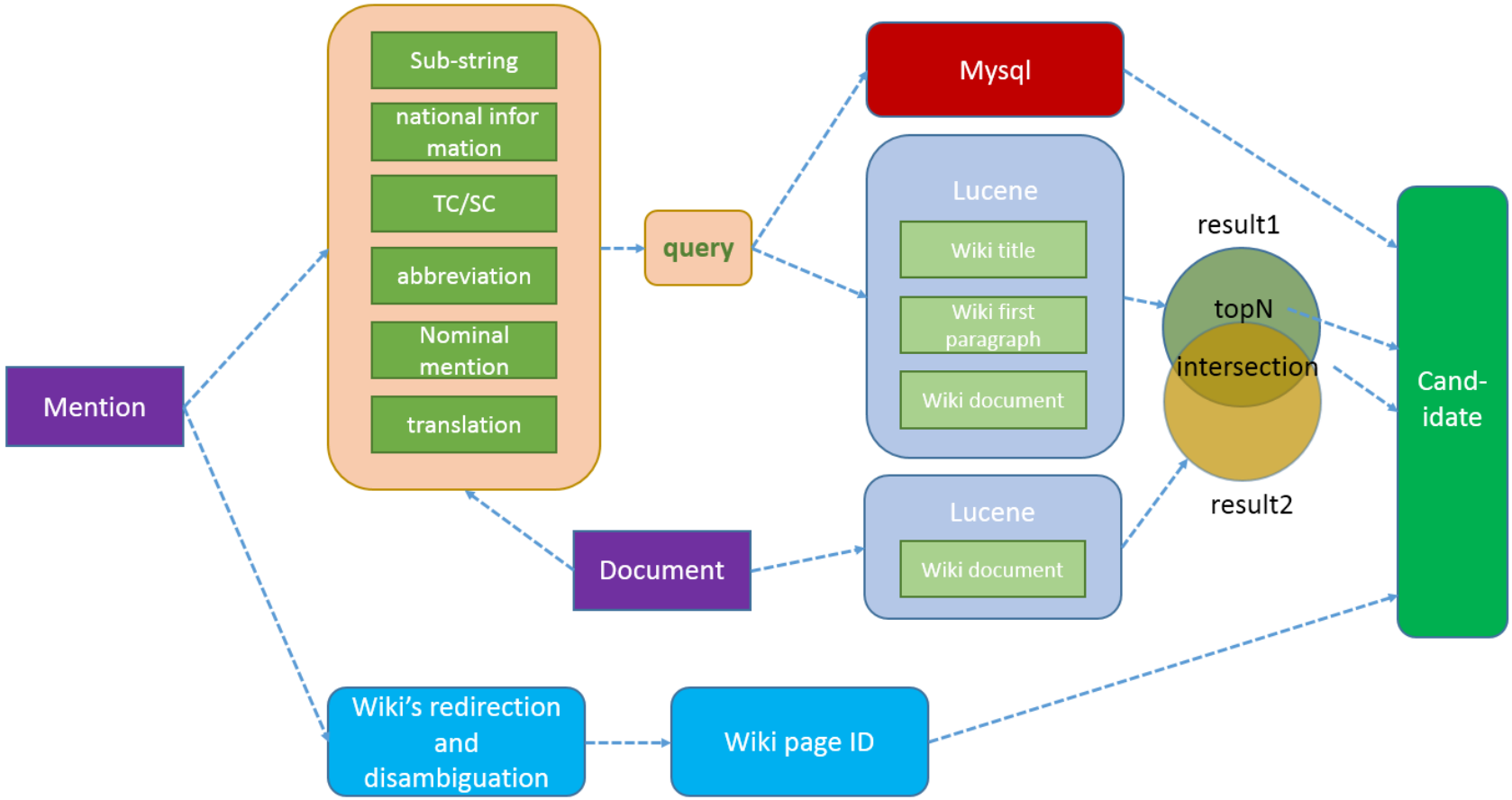


Entity Linking



Entity Linking: Candidate Generation

- **Rule-based Query Expansion**
- **Query search (mySQL) and fuzzy match (Lucene)**



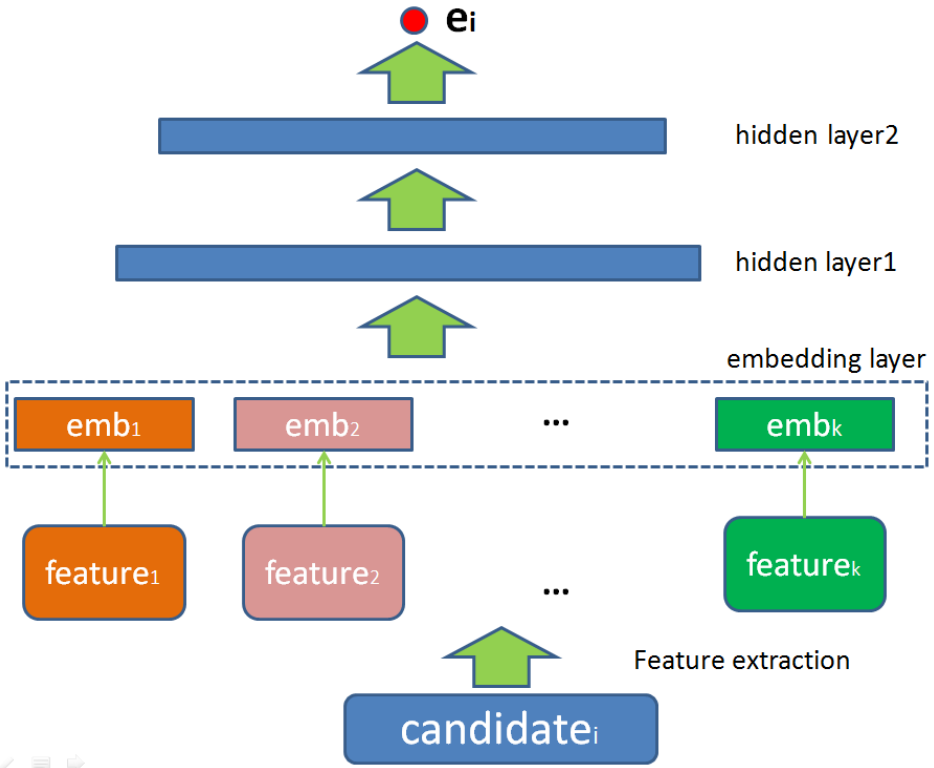
Candidate Generation: Performance

- **Quality of generated candidate lists**
- **Average count vs. coverage rate**

KBP2015 test set	ENG	CMN	SPA
avg. count	22.60	92.96	38.55
coverage rate	93%	92.1%	88.4%

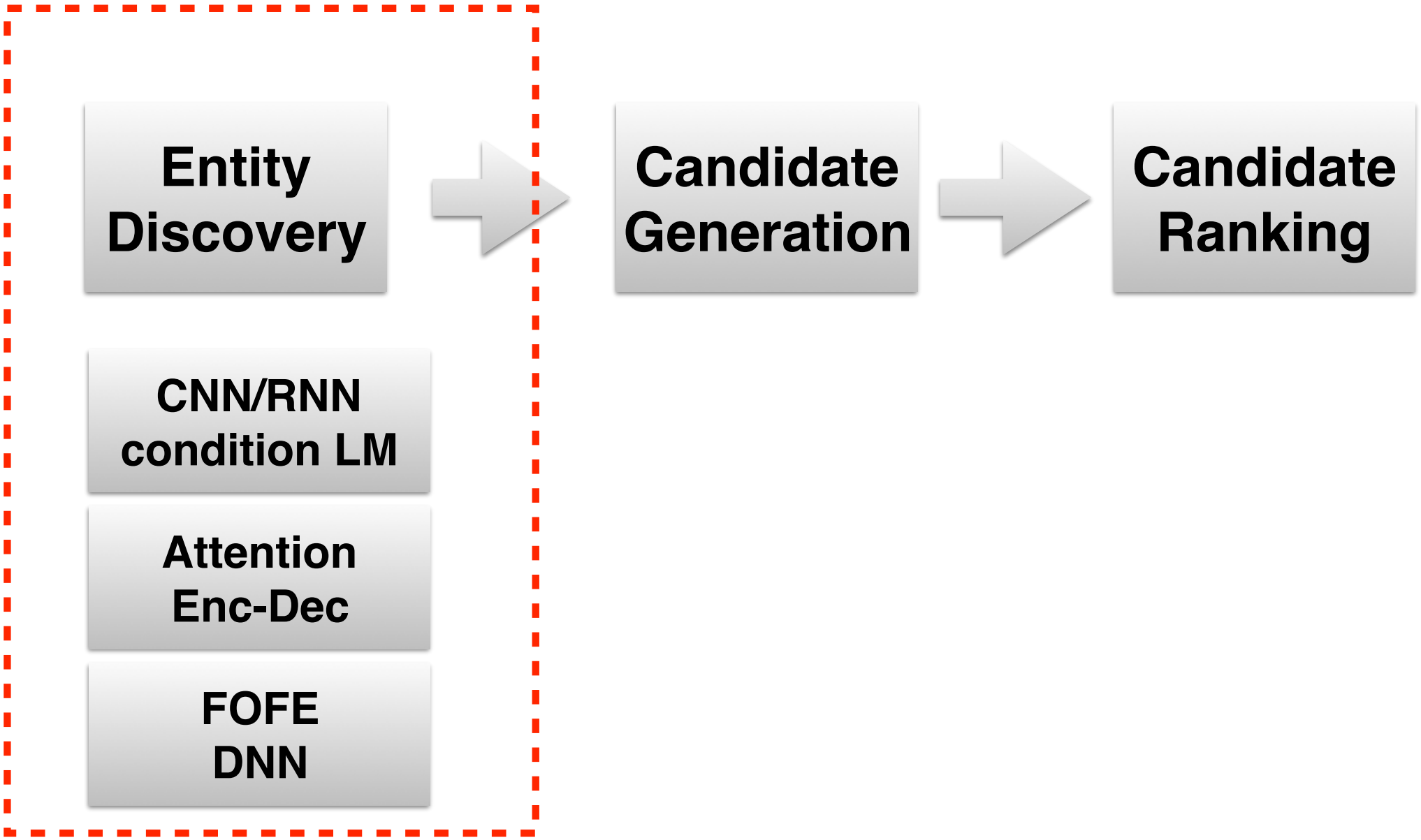
Entity Linking: NN-based Ranking

- Use some hand-crafted features as input
- Use feedforward DNNs to compute ranking scores
- NIL clustering based on string-match

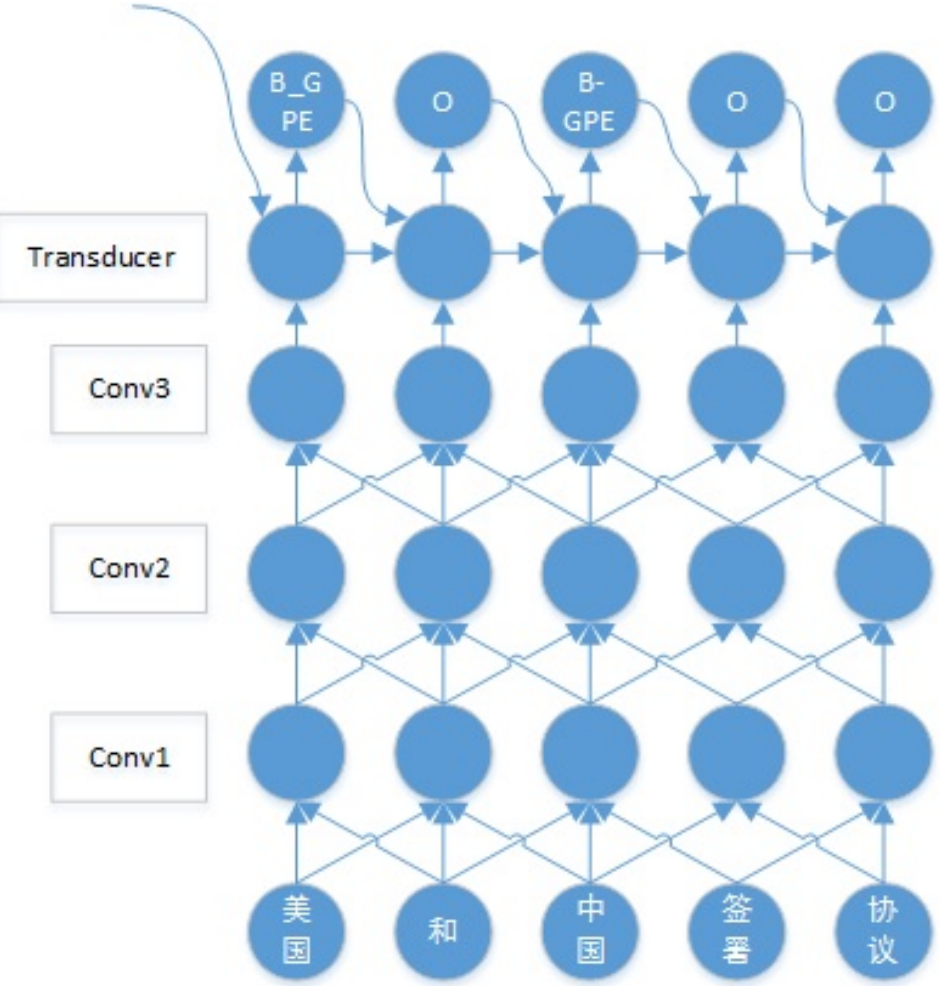


	dim	feature
$\mathbf{e}_1$	100	mention string embedding
$\mathbf{e}_2$	100	candidate name embedding
$\mathbf{e}_3$	10	mention type
$\mathbf{e}_4$	10	document type
$\mathbf{e}_5$	10	candidate hot value vector
$\mathbf{e}_6$	10	edit distance between mention string and candidate name
$\mathbf{e}_7$	10	cosine similarity of document and candidate description
$\mathbf{e}_8$	10	edit distance between translations of mention and candidate

Entity Discovery (ED)



USTC ED Model1

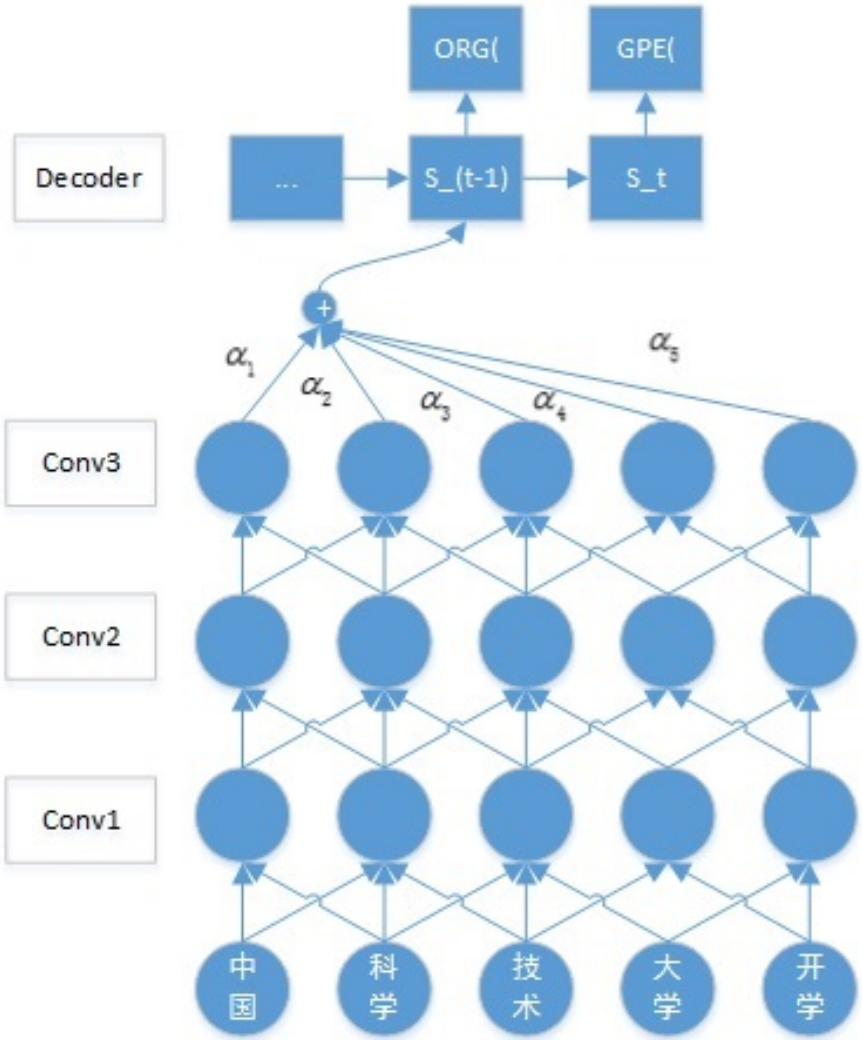


- **Mention Detection as Sequence Labelling**
- **Word sequence ==> BIO tags**

$$\Pr(Y|X) = \prod_{i=1}^{N} P(y_i \mid X, y_{i-1}, y_{i-2}, ...y_1)$$

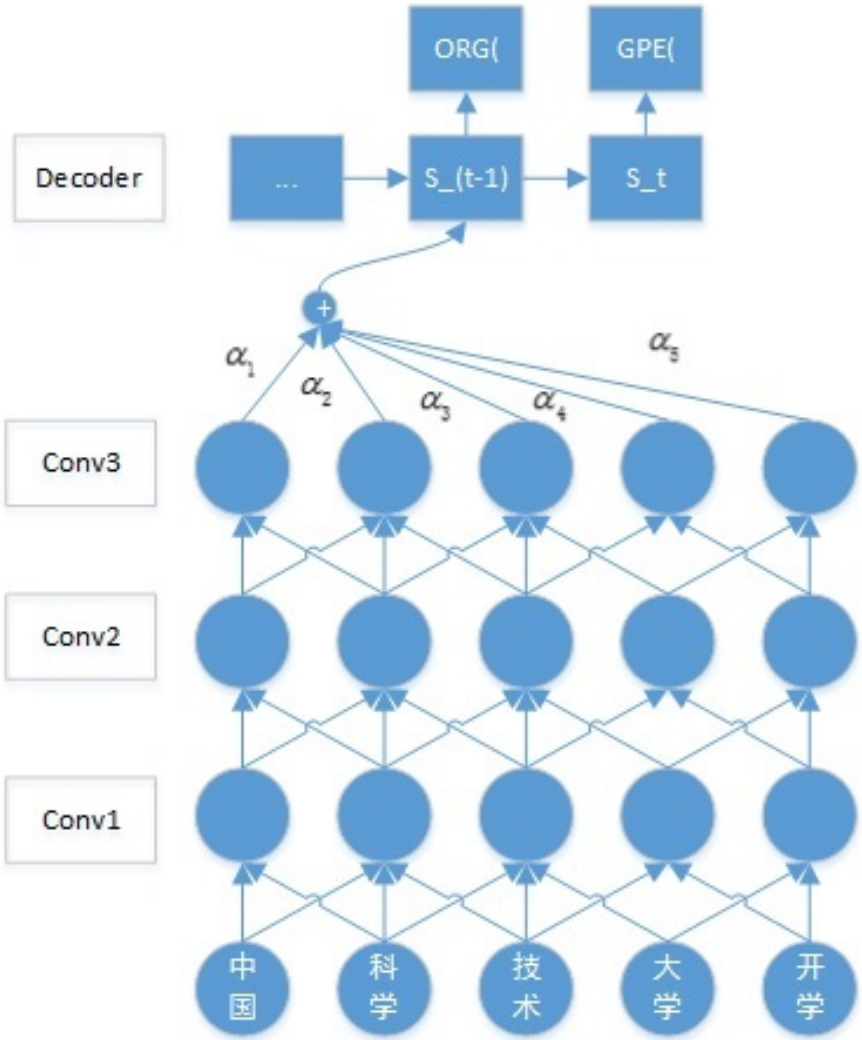
- **CNN: 5 layers of convolutional layers**
- **RNN: GRU-based model**
- **Viterbi decoding**

USTC ED Model2



- Introduce attention
- Tree-structured tags for nested entities

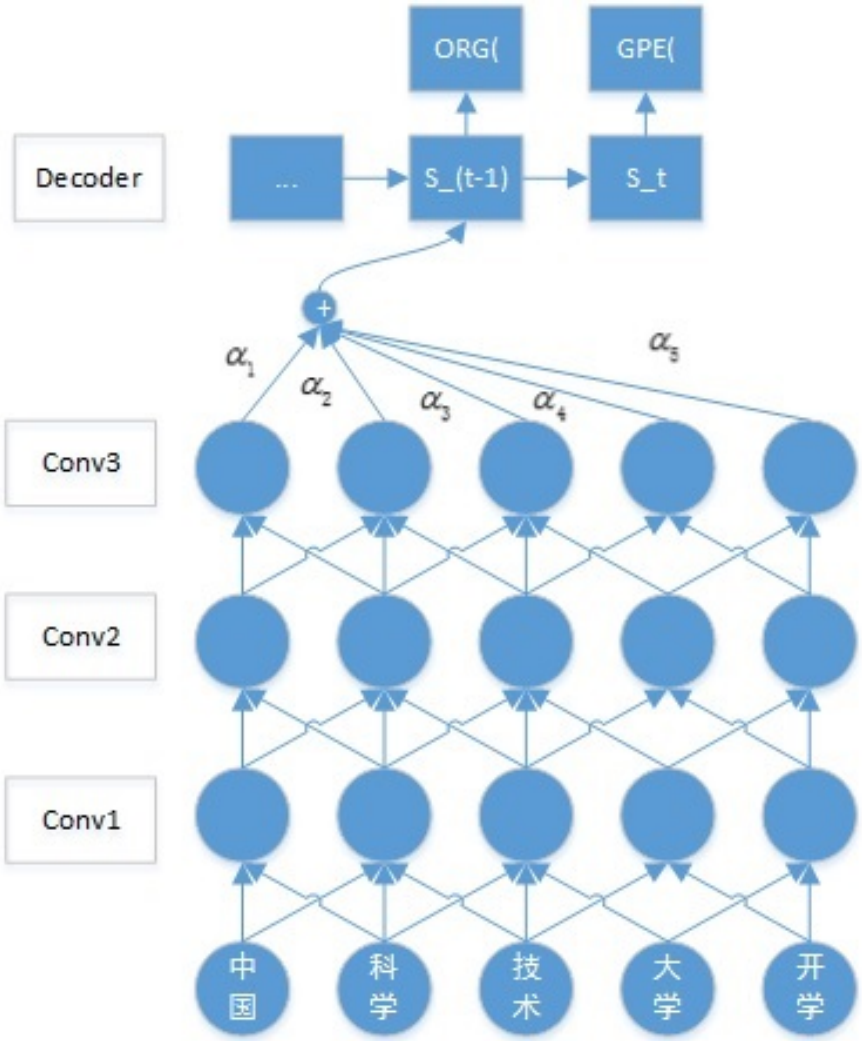
USTC ED Model2



- Introduce attention
- Tree-structured tags for nested entities

Kentucky Fried Chicken

USTC ED Model2

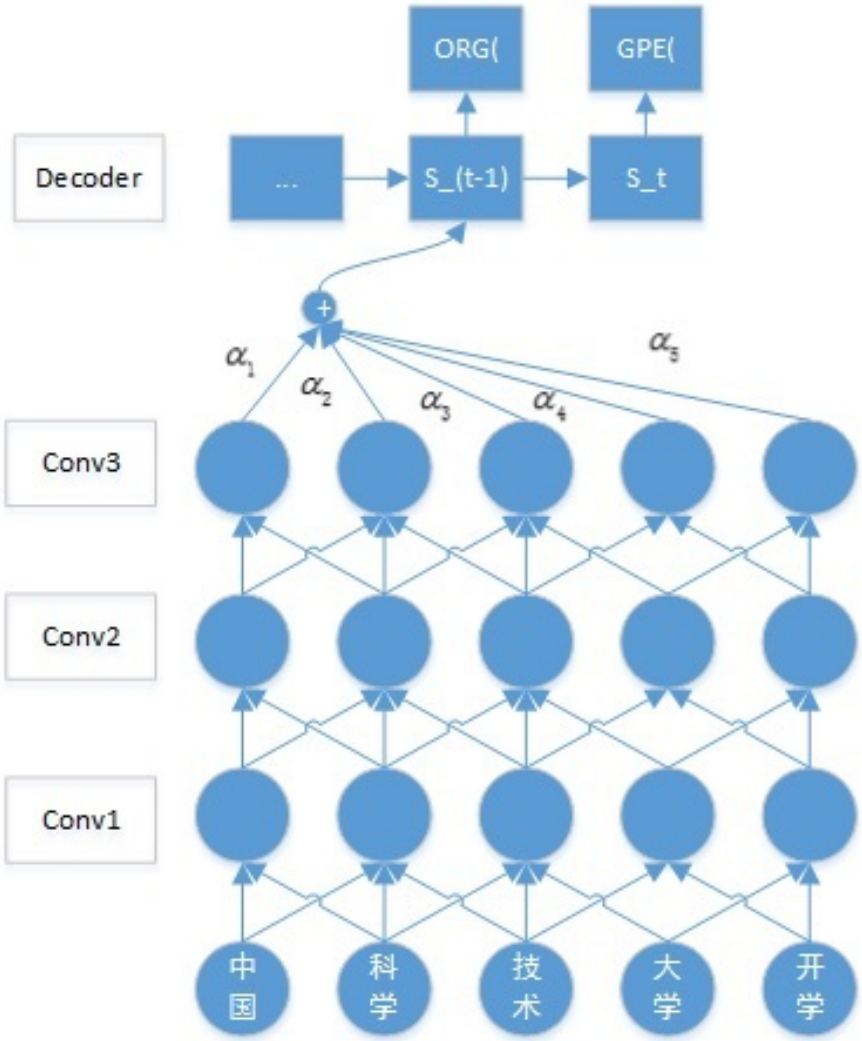


- **Introduce attention**
- **Tree-structured tags for nested entities**

Kentucky Fried Chicken

$$[_{FAC}\ [_{PER}\ \text{Kentucky}\ ]_{PER}\ \text{Fried}\ \ \text{Chicken}\ ]_{FAC}$$

USTC ED Model2



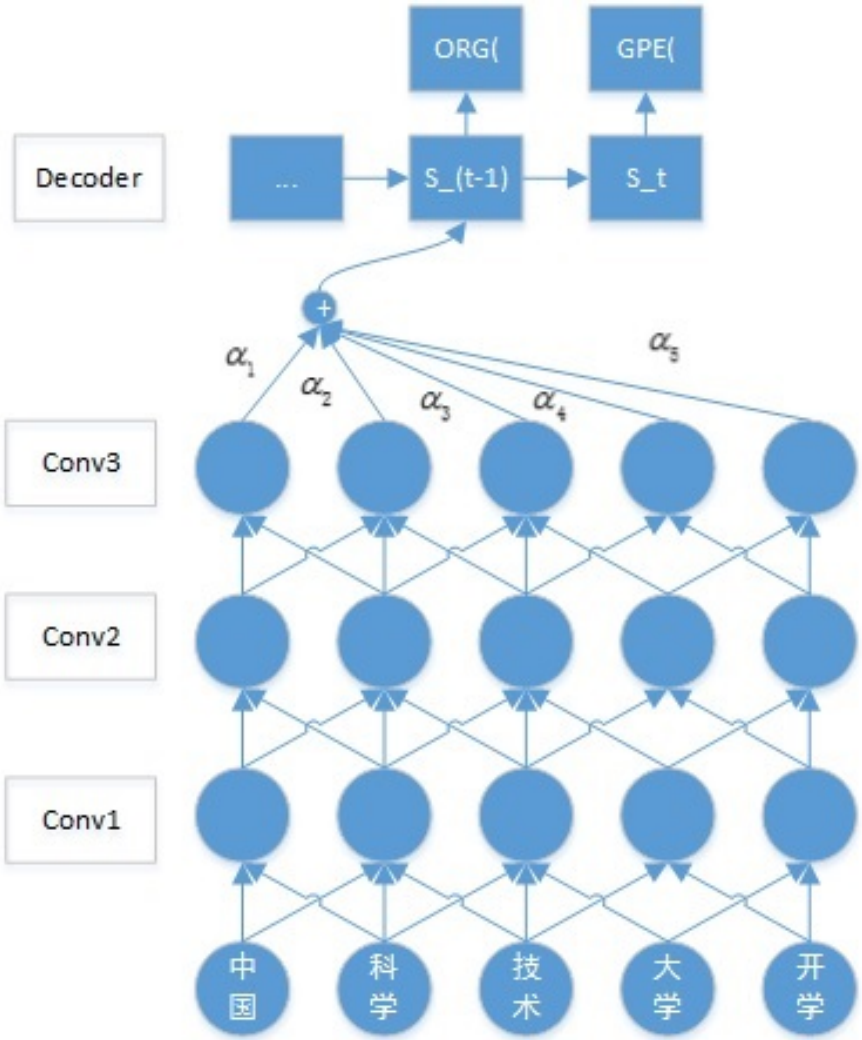
- Introduce attention
- Tree-structured tags for nested entities

Kentucky Fried Chicken

$$[_{FAC}\ [_{PER}\ \text{Kentucky}\ ]_{PER}\ \text{Fried}\ \ \text{Chicken}\ ]_{FAC}$$

$$[_{FAC}\ \ [_{PER}\ \ \mathbf{Z}\ \ ]_{PER}\ \ \mathbf{Z}\ \ \mathbf{Z}\ \ ]_{FAC}$$

USTC ED Model2



- **Introduce attention**
- **Tree-structured tags for nested entities**

Kentucky Fried Chicken

$$[_{FAC}\ [_{PER}\ \mathbf{Z}\ ]_{PER}\ \mathbf{Z}\ \mathbf{Z}\ ]_{FAC}$$

USTC ED Performance

- **Effect of various training data sets:**
  - **KBP15 training data**
  - **iFLYTEK in-house data (10,000 labelled Chinese and English doc)**

		P	R	F_1
KBP15	CMN	0.804	0.756	0.779
+ iFLYTEK		0.828	0.777	0.802
KBP15	ENG	0.807	0.698	0.749
+ iFLYTEK		0.802	0.815	0.751
KBP15	SPA	0.800	0.749	0.773
KBP15	ALL	0.805	0.727	0.764
+ iFLYTEK		0.817	0.759	0.787

Entity Discovery Performance on KBP2015 Test set

USTC ED Performance

- **Effect of various training data sets:**
  - **KBP15 training data**
  - **iFLYTEK in-house data (10,000 labelled Chinese and English doc)**

		P	R	F_1
KBP15	CMN	0.804	0.756	0.779
+ iFLYTEK		0.828	0.777	0.802
KBP15	ENG	0.807	0.698	0.749
+ iFLYTEK		0.802	0.815	0.751
KBP15	SPA	0.800	0.749	0.773
KBP15	ALL	0.805	0.727	0.764
+ iFLYTEK		0.817	0.759	0.787

1-2%

Entity Discovery Performance on KBP2015 Test set

USTC ED Performance

- **5-fold system combination (5SC)**
- **System fusion**

	P	**R**	F_1
model1	0.821	0.667	0.736
model1+5SC	0.836	0.694	0.758
model2	0.811	0.675	0.737
model2+5SC	0.821	0.699	0.755
fusion	0.805	0.727	0.764

Entity Discovery Performance on KBP2015 Test set

USTC ED Performance

- **5-fold system combination (5SC)**
- **System fusion**

	P	R	F_1
model1	0.821	0.667	0.736
model1+5SC	0.836	0.694	0.758
model2	0.811	0.675	0.737
model2+5SC	0.821	0.699	0.755
fusion	0.805	0.727	0.764

1.8-2.2%

Entity Discovery Performance on KBP2015 Test set

USTC ED Performance

- **5-fold system combination (5SC)**
- **System fusion**

	P	R	F_1
model1	0.821	0.667	0.736
model1+5SC	0.836	0.694	0.758
model2	0.811	0.675	0.737
model2+5SC	0.821	0.699	0.755
fusion	0.805	0.727	0.764

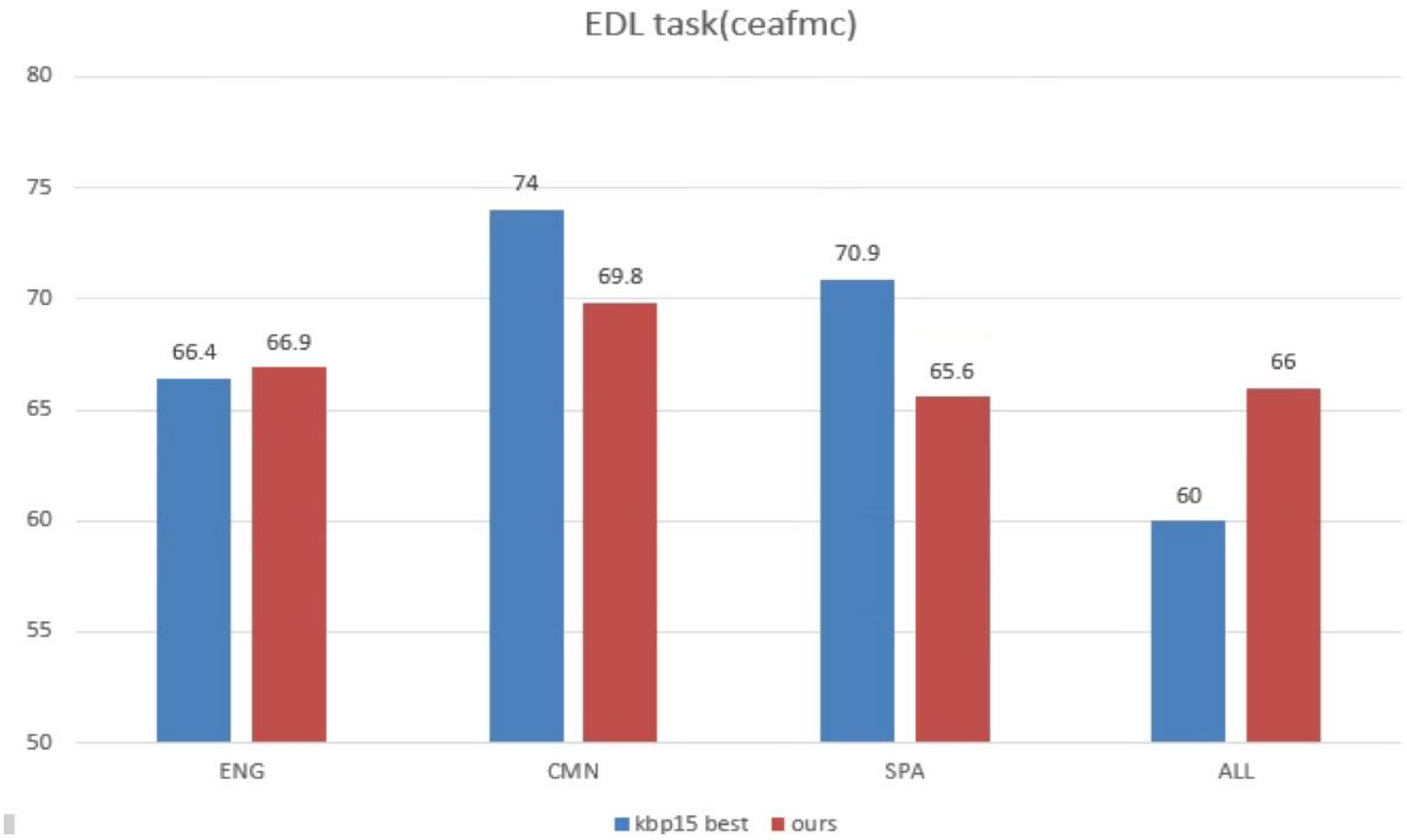
1.8-2.2%

0.6%

Entity Discovery Performance on KBP2015 Test set

USTC EDL Performance

- Trained with KBP2015 data
- 5SC + Fusion



Entity Linking Performance on KBP2015 Test set

USTC Official KBP2016 Results

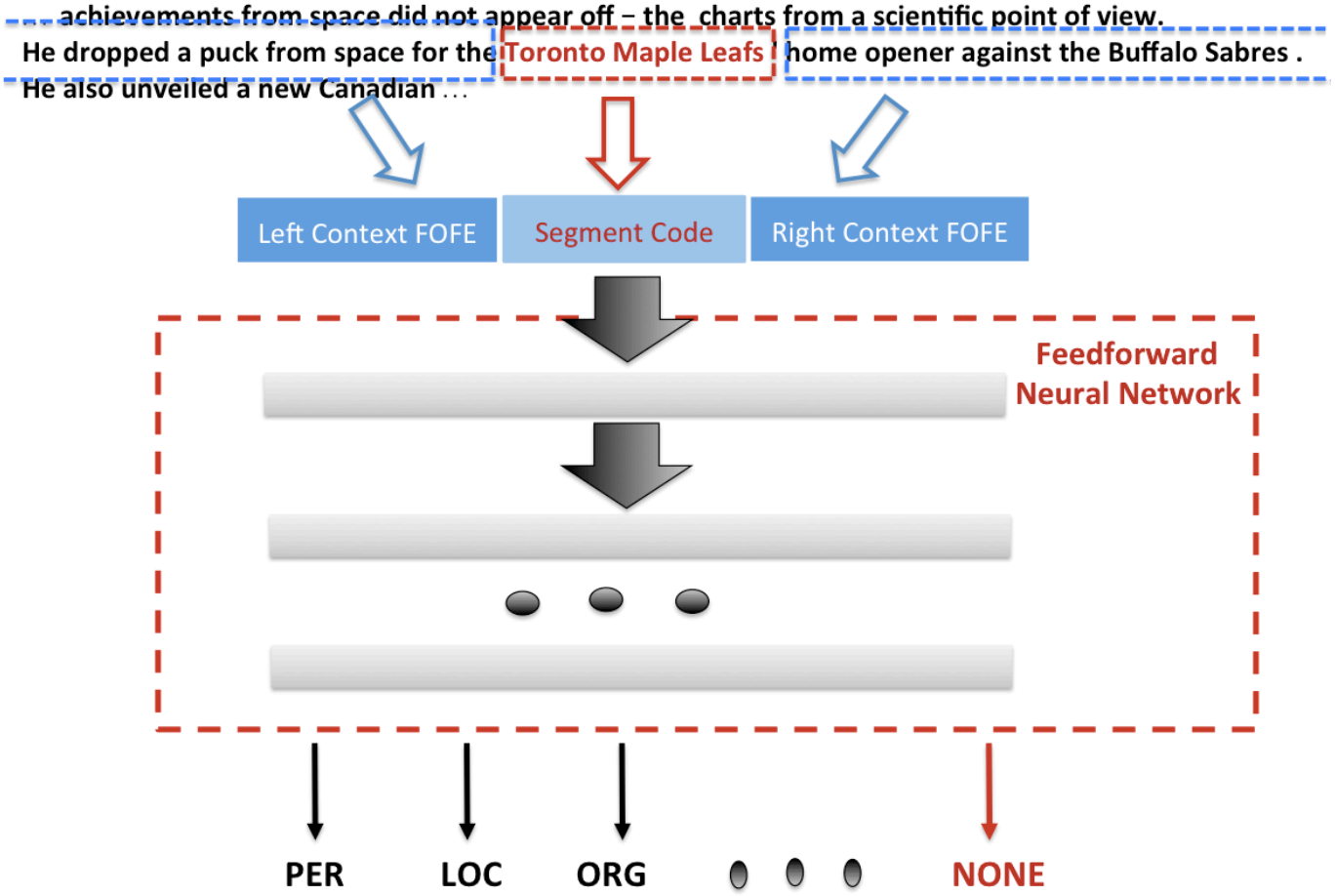
Entity Discovery Performance on KBP2016 EDL1 evaluation

System	P	R	F
system1 + 5SC	0.850	0.678	0.754
system2 + 5SC	0.836	0.681	0.751
fusion	0.822	0.704	**0.759**

Entity Linking Performance on KBP2016 EDL1 evaluation

KBP2016 Trilingual EDL	P	R	F
strong_all_match	0.720	0.617	**0.665**
typed_mention_ceaf_plus	0.676	0.579	**0.624**

York ED Model



- FOFE code for left context
- FOFE code for right context
- BoW vector
- Char FOFE code

- **Local detection**: no Viterbi decoding; Nested/Embedded entities
- No feature engineering: FOFE codes
- Easy and fast to train; make use of partial labels

York System ED Performance

- **Effect of various training data sets:**
  - **KBP2015 training set**
  - **Machine-labelled Wikipedia data**
  - **iFLYTEK in-house data**

training data	P	R	F_1
KBP2015	0.818	0.600	0.693
KBP2015 + WIKI	0.859	0.601	0.707
KBP2015 + iFLYTEK	0.830	0.652	**0.731**

English Entity Discovery Performance on KBP2016 EDL1 evaluation

York Official KBP2016 EDL Results

Entity Discovery Performance on KBP2016 EDL2 evaluation

	NAME			NOMINAL			OVERALL		
	P	R	F1	P	R	F1	P	R	F1
	RUN1 (our official ED result in KBP2016 EDL2)								
ENG	0.898	0.789	0.840	0.554	0.336	0.418	0.836	0.680	0.750
CMN	0.848	0.702	0.768	0.414	0.258	0.318	0.789	0.625	0.698
SPA	0.835	0.778	0.806	0.000	0.000	0.000	0.835	0.602	0.700
ALL	0.893	0.759	0.821	0.541	0.315	0.398	0.819	0.639	**0.718**
	RUN3 (system fusion of RUN1 + USTC)								
ENG	0.857	0.876	0.866	0.551	0.373	0.444	0.804	0.755	0.779
CMN	0.790	0.839	0.814	0.425	0.380	0.401	0.735	0.760	0.747
SPA	0.790	0.877	0.831	0.000	0.000	0.000	0.790	0.678	0.730
ALL	0.893	0.759	0.821	0.541	0.315	0.398	0.774	**0.735**	**0.754**

Entity Linking Performance on KBP2016 EDL2 evaluation

	RUN1			RUN3		
	P	R	F1	P	R	F1
strong_all_match	0.721	0.562	**0.632**	0.667	0.634	0.650
typed_mention_ceaf_plus	0.681	0.531	**0.597**	0.626	0.594	0.609

York Official KBP2016 EDL Results

Entity Discovery Performance on KBP2016 EDL2 evaluation

	NAME			NOMINAL			OVERALL		
	P	R	F1	P	R	F1	P	R	F1
	RUN1 (our official ED result in KBP2016 EDL2)								
ENG	0.898	0.789	0.840	0.554	0.336	0.418	0.836	0.680	0.750
CMN	0.848	0.702	0.768	0.414	0.258	0.318	0.789	0.625	0.698
SPA	0.835	0.778	0.806	0.000	0.000	0.000	0.835	0.602	0.700
ALL	0.893	0.759	0.821	0.541	0.315	0.398	0.819	0.639	**0.718**
	RUN3 (system fusion of RUN1 + USTC)								
ENG	0.857	0.876	0.866	0.551	0.373	0.444	0.804	0.755	0.779
CMN	0.790	0.839	0.814	0.425	0.380	0.401	0.735	0.760	0.747
SPA	0.790	0.877	0.831	0.000	0.000	0.000	0.790	0.678	0.730
ALL	0.893	0.759	0.821	0.541	0.315	0.398	0.774	**0.735**	**0.754**

Entity Linking Performance on KBP2016 EDL2 evaluation

	RUN1			RUN3		
	P	R	F1	P	R	F1
strong_all_match	0.721	0.562	**0.632**	0.667	0.634	0.650
typed_mention_ceaf_plus	0.681	0.531	**0.597**	0.626	0.594	0.609

York Official KBP2016 EDL Results

Entity Discovery Performance on KBP2016 EDL2 evaluation

	NAME			NOMINAL			OVERALL		
	P	R	F1	P	R	F1	P	R	F1
	RUN1 (our official ED result in KBP2016 EDL2)								
ENG	0.898	0.789	0.840	0.554	0.336	0.418	0.836	0.680	0.750
CMN	0.848	0.702	0.768	0.414	0.258	0.318	0.789	0.625	0.698
SPA	0.835	0.778	0.806	0.000	0.000	0.000	0.835	0.602	0.700
ALL	0.893	0.759	0.821	0.541	0.315	0.398	0.819	0.639	**0.718**
	RUN3 (system fusion of RUN1 + USTC)								
ENG	0.857	0.876	0.866	0.551	0.373	0.444	0.804	0.755	0.779
CMN	0.790	0.839	0.814	0.425	0.380	0.401	0.735	0.760	0.747
SPA	0.790	0.877	0.831	0.000	0.000	0.000	0.790	0.678	0.730
ALL	0.893	0.759	0.821	0.541	0.315	0.398	0.774	**0.735**	**0.754**

Entity Linking Performance on KBP2016 EDL2 evaluation

	RUN1			RUN3		
	P	R	F1	P	R	F1
strong_all_match	0.721	0.562	**0.632**	0.667	0.634	0.650
typed_mention_ceaf_plus	0.681	0.531	**0.597**	0.626	0.594	0.609

York Official KBP2016 EDL Results

Entity Discovery Performance on KBP2016 EDL2 evaluation

	NAME			NOMINAL			OVERALL		
	P	R	F1	P	R	F1	P	R	F1
	RUN1 (our official ED result in KBP2016 EDL2)								
ENG	0.898	0.789	0.840	0.554	0.336	0.418	0.836	0.680	0.750
CMN	0.848	0.702	0.768	0.414	0.258	0.318	0.789	0.625	0.698
SPA	0.835	0.778	0.806	0.000	0.000	0.000	0.835	0.602	0.700
ALL	0.893	0.759	0.821	0.541	0.315	0.398	0.819	0.639	**0.718**
	RUN3 (system fusion of RUN1 + USTC)								
ENG	0.857	0.876	0.866	0.551	0.373	0.444	0.804	0.755	0.779
CMN	0.790	0.839	0.814	0.425	0.380	0.401	0.735	0.760	0.747
SPA	0.790	0.877	0.831	0.000	0.000	0.000	0.790	0.678	0.730
ALL	0.893	0.759	0.821	0.541	0.315	0.398	0.774	**0.735**	**0.754**

Entity Linking Performance on KBP2016 EDL2 evaluation

	RUN1			RUN3		
	P	R	F1	P	R	F1
strong_all_match	0.721	0.562	**0.632**	0.667	0.634	0.650
typed_mention_ceaf_plus	0.681	0.531	**0.597**	0.626	0.594	0.609

Conclusions

- Exploring neural network models for EDL
- Proposed some new methods for EDL
  - Encoder-decoder model using CNN+RNN
  - Introduce attention mechanism
  - Extend for tree-structured tags
  - FOFE-based Local detection approach for NER and mention detection
- Achieved strong *(1st and 2nd)* performance in the KBP2016 EDL evaluations