The AI2 system at SemEval-2017 Task 10 (ScienceIE): semi-supervised end-to-end entity and relation extraction

2020.12.9

Abstract

- Task

ScienceIE shared task (SemEval-2017 Task 10)

Entity and relation extraction from scientific papers

- Model

Original: End-to-End Relation Extraction using LSTMs on Sequences and Tree Structures

Enhancements : semi-supervised learning via neural language models

character-level encoding

gazetteers extracted from existing knowledge bases

model ensembles

- Result

First in end-to-endentity and relation extraction

Second in the relation-only extraction

Original Model

- Traditional systems

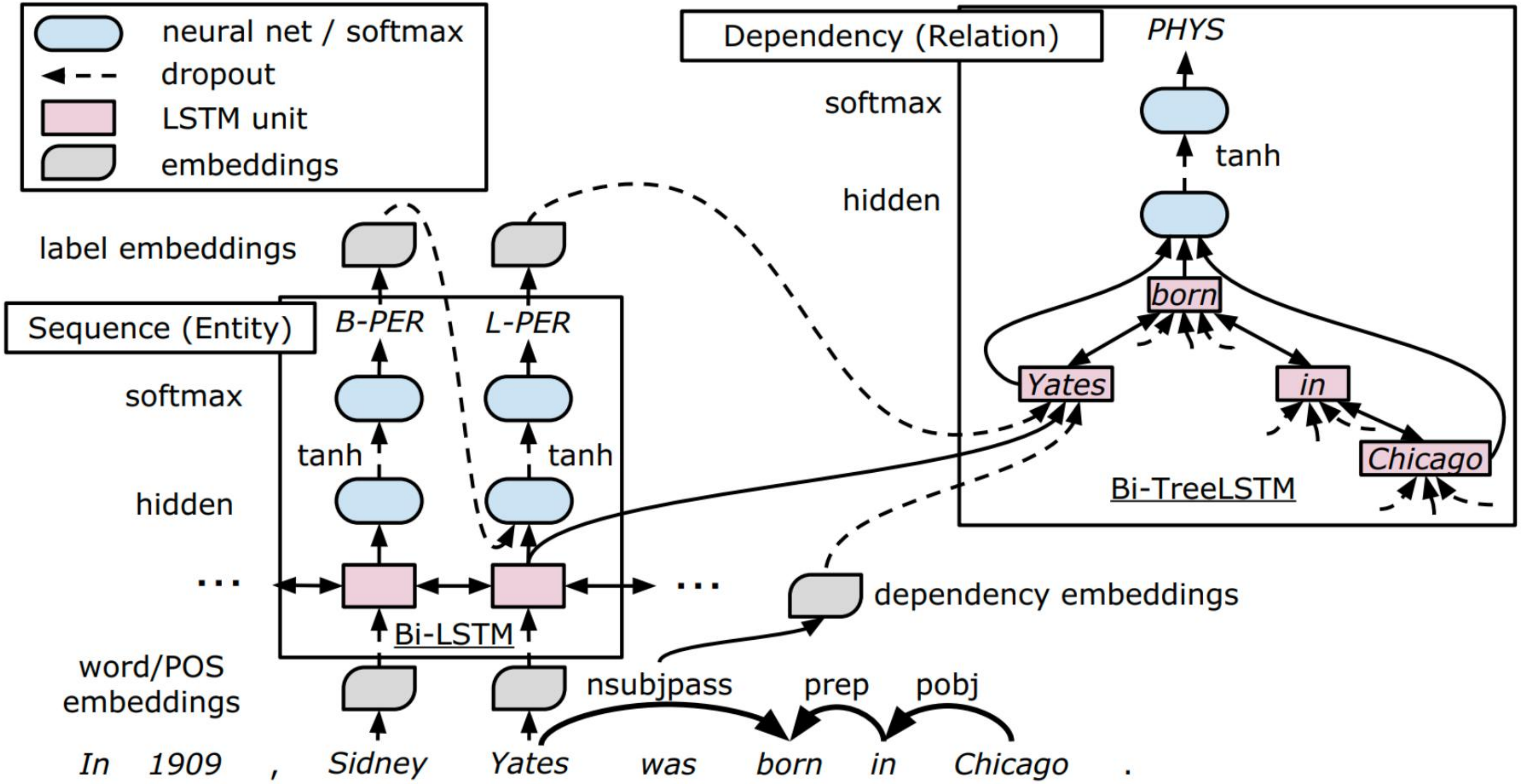
Seperate: named entity recognition (NER) + relation extraction

Joint: end-to-end modeling

Previous: feature-based structured learning

Information	Model
word sequence	bidirectional sequential LSTM-RNNs
dependency tree substructure	bidirectional tree-structured LSTM-RNNs

Original Model



Original Model

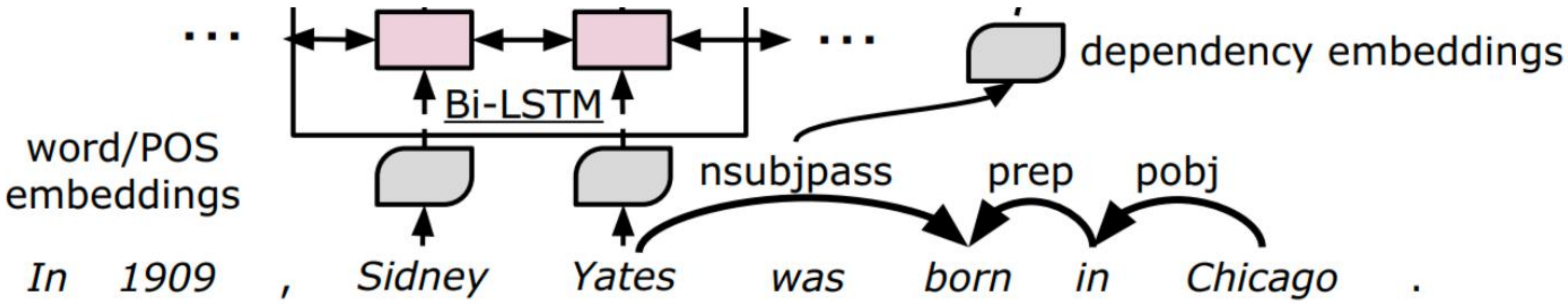
- Embedding Layer

$v^{(w)}$ word embedding

$v^{(p)}$ part-of-speech (POS) tags

$v^{(d)}$ dependency types

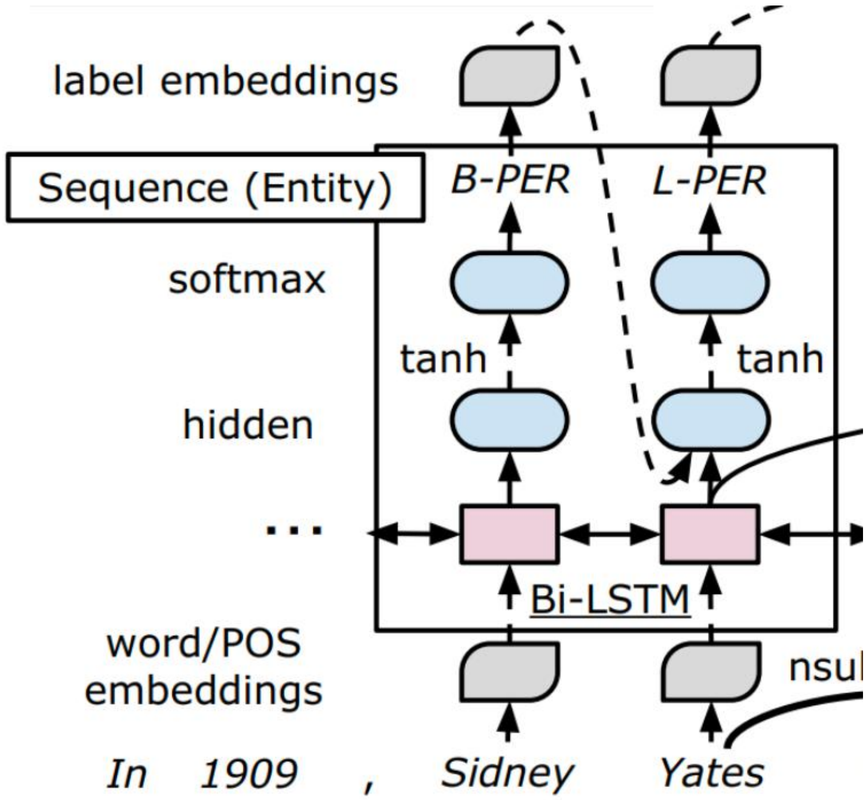
$v^{(e)}$ entity labels



Original Model

- Sequence Layer -- entity detection

$\begin{cases} \text{type} \\ \text{position} \end{cases}$



$$h_t^{(e)} = tanh(W^{(e_h)}[s_t; v_{t-1}^{(e)}] + b^{(e_h)})$$
$$y_t = softmax(W^{(e_y)} h_t^{(e)} + b^{(e_y)})$$

sequence labeling task (BILOU)

BiLSTM: $s_t = [\overrightarrow{h}_t; \overleftarrow{h}_t]$

input: $x_t = [v_t^{(w)}, v_t^{(p)}]$

Original Model

type
direction
(negative relation)

wrong entities
no relation

- Dependency Layer

-- relation classification

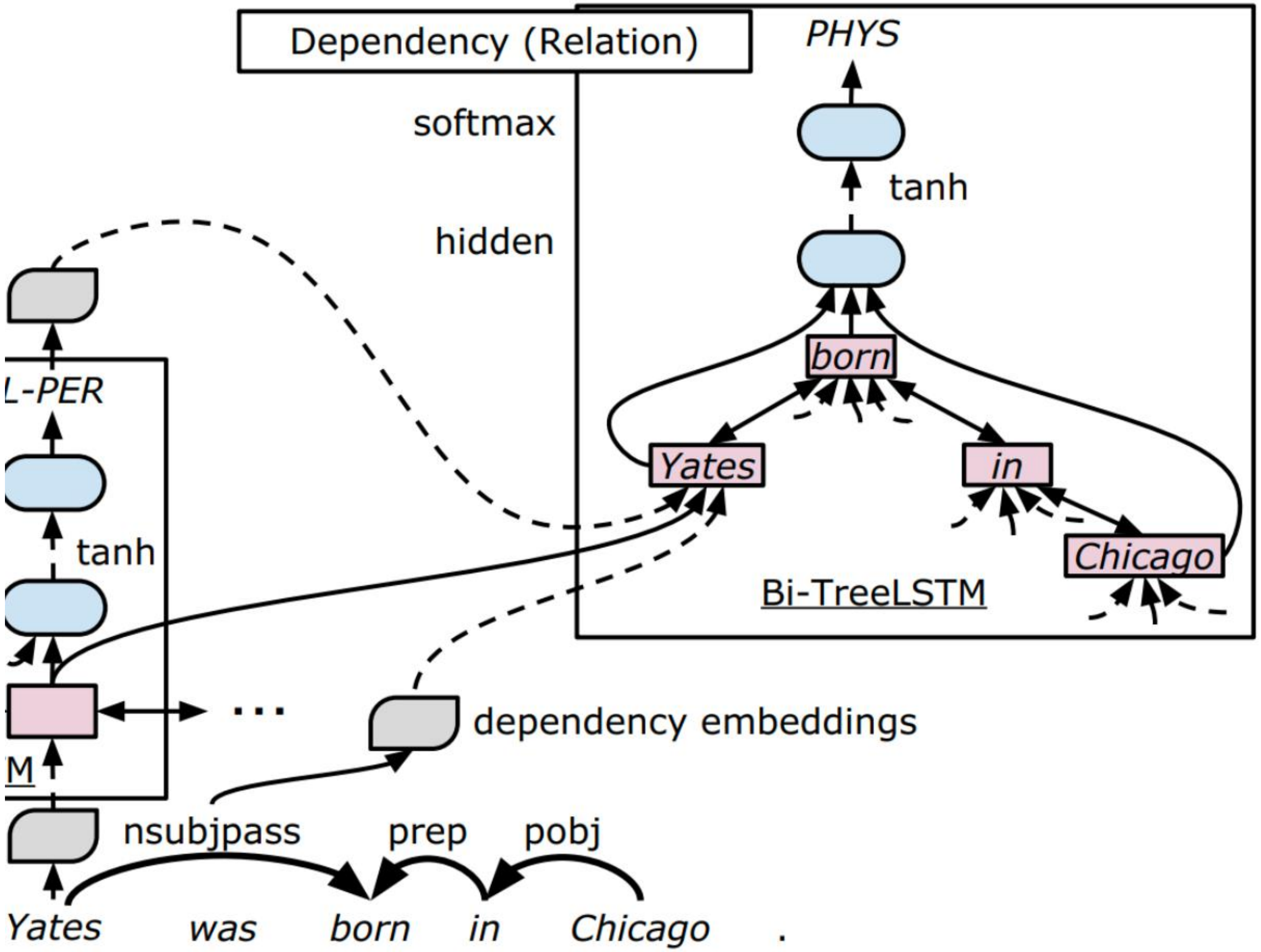
$$h_p^{(r)} = \tanh\left(W^{(r_h)} d_p + b^{(r_h)}\right)$$

$$y_p = \mathrm{softmax}\left(W^{(r_y)} h_t^{(r)} + b^{(r_y)}\right)$$

$$d'_p = [d_p; \frac{1}{|I_{p_1}|} \sum_{i \in I_{p_1}} s_i; \frac{1}{|I_{p_2}|} \sum_{i \in I_{p_2}} s_i]$$

$d_p = [\uparrow h_{pa}; \downarrow h_{p1}; \downarrow h_{p2}]$

input: $x_t=[s_t;v_t^{(d};v_t^{(e)}]$



Task overview

- Extraction

Typed entities： Task / Material / Process

Relations： Hyponym-of / Synonym-of

- Running example

“Here, we consider a radical pair in which the first electron spin is devoid of hyperfine interactions, while the second electron spin interacts isotropically with one spin-1 nucleus, e.g.nitrogen.”

在这里，我们考虑一个激进的对，其中第一个电子自旋没有超细相互作用，而第二个电子自旋与一个自旋- 1核（如氮）等向相互作用。

Process： electron spin

Material： spin-1 nucleus 、nitrogen

Hyponym-of： “nitrogen”,“spin-1 nucleus”

System description

- Text preprocessing

spaCy

- Label encoding：BILOU

B： 'beginning' (signifies beginning of an NE)

I ： 'inside' (signifies that the word is inside an NE)

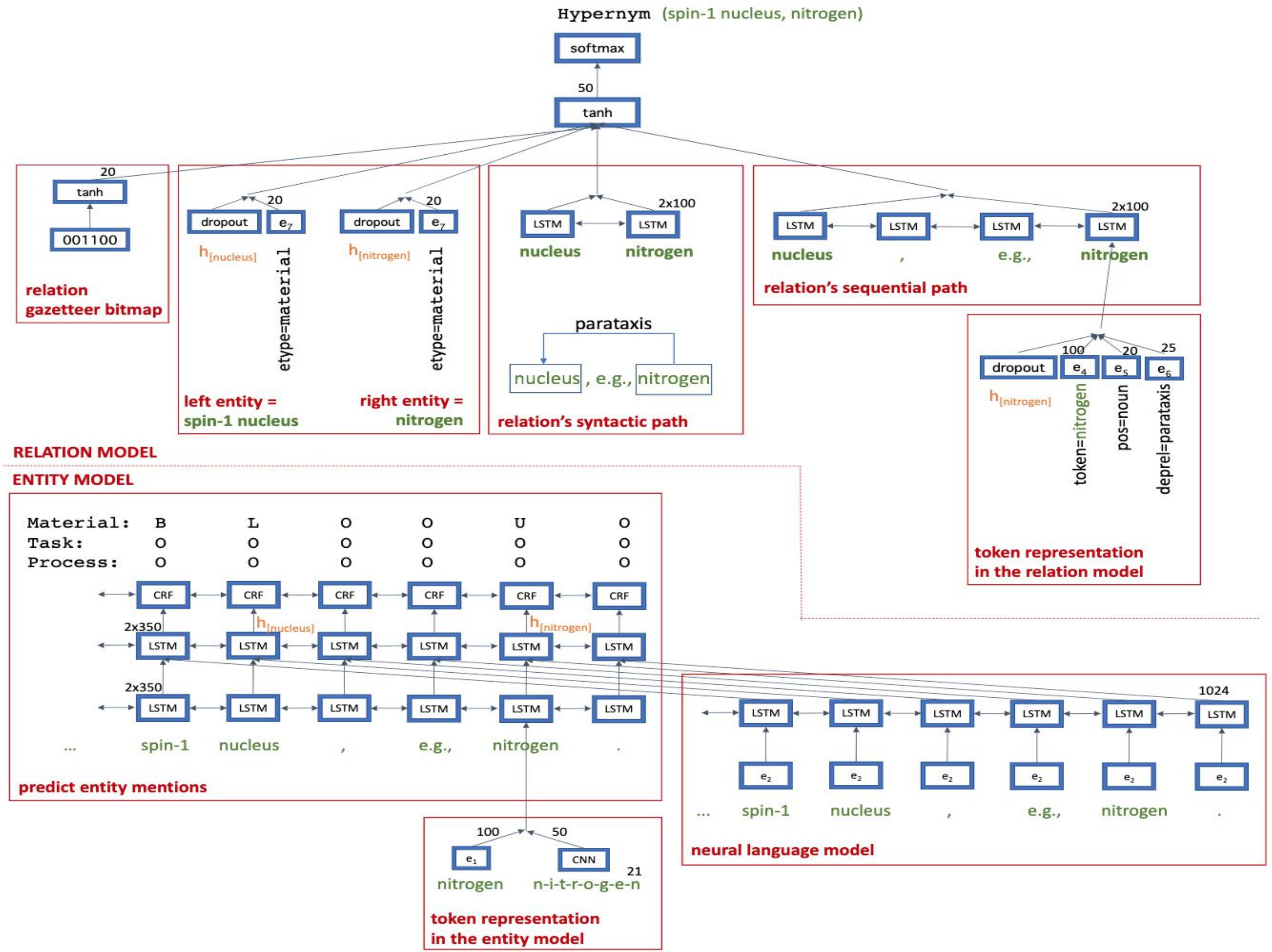
O： 'outside' (signifies that the word is just a regular word outside of an NE)

L ： 'last' (signifies that the last word of an NE)

U： 'unit'(signifies that the single word is an NE)

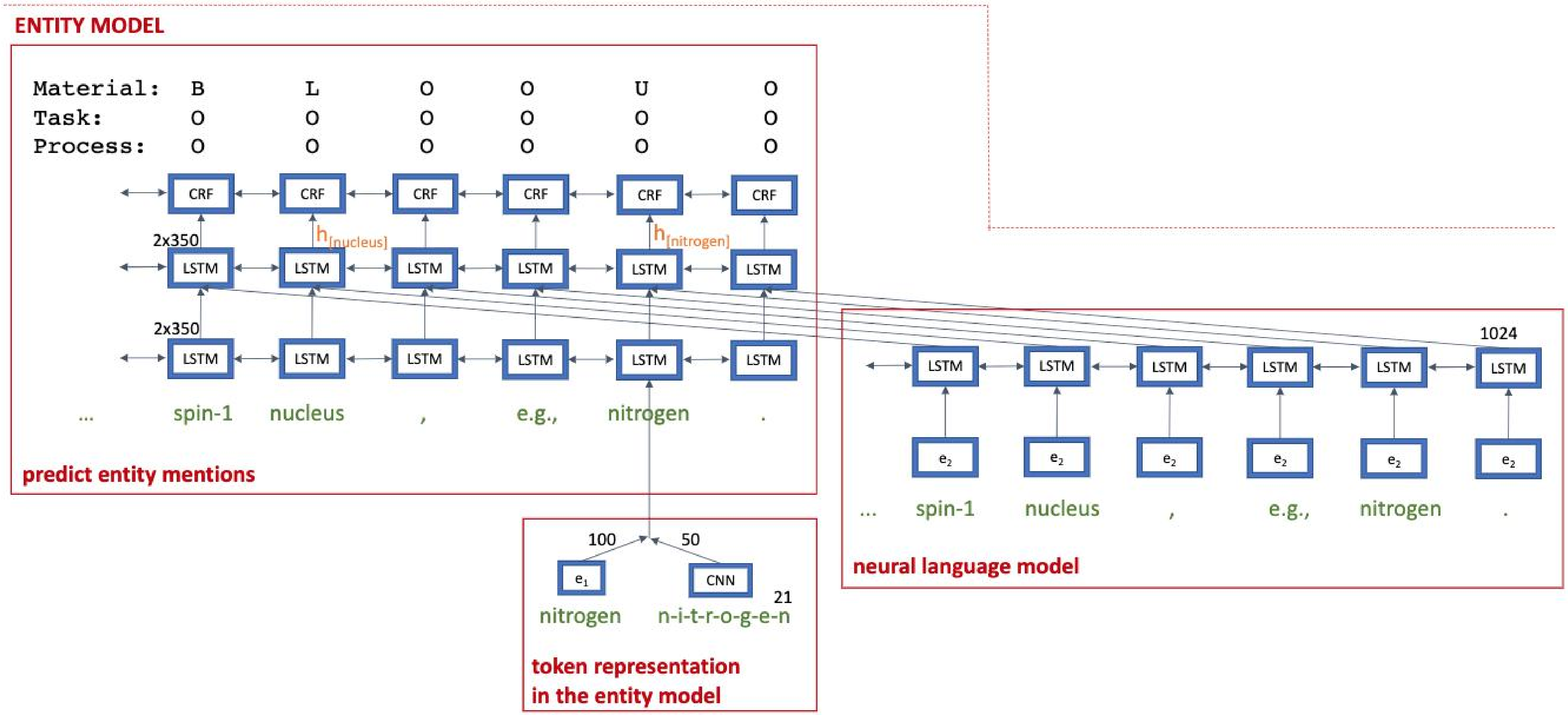
directional relation (Hyponym-of)

undirectional relation (Synonym-of)

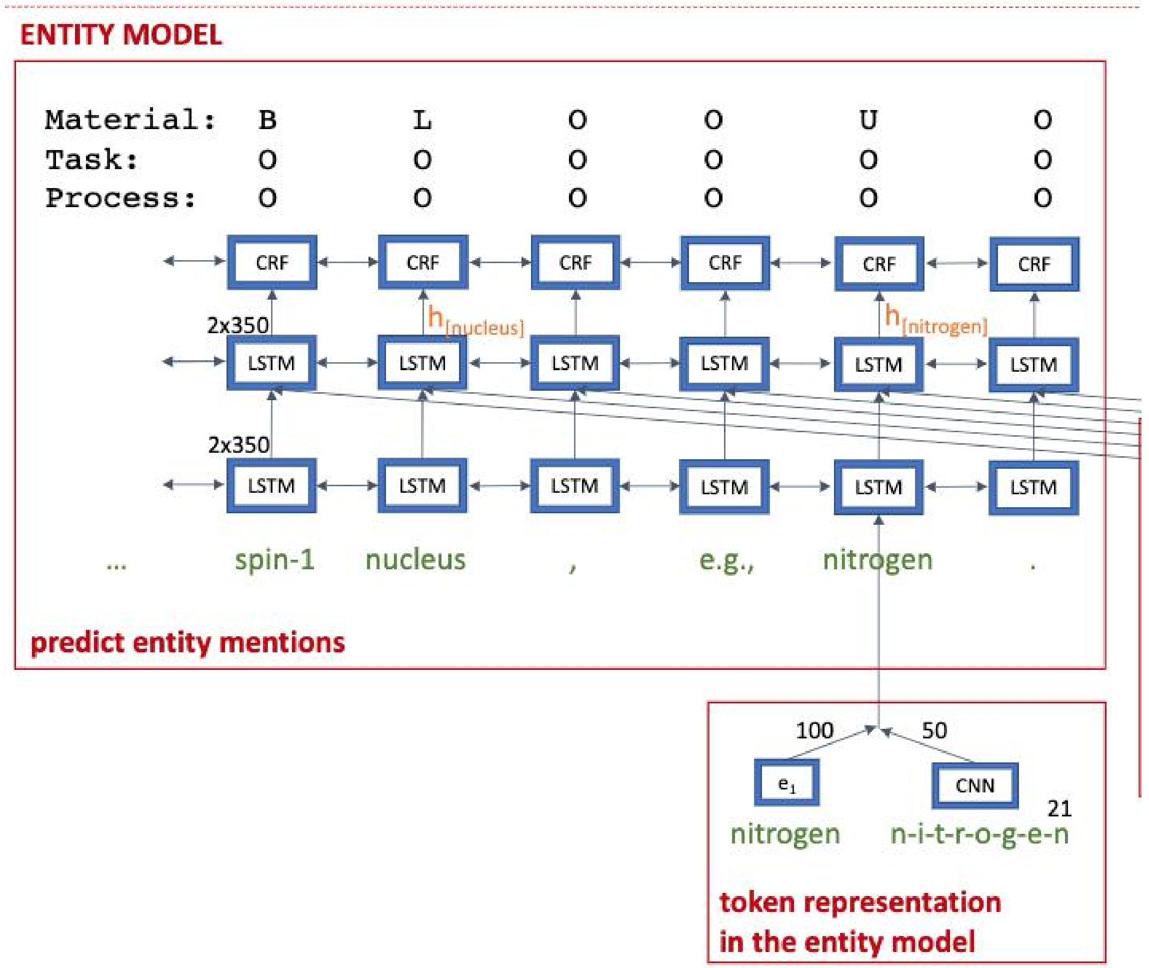


Hypernym (spin-1 nucleus, nitrogen)
softmax
50
tanh
20
tanh
001100
relation gazetteer bitmap
dropout
e7
20
h[nucleus]
etype=material
dropout
e7
20
h[nitrogen]
etype=material
left entity = spin-1 nucleus
right entity = nitrogen
LSTM
LSTM
2x100
nucleus
nitrogen
parataxis
nucleus , e.g., nitrogen
relation's syntactic path
LSTM
LSTM
LSTM
LSTM
2x100
nucleus
,
e.g.,
nitrogen
relation's sequential path
dropout
e4
e5
e6
100
20
25
h[nitrogen]
token=nitrogen
pos=noun
deprel=parataxis
token representation in the relation model
RELATION MODEL
ENTITY MODEL
Material: B L O O U O
Task: O O O O O O
Process: O O O O O O
CRF
2x350
h[nucleus]
h[nitrogen]
LSTM
2x350
... spin-1 nucleus , e.g., nitrogen .
predict entity mentions
100
50
e1
CNN
21
nitrogen
n-i-t-r-o-g-e-n
token representation in the entity model
1024
LSTM
e2
... spin-1 nucleus , e.g., nitrogen .
neural language model

Entity model



Entity model



- Token representation

using unlabeled data to learn feature representations of individual word types

$x_k = [c_k; w_k]$

c_k : character based representation (CNN) with a filter width of 3 characters

w_k : token embeddings
pretrained GloVe word embeddings

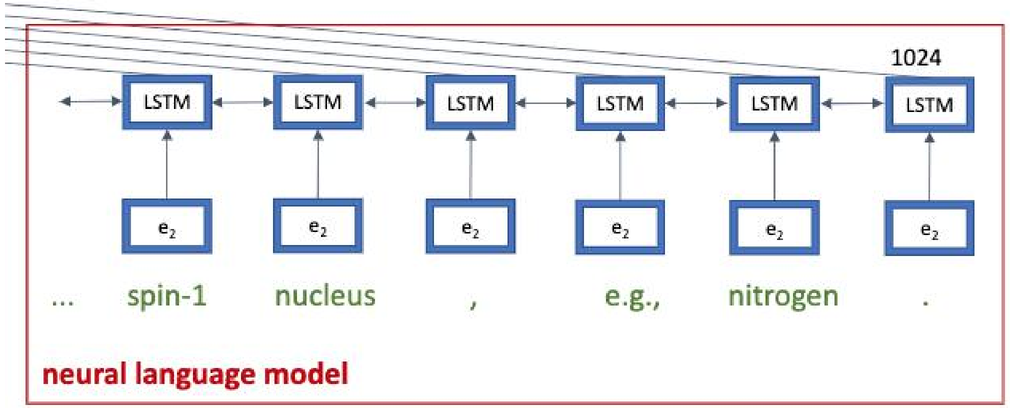
Entity model

- Neural language model

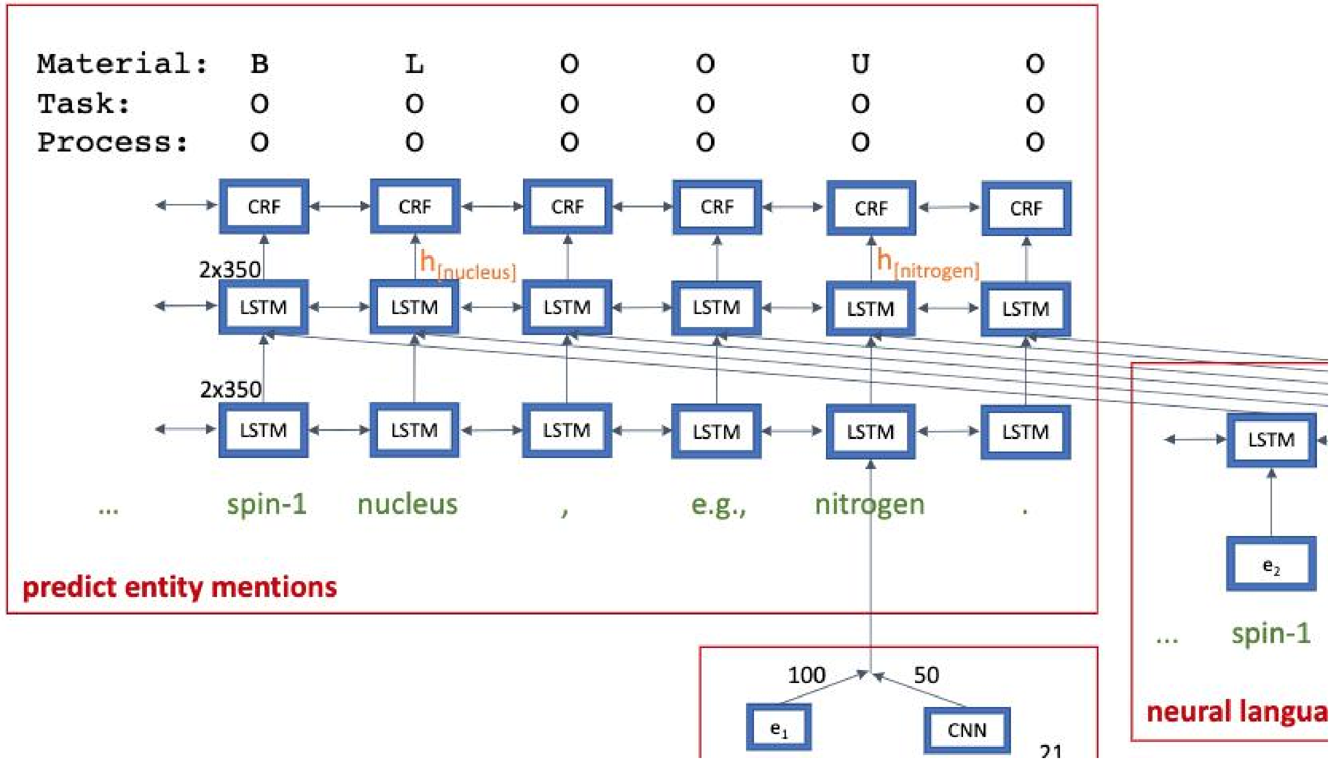
learn feature representations of words in a particular context

forward LM & backward LM trained seperately

$$h_k^{LM} = [\overrightarrow{\mathbf{h}}_k^{LM}; \overleftarrow{\mathbf{h}}_k^{LM}]$$



Entity model



- Sequence tagging model

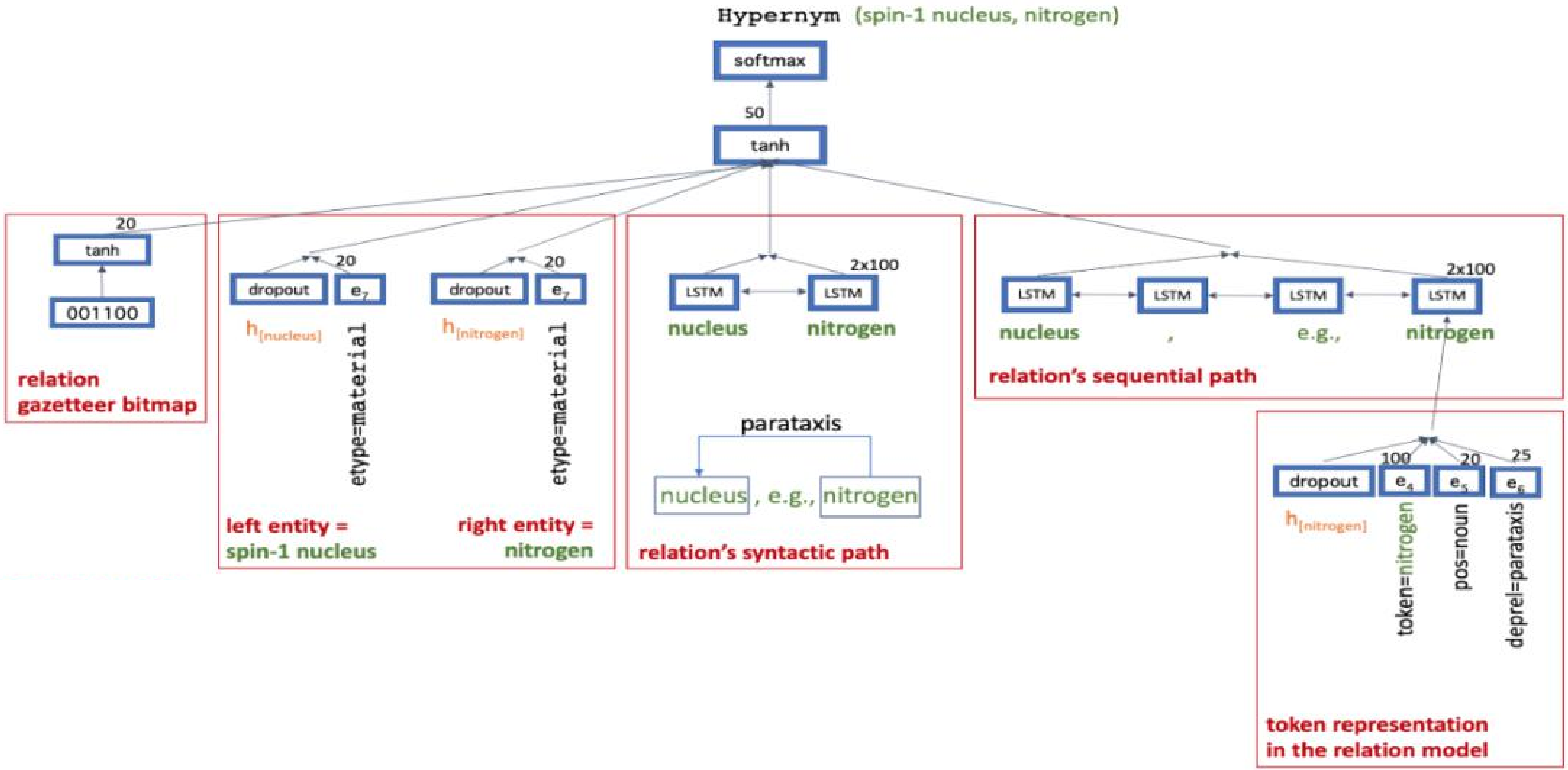
predict entity mentions of each type

Output of the 2nd LSTM layer:

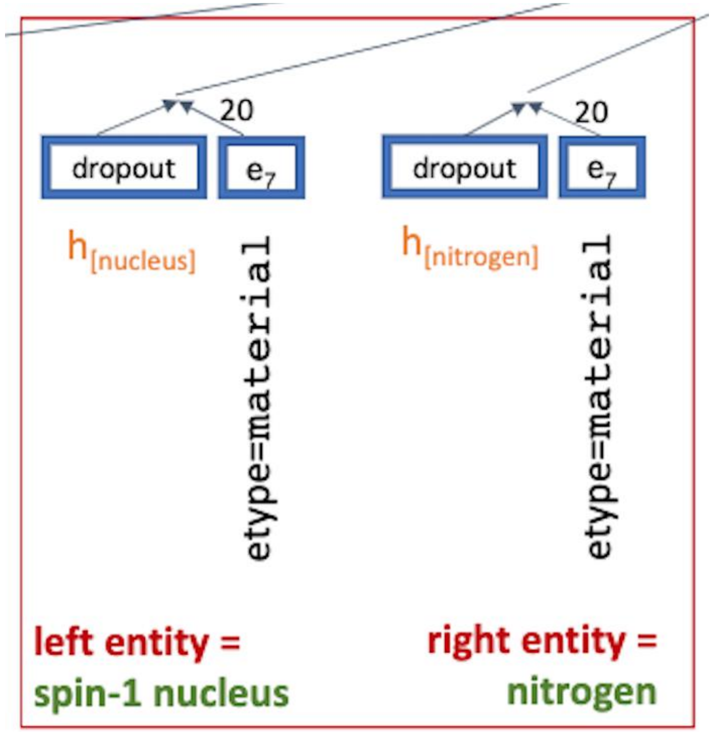
h_k : predict a score for each possible tag

CRF: conditional random field

Relation model



Relation model

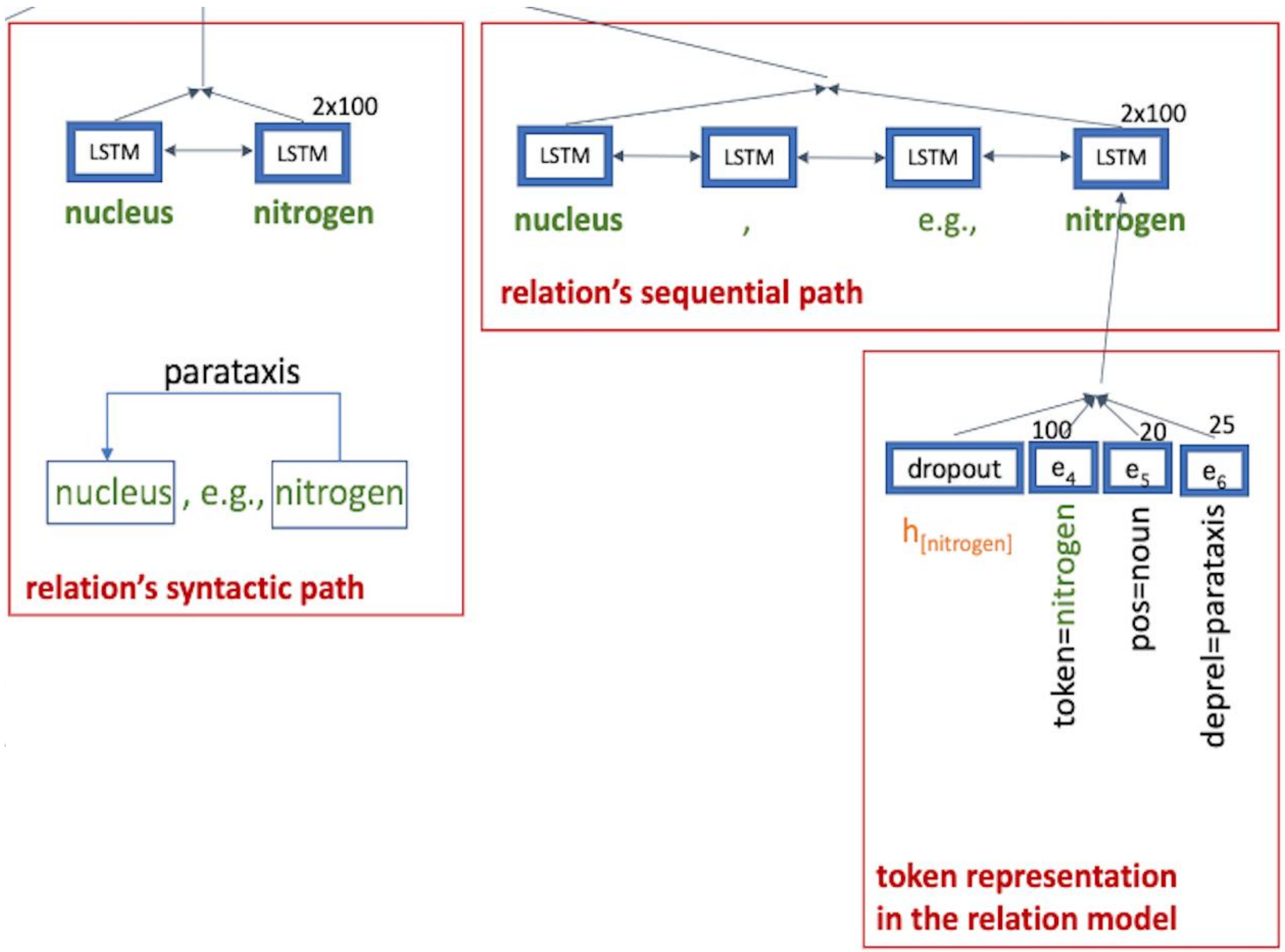


- Left and right entities

$x_t = [\mathrm{h}_k \; ; \; v_t^{(e)}]$

Obatain a fixed size encoding:
dependency tree
-- syntactic head of the entity

Relation model



- Syntactic and sequential path

1) Token representation

context-sensitive embedding h_k

context-insensitive embedding

POS tags

dependency types (node and its direct head)

2) Syntactic path

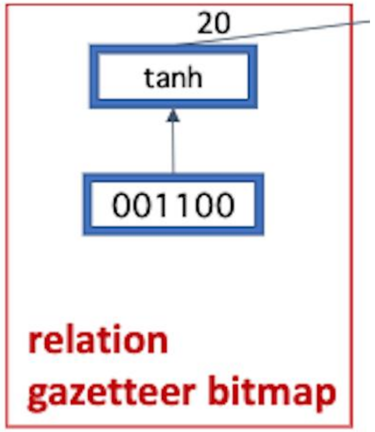
shortest path in a dependency tree

between the heads of the left and right entities

3) Sequential path

tokens between the two heads in the sentence

Relation model



- Relation gazetteer features

Wikipedia and freebase

Three features :

acronym

suffix

exact copy of the other

Ensembles

- Entity model ensemble

averagesthe label predictions at each position

15

+

- Relation model ensemble

only predicts a positive relation if 50% of the individual models predict the same relation

8

Result

Model	F_1
Our best model without language model	49.9
Our best model with language model	54.1
Our 15-model ensemble	55.2

Table 1: Development set entity only F_1 comparison.

Team	**End-to-end**	**Entities**	**Relations**
Ours	**0.43**	0.55	**0.28**
Team_24	0.42	**0.56**	-
Team_21	0.38	0.50	0.21
Team_19	0.37	0.51	0.19
Team_14	0.33	0.47	0.20

Table 2: Final test set F_1 for top five teams in Scenario 1, end-to-end extraction.