



Bert-BiLSTM-CRF for NER

Sun Xinyu, Wang Qian, Wang
Sijie, Wang Chenxi

1 Problem Definition

Named Entity Recognition (NER)

Jim bought 300 shares of Acme Corp. in 2006.

person value organization time

European NORP authorities fined Google ORG a record \$5.1 billion MONEY on Wednesday DATE for abusing its power in the mobile phone market and ordered the company to alter its practices

Locate and classify named entities in unstructured text into predefined categories, such as person name, organization, location, time, etc.

2 Dataset Introduction

CoNLL 2003

The CoNLL-2003 named entity data two languages: English and German . (we only use English)

For each of the languages there is a training file, a development file, a test file and a large file with unannotated data

2 Dataset Introduction

```
eng.train
EU NNP I-NP I-ORG
rejects VBZ I-VP 0
German JJ I-NP I-MISC
call NN I-NP 0
to TO I-VP 0
boycott VB I-VP 0
British JJ I-NP I-MISC
lamb NN I-NP 0
. . 0 0

Peter NNP I-NP I-PER
Blackburn NNP I-NP I-PER

BRUSSELS NNP I-NP I-LOC
1996-08-22 CD I-NP 0

The DT I-NP 0
European NNP I-NP I-ORG
Commission NNP I-NP I-ORG
said VBD I-VP 0
on IN I-PP 0
Thursday NNP I-NP 0
it PRP B-NP 0
disagreed VBD I-VP 0
with IN I-PP 0
German JJ I-NP I-MISC
advice NN I-NP 0
to TO I-PP 0
consumers NNS I-NP 0
to TO I-VP 0
```

The English data was taken from the Reuters Corpus, it means from news field.

4 categories:

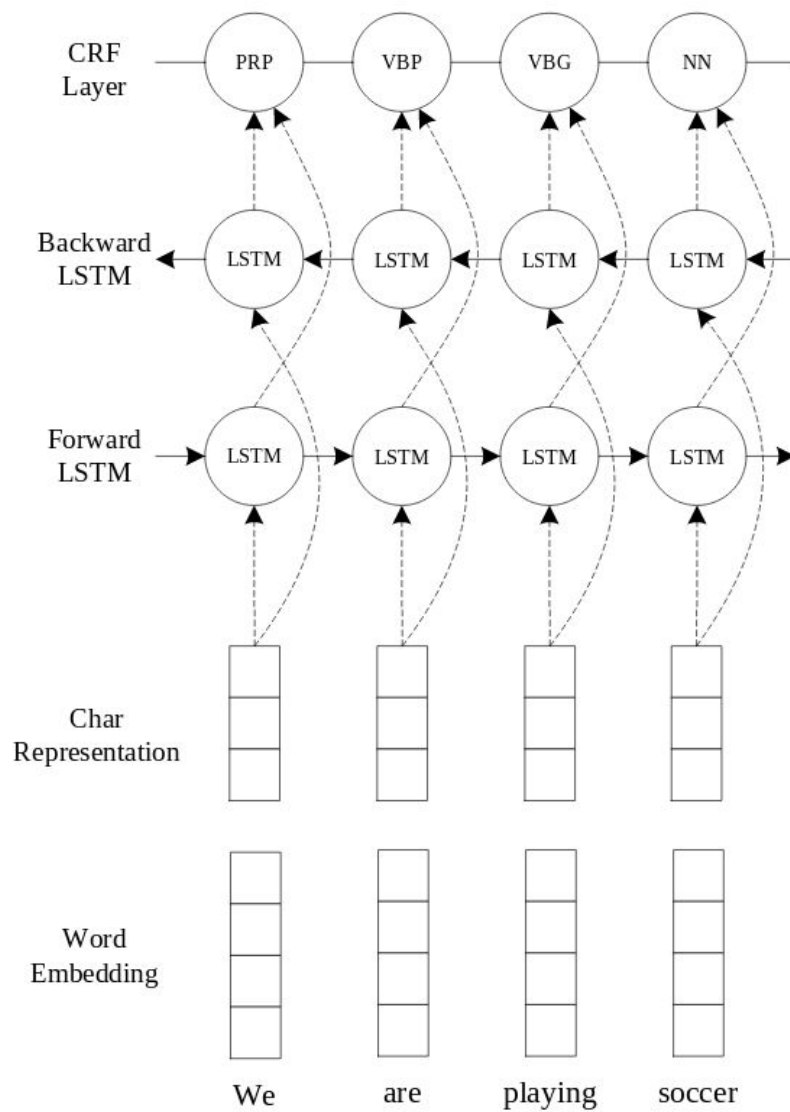
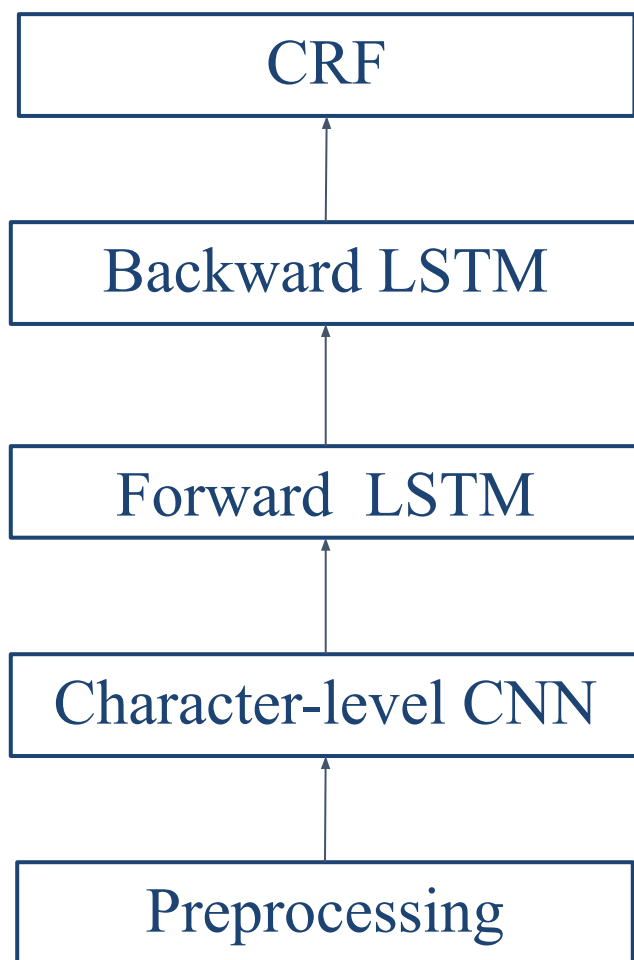
LOC(location)

MISC(miscellaneous)

ORG(Organization)

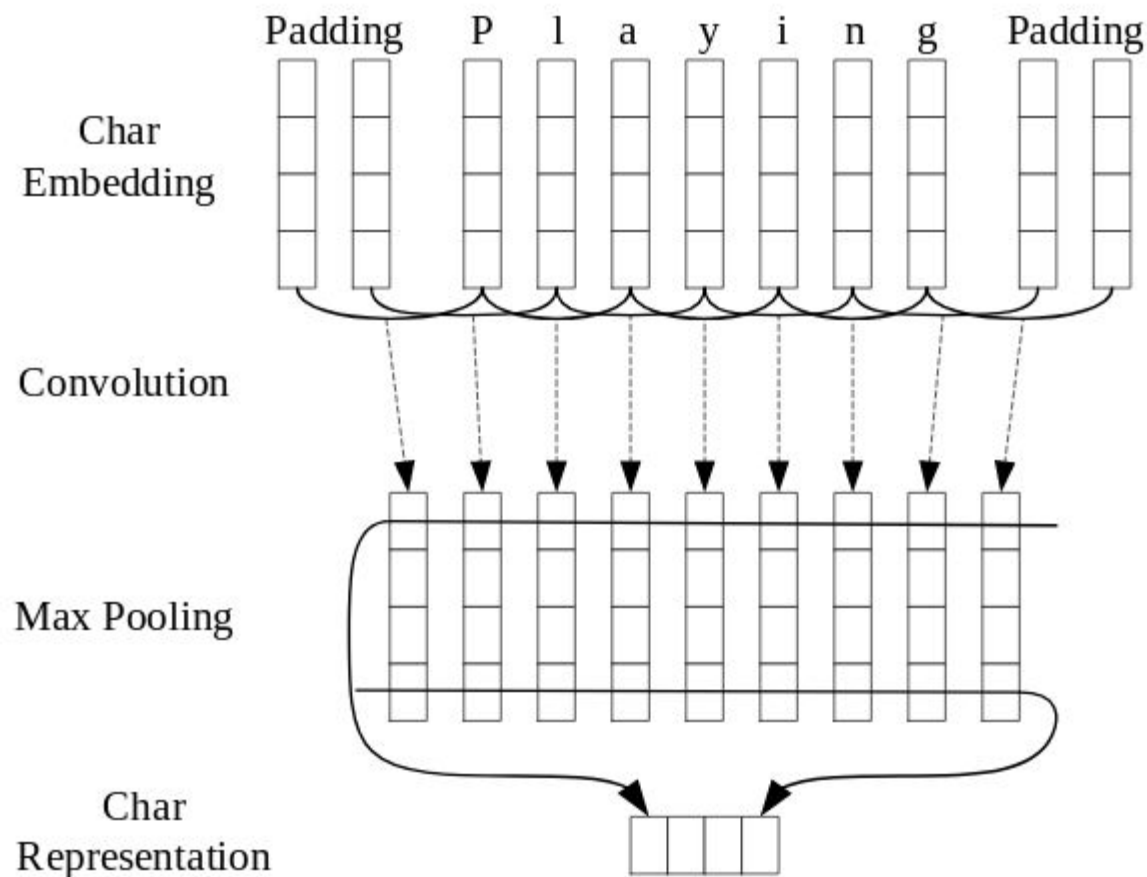
PER(Person)

3 NER Proposed Method



4 Character-level CNN

Why not LSTM?



4

BiLSTM

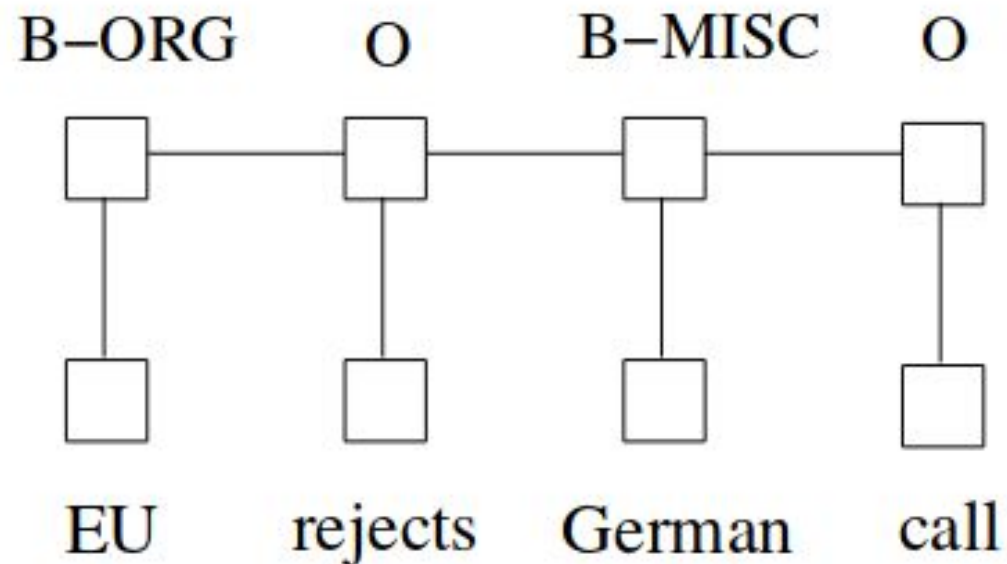
Difference between BiLSTM and normal LSTM:

Unidirectional LSTM only preserves information of the past while bidirectional one will run your inputs in two ways.

Running information backwards preserves information from the future. Two hidden states combined.

Forward LSTM: The boys went to

Backward LSTM: ... and then they got out of the pool



Using features extracted from BiLSTM and compute the transition probability.

6

What we proposed

Adding an extra layer of BiLSTM

Modifying word embedding with
Bert

7

Adding an extra layer of BiLSTM

The depth of neural networks is generally attributed to the success of the approach on a wide range of challenging prediction problems.

The additional hidden layers are understood to recombine the learned representation from prior layers and create new representations at high levels of abstraction. For example, from lines to shapes to objects.

More complex than a vanilla neural network. Time-consuming.

8

Word Embedding with Bert

What is Bert?

Bidirectional Encoder Representations from Transformers(BERT) is a language model that comes from a google paper.

BERT used the bidirectional training of Transformer to language model.

This new method can have a deeper sense of language context. So that BERT model can achieve state-of-the-art results in a wide variety of NLP tasks.

8

Word Embedding with Bert

“Fine-tuning” the word index

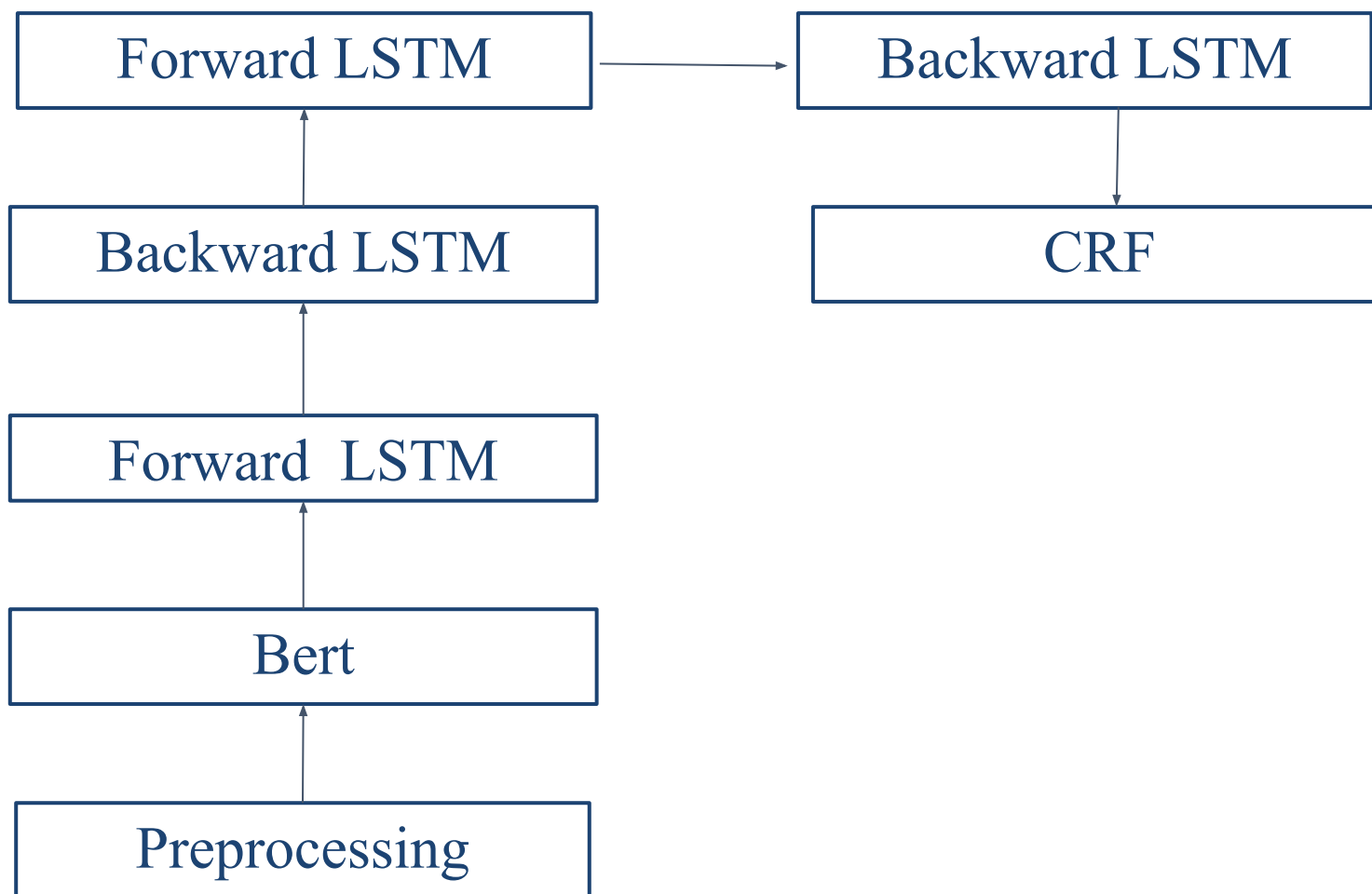
Original word index: fixed and manually inputted



OR



9 Our Proposed Method



10 Results

Comparison of different methods

Model	% P	% R	% F1
BLSTM	88.61	88.50	88.56
BLSTM-CRF	90.33	88.81	89.56
CNN-BLSTM	89.23	90.97	90.09
CNN-BLSTM-CRF	91.36	91.24	91.29
BERT-BLSTM-CRF(ours)	92.81	92.33	92.57

$$P = \frac{TP}{TP + FP}$$

$$R = \frac{TP}{TP + FN}$$

$$F1 = \frac{2PR}{P+R}$$

Donald is the president of USA

**Donald : PER
is : NA
the : NA
president : NA
of : NA
USA : LOC**

Jay is from India

**Jay : PER
is : NA
from : NA
India :
LOC**



THANK YOU