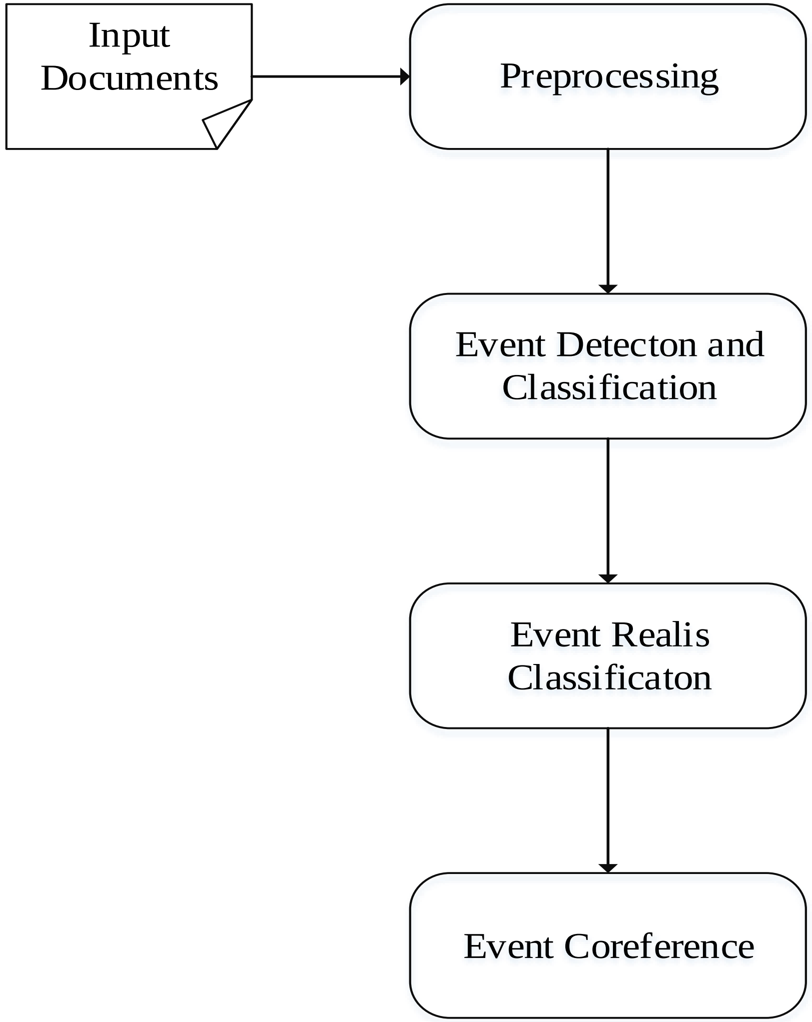
SRCB at TAC KBP 2017 Event Nugget Track

Data Mining Lab

Ricoh Software Research Center (Beijing) Co.,Ltd.

(TeamID: srcb)

Our system



Preprocessing includes:

- Sentence boundary
- Tokenization
- Vocabulary setup
- Word embedding

EN Detection and Classification

RICOH imagine. change.

- To detect explicit mentions of relevant events and identify event types and subtypes.
- A sequence labelling problem.

- An **ensemble model** which combines a **neural network model** and a **Conditional Random Fields (CRFs)** model.

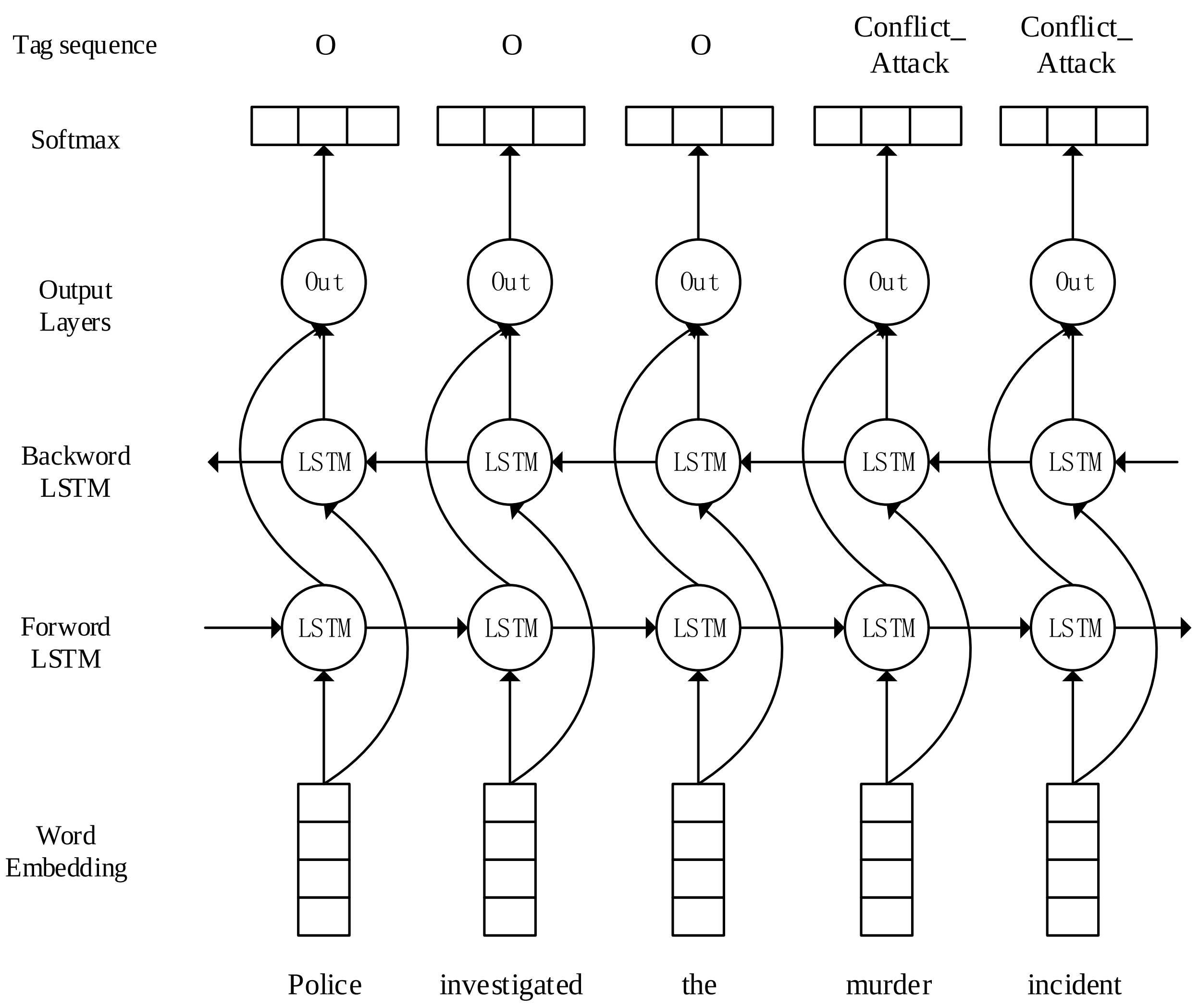
Contact_Contact Life_Die Movement_Transport-Person

He said the killings were in self-defense and he fled the state because he could not get a fair trial.

Contact_Broadcast Conflict_Attack Justice_Arrest-Jail

Leak information about US war crimes, and spend the rest of your life in a military prison.

Neural Network Model

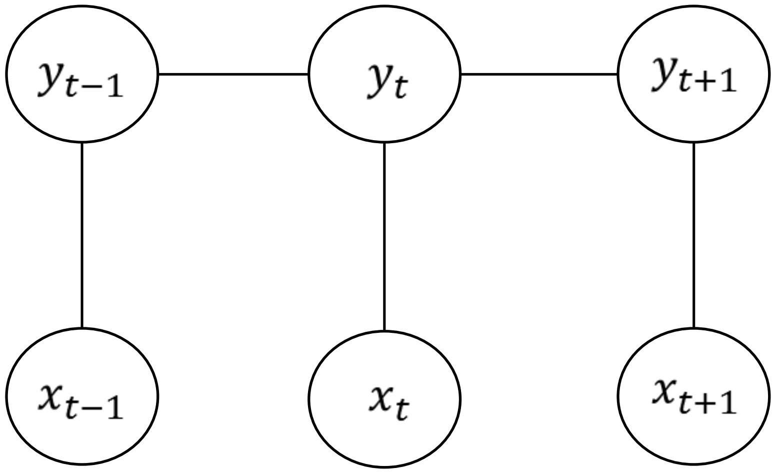


- **Bidirectional LSTM** is adopted to capture both past and future contexts for a given word.
- 18 event labels are defined according to 8 event types and 18 subtypes.
- Continuous words with the same type label are regarded as the same event mention with the specified type value.

CRFs Model

Features:

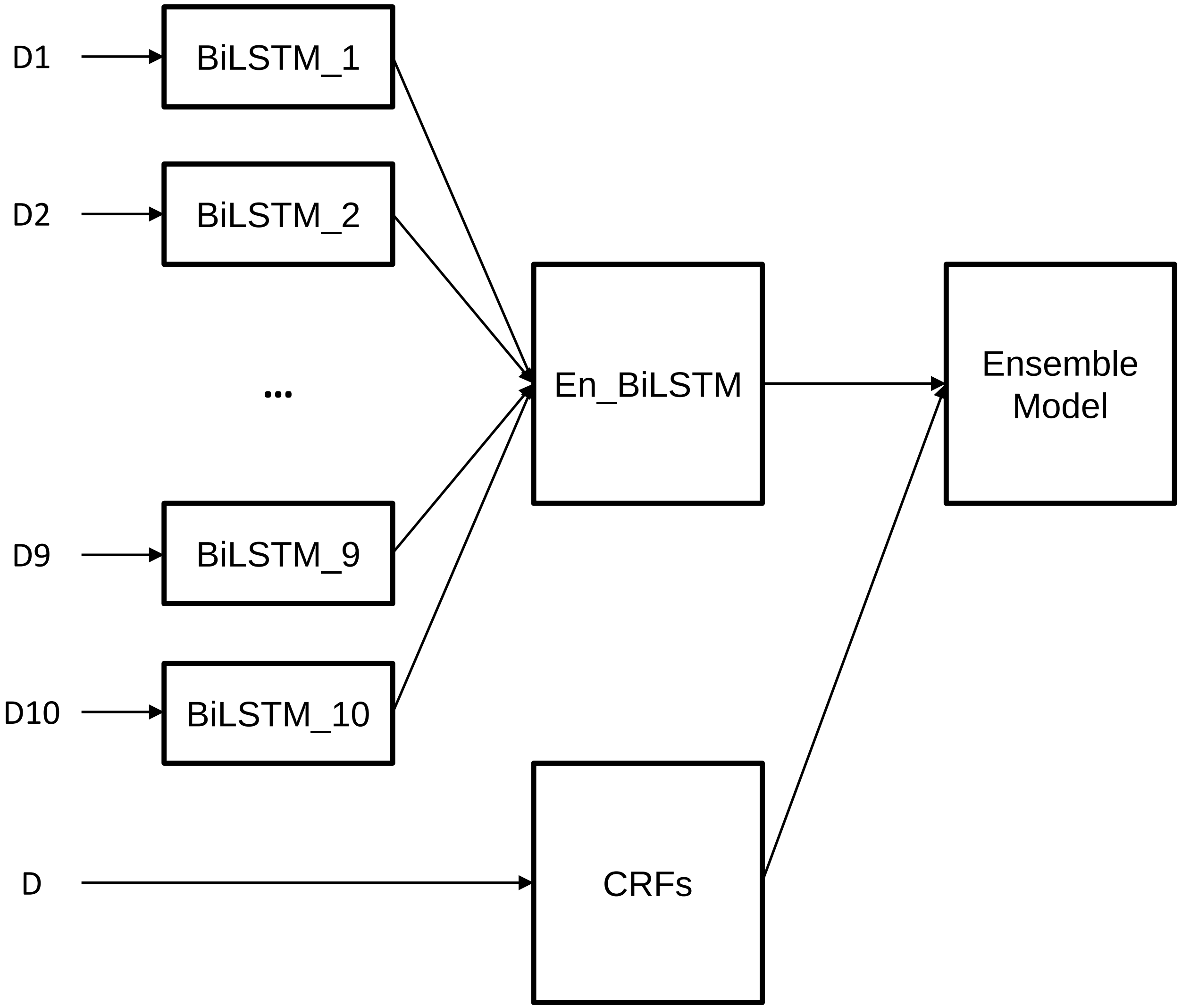
- Token
- Lemma
- Stemming
- POS tag
- Dependency type
- NER nearby
- Position in sentence
- Sentence position in document
- Trigger word dict
- WordNet



33 labels:

- 1 subtype
- 2 subtypes

Ensemble Model



The whole training dataset is denoted as D.

En_BiLSTM

- The dataset D is split into 10 parts. 10 BiLSTM models are trained separately using corresponding 9 of 10 as training data and the remaining as validation data. The training dataset used for model *i* is denoted as Di.
- For each Di, **over-sampling** technique is adopted to increase the number of event labels with fewer instances.
- **Voting strategy** is adopted to combine outputs of the 10 BiLSTM models.

Ensemble Model

- Combine: En_BiLSTM and CRFs outputs.
- Strategy: If conflict happens between two models, the results of CRFs are kept.

EN Realis Classification

To identify three REALIS values for event mentions (Actual, Generic, Other)

Contact_Contact (Actual) Life_Die (Actual) Movement_Transport-Person (Actual)

He said the killings were in self-defense and he fled the state because he could not get a fair trial.

Contact_Broadcast (Generic) Conflict_Attack (Generic) Justice_Arrest-Jail (Generic)

Leak information about US war crimes, and spend the rest of your life in a military prison.

EN Realis Classification

A **SVM** model is used to classify REALIS value for event mentions.

Features	Description
Token	Word itself
POS tag	POS tag of current word
NER nearby	NER tags for nearby words
Tense	Whether the word is ended with “ed” or not
WordNet	WordNet index

EN Coreference

To Identify the coreference links between event mention instances within a document.

Contact_Contact (Actual) Movement_Transport-Person (Other) Contact_Broadcast (Actual)

While U.S. space program is reduced to begging the Russians for rides Guest post by Investor's Business Daily Unexceptional …

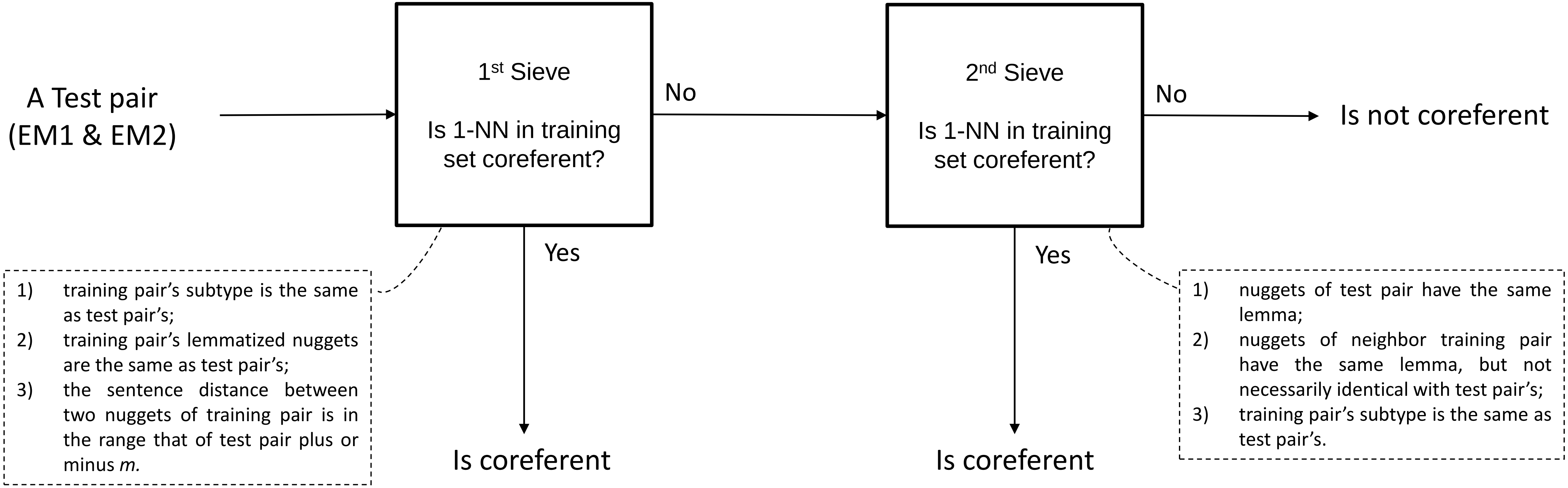
Coreference

Movement_Transport-Person (Actual)

Meanwhile, U.S. astronauts have to ride Russian spacecraft to fix toilets on the International Space Station.

EN Coreference

Classification problem: 2-pass sieve method.
The method was proved significantly efficient by UTD team @TAC2016.



Experiments

Datasets

- For development, LDC2017E02 (2014 and 2015), LDC2015E29 and LDC2015E68 are used as training data, and LDC2017E02 (2016) as testing data.
- For evaluation on 2017 datasets, LDC2016E31 and LDC2017E02 (2016) are further included as training data.

Neural network training

- Construct one vocabulary including most frequent words in documents. Words that are not in vocabulary are labeled by a special token “UNK”.
- Word embedding are pre-trained using Wikipedia English corpus.
- The training stage of each model took about 1.5 hours.

Experiments

Performance on development data

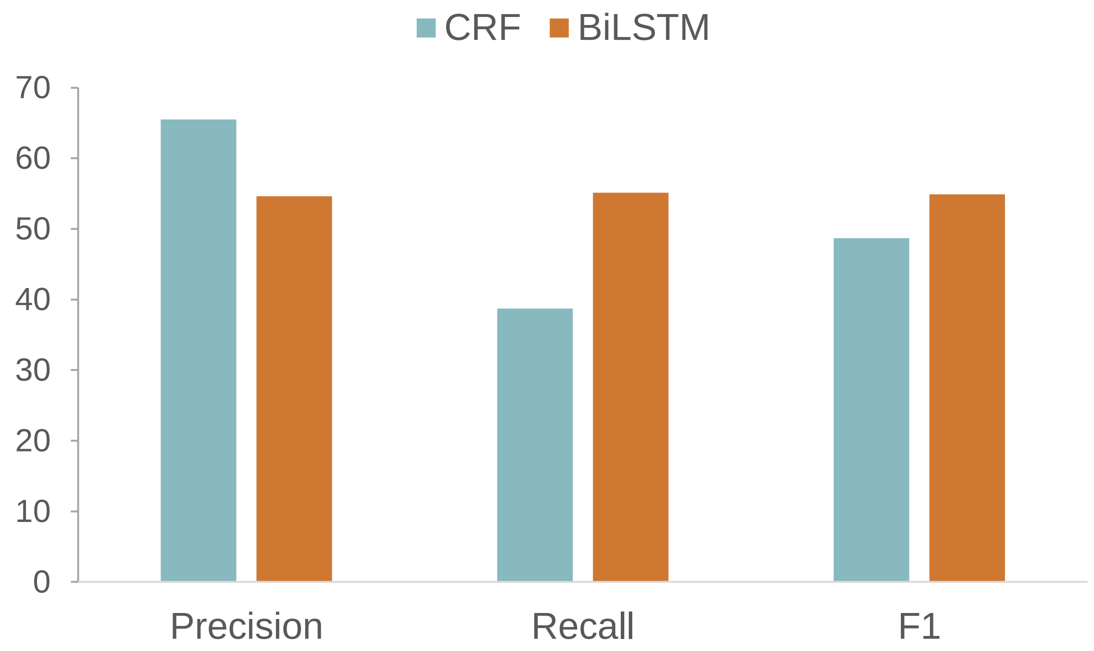
	Micro Average			Macro Average		
Attributes	Prec	Rec	F1	Prec	Rec	F1
plain	61.81	59.88	**60.83**	61.61	60.12	60.86
mention_type	51.92	51.02	**51.47**	52.12	52.10	52.11
realis_status	43.14	44.15	43.64	42.45	45.10	43.73
mention_type+realis_status	37.20	35.90	36.54	38.20	36.10	37.12
Overall Average CoNLL score	32.01					

Compared with other systems developed in 2016 TAC KBP, our model got better scores on plain and mention_type.

Experiments

On development data

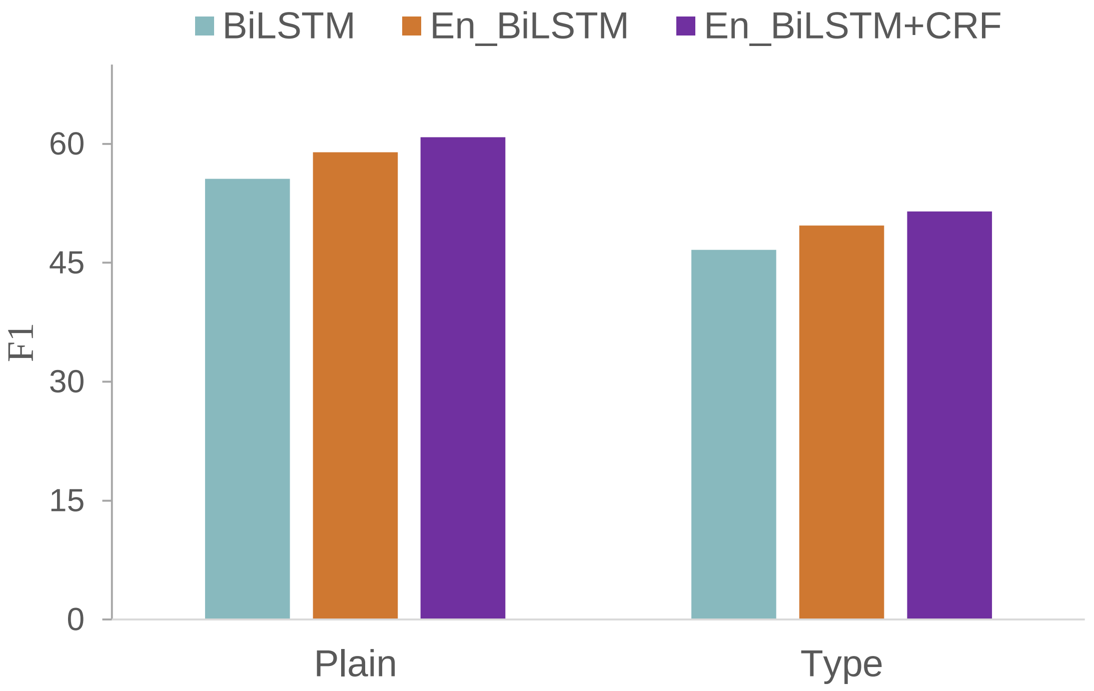
- BiLSTM model outperforms CRFs in F1 measure.
- However, CRFs can identify events in higher precision.



Experiments

On development data

- The ensemble of 10 BiLSTM models (En_BiLSTM) outperform one single BiLSTM model.
- Then ensemble of En_BiLSTM and CRF models outperform En_BiLSTM model.



Results

We submitted 3 runs to En detection and coreference.
The best performance of the 3 runs on 2017 official evaluation data are listed below.

	Micro Average			Macro Average		
Attributes	Prec	Rec	F1	Prec	Rec	F1
plain	68.04	66.53	**67.27**	68.07	68.04	68.06
mention_type	56.83	55.57	**56.19**	57.02	56.82	56.92
realis_status	47.95	46.89	**47.42**	48.77	48.73	48.75
mention_type+realis_status	39.69	38.81	**39.24**	40.47	40.17	40.32
Overall Average CoNLL score	**35.33**					

- Our ensemble model significantly outperforms other systems on EN plain and EN mention_type.
- For EN coreference, sieve-based method (srcb1) performs better than ME-based method (srcb2).

THANK YOU!