SRCB at TAC 2019 DDI Track

Data Mining Lab

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

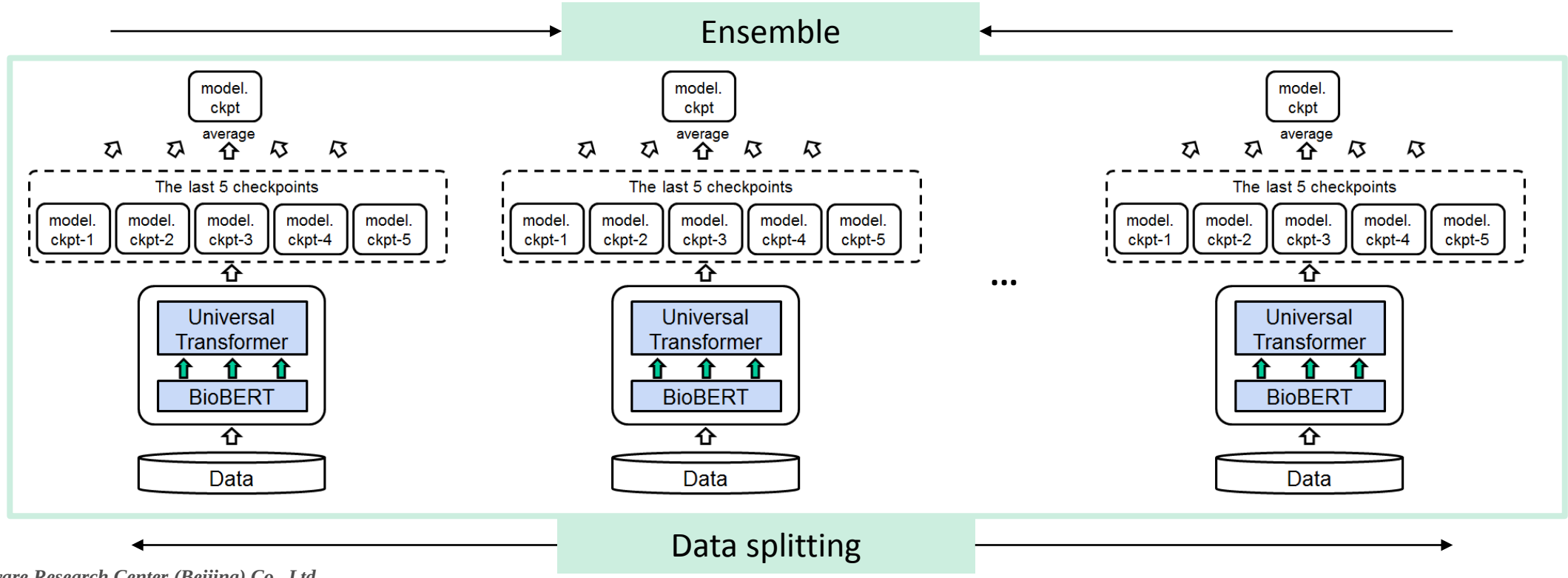
(TeamID: srcb)

Overview

- SRCB participated in three tasks in TAC 2019 DDI track
  - Task1: Entity recognition task. Extract Mentions of Interacting Drugs/Substances and specific interactions at sentence level.
    - Precipitant
    - SpecificInteraction
  - Task2: Relation identification task (sentence-level). Identify interactions at sentence level.
    - Interacting drugs
    - Interaction types: pharmacokinetic, pharmacodynamic or unspecified
    - Outcomes of pharmacokinetic and pharmacodynamic interactions.
  - Task3: Normalization task. The interacting substance should be normalized to standard codes.

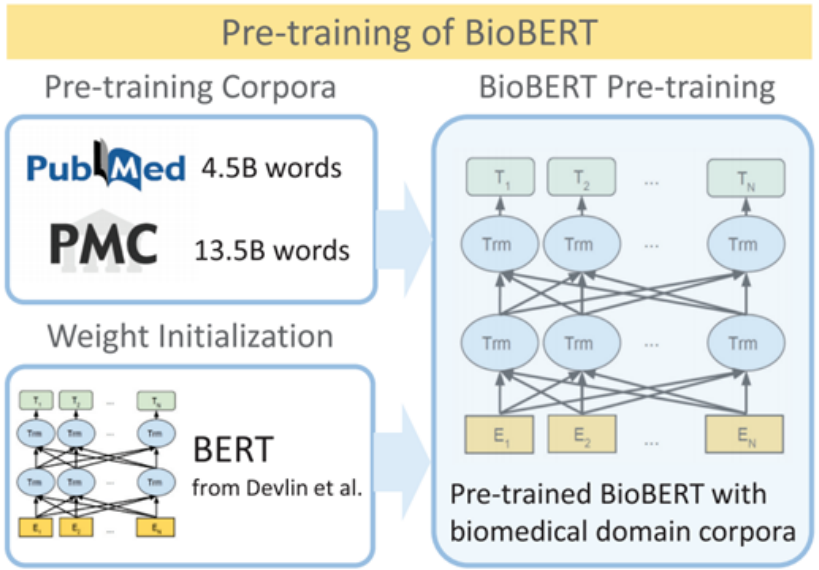
Task 1

- A BERT based NER model
  - BioBERT + Universal Transformer
  - Ensemble model based on models trained on different datasets
  - Data augmentation with unlabeled data



Task 1

- NER model
  - **BioBERT**[1] is a domain specific language representation model pre-trained on large-scale biomedical corpora.
  - **Universal Transformer**[2] layers are attached to the BioBERT model, which have the same structure with vanilla transformer, sharing the same parameters between different layers.



[1] Lee, Jinhyuk, et al. "Biobert: pre-trained biomedical language representation model for biomedical text mining." arXiv preprint arXiv:1901.08746 (2019).

[2] Dehghani, Mostafa, et al. "Universal transformers." arXiv preprint arXiv:1807.03819 (2018).

Task 1

RICOH
imagine. change.

- Data augmentation
  - Train a base model with the training data.
  - Automatically label the large-amount raw data with the base model.
  - Combine the predictions into the training data.
  - Train the target model.

- Model Ensemble
  - Train the models with different subset of the training data set (leave one-fold out).
  - Average confidents from these models for every time step.

Task 1

RICOH
imagine. change.

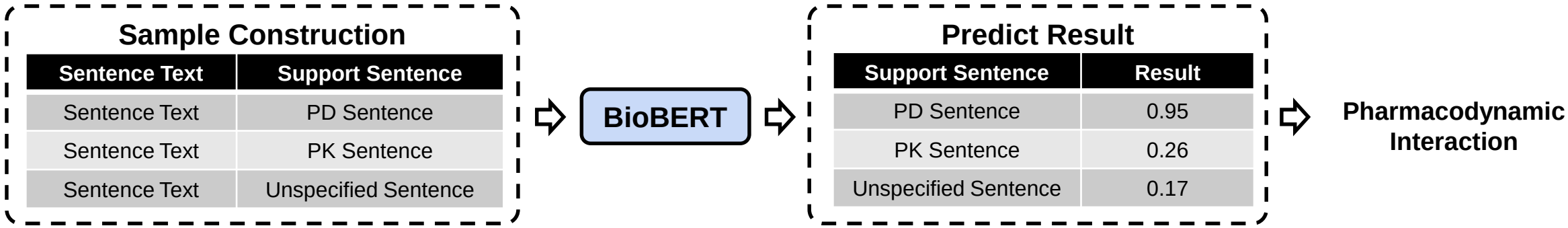
Experiment Results

- Training data
  - 22 SPLs used for training in 2018 (Training22)
  - 8,000 sentences from 180 SPLs re-annotated according to the 2018 guidelines (Additional66)
- Development data
  - Test_1: 57 SPLs used for testing in 2018
  - Test_2: 66 additional SPLs with only two sections annotated, used for testing in 2018

	Test_1			Test_2		
Technical point	P	R	F1	P	R	F1
BioBERT + Training22	30.65	41.56	35.28	32.45	40.39	35.99
+ Data Augmentation (DailyMed)	45.98	40.52	43.08	47.07	42.31	44.56
+ Additional66	61.75	64.16	62.93	68.01	65.02	66.48
+ Average Checkpoint	60.98	66.30	63.53	67.30	67.66	67.48
+ Universal Transformer	66.93	68.27	67.59	70.95	68.39	69.65
+ Ensemble Model	70.77	65.07	67.80	77.53	68.88	72.95

Task 2

- Relation Identification through **sentence pair classification**
  - Sample construction with support sentence
  - Sentence pair classification based on BioBERT
  - Data augmentation with NLM-180 data



Task 2

- Sample construction
  - Replace the Precipitant with unified word and surround the SpecificInteraction with curly braces.
  - Build Support Sentence
    - PD: “PRECIPITANT” + {SpecificInteraction}
    - PK: “PRECIPITANT” + keywords (increase, decrease, reduce, half time,…)
    - Unspecified: “PRECIPITANT” + “Unspecified”

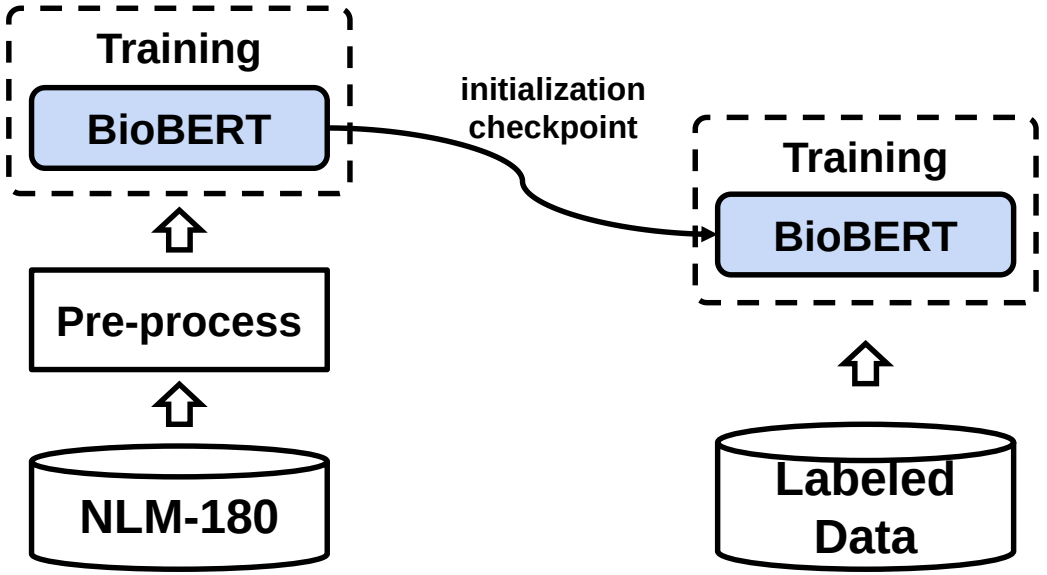
- **SentenceText**: Concomitant use of TARKA with other antihypertensive agents including diuretics, vasodilators, beta-adrenergic blockers, and alpha-antagonists may result in additive hypotensive effects.
- **Precipitant**: antihypertensive agents
- **SpecificInteraction**: additive hypotensive effects

⇨

Sentence Text	Support Sentence	
Concomitant use of TARKA with other PRECIPITANT including diuretics, vasodilators, beta-adrenergic blockers, and alpha-antagonists may result in {additive hypotensive effects}.	PRECIPITANT {additive hypotensive effects}	PD
	PRECIPITANT increase decrease reduce half time tmax ……	PK
	PRECIPITANT Unspecified	Unspecified

Task 2

- Model Training
  - A sentence pair classification model based on BioBERT is trained to judge whether the input sentence match the information in the support sentence or not.
  - Data augmentation with NLM-180
    - Field mapping
    - Predict the subtypes if PK interactions based on the model trained on labeled data



Task 2

RICOH
imagine. change.

Experiment Results

- Training data
  - 22 SPLs used for training in 2018
  - 8,000 sentences from 180 SPLs re-annotated according to the 2018 guidelines
- Development data
  - Test_1: 57 SPLs used for testing in 2018
  - Test_2: 66 additional SPLs with only two sections annotated, used for testing in 2018

	Test_1 (Task1 F1=67.80)			Test_2 (Task1 F1=72.95)		
Technical point	P	R	F1	P	R	F1
BioBERT (Base)	53.89	45.57	49.38	55.21	46.69	50.59
+ Support sentence	57.11	50.67	53.70	57.80	51.37	54.39
+ Data augmentation (NLM-180)	58.40	51.74	54.87	58.50	52.02	55.07

Task 3

- Code mapping through **retrieval** and **re-rank**
  - 30 candidates (Apache Solr)
  - Features
    - Query-document (relevancy)
    - Document-document (importance)
  - Similarity
    - Jaccard distance, Levenshtein distance
    - Longest Common Subsequence (LCS)
  - Ranking: empirical (linear) weights and Learning-to-rank (rankSVM)
  - Maximum span of the possible mention
  - Non-result threshold
  - Golden mapping dictionary

Task 3

RICOH
imagine. change.

Experiment Results

- Development data
  - 8,000 sentences from 180 SPLs re-annotated according to the 2018 guidelines (ADD-66)

Relaxed Match results on ADD-66 (**golden mentions**)

	Macro-F1
Retrieval (top1)	67.58
Retrieval (top30) + features (empirical)	80.87
Retrieval (top30) + features (learning to rank)	76.44
Retrieval (top30) + features (empirical) + non-result threshold	81.90
Retrieval (top30) + features (empirical) + non-result threshold + gold_dict	83.12

Relaxed Match results on ADD-66 (**extracted mentions**)

Task1 F1=58.80	Macro-F1
Retrieval (top1)	53.78
Retrieval (top30) + features (empirical)	59.13
Retrieval (top30) + features (learning to rank)	55.17
Retrieval (top30) + features (empirical) + span extension	60.93
Retrieval (top30) + features (empirical) + span extension + non-result threshold	62.05
Retrieval (top30) + features (empirical) + span extension + non-result threshold + gold_dict	64.03

Submissions

Runs	**Entity Recognition (Task1)**	**Relation Identification (Task2)**	**Normalization (Task3)**
srcb_1	BioBERT + universal transformer + data augmentation + model ensemble + average checkpoint	BioBERT + support sentence	Retrieval (top30) + features (empirical) + span extension + non-result threshold + gold_dict
srcb_2	BioBERT + universal transformer + data augmentation + model ensemble	BioBERT + support sentence + data augmentation	Retrieval (top30) + features (empirical) + span extension + non-result threshold + gold_dict
srcb_3	srcb_1 without the mentions have no interaction in Task2	BioBERT + support sentence	Retrieval (top30) + features (empirical) + span extension + non-result threshold + gold_dict

Submissions

RICOH
imagine. change.

Results

Entity Recognition (Task1)			
Run	Precision	Recall	F1-score
srcb_1	70.9276	56.5161	62.9070
srcb_2	71.3284	55.8109	62.6227
srcb_3	72.4608	55.5236	62.8715
Relation Identification (Task2)			
Run	Precision	Recall	F1-score
srcb_1	53.8435	41.3241	46.7603
srcb_2	54.7015	40.8436	46.7675
srcb_3	53.8435	41.3241	46.7603
Normalization (Task3)			
Run	Precision	Recall	F1-score
srcb_1	67.5508	59.3683	61.4320
srcb_2	65.7809	56.4855	59.4293
srcb_3	70.8757	58.4930	**62.3889**

RICOH
imagine. change.

THANK YOU!