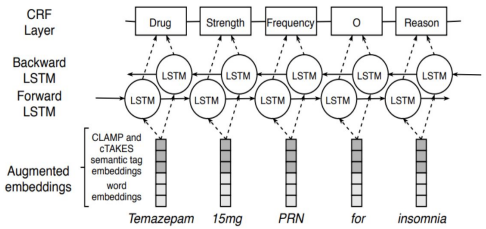
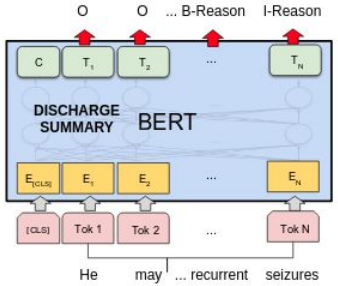
N2C2 2018 Track 2: Named Entity Extraction

Soujanya Ranganatha Bhat, Varun Chaudhary

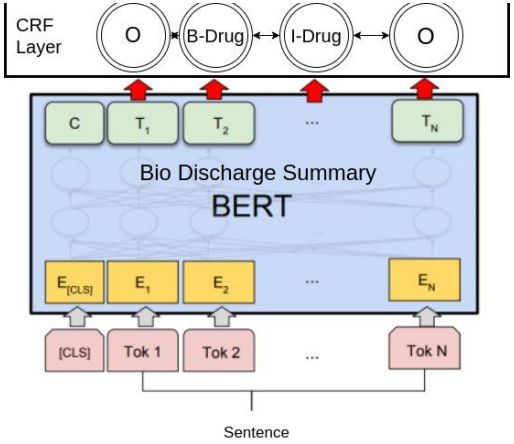
- The goal of this task is to extract concepts such as Drug, Reason and Adverse Drug Event(ADE) from patient discharge summaries.
- The previous methods used models based on BiLSTM-CRF for NER with feature augmentation using cTAKES and CLAMP pipeline.
The previous SOTA achieved class-wise scores of Drug(0.96), Reason(0.72) and ADE(0.58).
- We developed Bio+DischargeSummaryBERT(BDS) models- BDSBERT(our midterm model) and BDSBERT-CRF(our final model).
- Our model achieves better lenient F1 scores than SOTA(minutely) in case of ADE(**+0.05**) entities, and same for Drug and Reason concepts.



Existing Method



Our Midterm Method



Our Final Method

Type	Training	Validation	Test
Drug	14303	2045	10569
Reason	3258	445	2519
ADE	813	103	604
Total	18374	2593	13692

Preprocessed(filtered) Data Entity Count

Dataset usage(with data folders provided):
- train(track2-training data 2)
- validation(track2-training data 3)
- test(gold-standard-test-data)

Training/validation/test **split**- 53:7:40

Pre-processing- Efficient tag overlap filtering + Experimentally determined sentence length

Model is trained to maximise log-probability of correct tag sequence(obtained by Viterbi decoding)

Observation 1(see “Results on Test set”): Model performs significantly better for *Reason* and *ADE* concepts, compared to BERT baseline.
Observation 2(see “Confusion Matrix”): Model identifies more Reason and ADEs, but still is unable to detect a lot of ADE entities.
Conclusion: CRF’s better structure predictions via the state transition matrix in contrast to a simple softmax layer improve Reason and ADE performance significantly. Still, low Reason and especially ADE recall can be improved further possibly through feature augmentation.

.

	P	R	$F1_l$	$F1_s$
Drug	0.95	0.96	0.96	0.92
Reason	0.76	0.71	0.73	0.60
ADE	0.73	0.55	0.63	0.50

Results on Test set

Models	$Drug_{F1_l}$	$Reason_{F1_l}$	ADE_{F1_l}
BDSBERT(midterm)	0.68	0.21	0.06
BDSBERT(baseline)	0.92	0.57	0.26
BiLSTM-CRF(SOTA)	**0.96**	0.72	0.58
BDSBERT-CRF(ours)	**0.96**	**0.73**	**0.63**

Model Comparison

	Drug	Reason	ADE	O
Drug	**10153**	4	1	411
Reason	13	**1786**	23	697
ADE	2	40	**332**	230

Confusion Matrix

Type	Score
Micro $F1_l$	0.90
Macro $F1_l$	0.77
Weighted $F1_l$	0.90
Micro $F1_s$	0.85
Macro $F1_s$	0.68
Weighted $F1_s$	0.85

Overall F1 scores

Analysis	Category	Examples	Actual	Predicted	Explanation
Improvement	Better learnt Reason	recurrent disease	Reason	Reason(now) Unknown(earlier)	Reason are identified more often (even with only 18% entities as Reasons)
	Better learnt ADE	intractable diarrhea	ADE	ADE(now) Unknown(earlier)	ADEs are identified more often (even with only 4% entities as ADEs)
	Breaks in terms(hyphens, slashes)	beta-lactamase inhibitor combinations	Drug	Drug(now) Unknwon till hypen, then Drug(earlier)	Model now predicts hyphenated words much better
Error	Tag Overlap	vincristine vincristine toxic polyneuropathy	Drug ADE	Drug Drug	Same phrase consists of two tags, where the model identifies only one
	Abbreviations	SVC syndrome	Reason	Unknown	Superior vena cava (SVC) syndrome and other abbreviations is not identified
	Ambiguity	allergic reaction	Reason	ADE	Ambiguous use of Reasons and ADEs, without a distinct difference
	Complicated terms	mediastinal lymph nodes measuring up to 9 mm diffuse erythema/ulceration, in esophagus, stomach, dupdenum	Reason Reason	Unknown Unknown	Entity spans over multiple complicated words

Analysis (Improvements and Errors)