Effective Feature Representation for Clinical Text Concept Extraction

Yifeng Tao[1,2], Bruno Godefroy[1], Guillaume Genthial[1], Christopher Potts[1,3,*]

[1]Roam Analytics

[2]Carnegie Mellon University

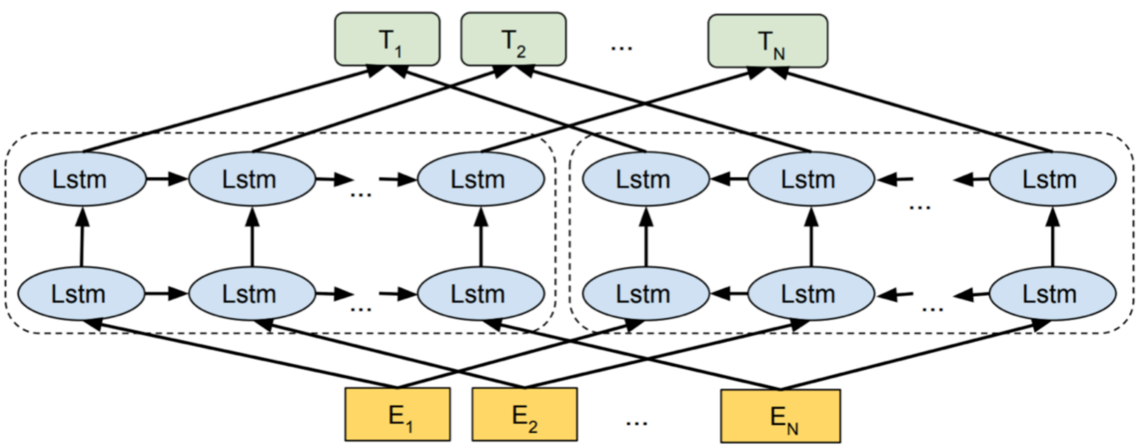
[3]Stanford University

Background

- Crucial information of healthcare is recorded only in free-form text
- Annotated clinical text datasets are scarce and expensive
  - privacy considerations
  - domain specialists
- Clinical text are different from general text

- SOTA methods:

ELMo



[Slide from https://arxiv.org/abs/1810.04805.]

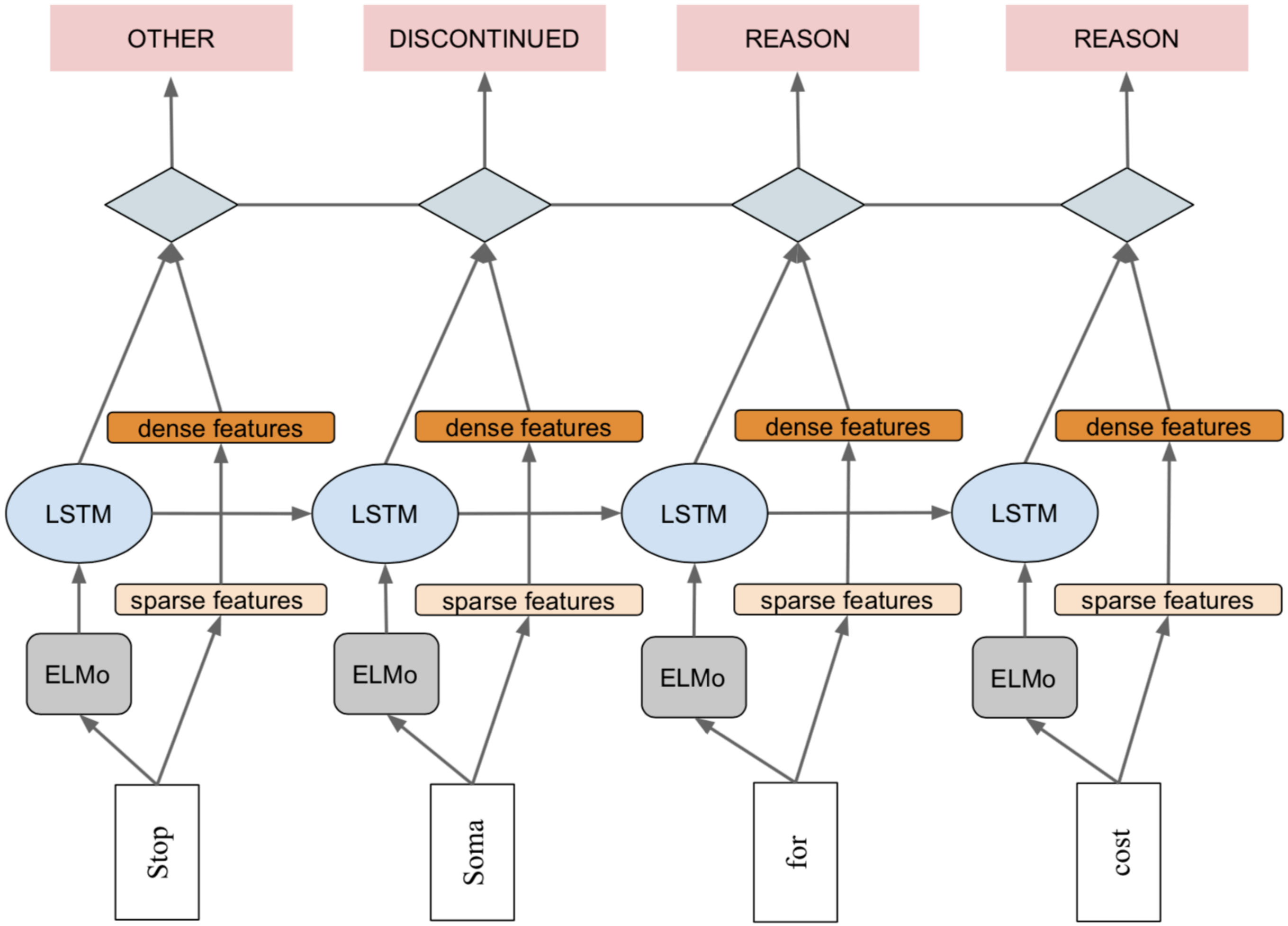
Task: Clinical Text Annotation

- Five clinical text datasets

Dataset	Example
Diagnosis Detection	Asymptomatic/POSITIVE bacteriuria/POSITIVE , could be neurogenic/CONCERN bladder/CONCERN disorder/CONCERN .
Prescription Reasons	I will go ahead and place him on Clarinex/PRESCRIBED for/REASON his/REASON seasonal/REASON allergic/REASON rhinitis/REASON .
Penn Adverse Drug Reactions (ADR)	#TwoThingsThatDontMixWell venlafaxine and alcohol- you'll cry/ADR and throw/ADR chairs/ADR at your mom's BBQ.
Chemical–Disease Relations (CDR)	Ocular/DISEASE and/DISEASE auditory/DISEASE toxicity/DISEASE in hemodialyzed patients receiving desferrioxamine/DRUG .
Drug–Disease Relations	Indicated for the management of active/TREATS rheumatoid/TREATS arthritis/TREATS and should not be used for rheumatoid/CONTRA arthritis/CONTRA in/CONTRA pregnant/CONTRA women/CONTRA .

Methods

- Combine the hand-built features and ELMo embeddings

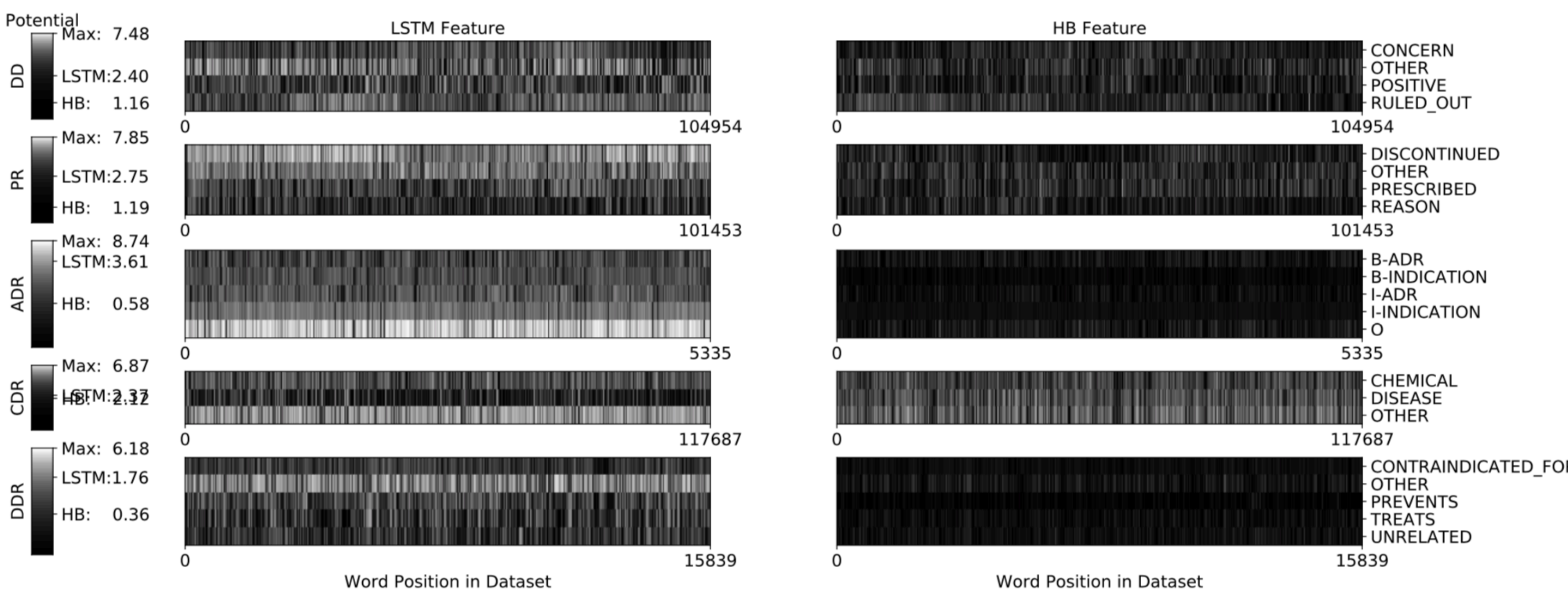


Results

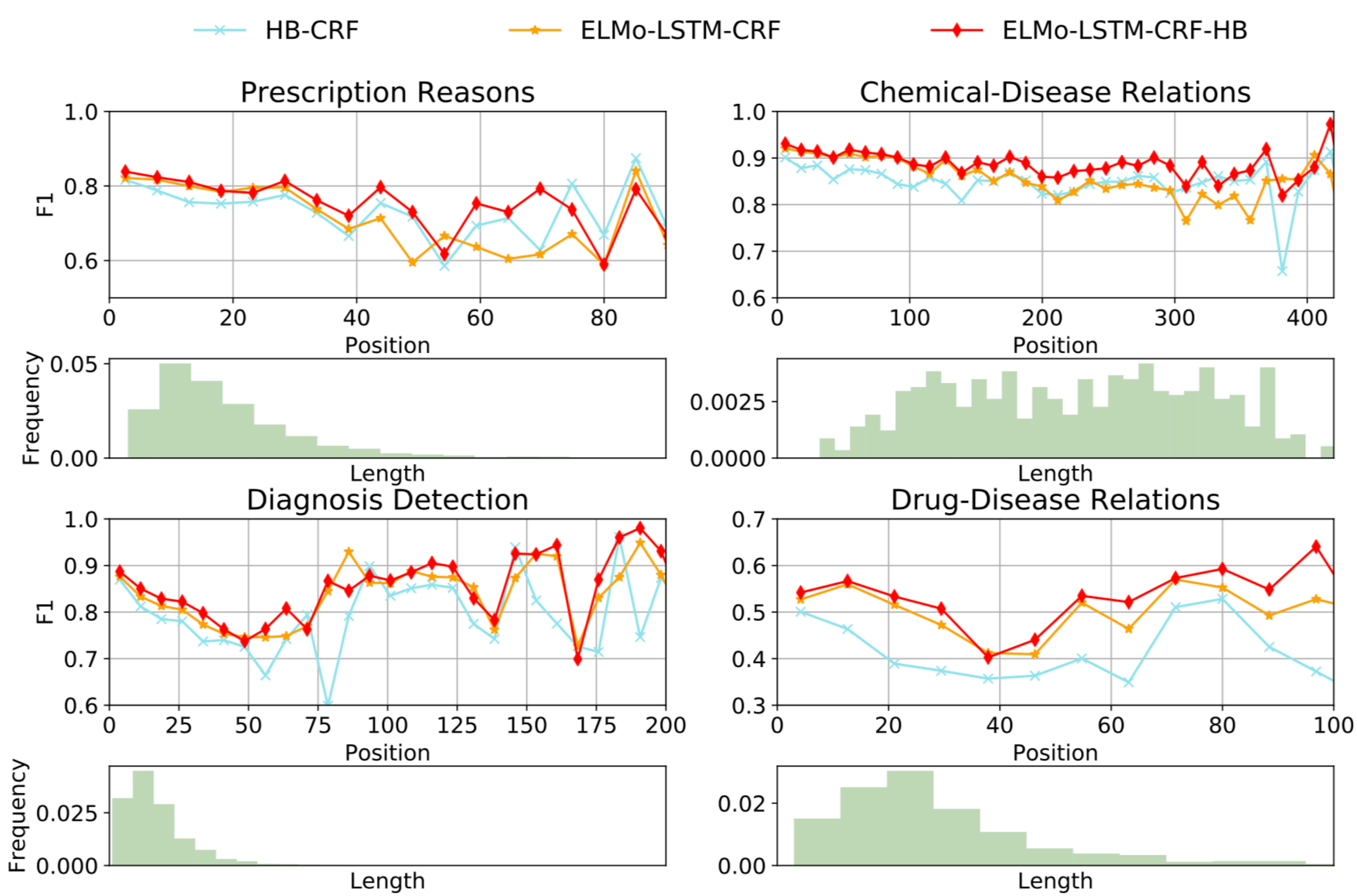
- Evaluated on four different models

	Diagnosis Detection	Prescription Reasons	Penn Adverse Drug Reactions (ADR)	Chemical–Disease Relations (CDR)	Drug–Disease Relations
rand-LSTM-CRF	77.3 ± 0.05	69.6 ± 0.25	53.8 ± 0.88	85.1 ± 0.10	48.2 ± 1.12
HB-CRF	82.0 ± 0.05	78.5 ± 0.01	58.8 ± 0.12	86.2 ± 0.02	42.3 ± 0.30
ELMo-LSTM-CRF	83.9 ± 0.35	81.0 ± 0.20	65.7 ± 0.35	88.2 ± 0.34	50.6 ± 0.64
ELMo-LSTM-CRF-HB	$\mathbf{85.3} \pm 0.24^{***}$	$\mathbf{82.0} \pm 0.03^{***}$	$\mathbf{68.5} \pm 1.67^{*}$	$\mathbf{89.9} \pm 0.12^{***}$	$\mathbf{51.9} \pm 0.52^{**}$

- Trade-off of two base models:



The Role of Text Length



Major Improvements in Minor Categories

Diagnosis Detection			
Label	Support	F1 score	Improvement
OTHER	74888	95.3	1.4%
POSITIVE	24489	86.1	4.4%
RULED-OUT	2797	86.4	3.6%
CONCERN	2780	72.1	5.6%

Prescription Reasons			
Label	Support	F1 score	Improvement
OTHER	83618	95.8	0.9%
REASON	9114	64.7	8.6%
PRESCRIBED	5967	84.7	4.4%
DISCONTINUED	2754	82.7	5.6%

Chemical–Disease Relations (CDR)			
Label	Support	F1 score	Improvement
OTHER	104530	98.3	0.5%
DISEASE	6887	84.2	6.3%
CHEMICAL	6270	87.0	6.7%

Drug–Disease Relations			
Label	Support	F1 score	Improvement
OTHER	10634	90.8	2.3%
TREATS	3671	76.0	5.7%
UNRELATED	1145	53.8	71.3%
PREVENTS	320	41.1	103.5%
CONTRAINDICATED-FOR	69	0	–

References

- Jacob Devlin et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding: https://arxiv.org/abs/1810.04805