Drug Drug Interaction

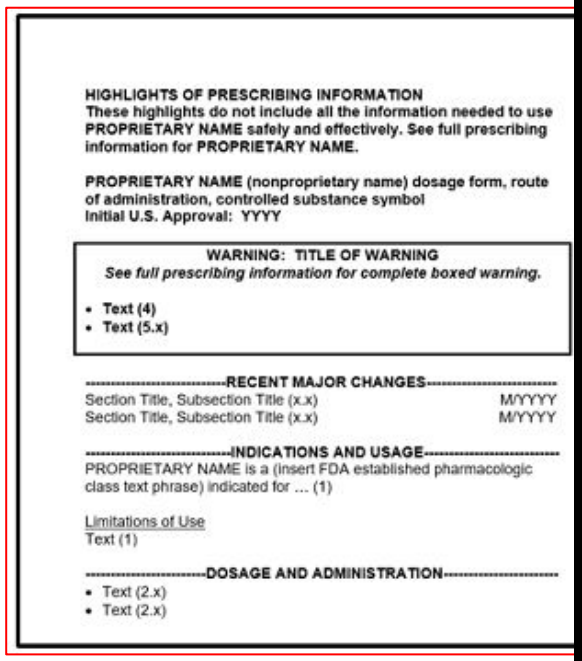
藥物交互作用

AIA-st3 Team 11
ET083010 林珮珺
ET083025 曾奕盛
ET083037 蘇芳毅

Outline

- Motivation & Goal
- Data resource & Model mechanism
- Code review
- Presentation
- Encountered problems & further workings

Motivation



HIGHLIGHTS OF PRESCRIBING INFORMATION
These highlights do not include all the information needed to use PROPRIETARY NAME safely and effectively. See full prescribing information for PROPRIETARY NAME.

PROPRIETARY NAME (nonproprietary name) dosage form, route of administration, controlled substance symbol
Initial U.S. Approval: YYYY

WARNING: TITLE OF WARNING
See full prescribing information for complete boxed warning.

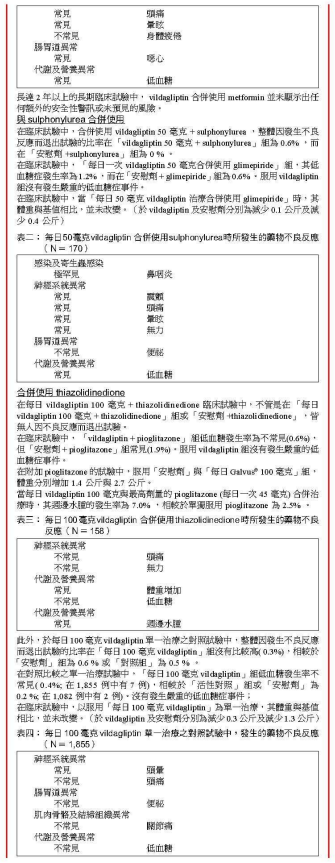
- Text (4)
- Text (5.x)

----------RECENT MAJOR CHANGES----------
Section Title, Subsection Title (x.x) M/YYYY
Section Title, Subsection Title (x.x) M/YYYY

----------INDICATIONS AND USAGE----------
PROPRIETARY NAME is a (insert FDA established pharmacologic class text phrase) indicated for … (1)

Limitations of Use
Text (1)

----------DOSAGE AND ADMINISTRATION----------
- Text (2.x)
- Text (2.x)



Goals

- Population : Elders who take multiple medicines
- Goal : Interactive program which captures drug information easily
- GUI : Chatbot

Data Resource

Journal of Biomedical Informatics 46 (2013) 914–920

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

The DDI corpus: An annotated corpus with pharmacological substances and drug–drug interactions

María Herrero-Zazo [a,*], Isabel Segura-Bedmar [a], Paloma Martínez [a], Thierry Declerck [b]

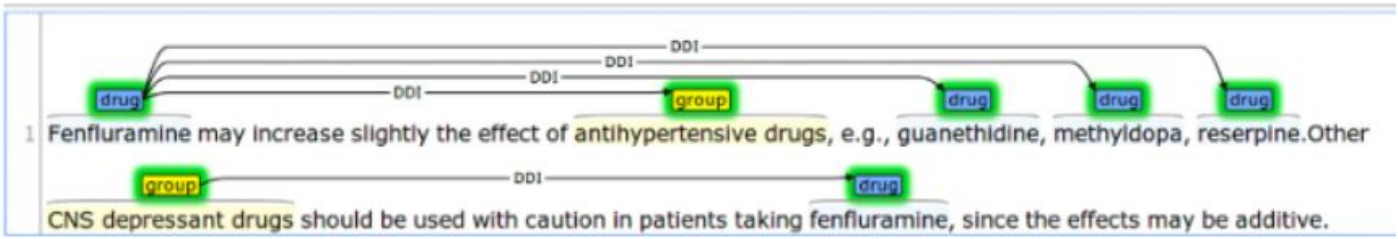
[a] Computer Science Department, Universidad Carlos III de Madrid, Leganés 28911, Madrid, Spain
[b] Language Technology Lab, DFKI GmbH, Saarbrücken D-66123, Germany

Data Resource

- Corpus resource:
  - DDI-DrugBank: 6795 sentences
  - DDI-MedLine: 2147 sentences

- 5 different types:

Interaction	Count
Negative	11036
Mechanism	1561
Effect	1949
Advice	1035
Interaction	280



Data Resource

- Ours (entity tagged)

  Train dataset : 12841 sentences

  □ train : valid = 9600 : 3241

  Test dataset : 3020 sentences

negative	The effects of <e1> DCG-IV </e1> and <e2> L-CCG-1 </e2> upon phencyclidine (PCP)-induced locomotion and behavioral changes in mice.
negative	The effects of <e1> DCG-IV </e1> and L-CCG-1 upon <e2> phencyclidine </e2> (PCP)-induced locomotion and behavioral changes in mice.
negative	The effects of <e1> DCG-IV </e1> and L-CCG-1 upon phencyclidine (<e2> PCP </e2>)-induced locomotion and behavioral changes in mice.
negative	The effects of DCG-IV and <e1> L-CCG-1 </e1> upon <e2> phencyclidine </e2> (PCP)-induced locomotion and behavioral changes in mice.
negative	The effects of DCG-IV and <e1> L-CCG-1 </e1> upon phencyclidine (<e2> PCP </e2>)-induced locomotion and behavioral changes in mice.

Model

BERT Family

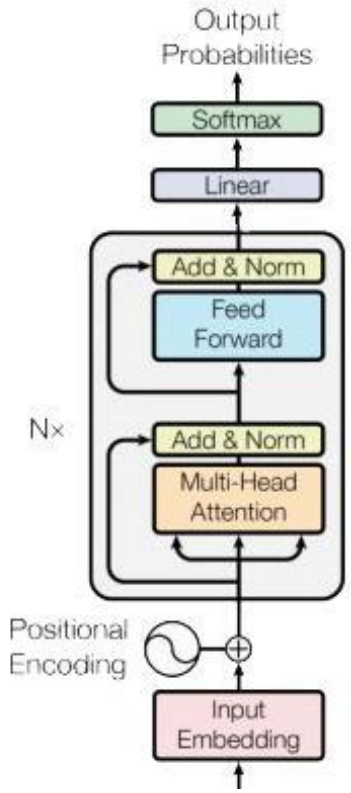
- For pretrained model :
  BioBERT (Lee et al., 2019)

- For relation extraction :
  R-BERT (Wu et al., 2019)

BERT

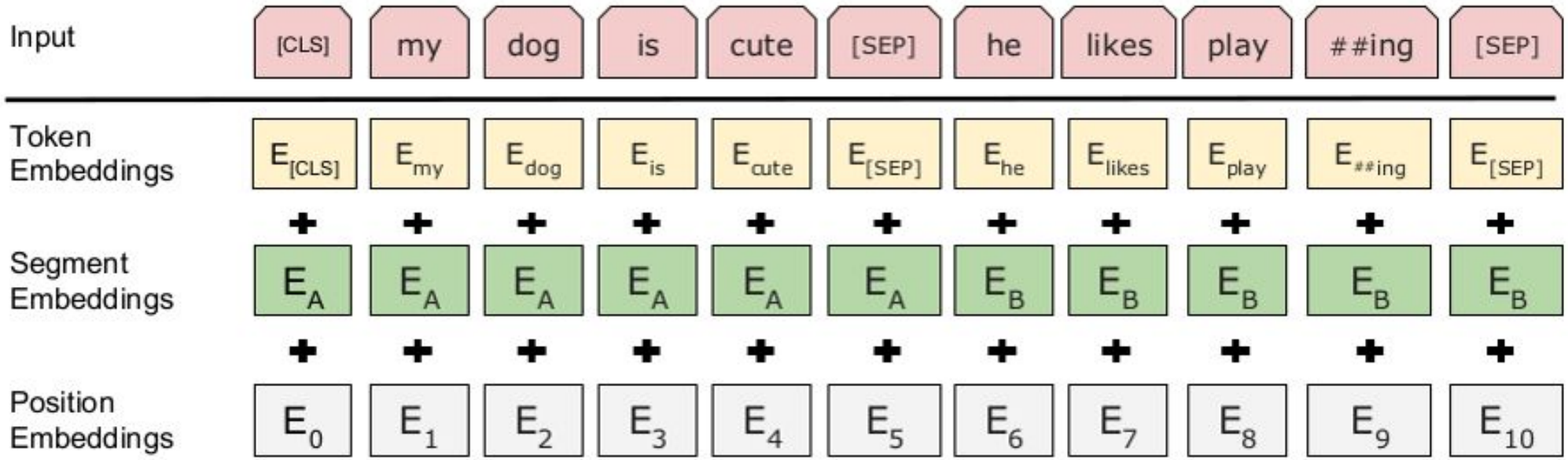
Bidirectional Encoder Representations from Transformers

- Encoder of transformer
- Pretraining & Finetuning
- Learned from a large amount of text without annotation



BERT – Input Embeddings

- Word-piece tokenization (Wu et al. 2016)
- Token / Segment / Position embeddings

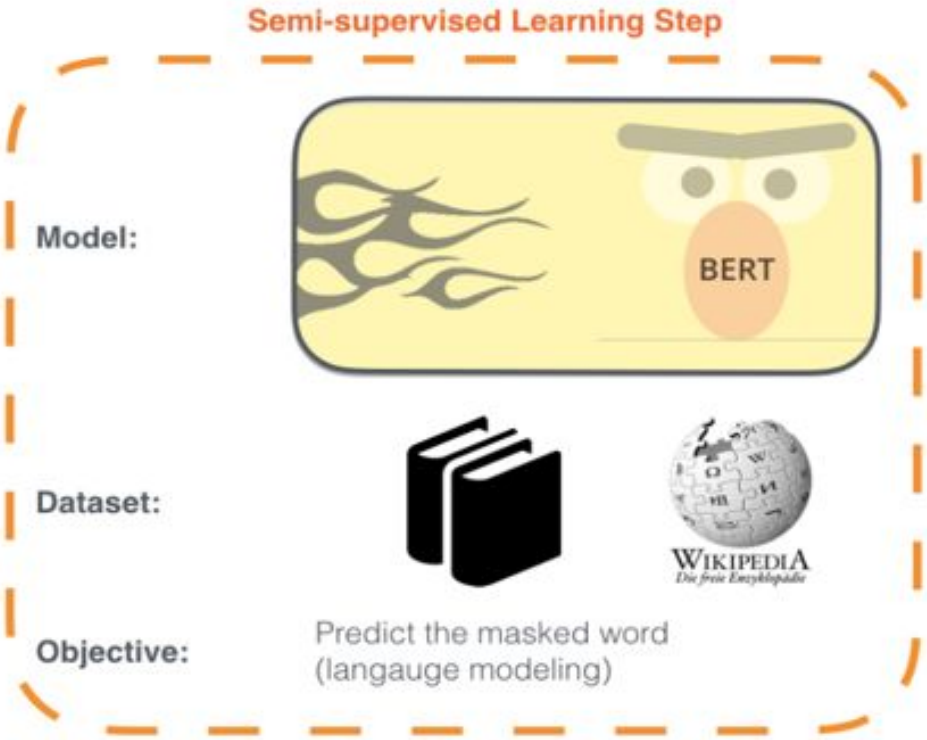


BERT – 2 Steps Implementation

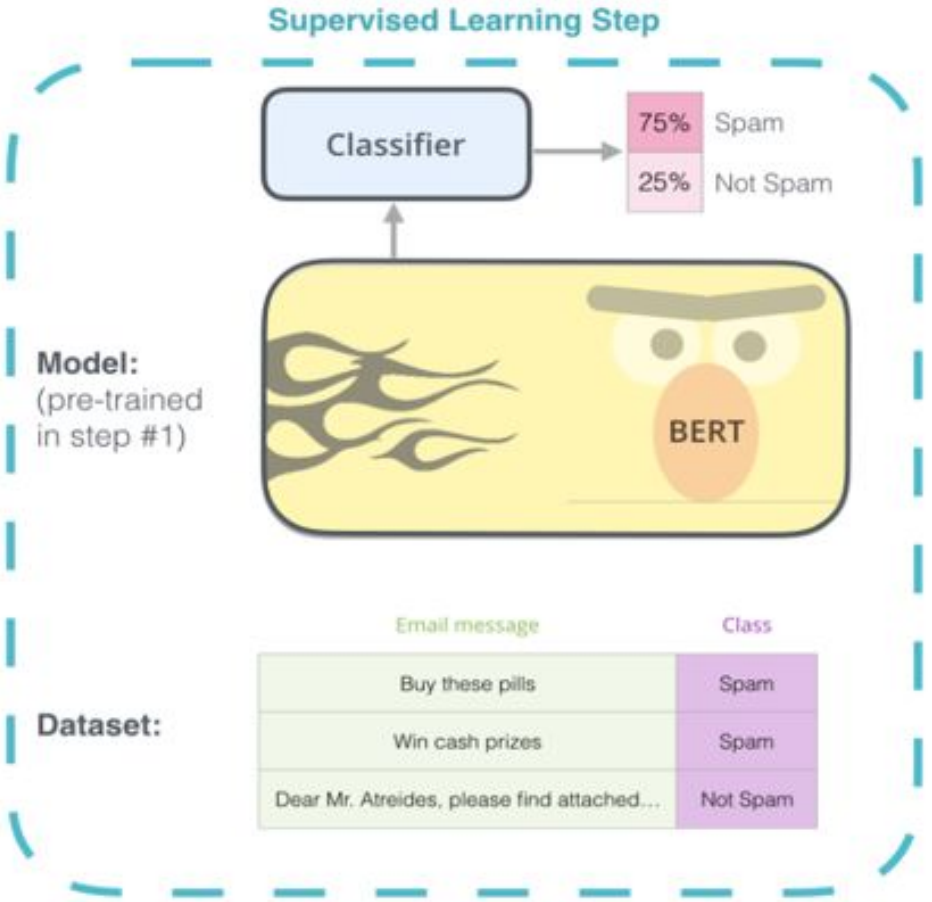
Reference: jalammar

1 - Semi-supervised training on large amounts of text (books, wikipedia..etc).

The model is trained on a certain task that enables it to grasp patterns in language. By the end of the training process, BERT has language-processing abilities capable of empowering many models we later need to build and train in a supervised way.

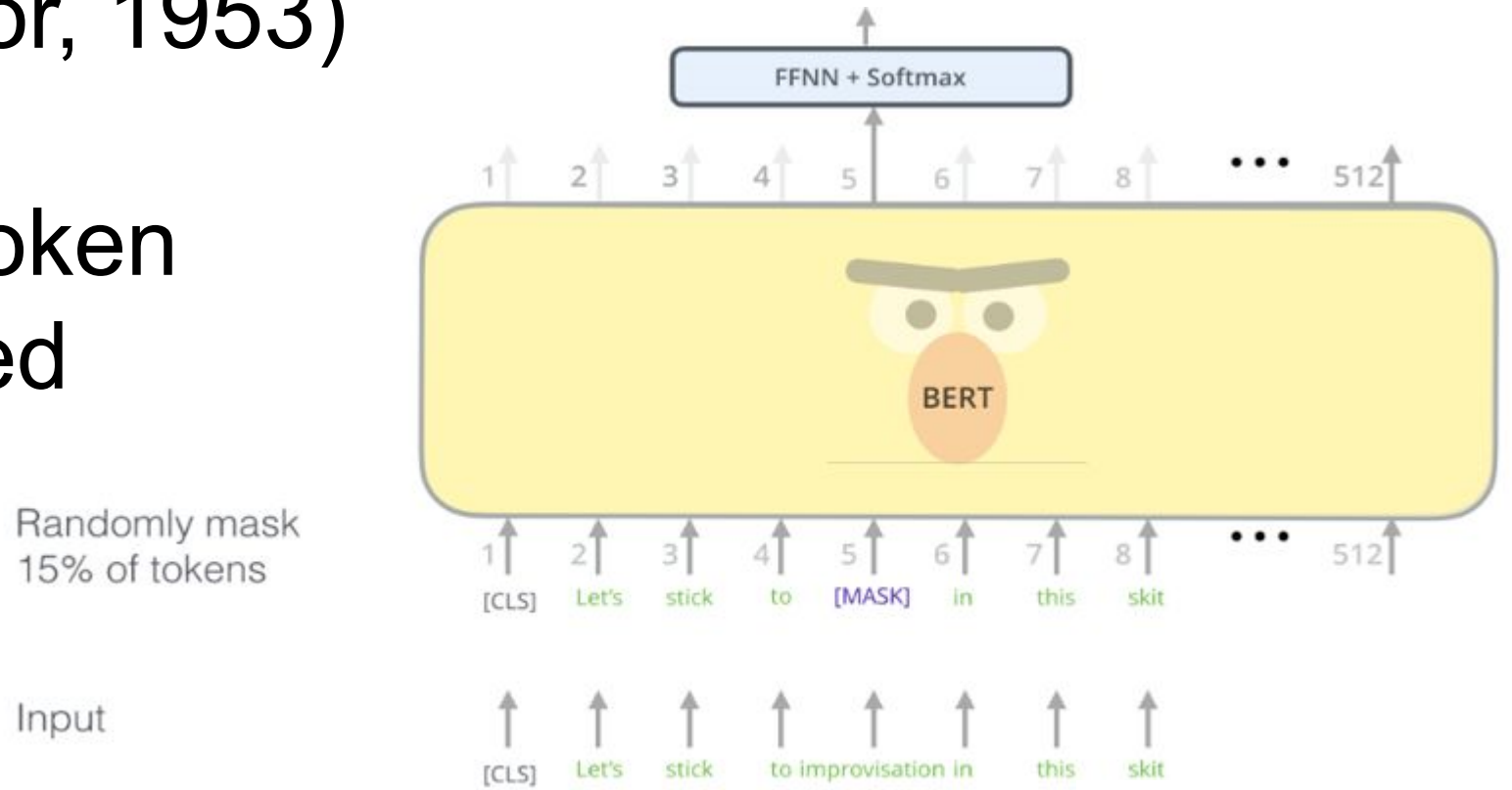


2 - Supervised training on a specific task with a labeled dataset.



BERT – Pretraining

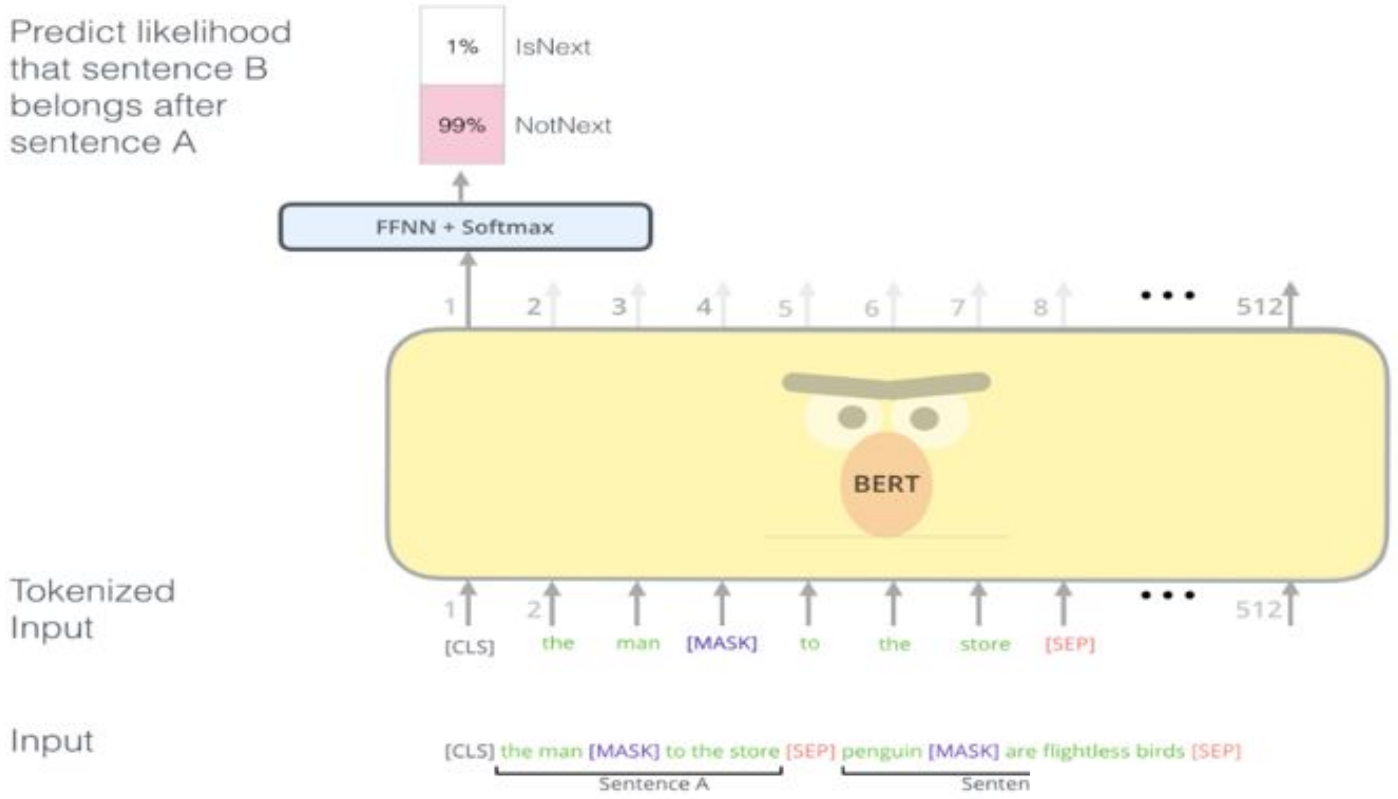
- Approach 1 : Masked Language Model (MLM)
- Cloze task (Taylor, 1953)
- 80 % - [MASK]
  10 % - random token
  10 % - unchanged



Reference: jalammar

BERT – Pretraining

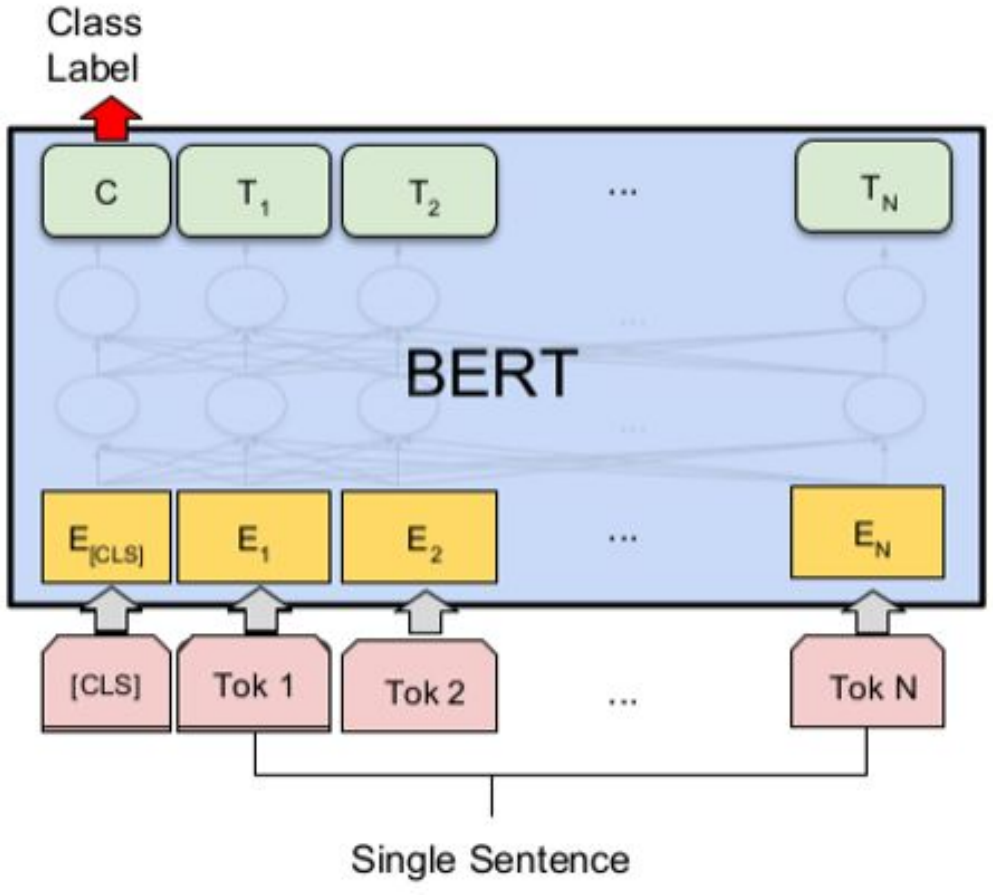
- Approach 2 : Next Sentence Prediction(NSP)
- For QA, NLI
- Document level



Reference: jalammar

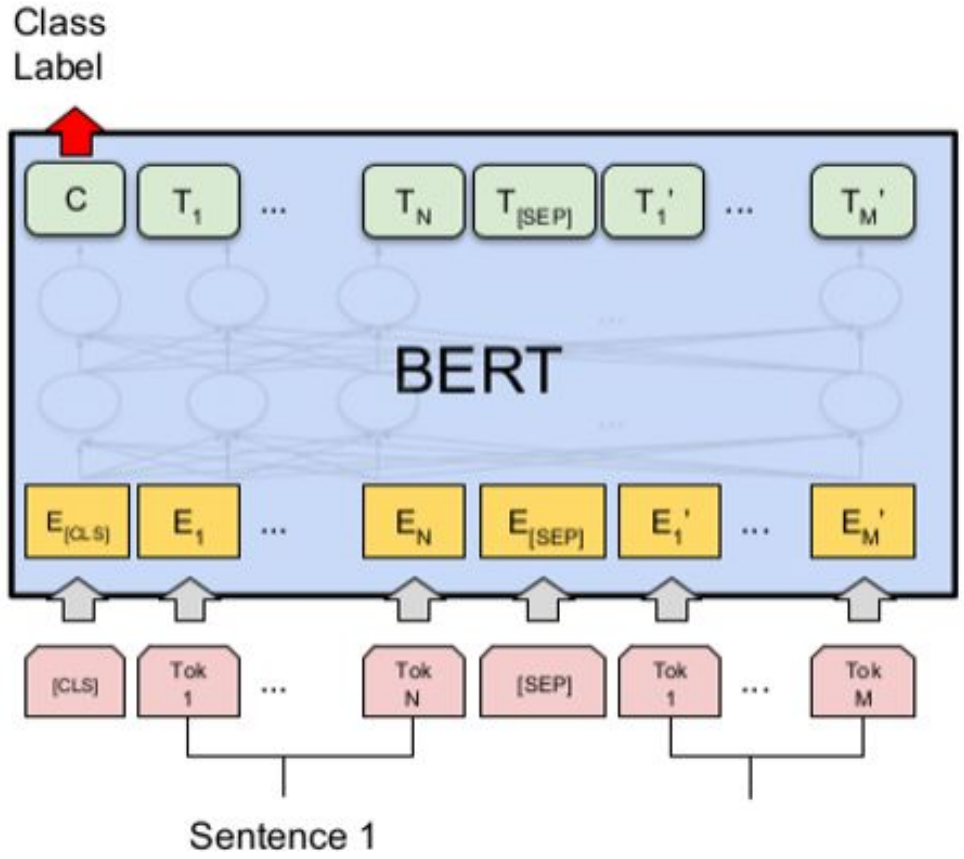
BERT – Finetuning

- Single sentence classification
- SST-2, COLA



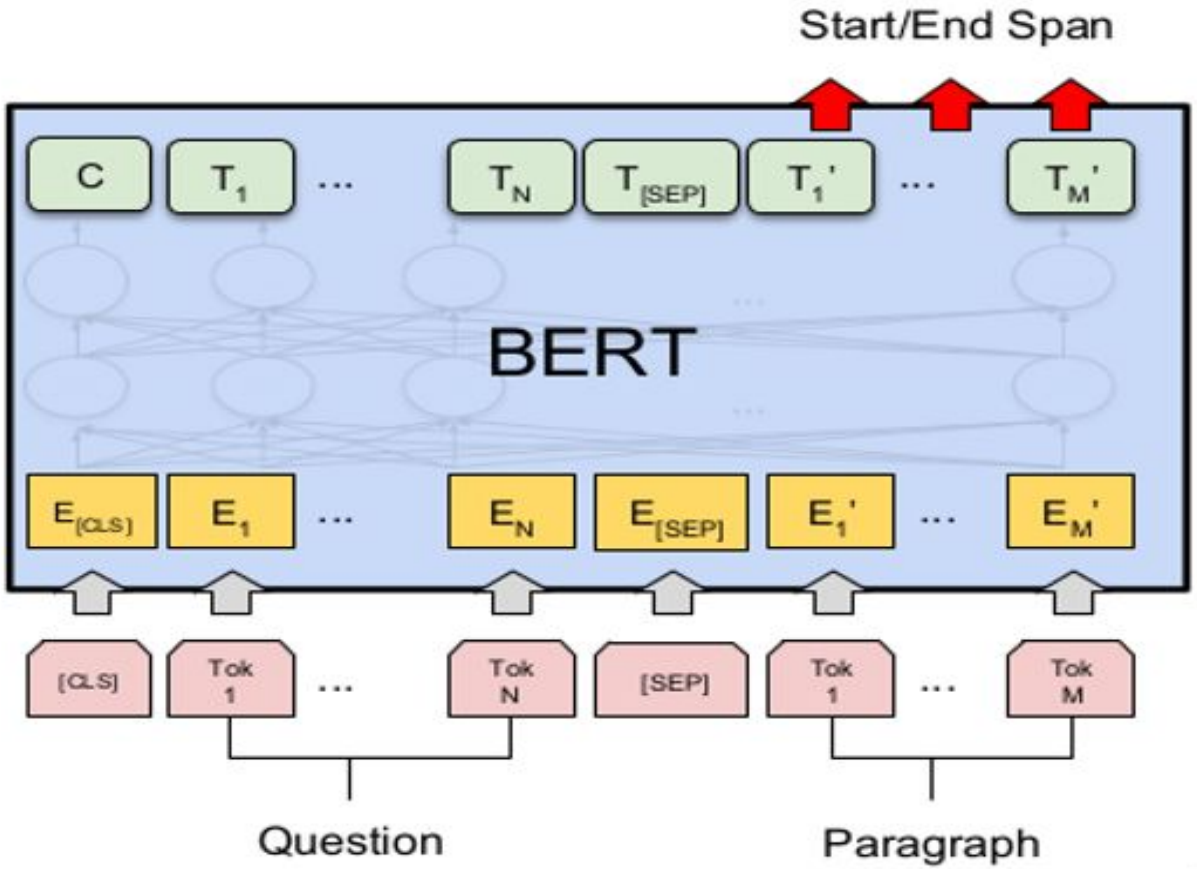
BERT – Finetuning

- Sentence-pair classification
- MNLI, QQP, QNLI, STS-B,MRPC, RTE, SWAG



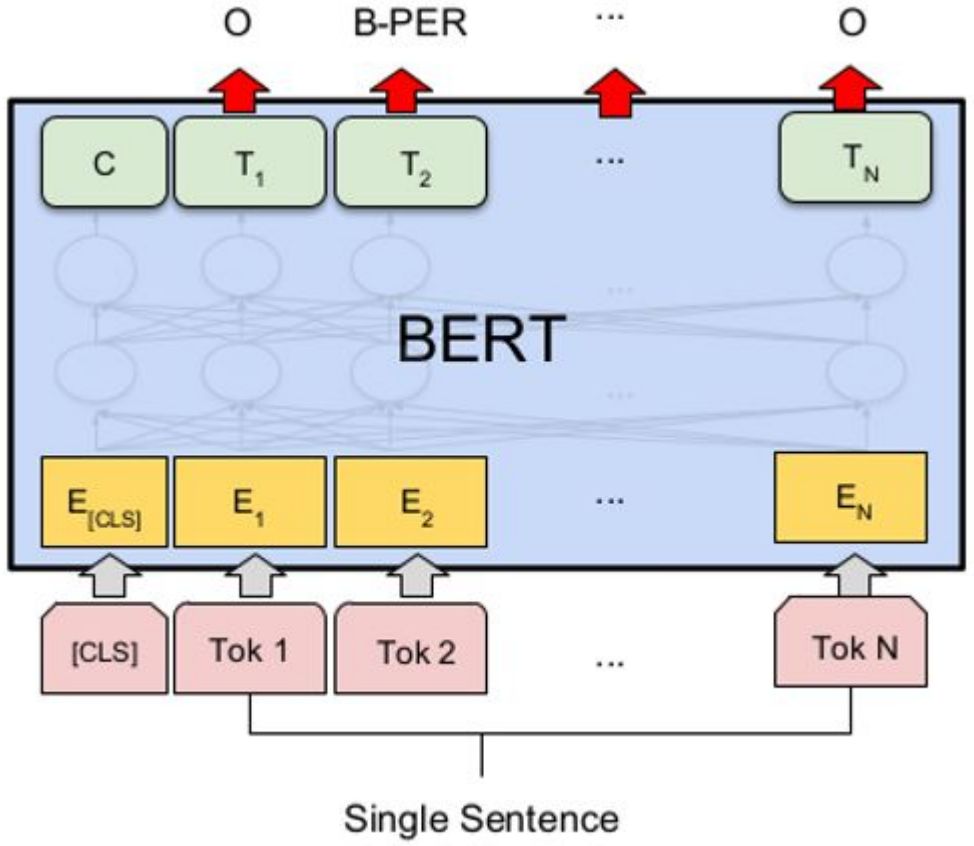
BERT – Finetuning

- Question answering
- SQuAD



BERT – Finetuning

- Single sentence tagging
- CoNLL-2003 NER



BioBERT

BioBERT: a pre-trained biomedical language representation model for biomedical text mining

Jinhyuk Lee [1,†], Wonjin Yoon [1,†], Sungdong Kim [2], Donghyeon Kim [1], Sunkyu Kim [1], Chan Ho So [3] and Jaewoo Kang [1,3,*]

[1]Department of Computer Science and Engineering, Korea University, Seoul 02841, Korea, [2]Clova AI Research, Naver Corp, Seong-Nam 13561, Korea and [3]Interdisciplinary Graduate Program in Bioinformatics, Korea University, Seoul 02841, Korea

*To whom correspondence should be addressed.

[†]The authors wish it to be known that the first two authors contributed equally.

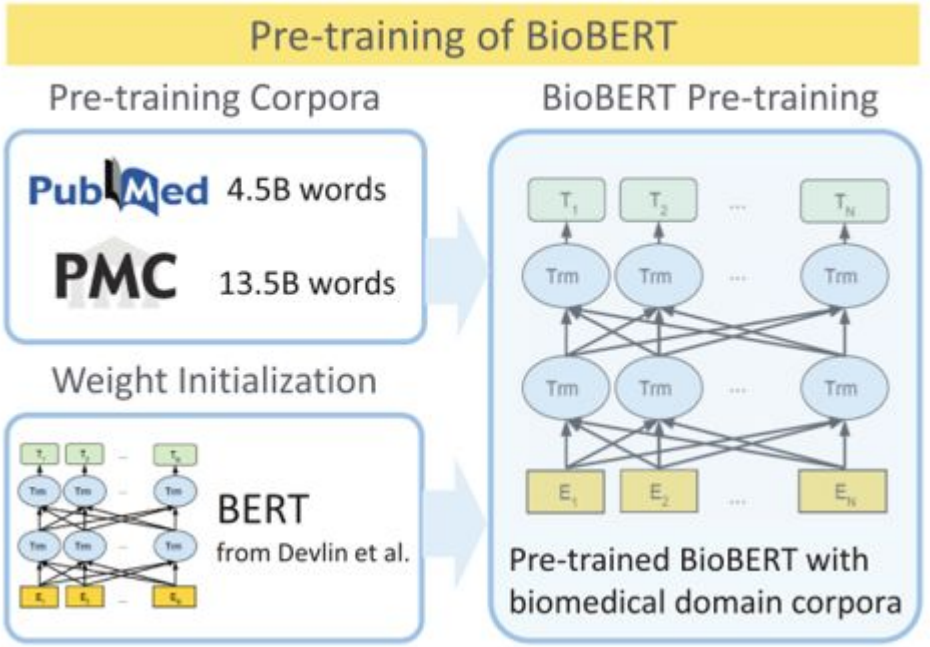
Associate Editor: Jonathan Wren

Received on May 16, 2019; revised on July 29, 2019; editorial decision on August 25, 2019; accepted on September 5, 2019

BioBERT

- Pretraining on biomedical texts

Corpus	Number of words
English Wikipedia	2.5B
BooksCorpus	0.8B
PubMed Abstracts	4.5B
PMC Full-text articles	13.5B



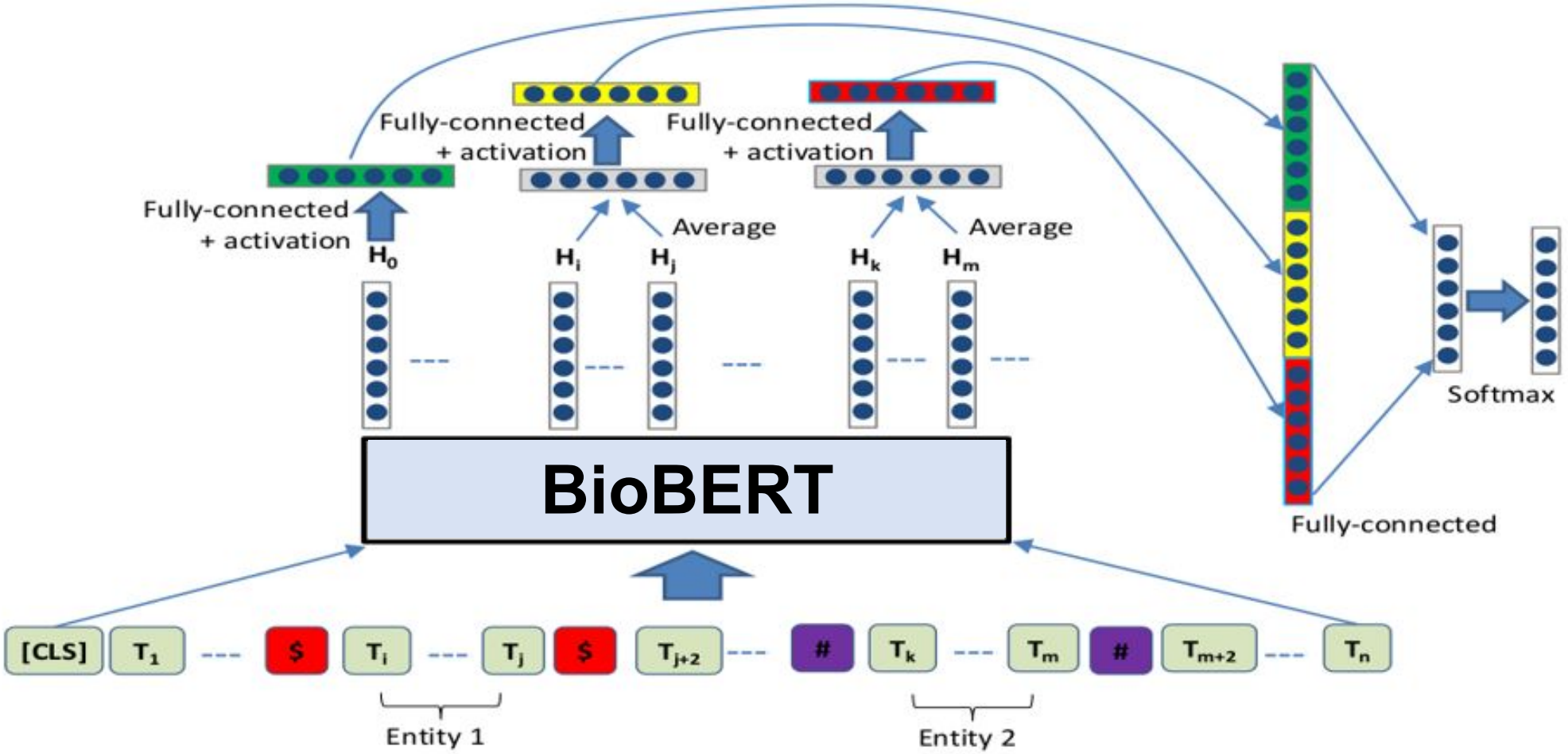
R-BERT

Enriching Pre-trained Language Model with Entity Information for Relation Classification

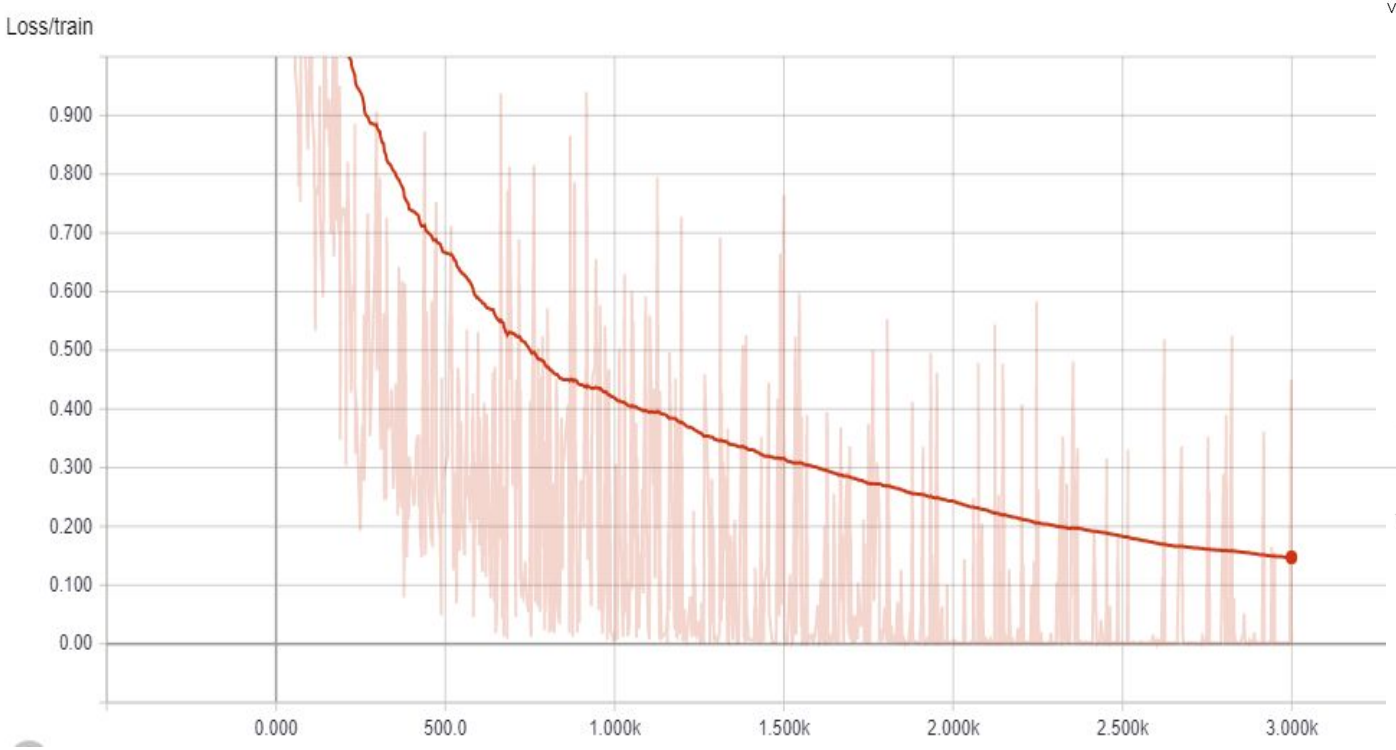
Shanchan Wu
Alibaba Group (U.S.) Inc., Sunnyvale, CA
shanchan.wu@alibaba-inc.com

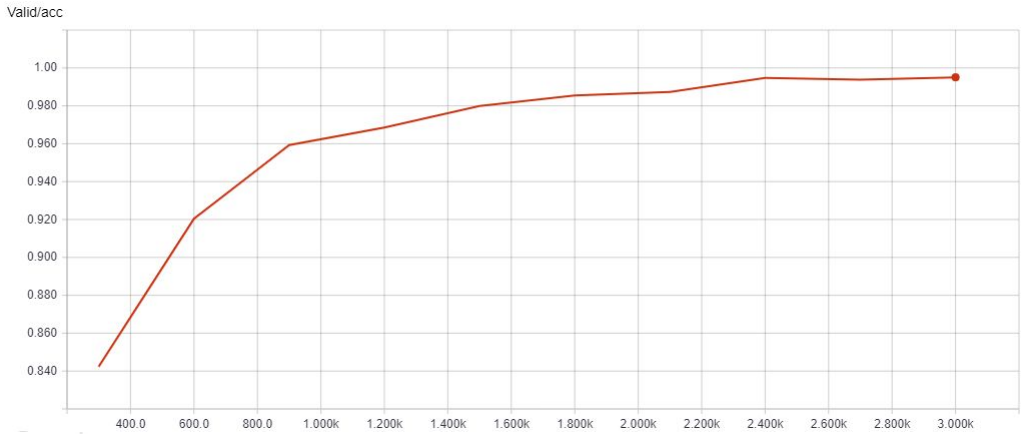
Yifan He
Alibaba Group (U.S.) Inc., Sunnyvale, CA
y.he@alibaba-inc.com

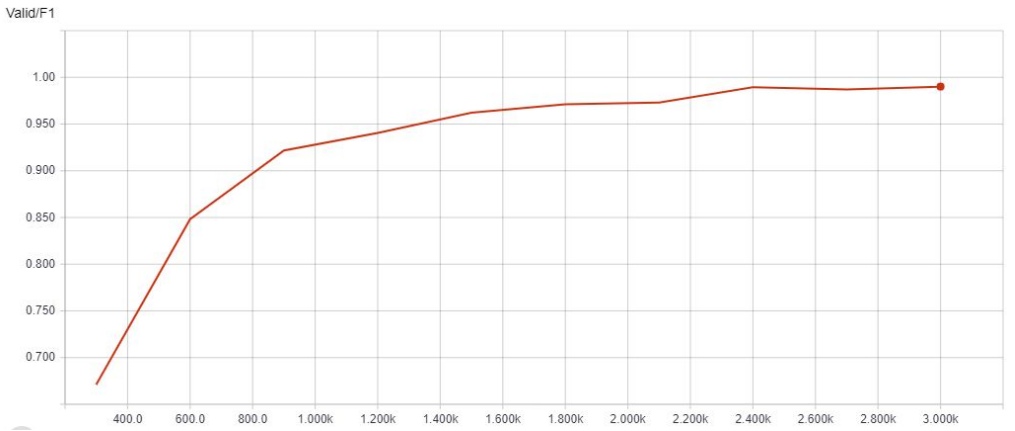
R-BERT



Results



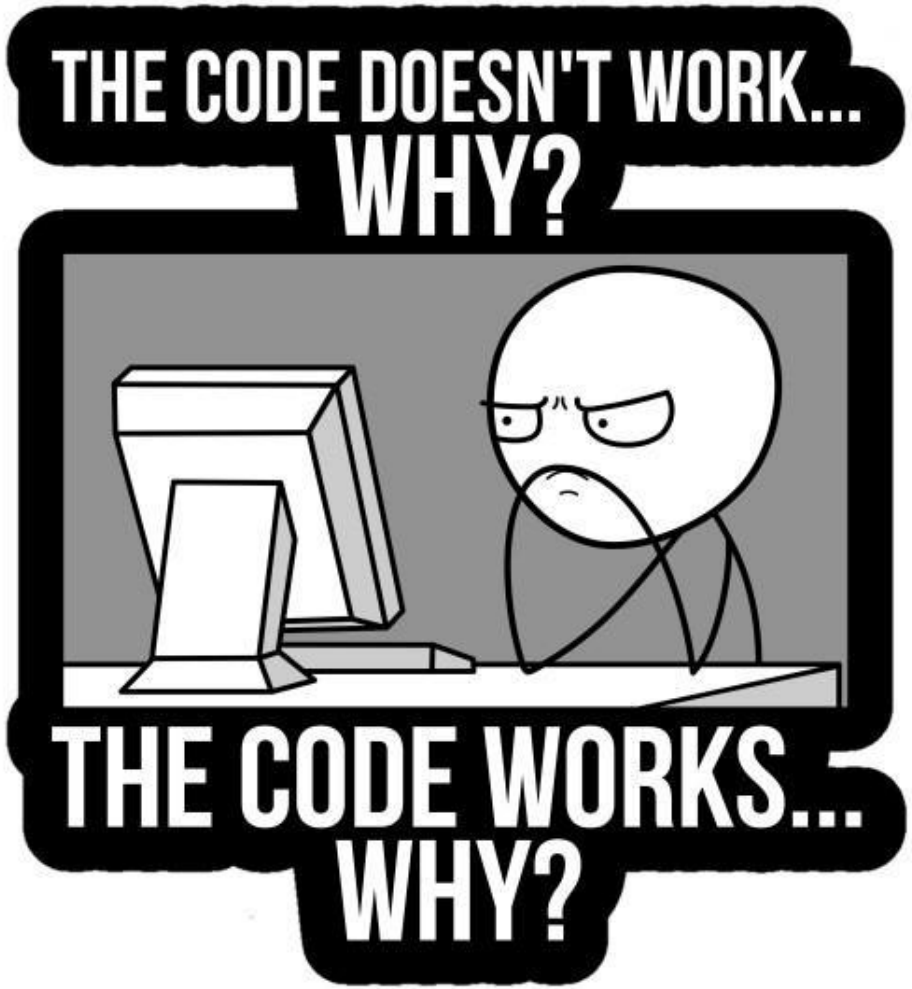




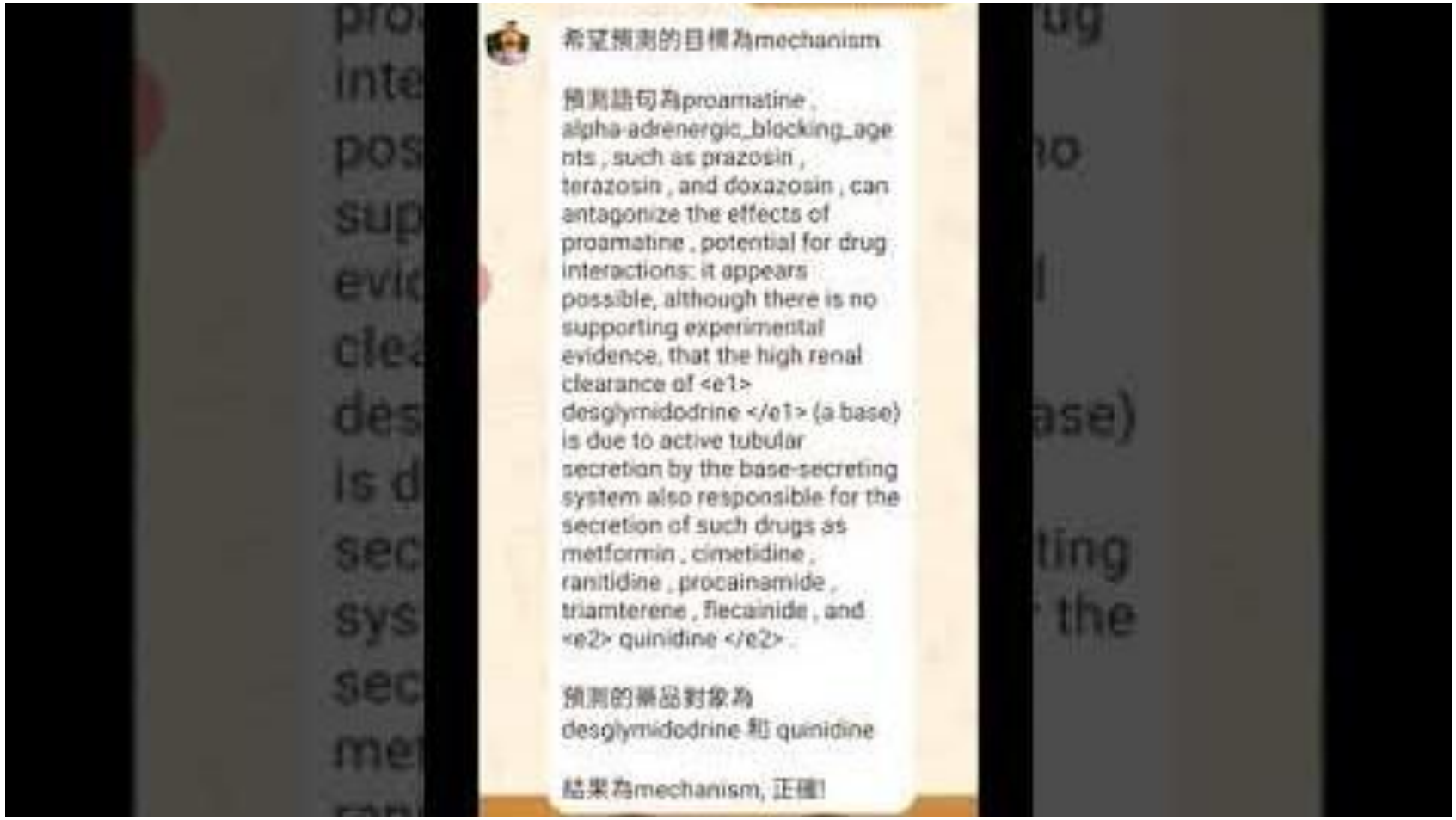
最終測試結果:

```
----------------------
loss: 0.570773
acc : 0.910927
F1   : 0.825545
----------------------
```

Code Review



Chatbot Representation



Encountered problems…

- PyTorch : By self-learning
- Low accuracy : To modify model
- Front-end : By worse way but works well

Further…

- Remove entity tags and training with more data for general usage
- Connect with OCR for information input
- Translation to target language

References

- Attention Is All You Need (Vaswani et al., 2017) arXiv:1706.03762
- BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (Devlin et al., 2018) arXiv:1810.04805
- BioBERT: a pre-trained biomedical language representation model for biomedical text mining (Lee et al., 2019) arXiv:1901.08746
- Enriching Pre-trained Language Model with Entity Information for Relation Classification (Wu et al., 2019) arXiv:1905.08284

Thanks for your attention

Chase excellence, success will follow