



The 14th European Summer School on Information Retrieval



MSCA ITN/ETN No.
860721

DoSSIER
Domain Specific Systems for
Information Extraction and Retrieval

Neural reranking

Georgios Peikos, Wojciech Kusa, Annisa Maulida Ningtyas, Oscar E. Mendoza

Agenda

1. Neural retrieval recap
2. Challenges in medical neural retrieval
3. Neural retrieval for clinical trials
4. Introduction to the TCRR model
5. Performance tracking
6. Hands-on #1: model training
7. Hands-on #2: inference

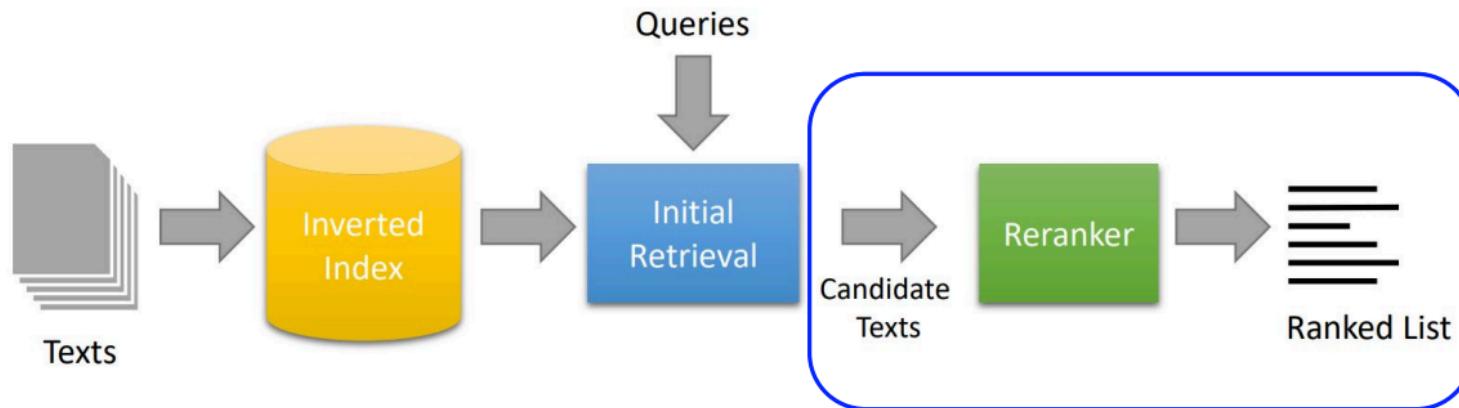


Agenda

1. Neural retrieval recap
2. Challenges in medical neural retrieval
3. Neural retrieval for clinical trials
4. Introduction to the TCRR model
5. Performance tracking
6. Hands-on #1: model training
7. Hands-on #2: inference



Neural Retrieval



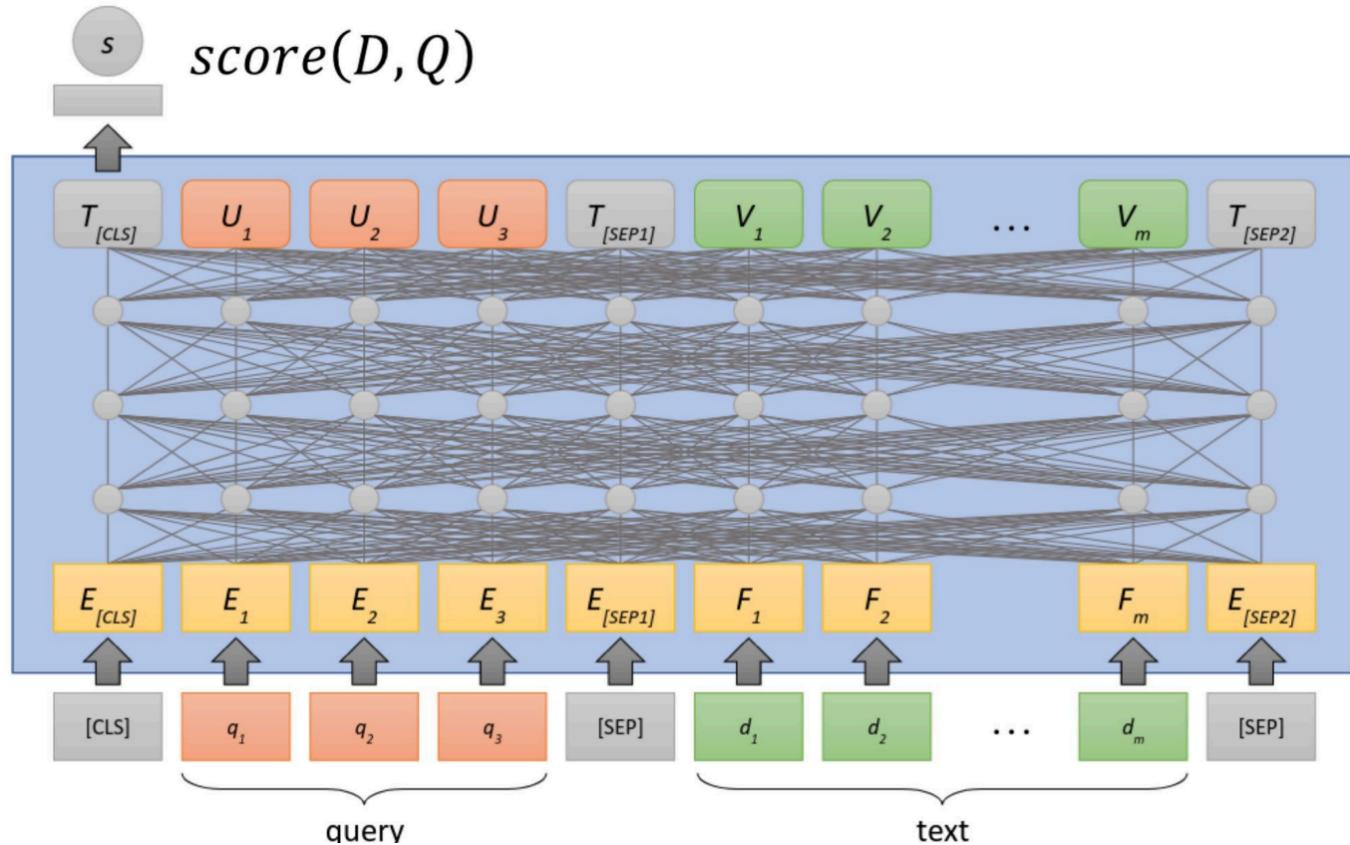
Three approaches to neural IR

1. Cross-encoder
2. Bi-encoder
3. Learned sparse retrieval



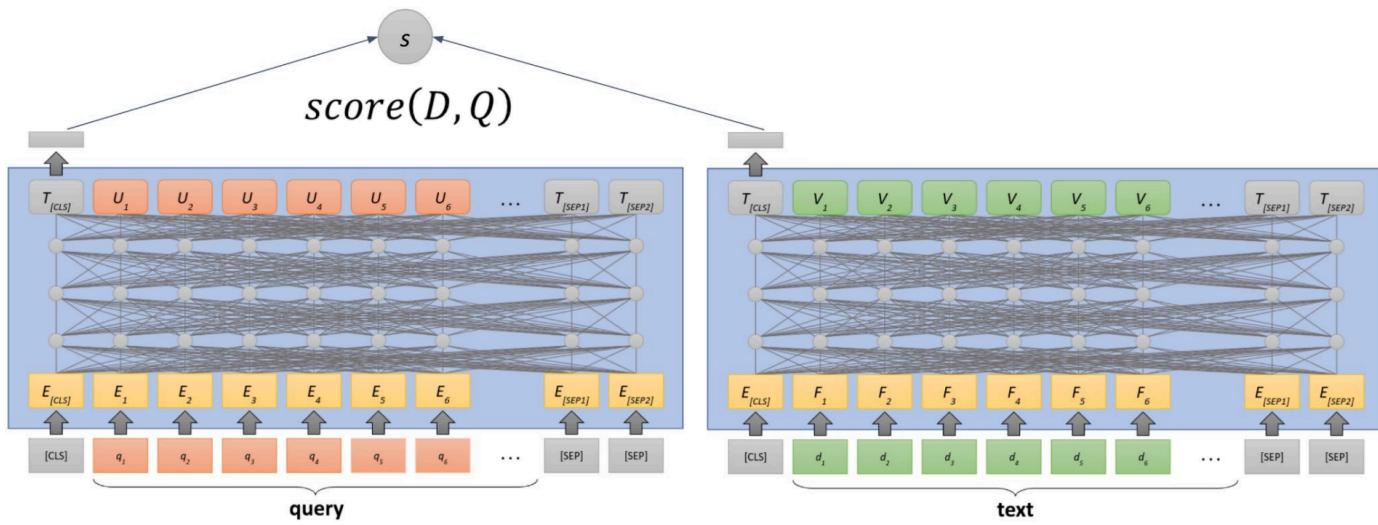
Cross-encoder

- Input: query and document
- Output: score
- Slow but robust



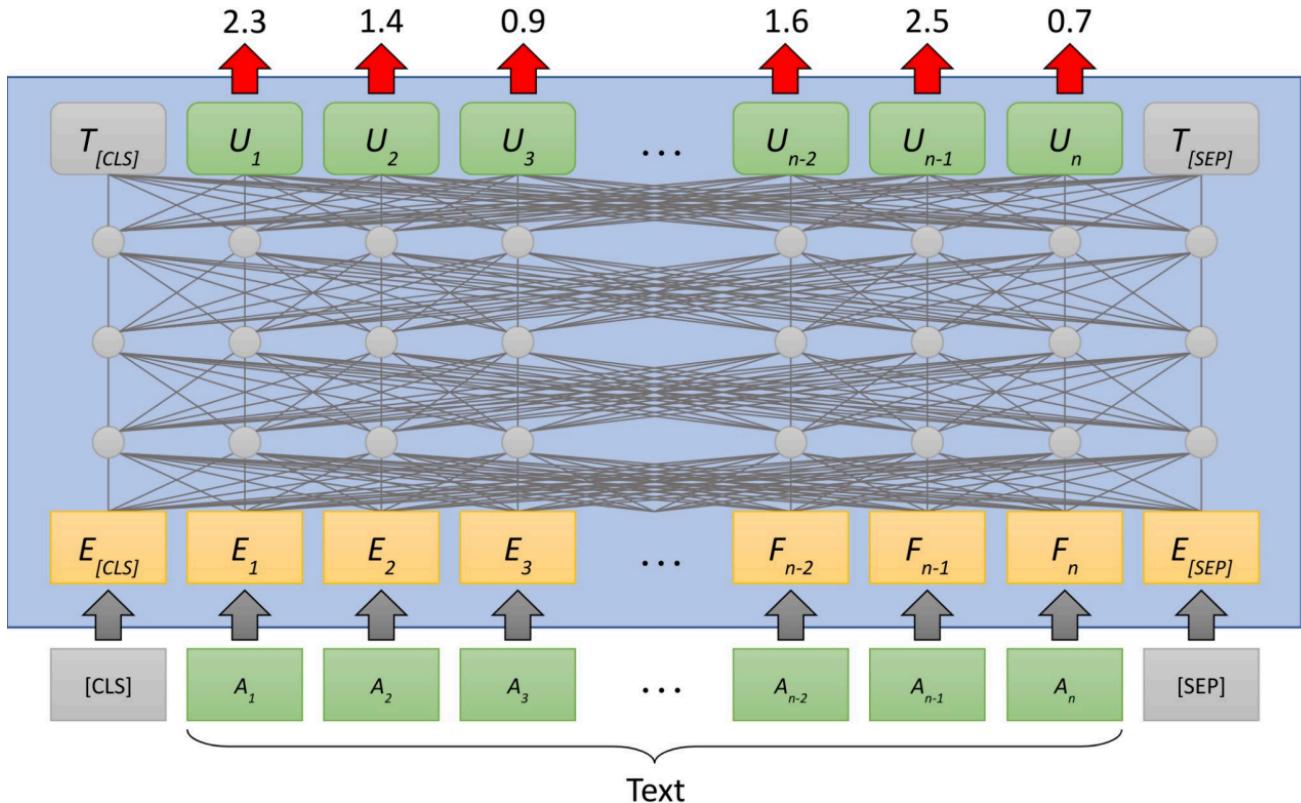
Bi-encoder

- Input: query or document
- Output: vectors
- Scores by comparing vectors
- Fast, but less efficient and less robust



Learned sparse models

- Input: query or document
- Output: new term weights
- “Newest” approach



Agenda

1. Neural retrieval recap
2. Challenges in medical neural retrieval
3. Neural retrieval for clinical trials
4. Introduction to the TCRR model
5. Performance tracking
6. Hands-on #1: model training
7. Hands-on #2: inference



Technical Challenges

- Data Quality
- Interpretability
- Scalability
- Integration



Ethical and Regulatory Challenges

- Patient Privacy
- Bias and Fairness
- Regulatory Hurdles
- Accountability



Medical datasets



Medical datasets

- BigBio

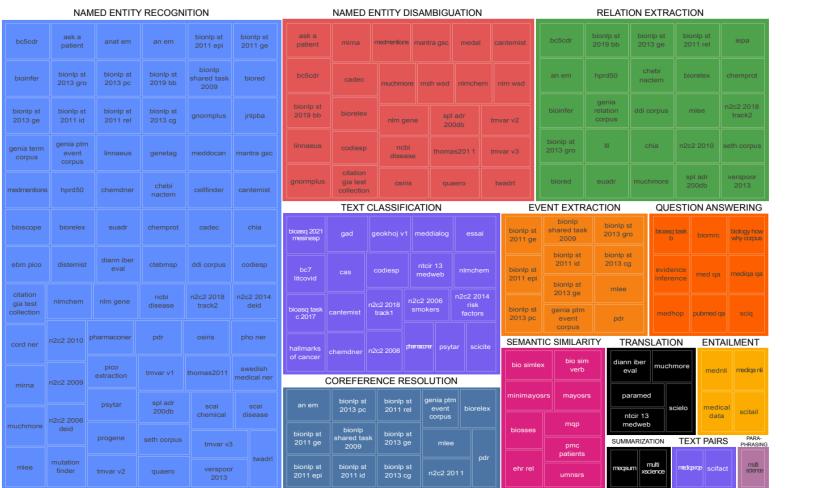


Figure 2: Treemap visualization of BIGBIO’s 126 datasets and 13 task categories, denoted by color.

BIGBIO: A Framework for Data-Centric Biomedical Natural Language Processing

Jason Alan Fries^{1*} Leon Weber^{2,3*} Natasha Seelam^{4*} Gabriel Altay^{5*}
 Debajyoti Datta^{6†} Ruiqi Su^{7†} Samuele Garda^{8†} Sunny MS Kang^{8†}
 Stella Biderman^{9,10†} Matthias Samwald^{11†} Stephen H. Bach^{12†} Wojciech Kusa^{13†}
 Samuel Cahyawijaya^{14†} Fabio Barth^{2†} Simon Ott^{11†} Mario Sänger^{2†} Bo Wang¹⁵
 Alison Callahan¹ Daniel León Periñán¹⁶ Théo Gigant⁷ Patrick Haller²
 Jenny Chim¹⁷ Jose Posada¹⁸ John Giorgi¹⁹ Karthik Rangasai Sivaraman²⁰
 Marc Pàmies²¹ Marianna Nezhurina²² Robert Martin² Moritz Freidank²³
 Nathan Dahlberg⁷ Shubhangshu Mishra²⁴ Shamik Bose⁷ Nicholas Broad²⁵
 Yanis Labrak²⁶ Shlok S Deshmukh²⁷ Sid Kiblawi²⁸ Ayush Singh⁷ Minh Chien Vu²⁹
 Trishala Neeraj³⁰ Jonas Golde² Albert Villanova del Moral²⁵ Benjamin Beilharz³¹

¹Stanford University ²Humboldt-Universität zu Berlin

³Max Delbrück Center for Molecular Medicine ⁴Sherlock Biosciences ⁵Tempus Labs Inc.
⁶University of Virginia ⁷BigScience ⁸Immuneering ⁹EleutherAI ¹⁰Booz Allen Hamilton
¹¹Institute of Artificial Intelligence, Medical University of Vienna ¹²Brown University

¹³TU Wien ¹⁴The Hong Kong University of Science and Technology

^{15–31}See Appendix B *Equal Contribution †Equal Contribution

Corresponding Authors: jason-fries@stanford.edu leonweber@posteo.de
 nseelam1@gmail.com gabriel.altay@gmail.com

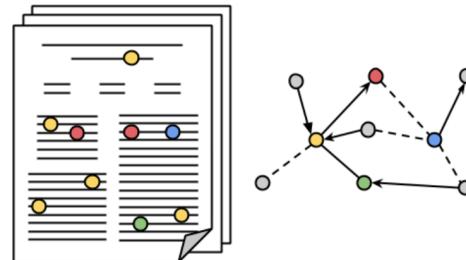


Medical datasets

- BigBio
- BLURB

Domain-Specific Language Model Pretraining for Biomedical Natural Language Processing

YU GU*, ROBERT TINN*, HAO CHENG*, MICHAEL LUCAS, NAOTO USUYAMA, XIAODONG LIU, TRISTAN NAUMANN, JIANFENG GAO, and HOIFUNG POON, Microsoft Research



BLURB

BLURB is the Biomedical Language Understanding and Reasoning Benchmark.



Medical datasets

- BigBio
- BLURB
- CBLUE

CBLUE: A Chinese Biomedical Language Understanding Evaluation Benchmark

Ningyu Zhang^{1*}, Mosha Chen^{2*}, Zhen Bi^{1*}, Xiaozhuan Liang^{1*}, Lei Li^{1*}, Xin Shang³, Kangping Yin², Chuanqi Tan², Jian Xu², Fei Huang², Luo Si², Yuan Ni⁴, Guotong Xie^{4,5,6}, Zhifang Sui^{7,13}, Baobao Chang^{7,13}, Hui Zong^{8,14}, Zheng Yuan⁹, Linfeng Li¹⁰, Jun Yan¹⁰, Hongying Zan^{11,13}, Kunli Zhang^{11,13}, Buzhou Tang^{12,13†}, Qingcai Chen^{12,13†}

¹ Alibaba-Zhejiang University Joint Research Institute of Frontier Technologies, Zhejiang University

² Alibaba Group, ³School of Mathematical Science, Zhejiang University, ⁴Pingan Health Technology

⁵Ping An Health Cloud Company Limited ⁶Ping An International Smart City Technology Co., Ltd

⁷Key Laboratory of Computational Linguistics, Ministry of Education, Peking University

⁸School of Life Sciences and Technology, Tongji University ⁹Tsinghua University,

¹⁰Yidu Cloud Technology Inc ¹¹School of Information Engineering, Zhengzhou University

¹²Harbin Institute of Technology (Shenzhen) ¹³Peng Cheng Laboratory, ¹⁴Philips Research China



Medical datasets

- BigBio
- BLURB
- CBLUE
- <https://ir-datasets.com>

Github: [allenai/ir_datasets](https://github.com/allenai/ir_datasets)

ir_datasets: Catalog

`ir_datasets` provides a common interface to many IR ranking datasets.

Getting Started

Install with pip:

```
| pip install --upgrade ir_datasets
```



Medical models

- BioBERT
- SciBERT
- BlueBERT
- ClinicalBERT
- ClinicalLongformer
- ClinicalBigBird
- ...



Agenda

1. Neural retrieval recap
2. Challenges in medical neural retrieval
3. Neural retrieval for clinical trials
4. Introduction to the TCRR model
5. Performance tracking
6. Hands-on #1: model training
7. Hands-on #2: inference



Neural retrieval for clinical trials

1. Neural query expansion with T5
2. Dense retrieval
3. Cross encoder



Neural Query Synthesis and Domain-Specific Ranking Templates for Multi-Stage Clinical Trial Matching

Authors:  Ronak Pradeep,  Yilin Li,  Yuetong Wang,  Jimmy Lin [Authors Info & Claims](#)

SIGIR '22: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval • July 2022 • Pages 2325–2330 • <https://doi.org/10.1145/3477495.3531853>

Published: 07 July 2022 [Publication History](#)



2 119



 Get Access

ABSTRACT

In this work, we propose an effective multi-stage neural ranking system for the clinical trial matching problem. First, we introduce NQS, a neural query synthesis method that leverages a zero-shot document expansion model to generate multiple sentence-long queries from lengthy patient descriptions. These queries are independently issued to a search engine and the results are fused. We find that on the TREC 2021 Clinical Trials Track, this method outperforms strong traditional baselines like BM25 and BM25 + RM3 by about 12 points in nDCG@10, a relative improvement of 34%. This simple method is so effective that even a state-of-the-art neural relevance ranking method trained on the medical subset of MS MARCO passage, when reranking the results of NQS, fails to improve on the ranked list. Second, we introduce a two-stage neural reranking pipeline trained on clinical trial matching data using tailored ranking templates. In this setting, we can train a pointwise reranker using just 1.1k positive examples and obtain effectiveness improvements over NQS by 24 points. This end-to-end multi-stage system demonstrates a 20% relative effectiveness gain compared to the second-best submission at TREC 2021, making it an important step towards better automated clinical trial matching.





Effective matching of patients to clinical trials using entity extraction and neural re-ranking

Wojciech Kusa^{a,1} , Óscar E. Mendoza^{c,1}, Petr Knoth^d, Gabriella Pasi^c, Allan Hanbury^{a,b}

Show more

Add to Mendeley Share Cite

<https://doi.org/10.1016/j.jbi.2023.104444>

Get rights and content

Under a Creative Commons license

open access

Highlights

- We conduct several experiments for the patient-to-trial matching retrieval problem.
- Inclusion criteria section has the biggest influence on the score in lexical models.
- Query and document enrichment techniques improve retrieval of relevant trials.
- Age and gender-based filtering helped remove 26% ineligible trials.
- Novel training strategy for re-ranking further increases retrieval effectiveness.





Matching a Patient from An Admission Note to Clinical Trials: Experiments with Query Generation and Neural-Ranking

Vincent Nguyen Maciej Rybinski Sarvnaz Karimi
CSIRO Data61
Sydney, Australia
firstname.lastname@csiro.au

ABSTRACT

Many clinical trials fail to attract enough eligible participants. The TREC 2022 Clinical Trials track set a task where patient data, in the form of clinical notes, can be used to match eligible patients to a relevant clinical trial. We explore a number of dense retrieval methods using Bidirectional Encoder Representations from Transformers (BERT). Our best method used BERT reranking using models based on monoBERT architecture. Our self-supervised monoBERT run achieved effectiveness competitive to that of a fully-tuned monoBERT run.

CCS CONCEPTS

- Information systems → Retrieval models and ranking; Language models; Decision support systems;
- Applied computing → Health informatics.

KEYWORDS

Clinical trials search; Medical information retrieval; Learning-to-rank; Evidence-based medicine

ACM Reference Format:

Vincent Nguyen Maciej Rybinski Sarvnaz Karimi. 2021. Matching a Patient from An Admission Note to Clinical Trials: Experiments with Query Generation and Neural-Ranking. In *TREC22: TREC, November, 2022*. ACM, New York, NY, USA, 4 pages.

tasks [4–6] in its 2017, 2018, and 2019 editions. The task of this year’s track is to link a synthetic patient’s electronic health record (EHR), in free text, to relevant clinical trials. TREC CT’s goal is to study the use of automatic retrieval systems to expose patients to relevant clinical trials to increase participation.

In our submission to this year in the TREC CT track, we build upon our last year’s submission [7]. Our experiments this year focus on neural ranking using resource-effective self-supervision and supervision signals from last year’s judgement pool. Our experiments with a neural reranking pipeline centered around resource-effective learning, used a reranker trained on labeled data (from last year’s edition of the track) compared with a self-supervised model trained using the target document corpus. We also experiment with efficient end-to-end neural ranking (where document representations can be pre-computed) with bi-encoders and with neural query expansion. Finally, we also probe the effect of a simple heuristic for matching the patient note with the demographic profile specified in the clinical trials, which we apply to one of the bi-encoder runs.

2 DATASET

The TREC 2022 CT dataset consists of 50 topics with 35,394 relevance judgments. The corpus for the task is a 2020 snap-





Matching a Patient from An Admission Note to Clinical Trials: Experiments with Query Generation and Neural-Ranking

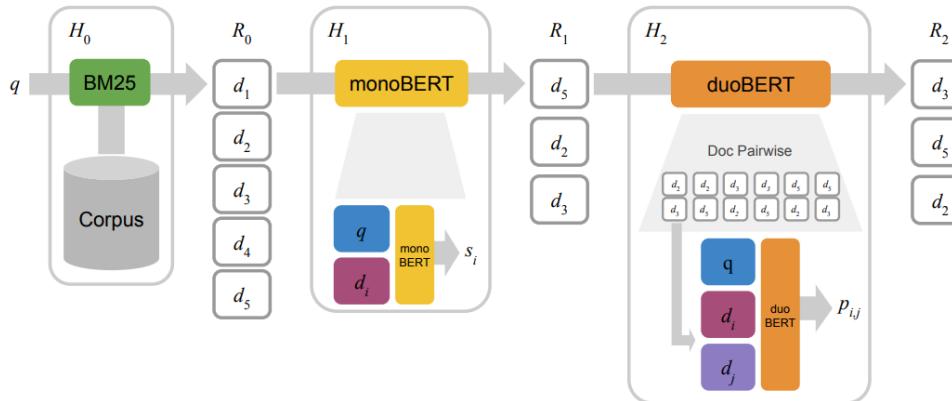
Vincent Nguyen Maciej Rybinski Sarvnaz Karimi
CSIRO Data61
Sydney, Australia
firstname.lastname@csiro.au

ABSTRACT

Many clinical trials fail to attract enough eligible participants. The TREC 2022 Clinical Trials track set a task where patient data, in the form of clinical notes, can be used to match eligible patients to a relevant clinical trial. We explore a number of dense retrieval methods using Bidirectional Encoder Representations from Transformers (BERT). Our best method

tasks [4–6] in its 2017, 2018, and 2019 editions. The task of this year’s track is to link a synthetic patient’s electronic health record (EHR), in free text, to relevant clinical trials. TREC CT’s goal is to study the use of automatic retrieval systems to expose patients to relevant clinical trials to increase participation.

In our submission to this year in the TREC CT track, we



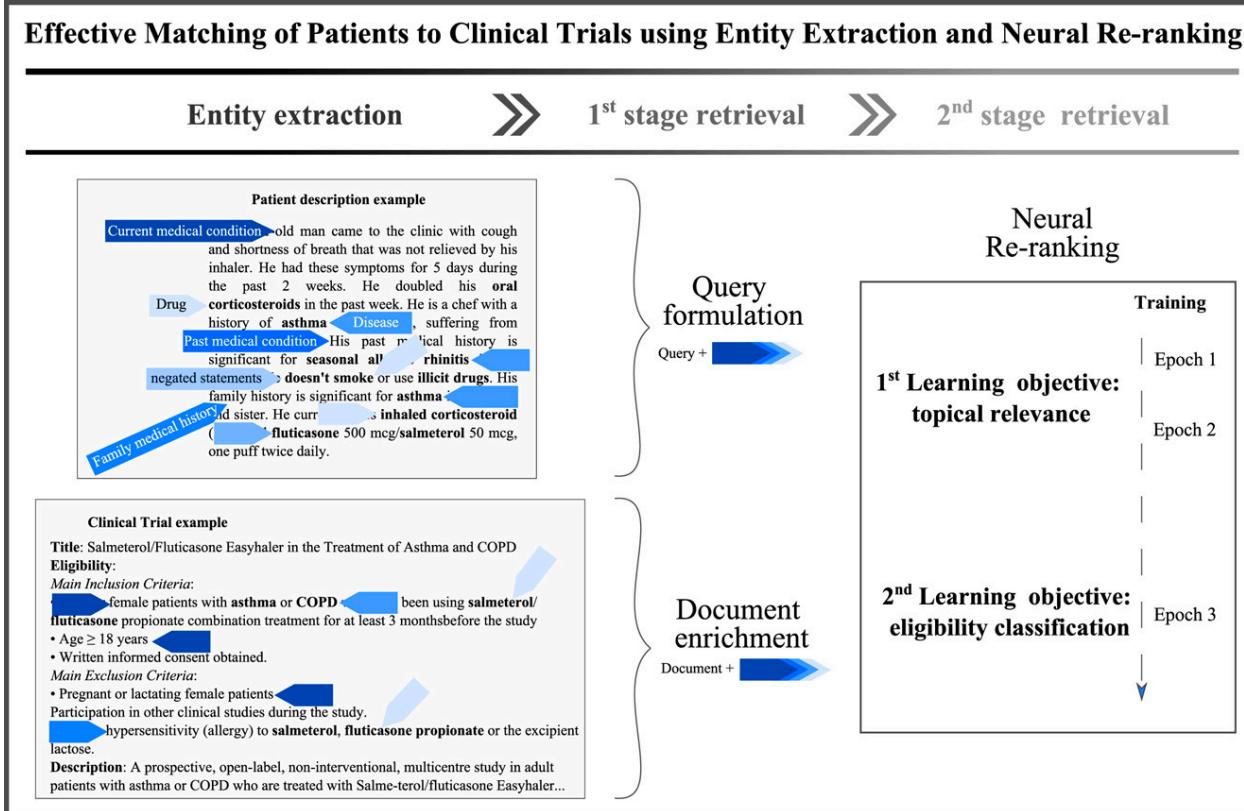
Agenda

1. Neural retrieval recap
2. Challenges in medical neural retrieval
3. Neural retrieval for clinical trials
4. Introduction to the TCRR model
5. Performance tracking
6. Hands-on #1: model training
7. Hands-on #2: inference

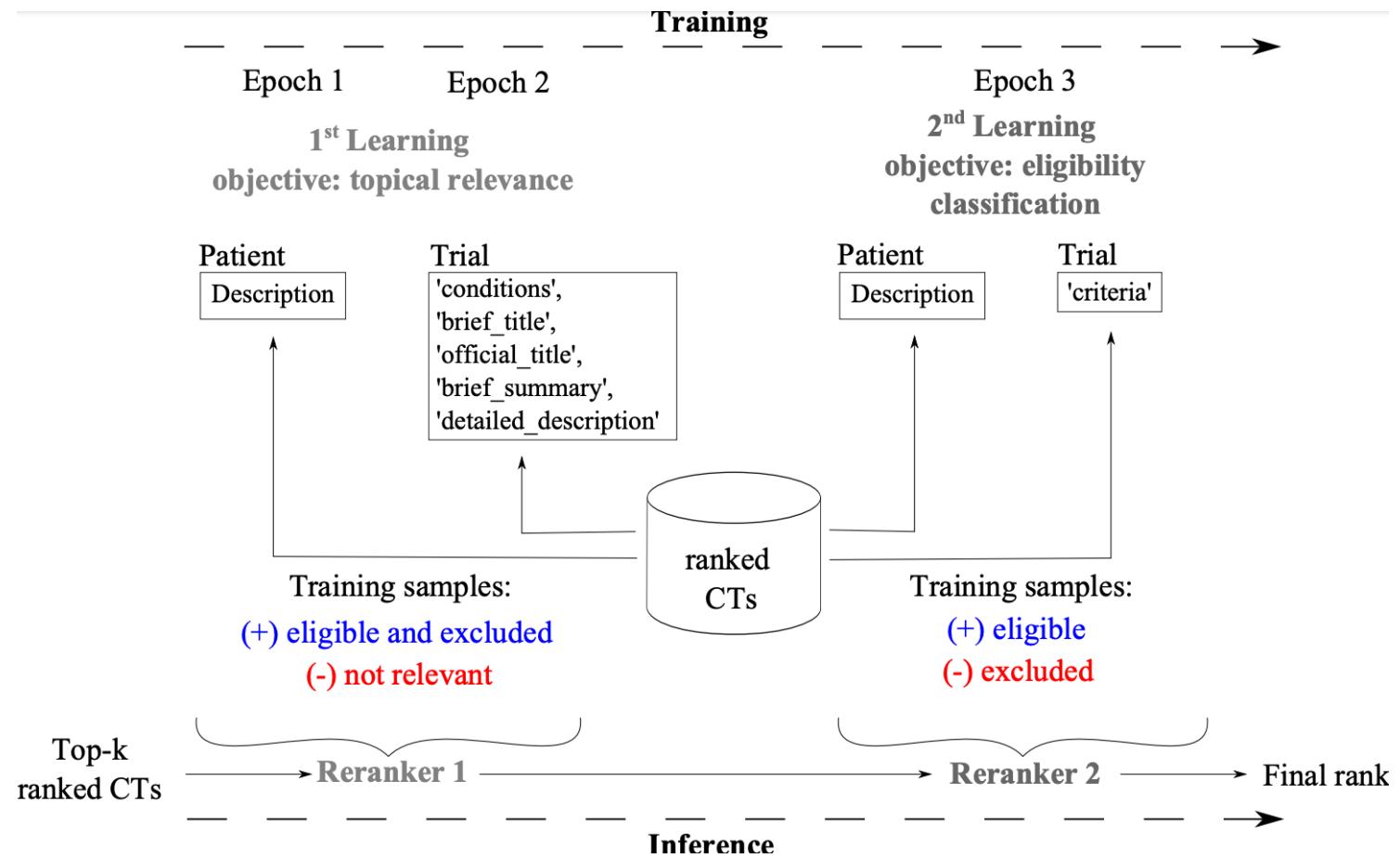




Topical and Criteria Re-Ranking – TCRR



Training Pipeline



Agenda

1. Neural retrieval recap
2. Challenges in medical neural retrieval
3. Neural retrieval for clinical trials
4. Introduction to the TCRR model
- 5. Performance tracking**
6. Hands-on #1: model training
7. Hands-on #2: inference



What can be visualised?

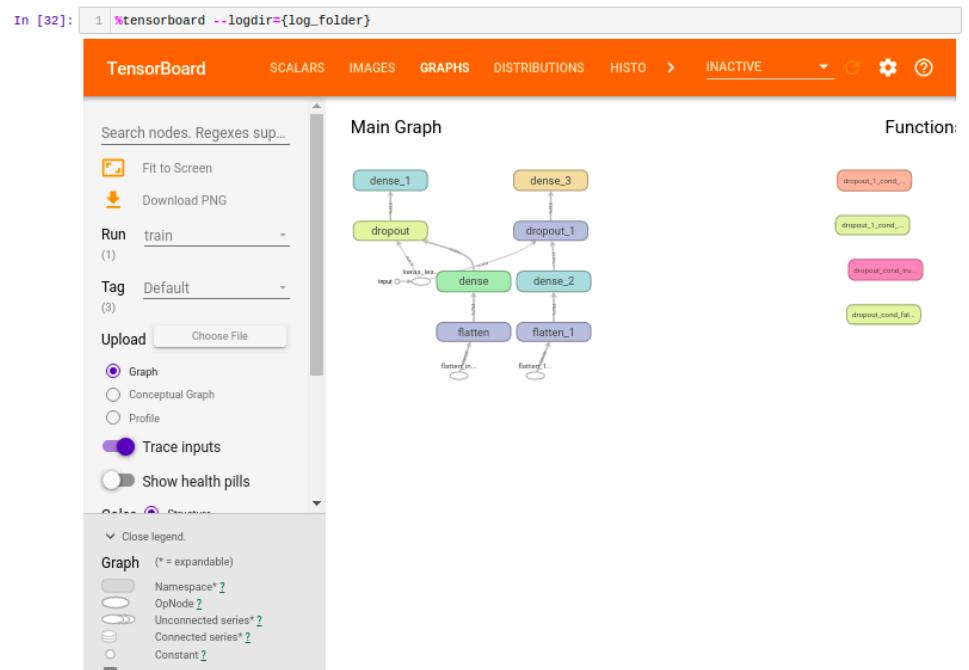
1. Model architecture
2. Model training
3. Model inference



What can be visualised?

1. Model architecture

TensorBoard



What can be visualised?

TensorBoard
Neptune
Weights & Biases

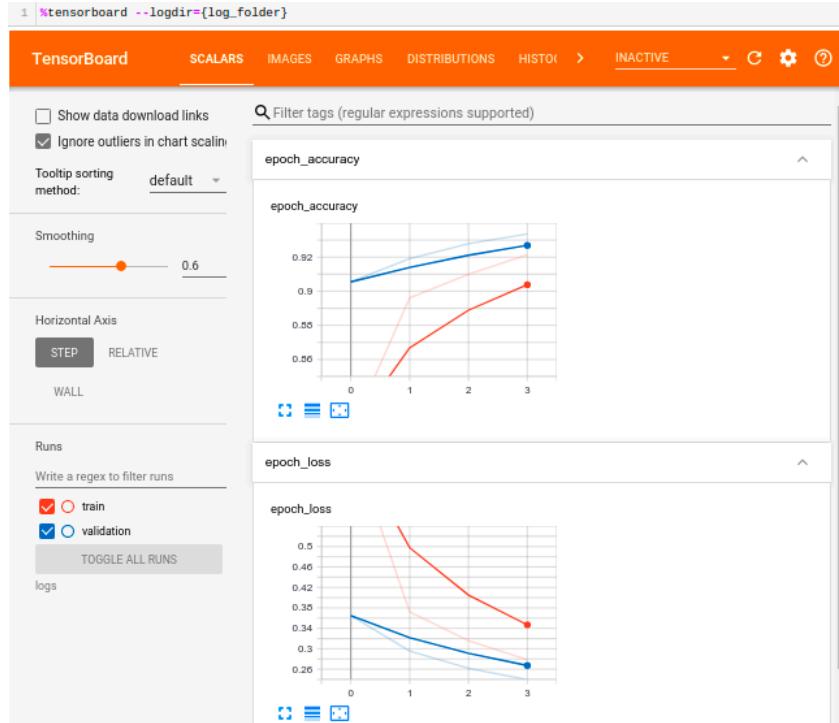
1. Model architecture
2. Model training



What can be visualised?

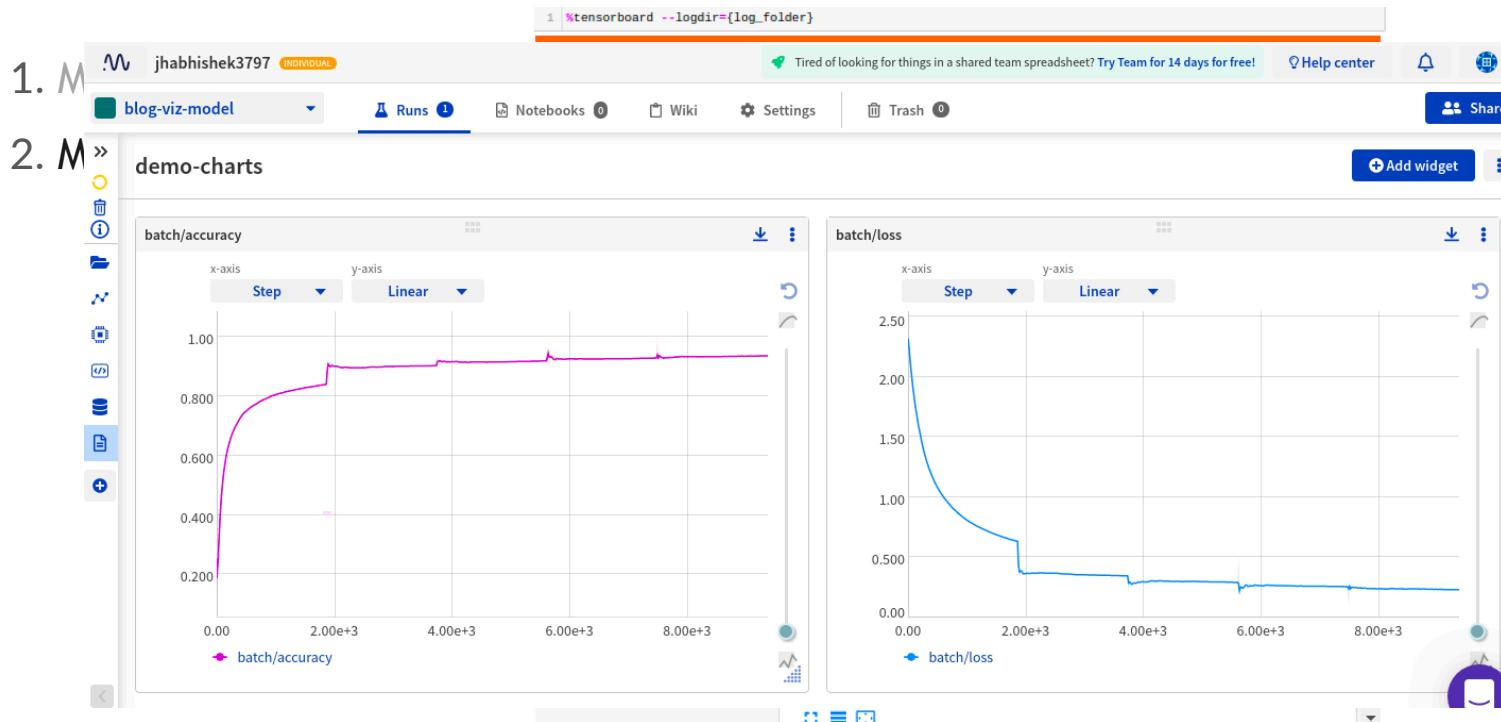
TensorBoard
Neptune
Weights & Biases

1. Model architecture
2. Model training



What can be visualised?

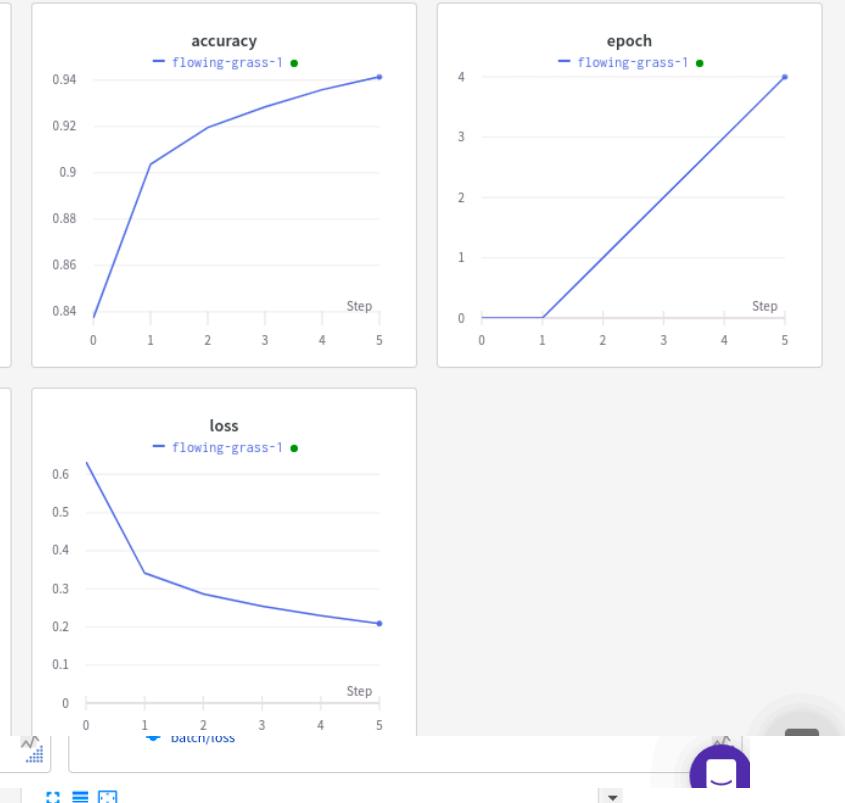
TensorBoard
Neptune
Weights & Biases



What can be visualized?



TensorBoard Next



What can be visualised?

1. Model architecture
2. Model training
3. Model inference

LIME
SHAP - SHapley Additive exPlanations

...



Wandb example

```
● ● ●

# train.py
import wandb
import random # for demo script

wandb.login()

epochs=10
lr=0.01

run = wandb.init(
    # Set the project where this run will be logged
    project="my-awesome-project",
    # Track hyperparameters and run metadata
    config={
        "learning_rate": lr,
        "epochs": epochs,
    })

offset = random.random() / 5
print(f"lr: {lr}")

# simulating a training run
for epoch in range(2, epochs):
    acc = 1 - 2 ** -epoch - random.random() / epoch -
    offset
    loss = 2 ** -epoch + random.random() / epoch + offset
    print(f"epoch={epoch}, accuracy={acc}, loss={loss}")
    wandb.log({"accuracy": acc, "loss": loss})

# run.log_code()
```



Wandb – PyTorch example

```
● ● ●

import wandb
wandb.init(config=args)

# set up your model
model = AutoModelForSequenceClassification.from_pretrained(
    model_name, num_labels=2
)

# Magic
wandb.watch(model, log_freq=100)

model.train()
for batch_idx, (data, target) in enumerate(train_loader):
    output = model(data)
    loss = F.nll_loss(output, target)
    loss.backward()
    optimizer.step()
    if batch_idx % args.log_interval == 0:
        wandb.log({"loss": loss})
```



Wandb – logging metrics

```
predictions = trainer.predict(  
    tokenized_datasets[test_data],  
)  
  
wandb.log(  
    {  
        f"{test_data}_precision": predictions.metrics["test_precision"],  
        f"{test_data}_recall": predictions.metrics["test_recall"],  
        f"{test_data}_f1": predictions.metrics["test_f1"],  
        f"{test_data}_loss": predictions.metrics["test_loss"],  
    }  
)
```



Agenda

1. Neural retrieval recap
2. Challenges in medical neural retrieval
3. Neural retrieval for clinical trials
4. Introduction to the TCRR model
5. Performance tracking
6. Hands-on #1: model training
7. Hands-on #2: inference



Hands-on Session 1: Training with TCRR

- Google Colab



Agenda

1. Neural retrieval recap
2. Challenges in medical neural retrieval
3. Neural retrieval for clinical trials
4. Introduction to the TCRR model
5. Performance tracking
6. Hands-on #1: model training
7. **Hands-on #2: inference**



Hands-on Session 2: Inference and Application

- Google Colab



Questions?

