COMPOSE: Cross-Modal Pseudo-Siamese Network for Patient Trial Matching

Junyi Gao[1], Cao Xiao[1], Lucas M. Glass[12], Jimeng Sun[3]

[1]Analytics Center of Excellence, IQVIA

[2]Department of Statistics, Temple University

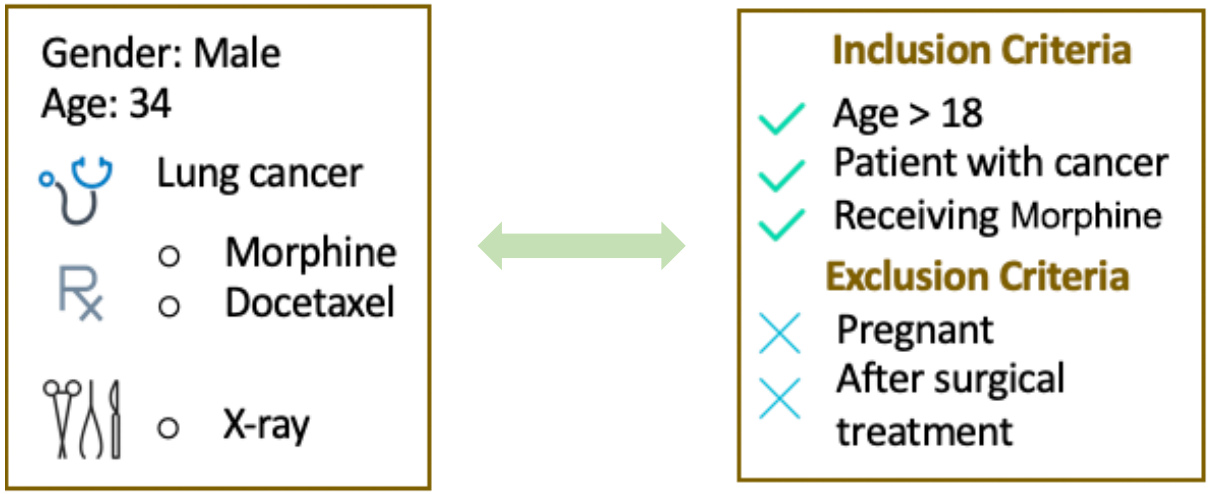
[3]Department of Computer Science, University of Illinois Urbana-Champaign

Content

- Clinical Background
- Challenges
- Method
- Experiment Results

Clinical Background 1: What is patient trial matching?

- Electronic Health Records (EHR): A type of high-dimensional sequence data
  - Procedures
  - Diagnosis
  - Drugs
- Clinical trials: Unstructured text data
  - Inclusion Criteria
  - Exclusion Criteria



Clinical Background 2: Why automated patient trial matching is important?

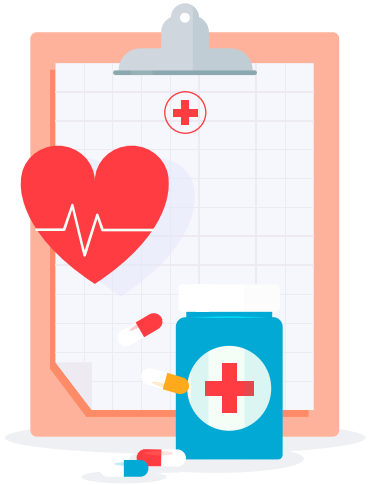
Essential	Annual market over $46 billion
Time Consuming	50% of trials delayed, 25% of cancer trials failed due to enrollment.
High Costs	High recruitment cost: $6000 to $7500 per patient.

Content

- **Clinical Background**
- **Challenges**
  - Multi-granularity medical concept
  - Many-to-many relationship between patient and trials
  - Explicit inclusion/exclusion criteria handling
- **Method**
- **Experiment Results**

Challenge 1: Multi-granularity medical concept

- Eligibility criteria encode more general disease
- EHRs use more specific medical codes

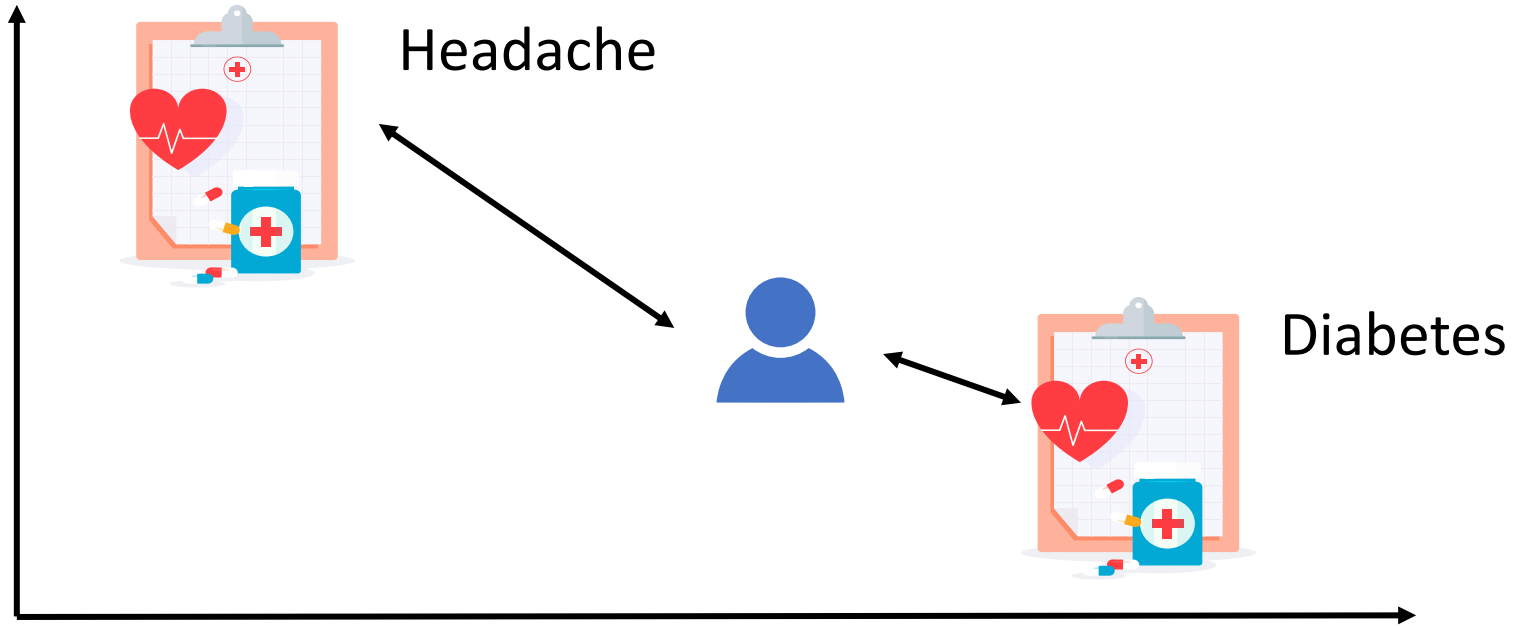


Trial of Cardiovascular Disesases

- ✓ Pleuropericardial adhesion
- ✓ Myocardial infraction
- ✓ Inflammatory cardiomyopathy

Challenge 2: Many-to-many relationship between patients and trials

- Each patient may enroll in more than one trial and vice versa



- Align the patient embedding to different trial embeddings may confuse the embed function

Challenge 3: Explicit inclusion/exclusion criteria handling

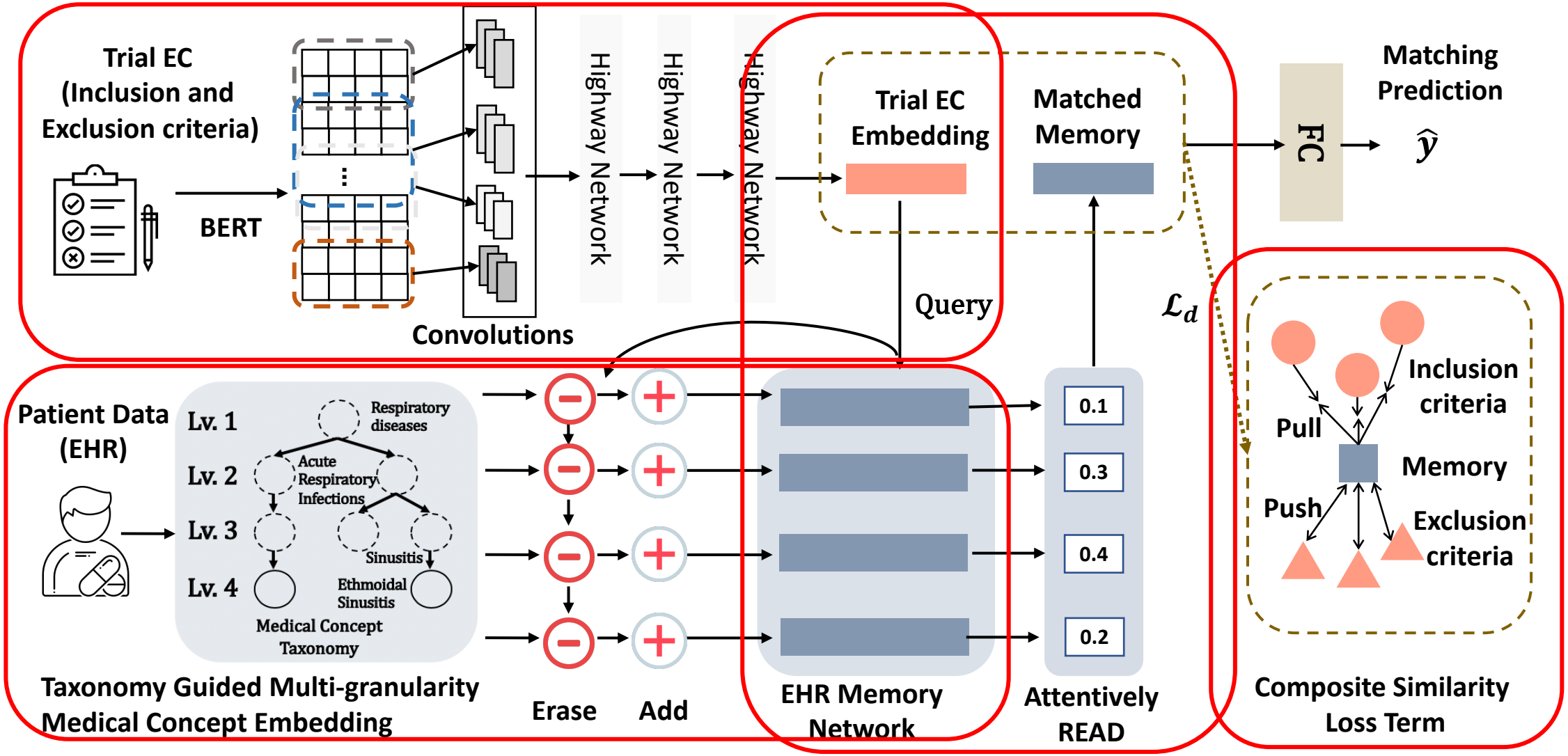
- Inclusion and Exclusion criteria describe desired and unwanted from the targeted patients

Inclusion criteria ⟵ Age > 18 ⟶ Exclusion criteria

Content

- **Clinical Background**
- **Challenges**
- **Method**
  - Trial eligibility criteria embedding
  - Taxonomy guided patient embedding
  - Attentional record alignment and dynamic matching
  - Explicit inclusion/exclusion criteria handling
- **Experiment Results**

Method Overview: COMPOSE



Method: Trial eligibility criteria embedding

- Use BERT to learn contextual embeddings for EC sentence $[w_1, \ldots, w_N]$

$$\widetilde{\boldsymbol{c}} = [\widetilde{\boldsymbol{w}}_1, \ldots, \widetilde{\boldsymbol{w}}_N] = BERT([\boldsymbol{w}_1, \ldots, \boldsymbol{w}_N])$$

- Use different kernel sizes to capture different granularity semantics

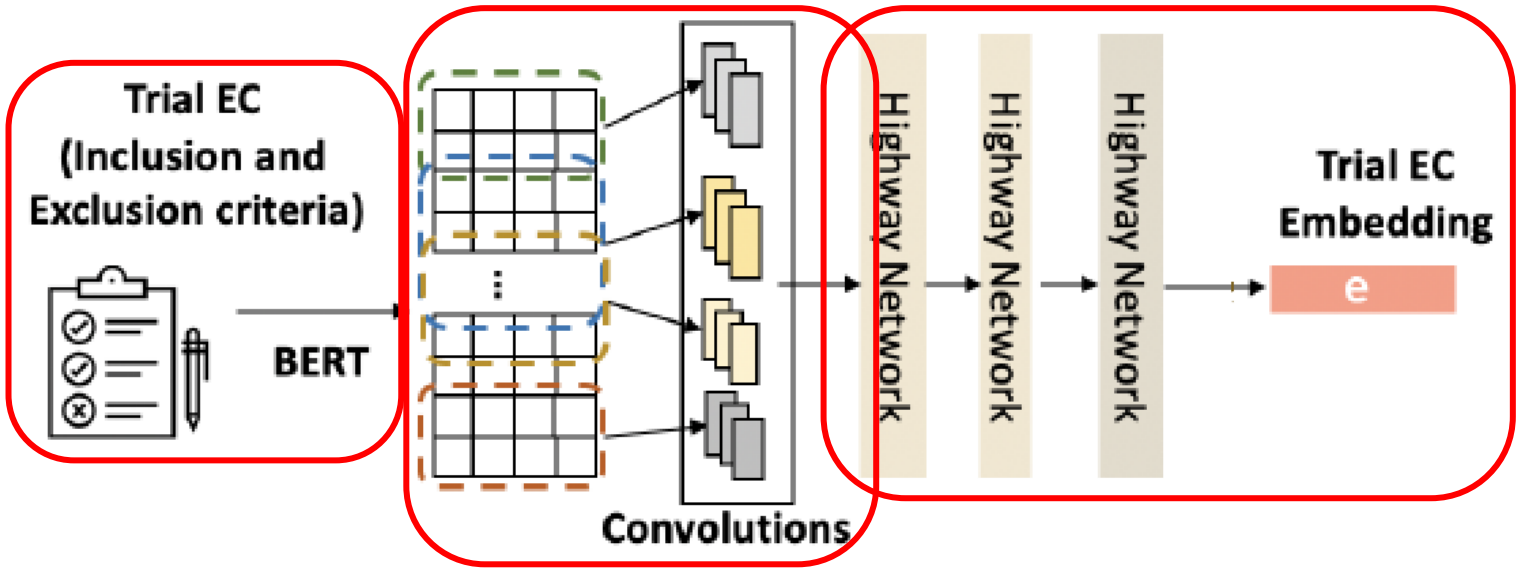
$$\boldsymbol{x} = [Conv(\widetilde{\boldsymbol{c}}, k_1), Conv(\widetilde{\boldsymbol{c}}, k_2), Conv(\widetilde{\boldsymbol{c}}, k_3), Conv(\widetilde{\boldsymbol{c}}, k_4)]$$

- Use highway network and max pooling to obtain the final EC embedidng

$$\boldsymbol{u} = \sigma(Conv(\boldsymbol{x}, k))$$

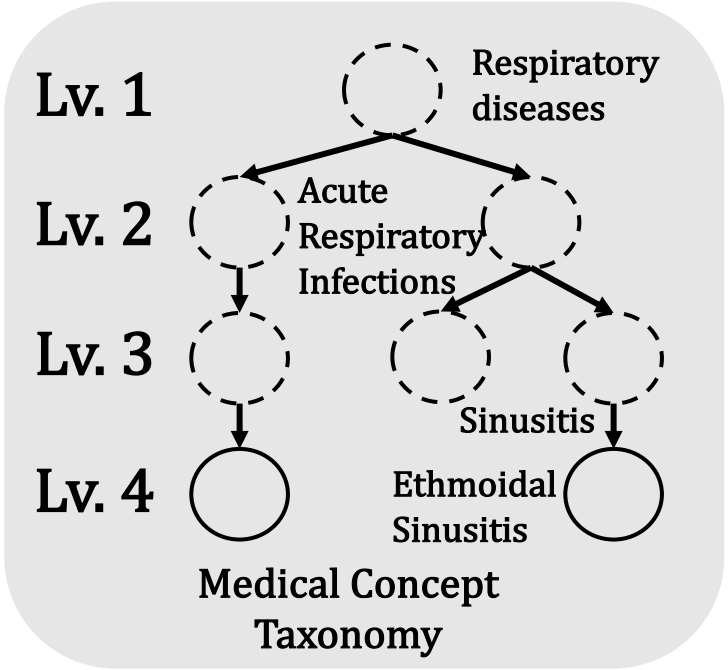
$$\boldsymbol{v} = \boldsymbol{u} \cdot Conv(\boldsymbol{x}, k) + \boldsymbol{x} \cdot (1 - \boldsymbol{u})$$

$$\boldsymbol{e} = MaxPool(\boldsymbol{v})$$

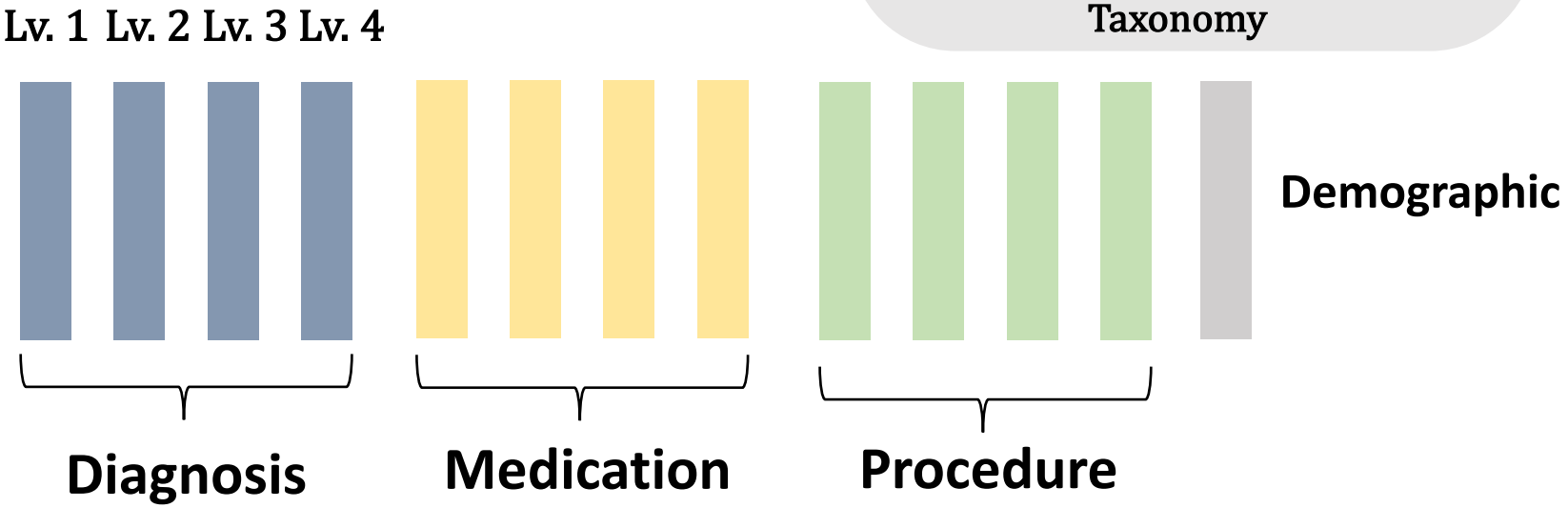


Method: Taxonomy guided patient embedding

- Use medical concept taxonomy to divide each concept into four levels
  - the Uniform System of Classification (USC)
- Three memory networks to store diagnosis, medications and procedures



$$\mathbf{m} = [\mathbf{m}_{\mathcal{D}}, \mathbf{m}_{\mathcal{O}}, \mathbf{m}_{\mathcal{P}}]$$
$$= [\mathbf{m}^1_{\mathcal{D}}, \ldots, \mathbf{m}^4_{\mathcal{D}}, \mathbf{m}^1_{\mathcal{O}}, \ldots, \mathbf{m}^4_{\mathcal{O}}, \mathbf{m}^1_{\mathcal{P}}, \ldots, \mathbf{m}^4_{\mathcal{P}}]$$



Method: Taxonomy guided patient embedding

- Augment medical codes with textual description:
  - Code 692.9 -> “Contact dermatitis and other eczema”

$$\widetilde{g}_t = MaxPool(BERT([w_1, ..., w_L]))$$

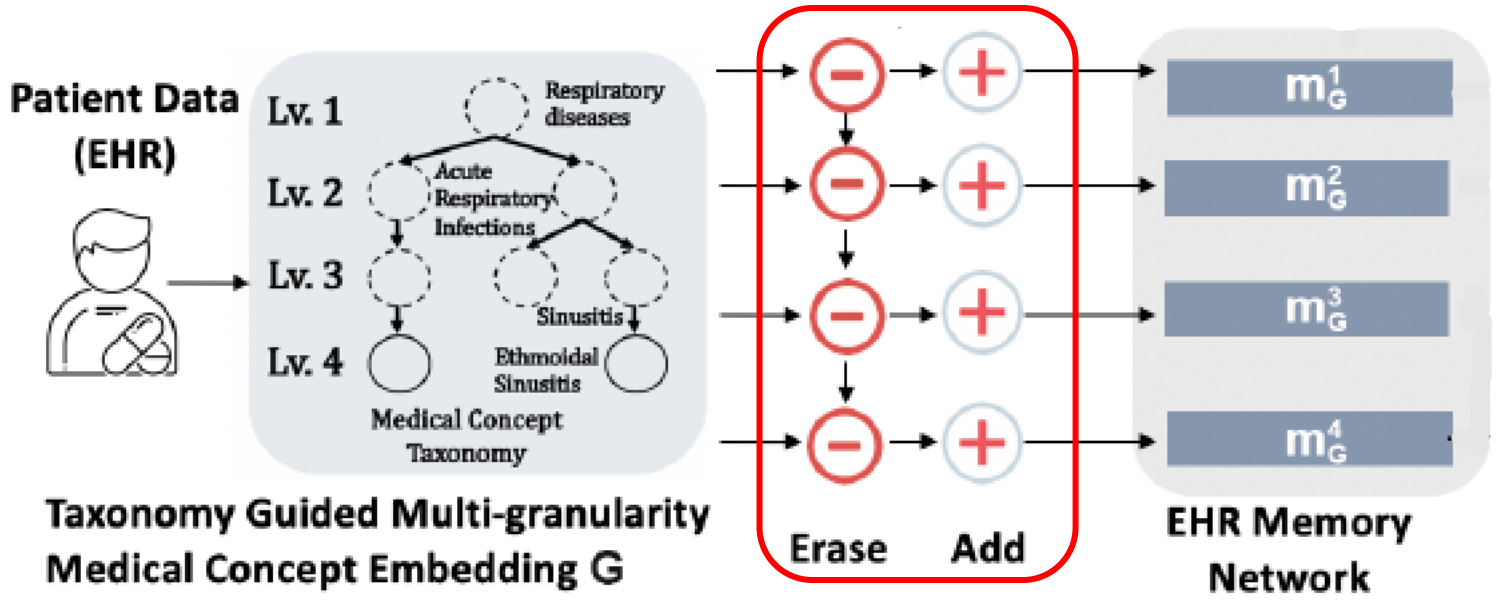
- Update memories at each visit
  - Erase-followed-by-add:

$$\mathbf{erase}_t = \sigma(\mathbf{W}_e \widetilde{g}_t^k + \mathbf{b}_e),$$

$$\mathbf{add}_t = tanh(\mathbf{W}_a \widetilde{g}_t^k + \mathbf{b}_a)$$

  - Update slot:

$$m_G^k \leftarrow m_G^k \odot (1 - \mathbf{erase}_t) + \mathbf{add}_t$$



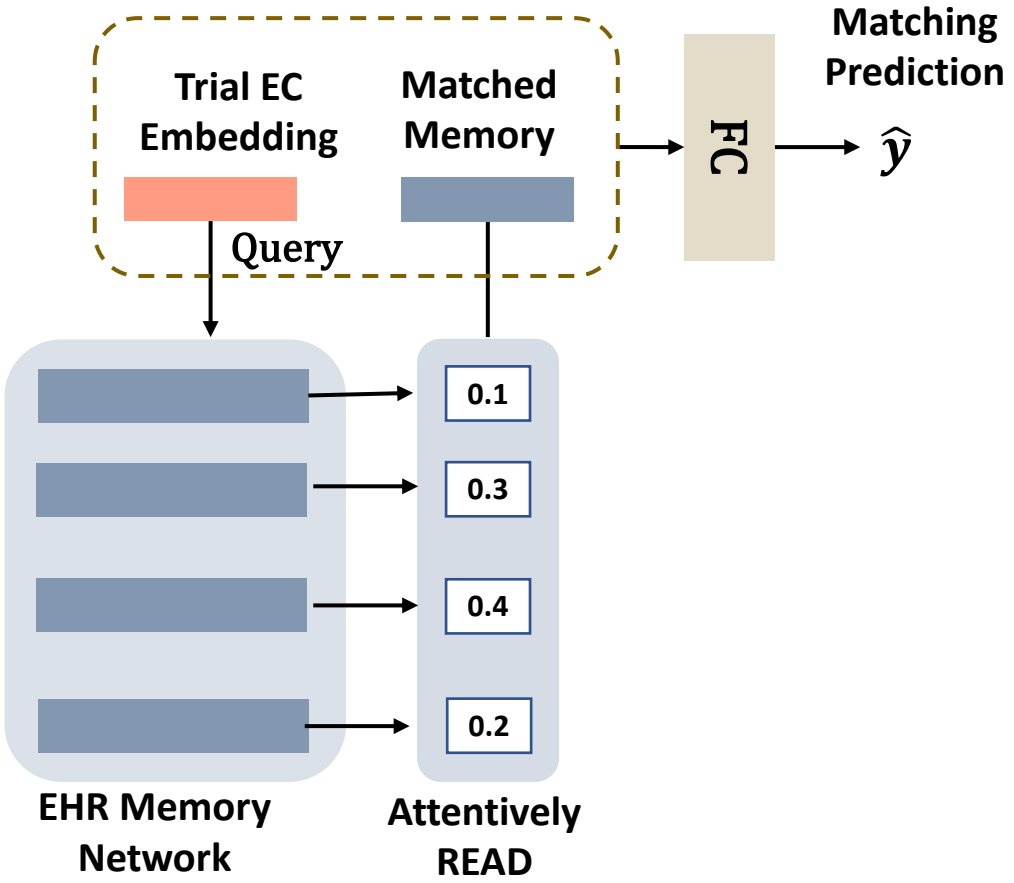
Method: Attentional record alignment and dynamic matching

- Let each EC correspond to the sub-memories

- Attentional matching
  - Trial EC embedding -> Query
  - Matched memory -> Response

$$a_{k,G} = \frac{exp(\boldsymbol{m}_G^{k\,\mathrm{T}} MLP(\boldsymbol{e}))}{\sum_{x\in\{\mathcal{D},\mathcal{O},\mathcal{P}\}} \sum_{i=1}^{4} exp(\boldsymbol{m}_x^{i\,\mathrm{T}} MLP(\boldsymbol{e}))}$$

$$\widetilde{\boldsymbol{m}} = \sum_{x\in\{\mathcal{D},\mathcal{O},\mathcal{P}\}} \sum_{i=1}^{4} a_{i,x} \boldsymbol{m}_x^i$$



Method: Explicit inclusion/exclusion criteria handling

- Classification loss:

$$\mathcal{L}_c = -(\boldsymbol{y}^{\mathrm{T}} log(\hat{\boldsymbol{y}}) + (1-\boldsymbol{y})^{\mathrm{T}} log(1-\hat{\boldsymbol{y}}))$$

- Inclusion/Exclusion loss:

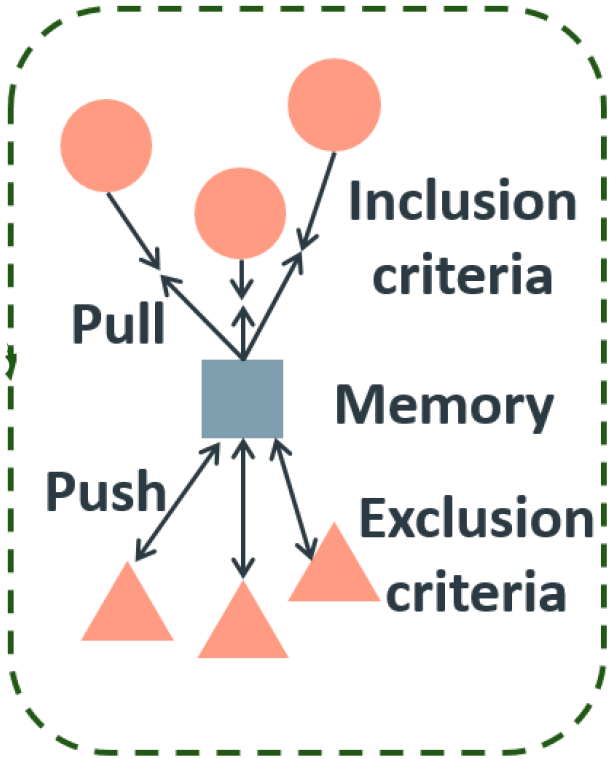
$$\mathcal{L}_d = \begin{cases} 1 - d(\boldsymbol{e}, \widetilde{\boldsymbol{m}}_I)), & if\ \boldsymbol{e}\ is\ \boldsymbol{e}_I \\ max(0, d(\boldsymbol{e}, \widetilde{\boldsymbol{m}}_E) - \alpha), & if\ \boldsymbol{e}\ is\ \boldsymbol{e}_E \end{cases}$$

-> 0

>= α

- Final loss:

$$\mathcal{L} = \mathcal{L}_c + \mathcal{L}_d$$



Composite Similarity
Loss Term

Content

- **Background & Motivation**
- **Problem Formulation**
- **Insights**
- **Solution**
- **Experiment**
  - Patient trial matching
  - Discussions
  - Case studies

Experiment

- **Dataset**
  - Clinical trial data
    - 590 trials from publicly available data source (clinicaltrials.gov)
    - 12,445 criteria-level EC statements
  - Patient EHR data
    - 83,371 patients from 2002 to 2018

Experiment: Patient trial matching

- Outperforms all baseline models across both trial level and criteria level matching in all evaluation metrics.

- 24.3% higher accuracy for trial level matching
- 8.8% higher accuracy and 4.7% higher AUROC for criteria level matching

	Model	Accuracy
Baselines	LSTM+GloVe	0.4294±0.010
	LSTM+BERT	0.5460±0.008
	Criteria2Query	0.6147±-
	DeepEnroll	0.6737±0.021
Reduced	COMPOSE-MN	0.7833±0.011
	COMPOSE-Highway	0.8102±0.009
	COMPOSE-$\mathcal{L}_1$	0.8212±0.010
Proposed	COMPOSE	**0.8373±0.012**

	Model	Accuracy	AUROC	AUPRC
Baselines	LSTM+GloVe	0.722±0.010	0.789±0.009	0.784±0.009
	LSTM+BERT	0.834±0.008	0.845±0.007	0.840±0.007
	DeepEnroll	0.869±0.012	0.936±0.013	0.947±0.011
Reduced	COMPOSE-MN	0.899±0.012	0.955±0.013	0.960±0.010
	COMPOSE-Highway	0.912±0.007	0.965±0.007	0.967±0.009
	COMPOSE-$\mathcal{L}_d$	0.939±0.010	0.976±0.009	0.973±0.007
Proposed	COMPOSE	**0.945±0.008**	**0.980±0.007**	**0.979±0.008**

Discussion: Varying length of patient record

- How COMPOSE performs in matching trials with patients who have short or long records?
  - Short (1 visit), Medium (2-3 visits), Long (≥ 4 visits)

- COMPOSE have robust performance

Model	Short	Medium	Long
LSTM+GloVe	0.4906	0.4328	0.0000
LSTM+BERT	0.5484	0.5512	0.5338
Criteria2Query	0.6833	0.5989	0.5172
DeepEnroll	0.6779	0.6797	0.6443
COMPOSE	**0.8420**	**0.8389**	**0.8350**

Discussion: Varying disease types

- How COMPOSE performs on different types of diseases?
  - Chronic, Oncology, Rare diseases

- Achieves 77.3% higher accuracy for chronic diseases
- Most baseline models fail to match correct patients for oncology and rare diseases

Model	Chronic Diseases	Oncology	Rare Diseases
LSTM+GloVe	0.1793	0.0000	0.0000
LSTM+BERT	0.2062	0.0000	0.0000
Criteria2Query	0.5103	0.2722	0.2292
DeepEnroll	0.3345	0.0000	0.0000
COMPOSE	**0.5931**	**0.6370**	**0.6875**

Discussion: Varying trial phases

- How COMPOSE performs on different phases?
  - Phase I, II, III

- 155% higher accuracy for phase I trials
- 19% higher accuracy for phase II trials
- 27% higher accuracy for phase III trials

Model	Phase I	Phase II	Phase III
LSTM+GloVe	0.0008	0.5865	0.3743
LSTM+BERT	0.0025	0.6045	0.4862
Criteria2Query	0.3025	0.6433	0.5870
DeepEnroll	0.2034	0.7493	0.6329
COMPOSE	**0.5189**	**0.8939**	**0.8005**

Discussion: Varying threshold of matching

- Some inclusion or exclusion criteria can be too strict to prevent finding patients
- How COMPOSE performs on varying thresholds?
  - 70%, 80%, 90%

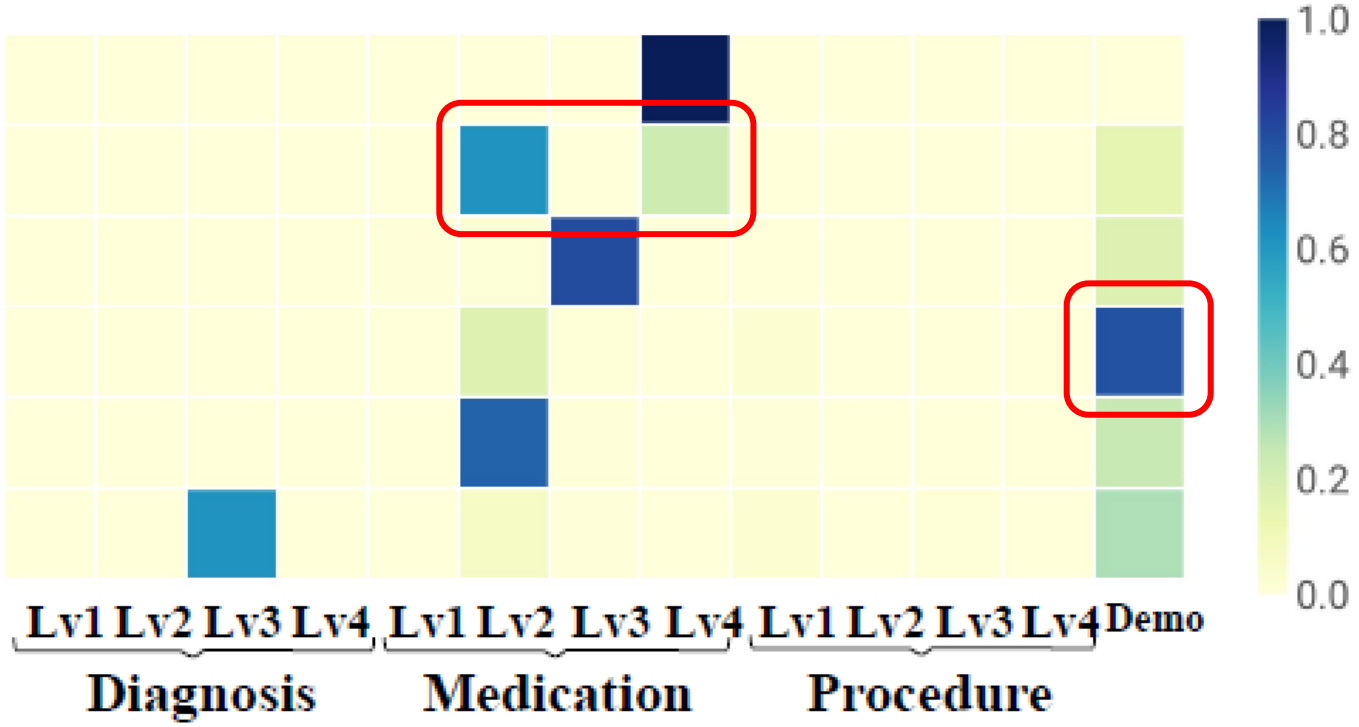
- COMPOSE have robust performance under all thresholds

Model	70% Matching	80% Matching	90% Matching
LSTM+GloVe	0.6218	0.5862	0.5057
LSTM+BERT	0.7231	0.6861	0.6238
DeepEnroll	0.8225	0.7963	0.7422
COMPOSE	**0.9334**	**0.9193**	**0.8915**

Case study: Attention weights on memory slots

- A trial on Cabozantinib which treats grade IV astrocytic tumors

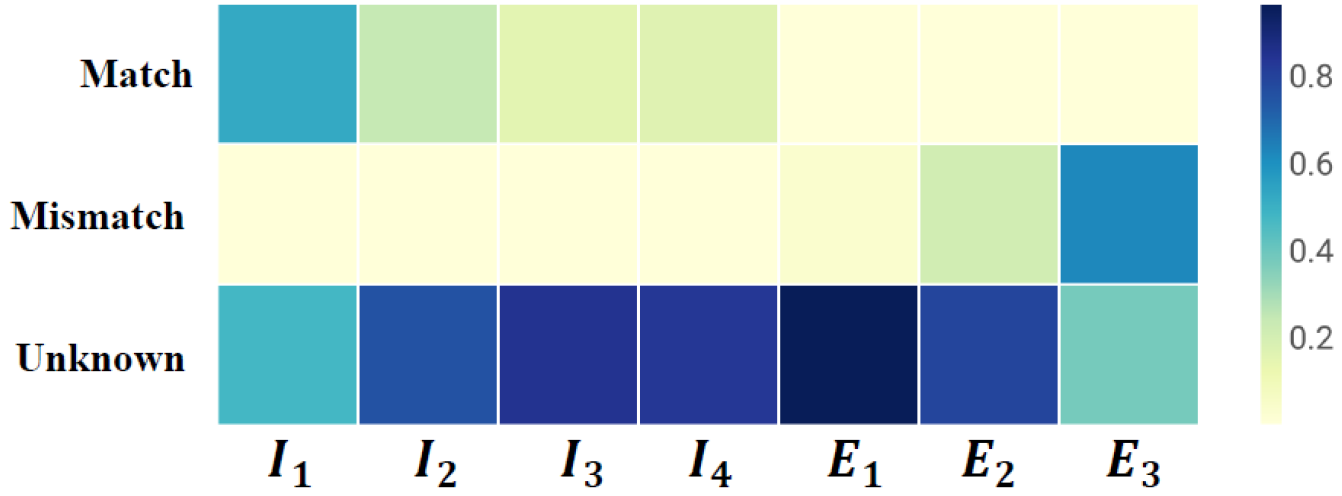
1. received temozolomide therapy
2. receiving warfarin (or other coumarin derivatives)
3. acute intracranial/ intratumoral hemorrhage.
4. pregnant or breast-feeding
5. serious intercurrent illness
6. inherited bleeding diathesis or coagulopathy



Case study: Failed case

- A trial for *Early Stage Non-Small Cell Lung Cancer*

- I2: Lung function capacity capable of tolerating the proposed lung surgery
- I3: Eastern Cooperative Oncology Group (ECOG) Performance Status of 0-1
- I4: Available tissue of primary lung tumor



Thank you!

COMPOSE: Cross-Modal Pseudo-Siamese Network for Patient Trial Matching

https://github.com/v1xerunt/COMPOSE

Junyi Gao[1], Cao Xiao[1], Lucas M. Glass[12], Jimeng Sun[3]

[1]Analytics Center of Excellence, IQVIA

[2]Department of Statistics, Temple University

[3]Department of Computer Science, University of Illinois Urbana-Champaign