A Neural Candidate-Selector Architecture for Automatic Structured Clinical Text Annotation

1

Author: Gaurav Singh, Iain J. Marshall, James Thomas, John Shawe-Taylor,
Byron C. Wallace

Sourse: CIKM ‘17

Advisor: Jia-Ling Koh

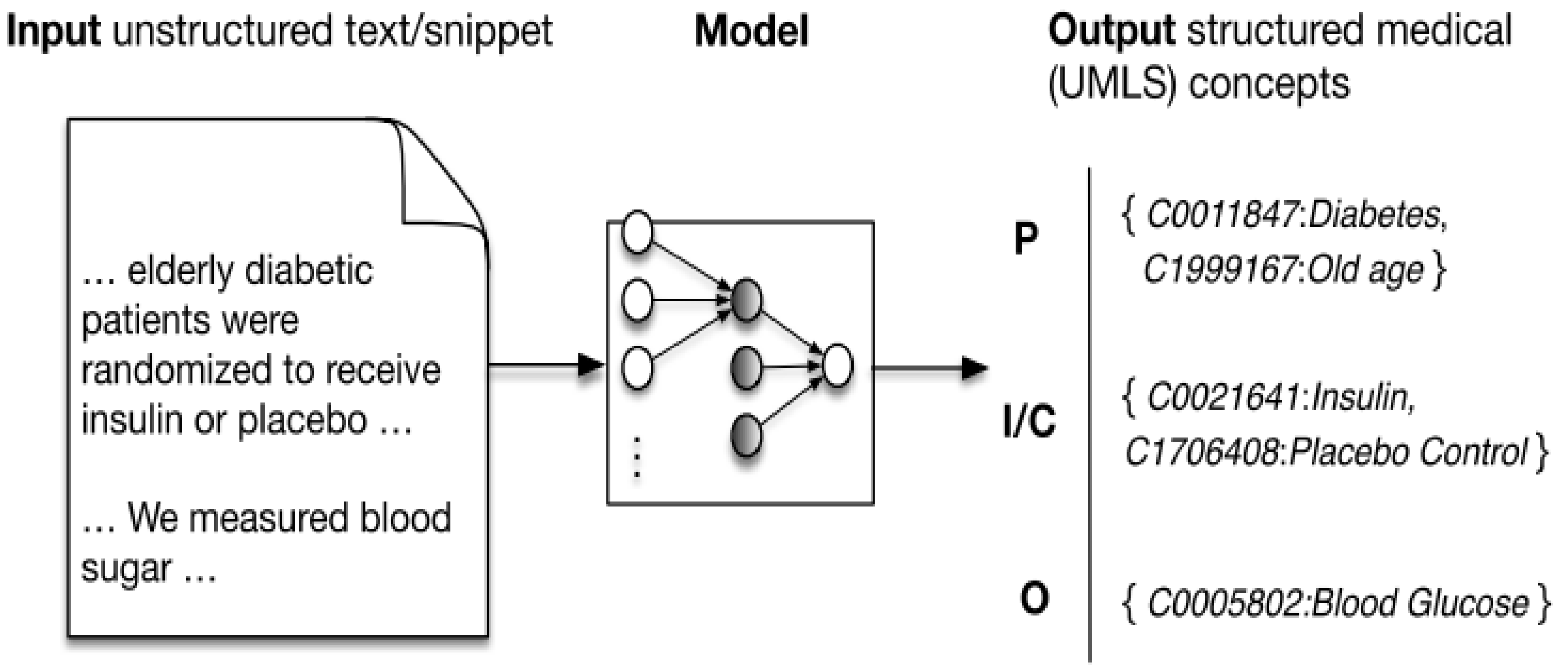
Speaker: Bo-wen Jiang

Date: 2017/12/26

outline

- <u>Introduction</u>
- Method
- Experiment
- Conclusion

Introduction



- Figure 1: Illustration of the annotation task. The output comprises **concepts drawn from the UMLS controlled medical vocabularly**, grouped into **terms that describe** the study Population, Interventions/Comparators and Outcomes.

Introduction - PICO

- PICO element
  - P : Population (**P**atient **or** **P**opulation or **P**roblem)
  - I/C : Interventions/Comparators
  - O: Outcomes

- Example :

  P : {糖尿病患者}

  I/C : {胰島素治療}

  O: {血糖降低}

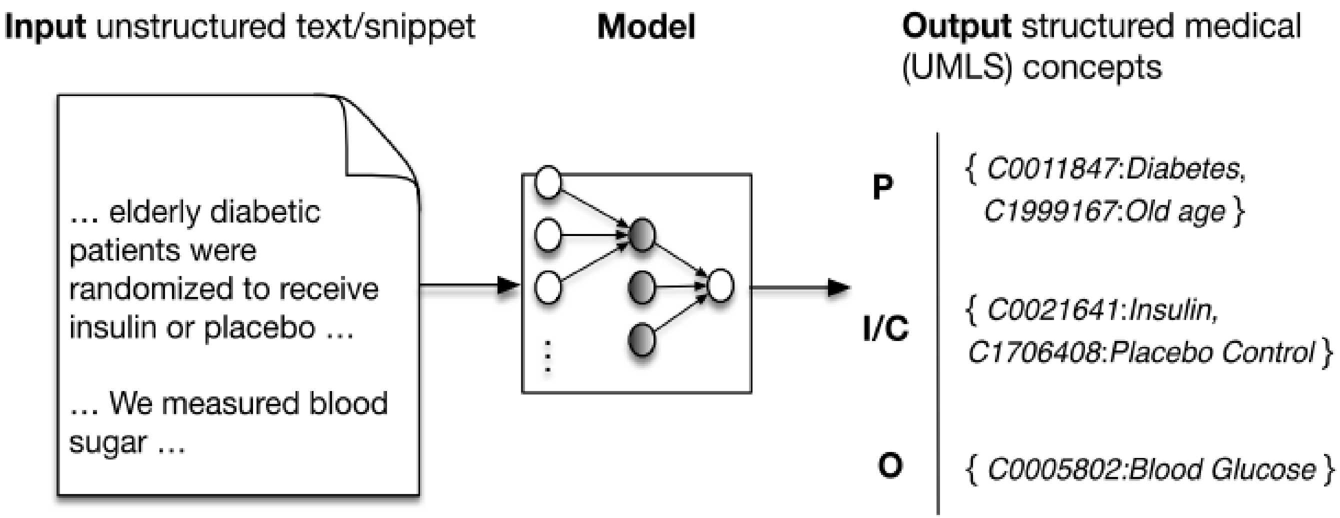
Introduction

- Motivation
  - To infer distinct sets of (ontological) concepts describing complementary clinically salient aspects of the underlying trials
  - i.e., the PICO

- Example :
  - P : {糖尿病患者}
  - I/C : {胰島素治療}
  - O: {血糖降低}

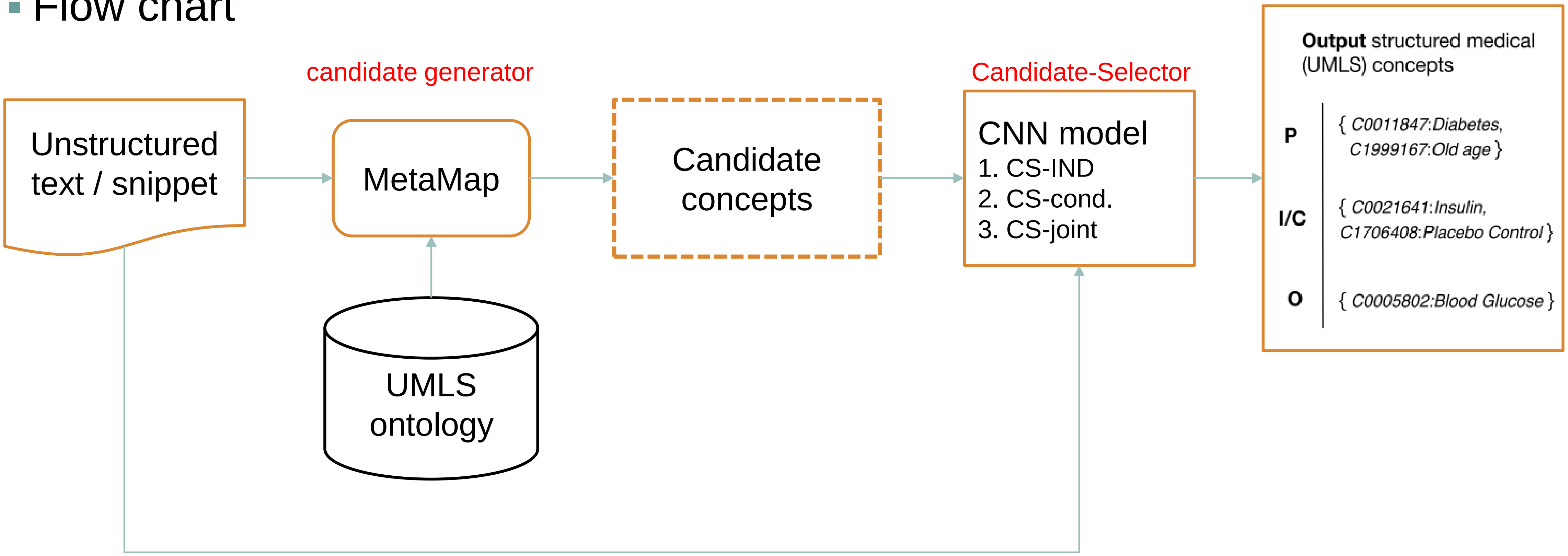
Introduction

- Research's purpose
  - aim to develop an automated approach to mapping from free-texts to distinct sets of terms from the UMLS corresponding to each PICO element.



Introduction

- Flow chart



outline

- Introduction
- Method
- Experiment
- Conclusion

Method

- **Identify**

Formally, denote an input text by x.

Then we run through this our candidate generator, g:

$$\mathcal{C} = g(\mathbf{x}) \qquad (1)$$

and the outputs are consumed by the selector, s:

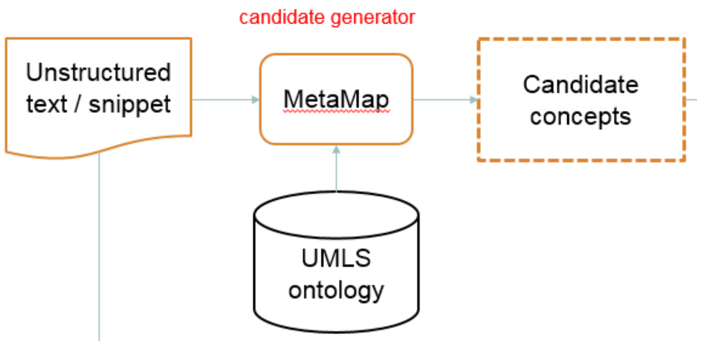
$$\mathcal{Y} = s(\mathcal{C}, \mathbf{x}) = s(g(\mathbf{x})). \qquad (2)$$

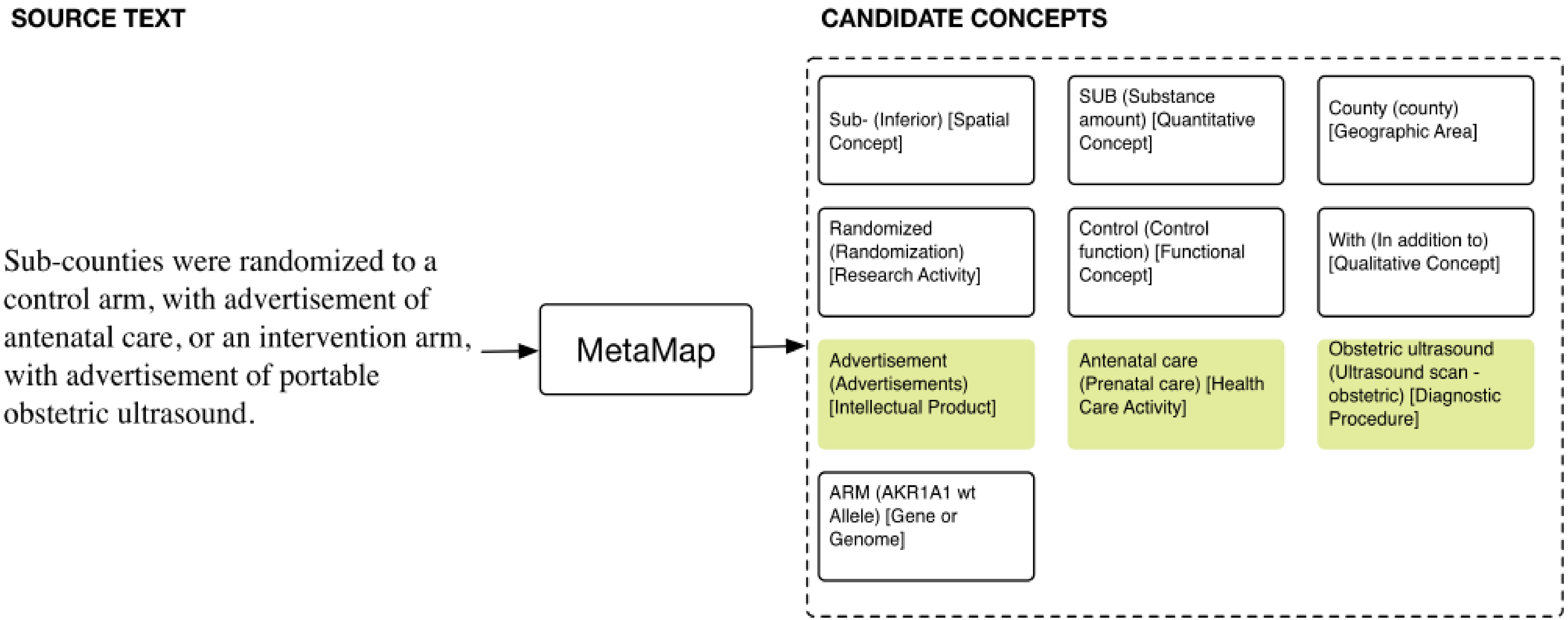
Here $\mathcal{Y}$ is assumed to be structured, i.e., include particular concepts corresponding to the PICO elements. Thus $\mathcal{Y} = \{\mathcal{Y}_{\mathrm{P}}, \mathcal{Y}_{\mathrm{I/C}}, \mathcal{Y}_{\mathrm{O}}\}$.

Method

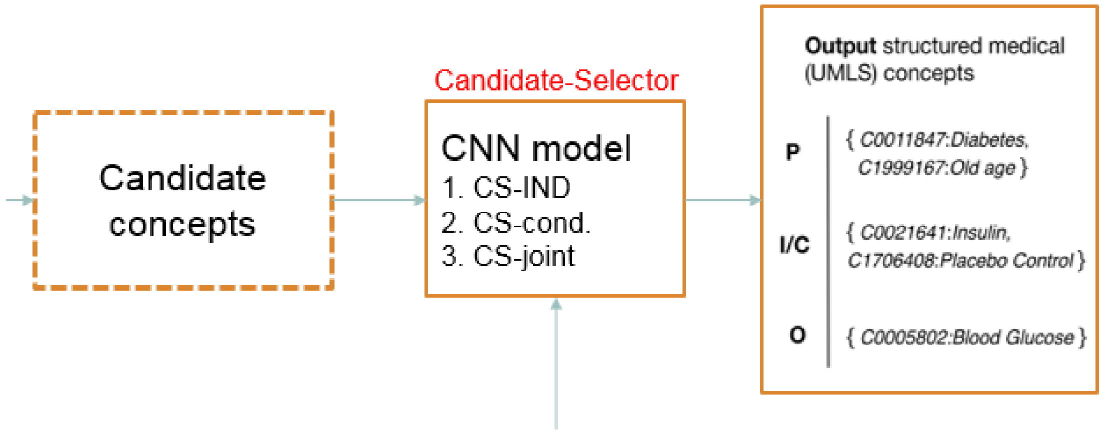
- Candidate Concepts Generation
  - MetaMap

    Use UMLS ontology





Method



- Candidate-Selector
  - CNN model
  - 1. CS-IND
    - PICO elements as independent
  - 2. CS-cond.
    - The population is not independent of the interventions and outcomes considered, as the former will clearly influence the latter.
  - 3. CS-joint
    - a fully joint approach to selecting P, I/C and O candidate terms.
    - a triplet of candidate concepts $(c_P, c_{I/C}, c_O)$

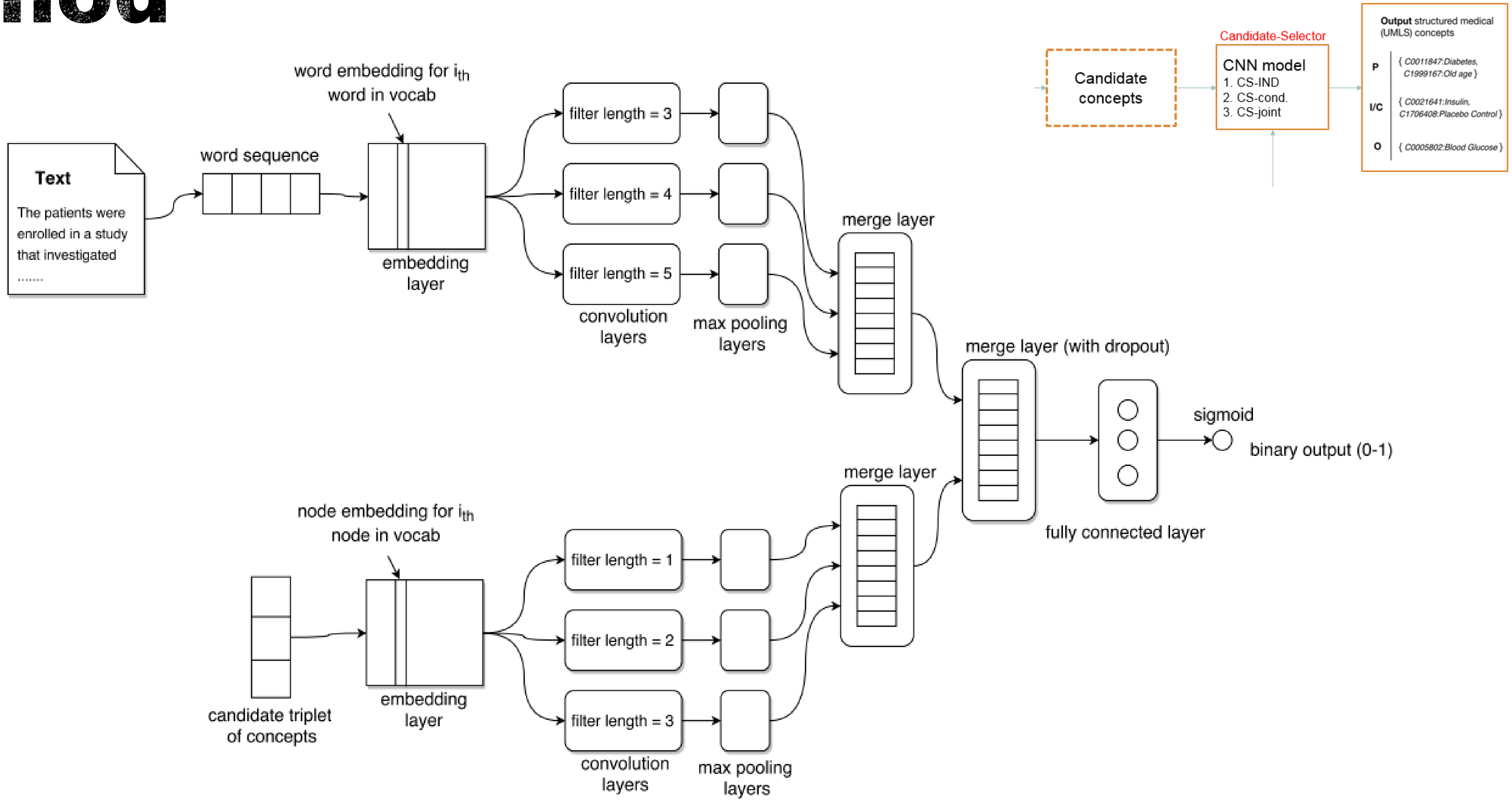
- Example :

  P :　{糖尿病患者}

  I/C : {胰島素治療}

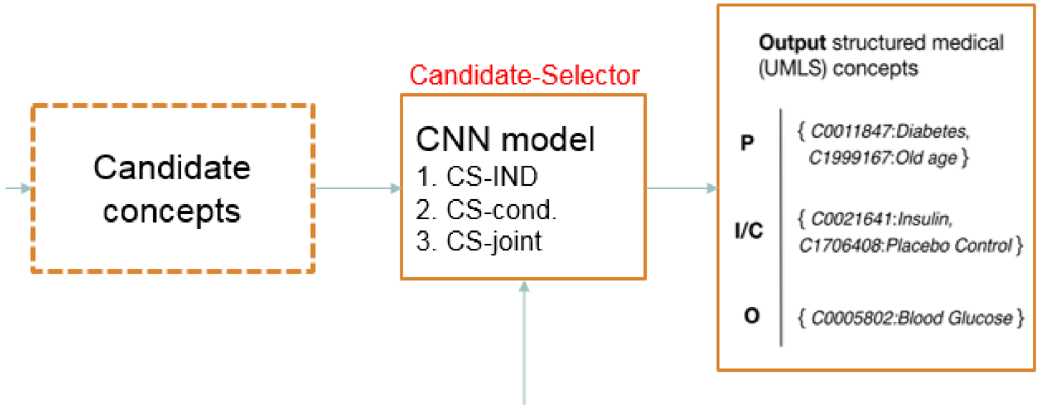
  O:　{血糖降低}

Method



- Figure 2: A schematic of our **selector network** variant CS-joint.

Method



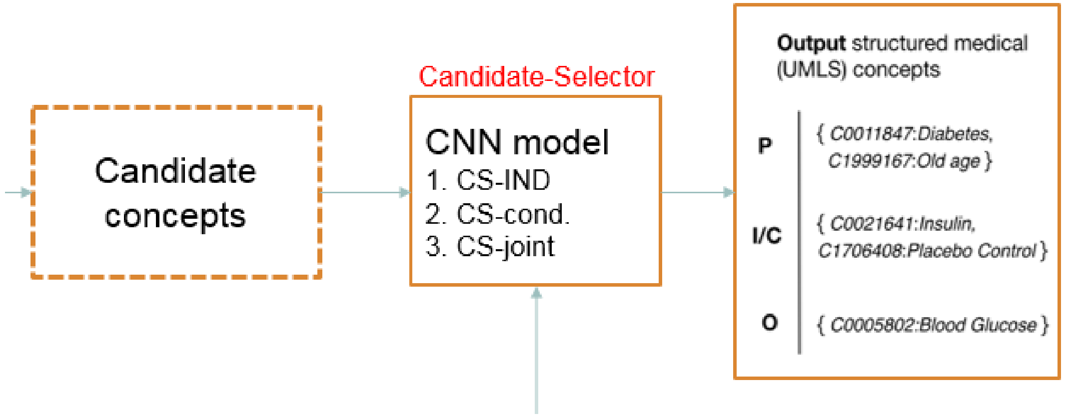
- Candidate-Selector
  - 1. CS-IND

    PICO elements as independent

$$P_e(\text{concept } j|\mathbf{x}_i^{(e)}, \mathbf{w}_o^{(e)}, \mathbf{C}^{(e)}, \mathbf{W}_h^{(e)}) = \sigma(\mathbf{w}_o^{(e)} \cdot \mathbf{v}_h)$$
$$\mathbf{v}_h = \lambda(\mathbf{W}_h^{(e)}[\mathbf{x}_i^{(e)} \oplus \mathbf{C}_j^{(e)}]) \quad (3)$$

- e: indexes PICO elements
- $w_o^{(e)}$: a weight vector parameterizing the output probability model
- $C^{(e)}$: concept embedding matrix
- $W_h^{(e)}$: a weight matrix for a hidden layer
- $xi^{(e)}$: a vector representation of input text i (induced via a CNN)
- $\lambda(.)$: activation function
- $\oplus$: denotes vector concatenation

Method



- Candidate-Selector
  - 2. CS-cond.

    The population is not independent of the interventions and outcomes considered, as the former will clearly influence the latter.

$$
\begin{aligned}
P_{\mathrm{I/C}}(\text{concept } j|\mathbf{x}_i^{(\mathrm{I/C})}, \mathbf{w}_o^{(\mathrm{I/C})}, \mathbf{C}^{(\mathrm{I/C})}, \mathbf{W}_h^{(\mathrm{I/C})}, \boxed{\mathbf{z}_i^{(P)}}) &= \sigma(\mathbf{w}_o^{(\mathrm{I/C})} \cdot \mathbf{v}_h) \\
\mathbf{v}_h = \lambda(\mathbf{W}_h^{(\mathrm{I/C})}[\mathbf{x}_i^{(\mathrm{I/C})} \oplus \boxed{\mathbf{z}_i^{(P)}} &\oplus \mathbf{C}_j^{(\mathrm{I/C})}])
\end{aligned} \tag{4}
$$

- $z_i^{(P)}$: vector representations of the population concepts selected by the model

outline

- Introduction
- Method
- Experiment
- Conclusion

Experiment

- Dataset
  - We use a real-world dataset provided by the Cochrane Collaboration.
  - http://www.cochrane.org/

samples (clinical trials)	4306
distinct population concepts	875
distinct intervention concepts	1115
distinct outcome concepts	1731
population concepts	9387
intervention concepts	5458
outcome concepts	13800

Table 1: Dataset statistics.

Experiment

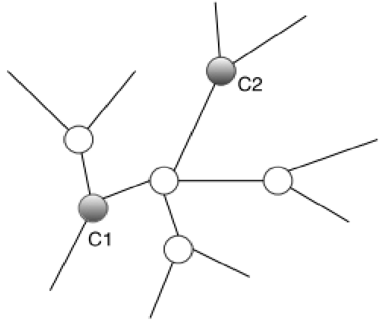
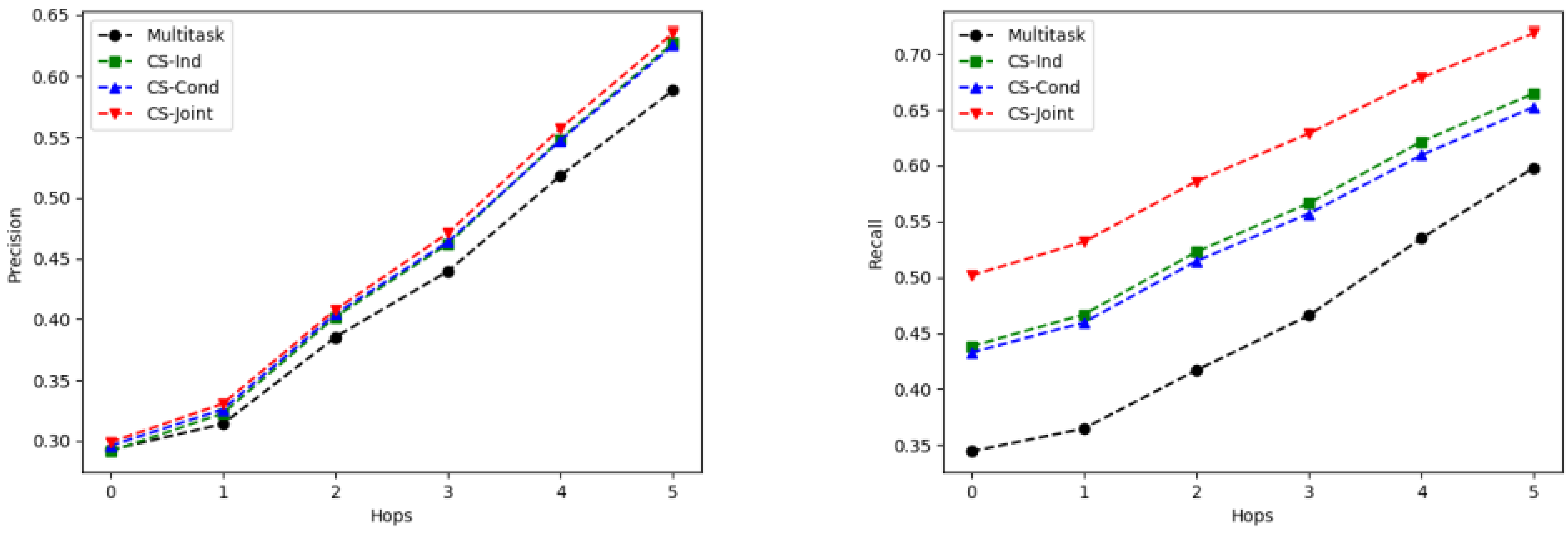


Figure 5: We consider two nodes at a distance of less than r hops as an 'r-hop match'; with this we compute the precision@r-hops and recall@r-hops metrics.

Category	Model	Precision	Recall	F1-score	Pr-2hops	Re-2hops	F1-2hops
Population	MetaMap	0.134	0.280	0.181	0.262	0.489	0.341
	Multitask	0.358	0.383	0.370	0.501	0.502	0.501
	CS-Ind	**0.385**	0.529	0.446	**0.557**	0.636	**0.594**
	CS-Cond	0.384	0.535	**0.447**	0.553	0.640	0.593
	CS-Joint	0.318	**0.594**	0.415	0.485	**0.709**	0.576
Interventions/Comparator	MetaMap	0.108	0.288	0.157	0.163	0.387	0.230
	Multitask	0.248	0.245	0.246	0.264	0.262	0.263
	CS-Ind	0.226	0.272	0.247	0.274	0.322	0.296
	CS-Cond	0.225	0.282	0.250	0.275	0.331	0.300
	CS-Joint	**0.265**	**0.421**	**0.326**	**0.314**	**0.473**	**0.378**
Outcomes	MetaMap	0.209	0.391	0.273	0.314	0.518	0.391
	Multitask	0.198	0.211	0.204	0.283	0.290	0.286
	CS-Ind	0.272	0.497	0.352	0.380	0.593	0.464
	CS-Cond	0.268	0.497	0.348	0.378	0.591	0.461
	CS-Joint	**0.279**	**0.503**	**0.359**	**0.38**	**0.595**	**0.468**

Table 2: Precisions, recalls and f1 measures realized by different models on the respective PICO elements. Best result for each element and metric are bolded. Models with prefix 'CS' (below the dotted lines) are variants of the Candidate-Selector approach we have proposed in this work. We should mention that r-hop refers to the case when we consider a match between two concepts that are at a distance of $\leq r$ hops.

Experiment



(a) Precision

(b) Recall

Figure 6: Average (over PICO elements) r-precisions (a) and recalls (b) for each method as a function of r (i.e., using increasingly relaxed metrics; r-precision) counts a predicted concept as matching the truth concept when it is $\leq r$ hops away.

outline

- Introduction
- Method
- Experiment
- Conclusion

Conclusion

- We then use a neural discriminative model to infer plausible triplets of concepts from the unstructured candidate set.
- We demonstrate that this model improves performance (compared to relevant baselines) on the important **task of automatically annotating biomedical literature with structured UMLS concepts**.