Knowledge-augmented Graph Neural Networks with Concept-aware Attention for Adverse Drug Event Detection

Ya Gao*, Shaoxiong Ji*, Pekka Marttinen

Aalto University

HELSINGIN YLIOPISTO

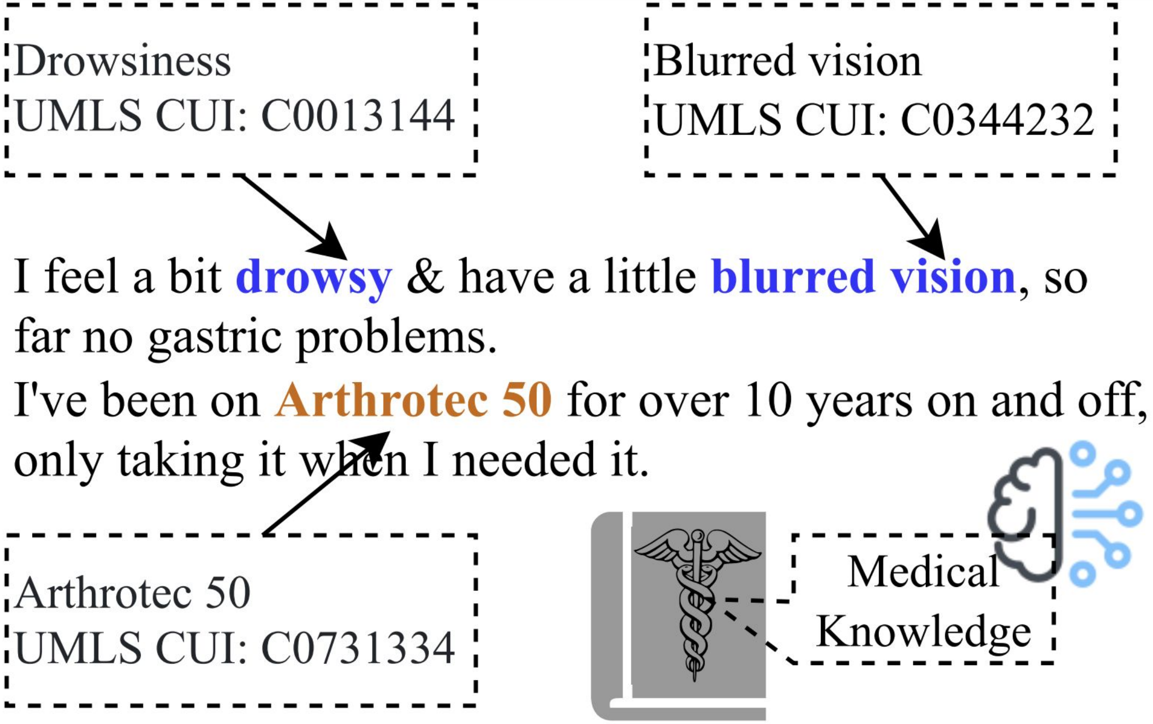
Adverse Drug Events (ADEs)

ADE: An adverse reaction resulting from improper drug use

How to detect ADEs?

- A clinical trial
- From users' voluntary ADE reports
- **Detect ADEs automatically using NLP techniques**

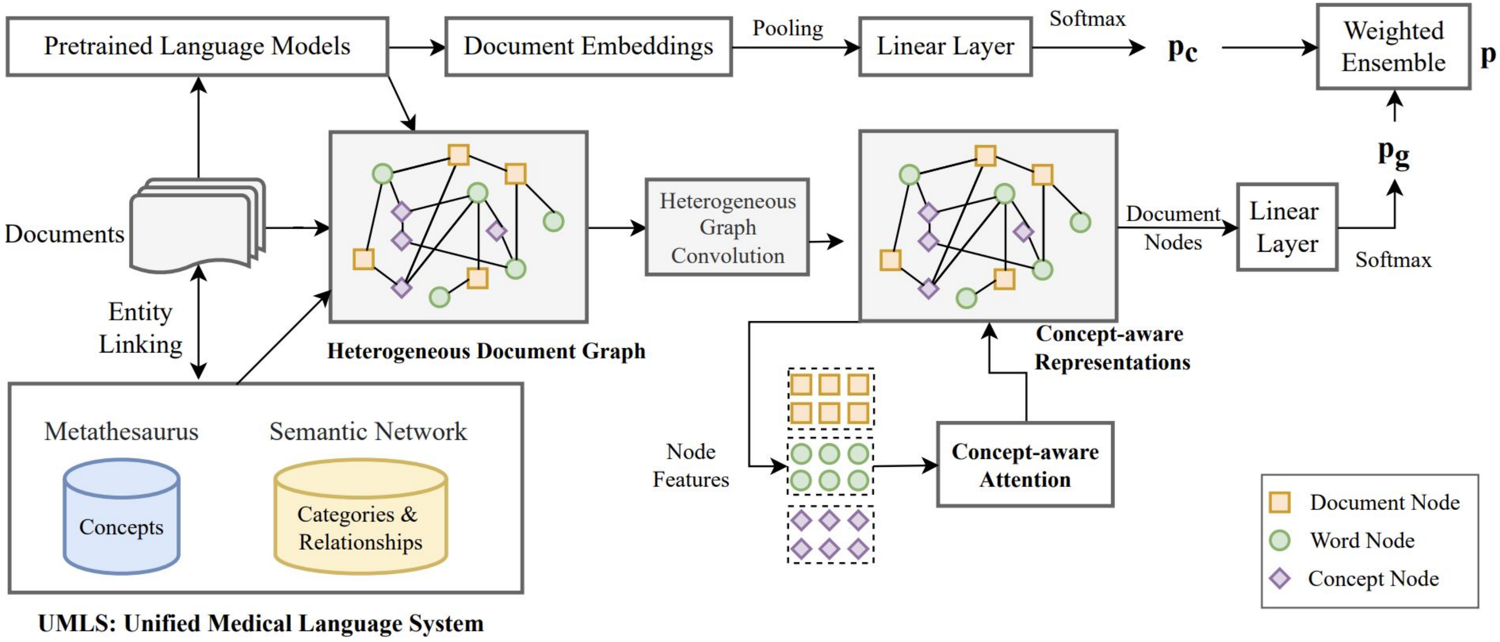
ADEs detection



Text mentions of adverse drug events include a plethora of drug names and adverse reactions, which can be mapped to **Unified Medical Language System (UMLS) concepts.**

Methods

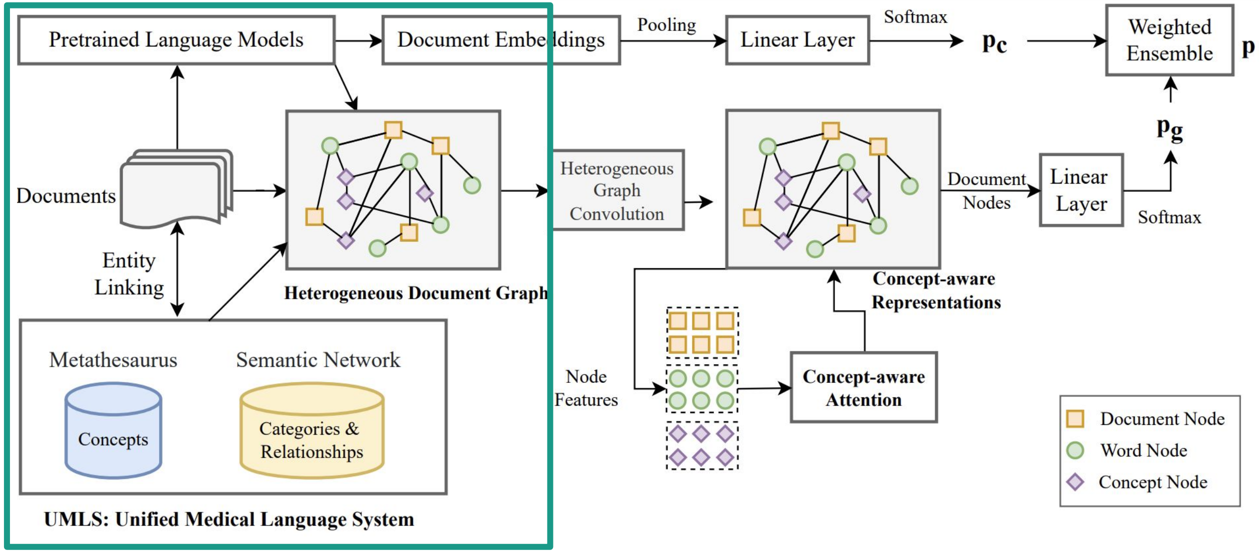
Knowledge-augmented Concept-Aware Graph Embeddings (KnowCAGE)



Methods

Word Nodes, Document Nodes and **Concept Nodes:**

- Map words or phrases in the document to Concept Unique Identifiers (CUIs);
- Use "preferred name", a short description or a synonym of this concept, to represent each CUI



Knowledge-augmented Graph Construction

Methods

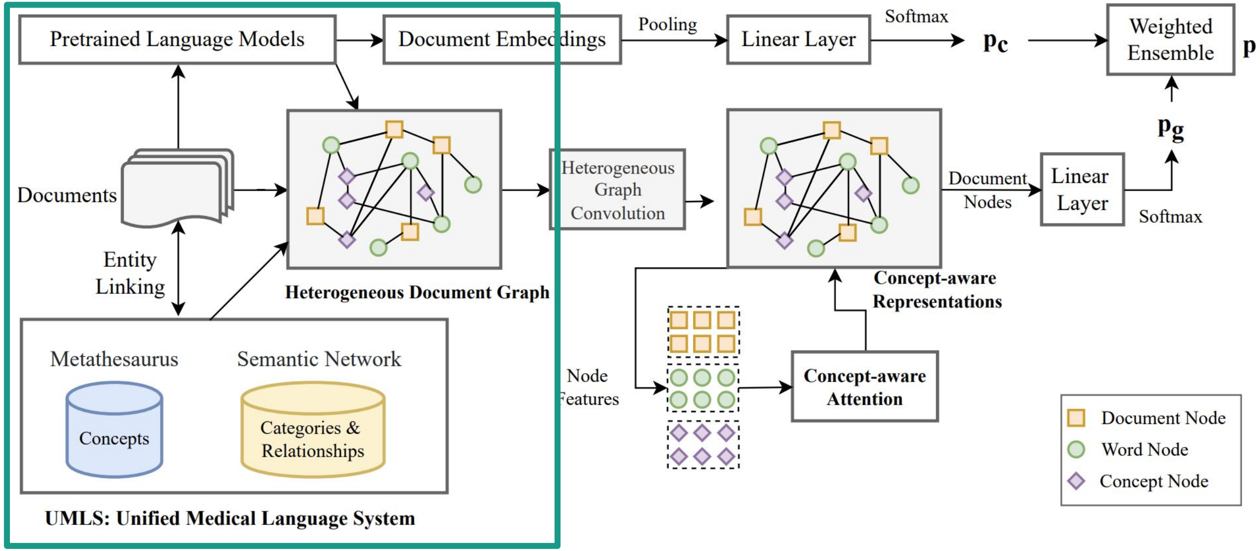
Adjacent Matrix A

$$\mathbf{A}_{ij} = \begin{cases} \mathrm{SIM}(i,j), & \mathrm{SIM} > 0;\ \text{i, j: word/concept} \\ \text{TF-IDF}_{\mathrm{ij}}, & \text{i: document, j: word/concept} \\ 0, & \text{otherwise} \end{cases}$$

We explore different measurement methods for SIM(i,j): L1 distance, L2 distance, Cosine distance, and Pointwise Mutual Information.

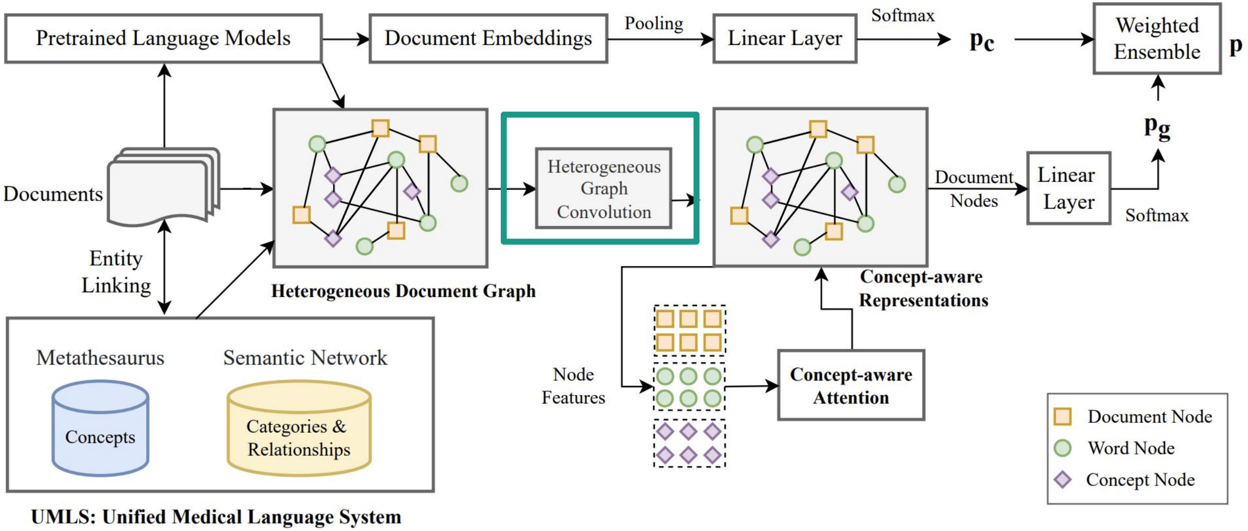
Initial feature matrix H

$$\mathbf{H}^{[0]} = \begin{pmatrix} \mathbf{H}_{doc} \\ \mathbf{0} \end{pmatrix}$$



Knowledge-augmented Graph Construction

Methods



Heterogeneous Graph Convolution

Different choices: GCN, GAT, DGCNN

Methods

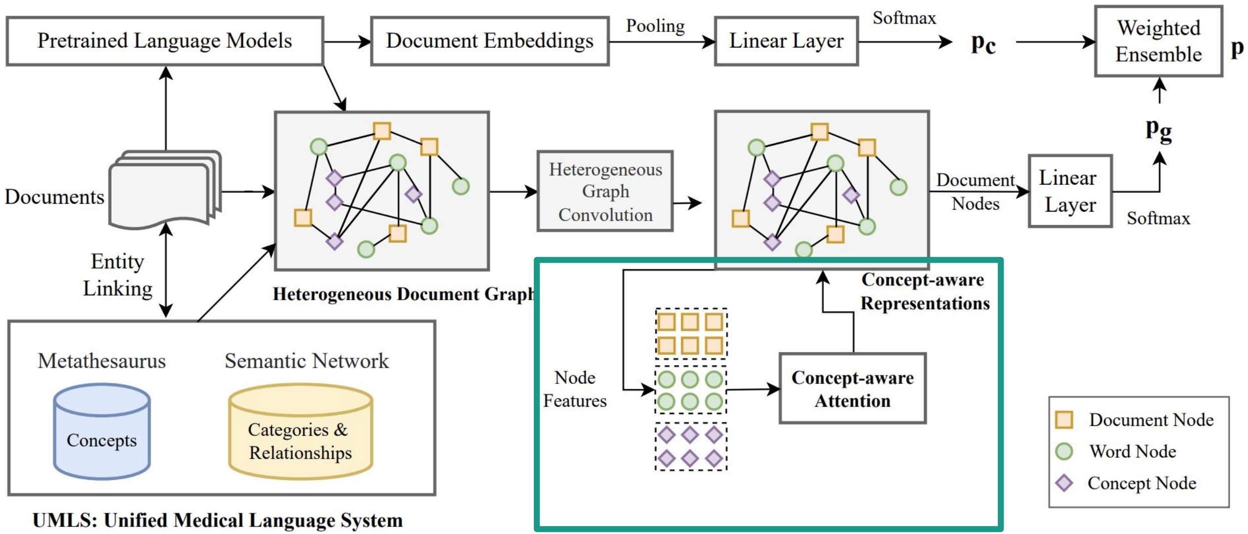
Different **query matrices** for different types of nodes accordingly:

$$Q_{ww}, Q_{dd}, Q_{cc}, Q_{wc}, Q_{wd}, Q_{dw}, Q_{dc}, Q_{cw}, Q_{cd}$$

For i-th document and j-th concept nodes:

$$\alpha_{ij} = \mathrm{Softmax}\left(\frac{(\mathbf{K}\mathbf{x}_j)^\top \mathbf{Q}_{cd}\mathbf{x}_i}{\sqrt{l}}\right)$$

$$\mathbf{h}_i = \sum_{j=1}^{n} \alpha_{ij}\mathbf{V}\mathbf{x}_j$$



Concept-aware Attention Mechanism

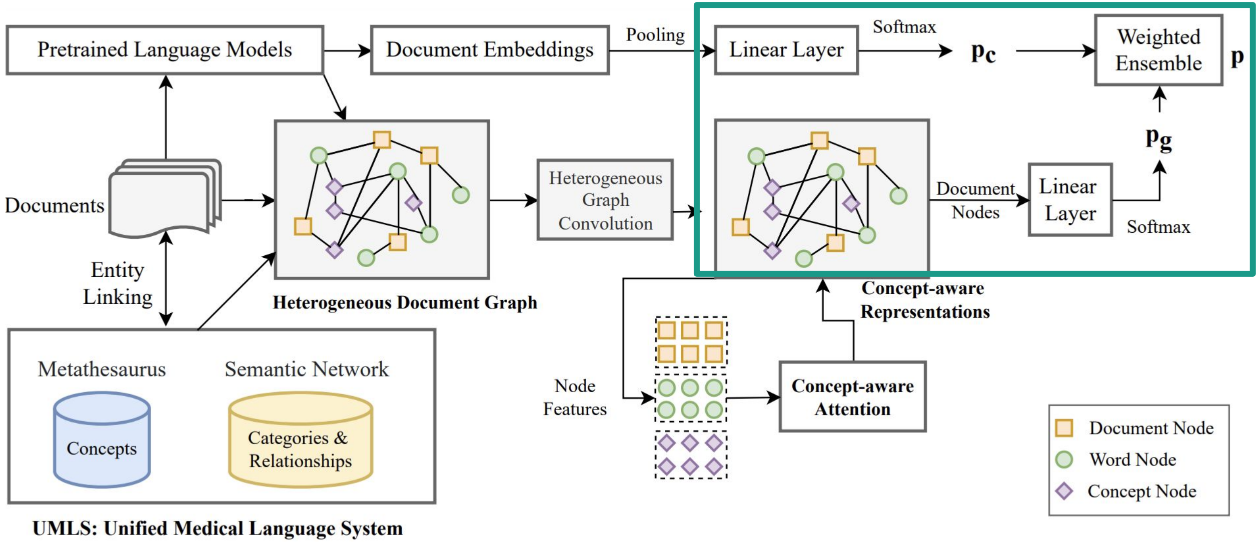
Methods

probability of classifying the document: p_g and p_c

p_g: from concept-aware document embeddings

p_c: from contextualized embeddings from the pretrained language model

$$\mathbf{p} = \lambda \mathbf{p}_g + (1 - \lambda)\mathbf{p}_c$$



Classification Layers and Model Training

Dataset

Table 1: A statistical summary of datasets

Dataset	Documents	ADE	non-ADE
SMM4H	2,418	1,209	1,209
TwiMed-Pub	1,000	191	809
TwiMed-Twitter	625	232	393
CADEC	7,474	2,478	4,996

- TwiMed-Twitter: from social media
- TwiMed-Pub: from biomedical publications
- SMM4H dataset: from Social Media Mining for Health Applications (#SMM4H) shared tasks
- CADEC dataset: patient-reported posts from a medical forum

Results

Results for two TwiMed datasets:

Models	TwiMed-Pub			TwiMed-Twitter		
	P (%)	R (%)	F1 (%)	P (%)	R (%)	F1 (%)
HTR-MSA (Wu et al., 2018)	75.0	66.0	70.2	60.7	61.7	61.2
IAN (Alimova and Solovyev, 2018)	87.8	73.8	79.2	83.6	81.3	82.4
CNN-T (Li et al., 2020)	81.3	63.9	71.6	61.8	60.0	60.9
MSAM (Zhang et al., 2019)	85.8	85.2	85.3	74.8	**85.6**	79.9
ATL (Li et al., 2020)	81.5	67.0	73.4	63.7	63.4	63.5
CGEM (Gao et al., 2022)	88.4	85.0	86.7	84.2	83.7	83.9
GPT-4 (OpenAI, 2023)	89.2	85.4	87.0	76.1	85.3	80.1
KnowCAGE (GCN)	88.8	**85.8**	**87.3**	84.1	84.0	84.0
KnowCAGE (GAT)	**89.6**	83.4	86.4	**84.8**	84.1	**84.4**
KnowCAGE (DGCNN)	88.7	83.7	86.1	83.5	84.1	83.8

Results for CADEC dataset:

Models	P (%)	R (%)	F1 (%)
HTR-MSA (Wu et al., 2018)	81.8	77.6	79.7
CNN-T (Li et al., 2020)	84.8	79.4	82.0
ATL (Li et al., 2020)	84.3	81.3	82.8
ANNSA (Zhang et al., 2021)	82.7	83.5	83.1
GPT-4 (OpenAI, 2023)	68.6	83.0	75.1
KnowCAGE (GCN)	86.6	90.8	88.7
KnowCAGE (GAT)	87.1	89.7	88.4
KnowCAGE (DGCNN)	**87.1**	**93.9**	**90.4**

Results for SMM4H dataset:

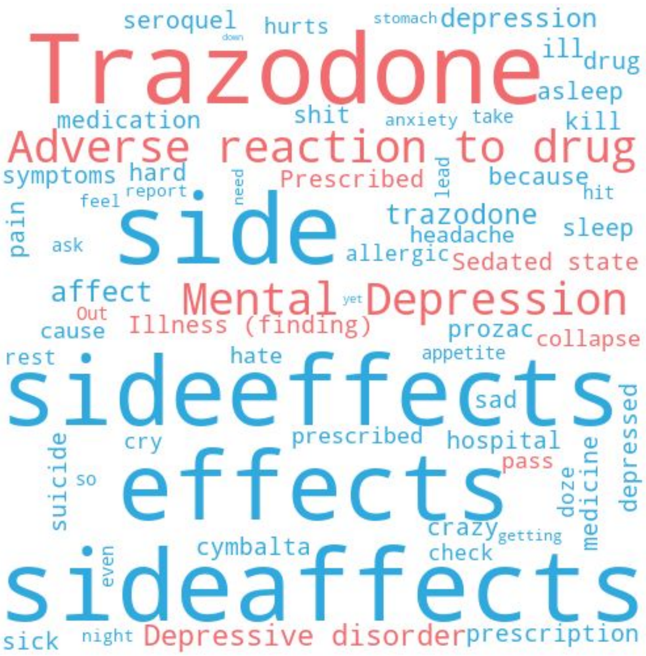
Models	P (%)	R (%)	F1 (%)
BERTweet-LSTM (Kayastha et al., 2021)	81.2	86.2	83.6
RoBERTa-aug (Pimpalkhute et al., 2021)	82.1	85.7	84.3
BERT-LSTM (Yaseen and Langer, 2021)	77.0	72.0	74.0
CGEM (Gao et al., 2022)	86.7	93.4	89.9
GPT-4 (OpenAI, 2023)	62.4	96.7	75.9
KnowCAGE (GCN)	85.3	95.9	90.3
KnowCAGE (GAT)	85.4	94.6	89.8
KnowCAGE (DGCNN)	**87.2**	**97.0**	**91.8**

Effectiveness of the Concept-Aware Attention

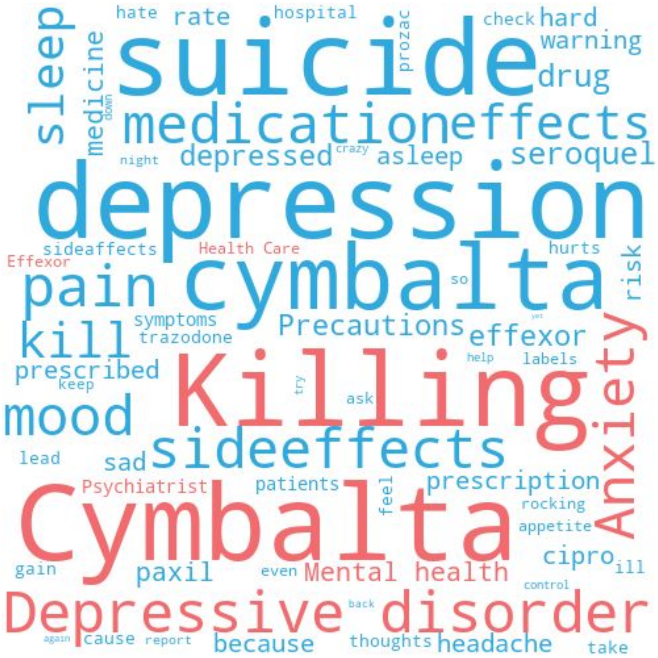
Datasets	dot-product attention			structured attention			concept-aware attention		
	P (%)	R (%)	F1 (%)	P (%)	R (%)	F1 (%)	P (%)	R (%)	F1 (%)
SMM4H	86.0	94.2	89.9	85.0	95.4	89.3	89.9	97.0	**91.8**
TwiMed-Pub	87.9	84.5	86.2	88.9	82.9	85.8	88.8	85.8	**87.3**
TwiMed-Twitter	84.5	82.2	83.4	83.0	81.8	82.4	84.8	84.1	**84.4**
CADEC	86.7	89.1	87.9	84.3	88.2	86.2	87.1	93.9	**90.4**

Concept-aware attention consistently achieves the best F1 score on four datasets.

Case Study - Node Cloud



positive example (with ADE): "debating on taking a trazodone and literally passing out for the day"



negative example ('without ADE'): "i don't get why they make stuff like cymbalta to decrease the suicide rate, when there are bunch of precautions about how it may kill you."

Red:
Concept Nodes

Blue:
Word Nodes

Node sizes: contribution when classifying the document

Conclusion

- Our work focuses on the pragmatic application of medical KG (UMLS), and proposes a simple yet effective integration method of external medical information;
- We introduce concept-aware attention, which is designed to emphasize information in various levels that holds significant connections to the original text. This also contributes to improved model explainability.

Thanks you!