Medical Recommendation on Knowledge Graph Diffusion Model

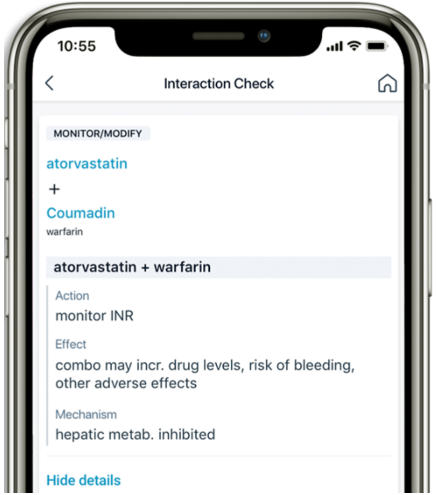
Yang Zhao
PhD
CIS Department

Background

- Medical Recommendation System
  - ✓ Drug Recommendation System
  - ✓ Diagnostic Recommendation System
  - ✓ Lifestyle and Preventative Health System
  - ✓ Clinical Decision Support Systems

https://images.app.goo.gl/ixRJkQ7ZEnmX49vJ8

https://images.app.goo.gl/LofUbVs1zMTWh7CG8



https://images.app.goo.gl/hrNcEY2McPfizaTK8

Epocrates: https://www.epocrates.com/

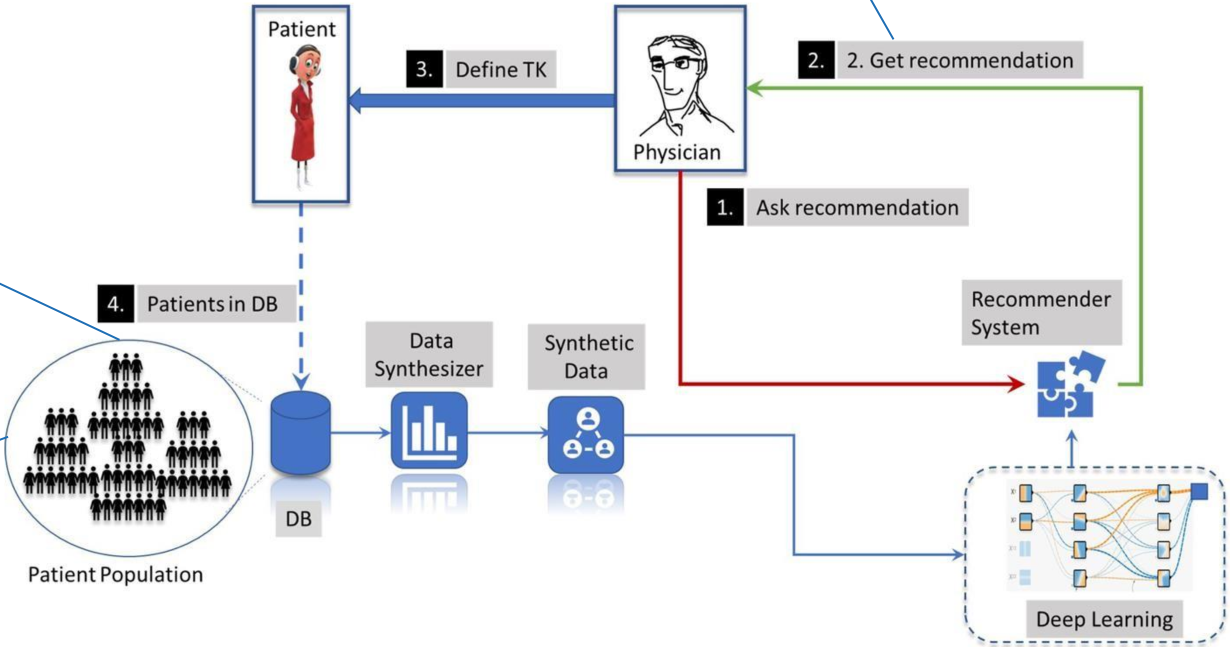
Background

- Medical Recommendation System

(P-I)Graph models, LLMs

MIMIC 3: Electronic health records (EHRs), they are often fragmented, unstructured, sparse, and difficult to analyze due to the heterogeneity of the data sources and the sheer volume of information.

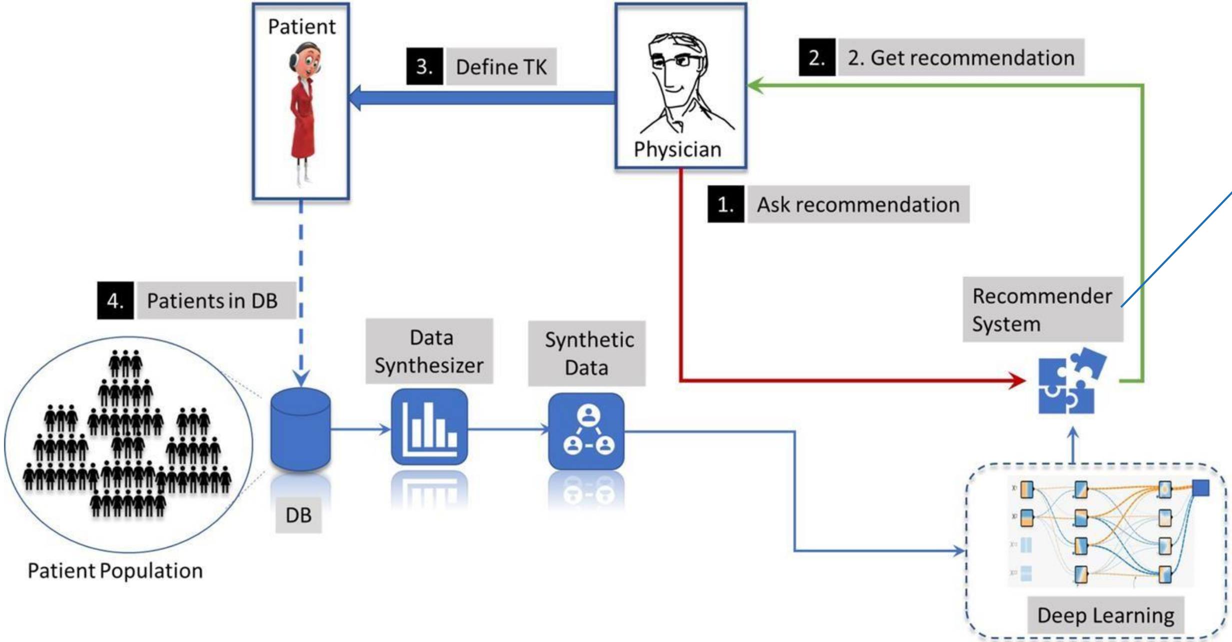
Drug-Drug interaction



[Medical recommender systems based on continuous-valued logic and multi-criteria decision operators, using interpretable neural networks https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-021-01553-3/figures/1](https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-021-01553-3/figures/1)

Background

- Medical Recommendation System



- ✓ noisy implicit feedback between user and KG; personal recommendation
  - ✓ alleviate the noise issues from data cleaning perspective (resampling or reweighting)
- ✓ from model view: denoising by diffusion models
- ✓ Idea: controlled Gaussian noises in the forward process and iteratively removes noises in the reverse denoising process

[Medical recommender systems based on continuous-valued logic and multi-criteria decision operators, using interpretable neural networks https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-021-01553-3/figures/1](https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-021-01553-3/figures/1)

Related Works

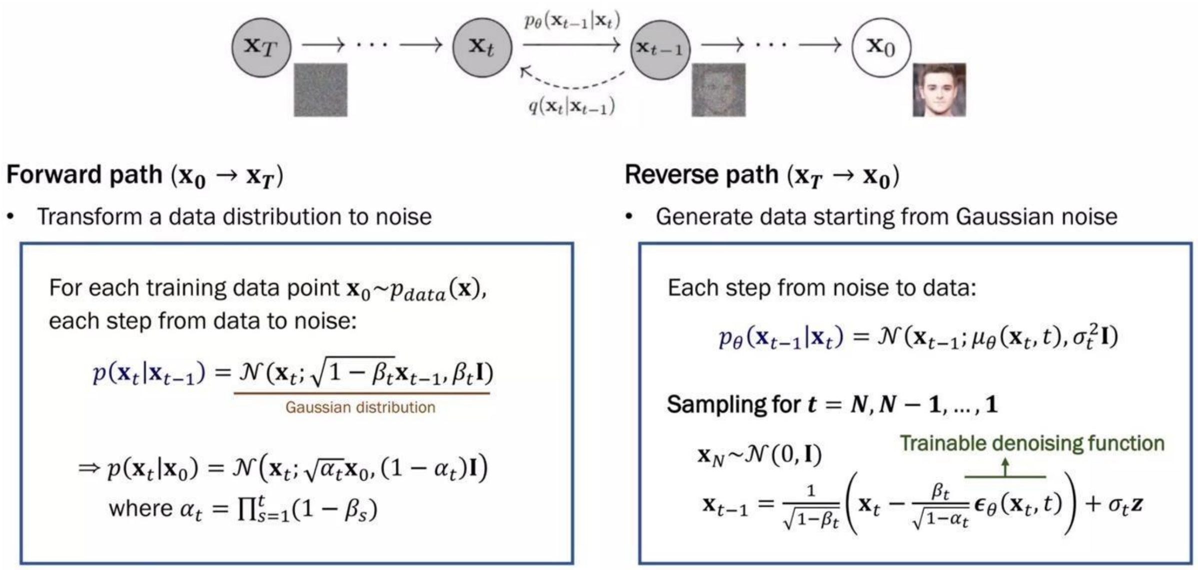
- Knowledge-aware Recommender System
  - ✓ Embedding-based
  - ✓ Path-based
  - ✓ GNN-based: KGCN: consider items' fixed number of neighborhoods for aggregation

    KGAT: assign weights to important neighborhoods

    KGIN: integrate user's preferences as embedding
- Self-supervised learning + data augmentation
  - ✓ To address data sparsity and improve recommendation performance
  - ✓ Generate new view of use/item representations by maximize the differences between positive and negative pairs
  - ✓ BERT in NLP, mask items to predict, enforce model learn contextual relations

Related Works

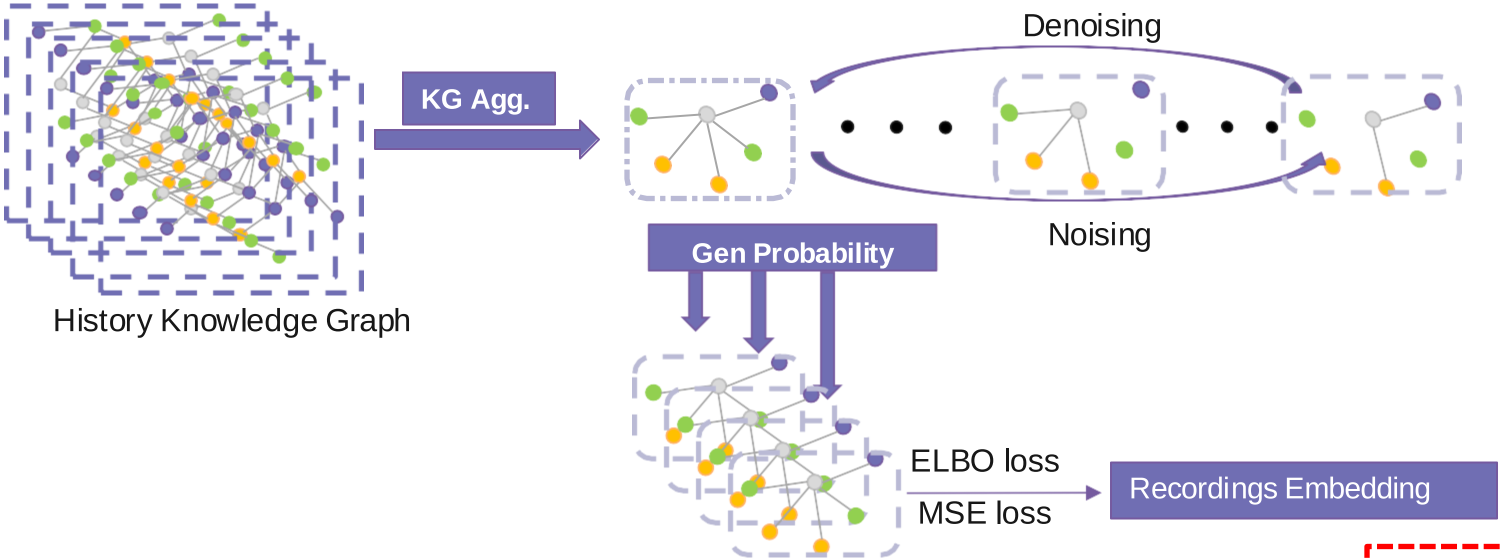
- Diffusion models

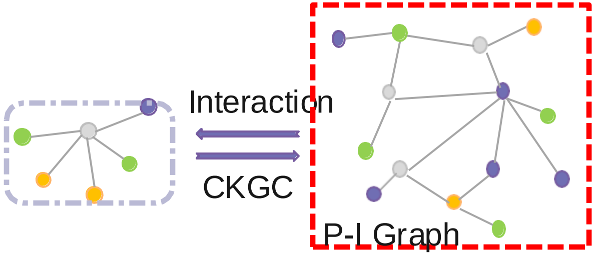


https://www.slideshare.net/slideshow/diffusion-deformable-model-for-4d-temporal-medical-image-generation-253730447/253730447#4

- Applications
  - ✓ GraphGDP: generate new graphs by continuous-time noise;
  - ✓ Diffusion probabilistic model: Recommendation systems

Method





- Entity knowledge aggregation module
- Knowledge graph diffusion model
- KG data augmentation

Method

Entity Knowledge Aggregation Module

For knowledge graph G_k, we have:

$$\mathbf{x}_i = \text{Drop}\left(\text{Norm}\left(\mathbf{x}_i + \sum_{e \in \mathcal{N}_i} \alpha(e, r_{e,i}, i)\mathbf{x}_e\right)\right),$$

$$\alpha(e, r_{e,i}, i) = \frac{\exp\left(\text{LeakyReLU}\left(r_{e,i}^\top W[\mathbf{x}_e \| \mathbf{x}_i]\right)\right)}{\sum_{e \in \mathcal{N}_i} \exp\left(\text{LeakyReLU}\left(r_{e,i}^\top W[\mathbf{x}_e \| \mathbf{x}_i]\right)\right)},$$

- N_i: the neighboring entities of record i
- $X_i \in R^d$: embedding of record
- $\alpha\ \ e, r_{e,i}, i$: the estimated records-specific and relation specific attentive relevance during knowledge aggregation process, to capture distinct semantics of relationships between i and e.
- $r_{e,i}$: relation type
- $X_e \in R^d$: embedding of patient
- $Norm$: normalization operation
- $W \in R^{d\times 2d}$: customize the input i and e.
- $LeakyReLU$: non-linear activation function

- Random dropout before aggregation: sparse KG has potential to significantly enhance the performance of recommender system.

Method

Diffusion with KG

Objective: Generate G'_k from G_k

Setup: patient i has relations $z_i = [z_i^0, z_i^0, \cdots, z_i^{|\mathcal{E}|-1}]$ with records set $\mathcal{E}$, where $z_i^e = 0 \ or \ 1$

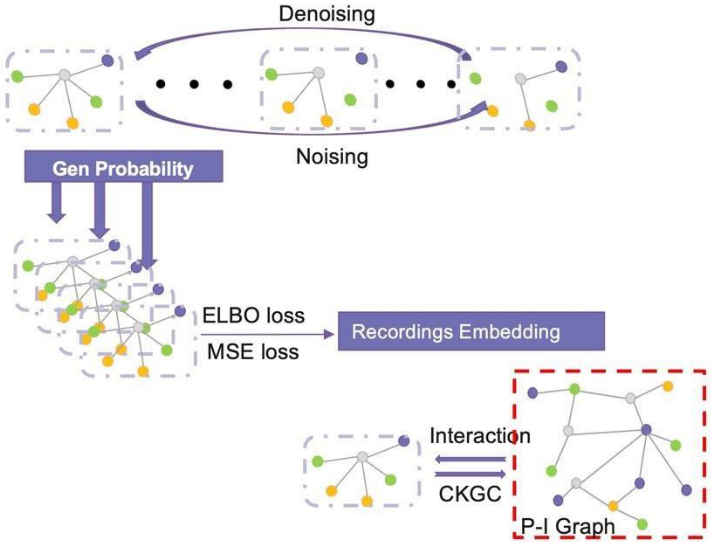
✓ Forward Process

$$q(\chi_t|\chi_{t-1}) = \mathcal{N}(\chi_t; \sqrt{1-\beta_t}\,\chi_{t-1}, \beta_t \boldsymbol{I})$$

Re-parameterize by two independent Gaussian noise

$$q(\chi_t|\chi_0) = \mathcal{N}(\chi_t; \sqrt{\bar{\alpha}_t}\,\chi_0, (1-\bar{\alpha}_t)\boldsymbol{I}), \bar{\alpha}_t = \prod_{t'=1}^{t}(1-\beta_{t'})$$

$$\chi_t = \sqrt{\bar{\alpha}_t}\,\chi_0 + \sqrt{1-\bar{\alpha}_t}\,\epsilon, \epsilon \sim \mathcal{N}(0, \boldsymbol{I}).$$



- x_0: initial state; original adjacency matrix z_i of the record
- $x_{1:T}$: in a Markov chain by gradually adding Gaussian noise in T steps
- N: Gaussian noise distribution
- $\beta_t \in (0,1)$: control the scale of Gaussian noise
- Linear noise scheduler:

Method

Diffusion with KG:

✓ Reverse Process

$$p_\theta(\chi_{t-1}|\chi_t) = \mathcal{N}(\chi_{t-1}; \mu_\theta(\chi_t, t), \Sigma_\theta(\chi_t, t)).$$

- $\mu_\theta(x_t, t)$: neural network parameterized by θ
- $\sigma\theta(xt,\ t)$: covariance of Gaussian distribution

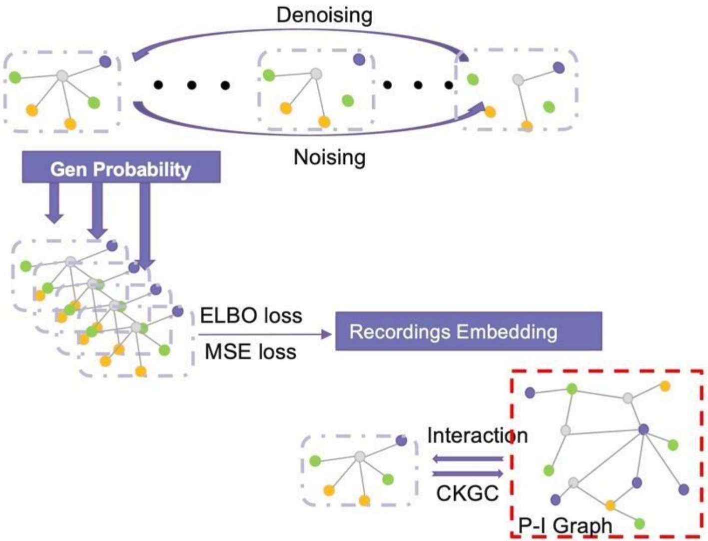
✓ Optimization of KG Diffusion Process

Maximize the Evidence Lower Bound (ELBO):

$$\log p(\chi_0) \geq \mathbb{E}_{q(\chi_1|\chi_0)}[\log p_\theta(\chi_0|\chi_1)] - \sum_{t=2}^{T} \mathbb{E}_{q(\chi_t|\chi_0)}[D_{KL}(q(\chi_{t-1}|\chi_t, \chi_0)||p_\theta(\chi_{t-1}|\chi_t))] \longrightarrow \mathcal{L}_{elbo} = \mathbb{E}_{t\sim\mathcal{U}(1,T)}\mathcal{L}_t$$

- The first term: Gaussian log-likelihood $logp(x_0|x_1)$

$$\mathcal{L}_{first} \triangleq -\mathbb{E}_{q(\chi_1|\chi_0)}[\log p_\theta(\chi_0|\chi_1)] = \mathbb{E}_{q(\chi_1|\chi_0)}\left[||\hat{\chi}_\theta(\chi_1, 1) - \chi_0||_2^2\right]$$



- The second term: make the distribution $p_\theta(x_{t-1}|x_t)$ approximate the tractable distribution $q(x_{t-1}|x_t, x_0)$ through the KL divergence $D_{KL}(\ )$.

$$\mathcal{L}_t = \mathbb{E}_{q(\chi_t|\chi_0)}\left[\frac{1}{2}\left(\frac{\bar{\alpha}_{t-1}}{1-\bar{\alpha}_{t-1}} - \frac{\bar{\alpha}_t}{1-\bar{\alpha}_t}\right)||\hat{\chi}_\theta(\chi_t, t) - \chi_0||_2^2\right]$$

Method

Collaborative Knowledge Graph Convolution (CKGC)

Objective: Aggregate the user-item interaction data into denoised KG

Enhance its relevance to recommendation tasks

↓

Introduce a dimension to KG's diffusion optimization

↓

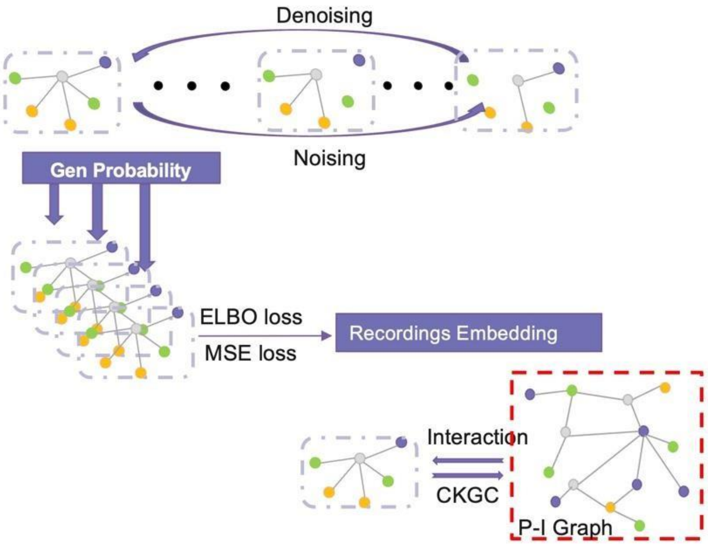
Construct the loss (MSE) to CKGC L_{ckgc}

$$\mathcal{L}_{\mathrm{ckgc}} = \left\| \left[\mathcal{A} \cdot \hat{\mathbf{x}}_0^\top \right]^\top \cdot \mathbf{E}_p - \mathbf{E}_i \right\|_2^2$$

↓

Optimizing ELBO and CKGC loss simultaneously:

$$\mathcal{L}_{\mathrm{kgdm}} = (1 - \lambda_0)\mathcal{L}_{\mathrm{elbo}} + \lambda_0 \mathcal{L}_{\mathrm{ckgc}}$$



- A: item-user interaction
- $\widehat{x_0}$: denoised KG's predicted relation probability
- $\mathcal{A}$: aggregation operation
- E_p: patient's embedding
- E_i: record's embedding
- λ_0 : hyperparameter of strength

Method

KG-enhanced Data Augmentation

Objective: reconstruct G_k' from G_k, only contains the relationships relevant to the downstream recommendation tasks.

graph embedding propagation layer

$$\mathbf{x}_p^{(l+1)} = \sum_{i \in \mathcal{N}_p} \frac{\mathbf{x}_i^{(l)}}{\sqrt{|\mathcal{N}_p| \cdot |\mathcal{N}_i|}}, \quad \mathbf{x}_i^{(l+1)} = \sum_{p \in \mathcal{N}_i} \frac{\mathbf{x}_p^{(l)}}{\sqrt{|\mathcal{N}_i| \cdot |\mathcal{N}_p|}}$$

- x_i^l: the encoded representations of item i
- x_p^l: the encoded representations of patient p
- N_i: the neighboring entities of item i
- N_p: the neighboring entities of patient p

Graph-based collaborative filtering (CF) capture collaborative signals of higher order.

Contrastive loss: maximize the agreement among
Positive pairs and minimize the agreement among negative pairs.

$$\mathcal{L}_{cl}^{user} = \sum_{u \in \mathcal{U}} -\log \frac{\exp(s(\mathbf{x}_u', \mathbf{x}_u'')/\tau)}{\sum_{v \in \mathcal{U}} \exp(s(\mathbf{x}_u', \mathbf{x}_v'')/\tau)}$$

$$\mathcal{L}_{\mathrm{cl}} = \mathcal{L}_{\mathrm{cl}}^{\mathrm{patient}} + \mathcal{L}_{\mathrm{cl}}^{\mathrm{item}}$$

- $s(\cdot)$: cosine similarity
- τ: hyper-parameter
- $(x_u', x_v'') \mid u, v \in U, u \neq v$: negative pairs (the different nodes pairs)
- $(x_u', x_u'') \mid u, u \in U$: positive pairs (the same node pairs)
- L_{cl}^{user}: contrastive loss of patient, L_{cl}^{item}: contrastive loss of item

Method

- The overall loss of DiffKG

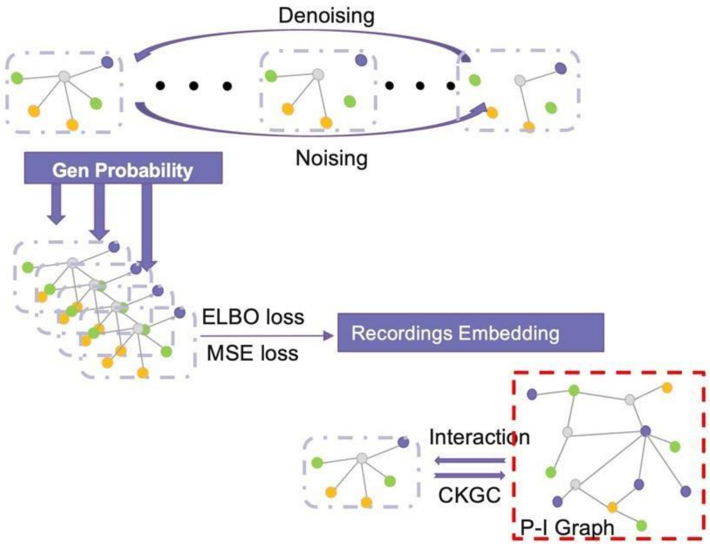
Optimizing recommendation task by Bayesian personalized ranking (BPR):

$$\mathcal{L}_{\text{bpr}} = \sum_{(u,i,j)\in\mathcal{O}} -\log\sigma(\hat{y}_{ui} - \hat{y}_{uj})$$

$\downarrow$

Integrative Optimization:

$$\mathcal{L}_{\text{rec}} = \mathcal{L}_{\text{bpr}} + \lambda_1\mathcal{L}_{\text{cl}} + \lambda_2\|\Theta\|_2^2$$



- λ_1, λ_2 : hyperparameter
- o^+ : observed interaction from the Cartesian product of patient and record set
- o^-: observed interaction from the Cartesian product of patient and record set
- Θ: learnable parameters set of model

Method

- Training and Inference

Algorithm 1: Pipeline for Knowledge Graph Modification

Original KG relations x_0, items, entities, k Updated Knowledge Graph G'_k

Procedure `AddNoiseToKG(`x_0`):`
- Apply noise to x_0;
- **return** $x_{T'}$;

Procedure `ReverseDenoise(`$x_{T'}$`):`
- Initialize $\hat{x}_T = x_{T'}$;
- **for** $t = T$ **to** 1 **do**
  - $\hat{x}_{t-1} = \mu_\theta(\hat{x}_t, t)$ `// Deterministic denoising;`
- **return** $\hat{x}_T$;

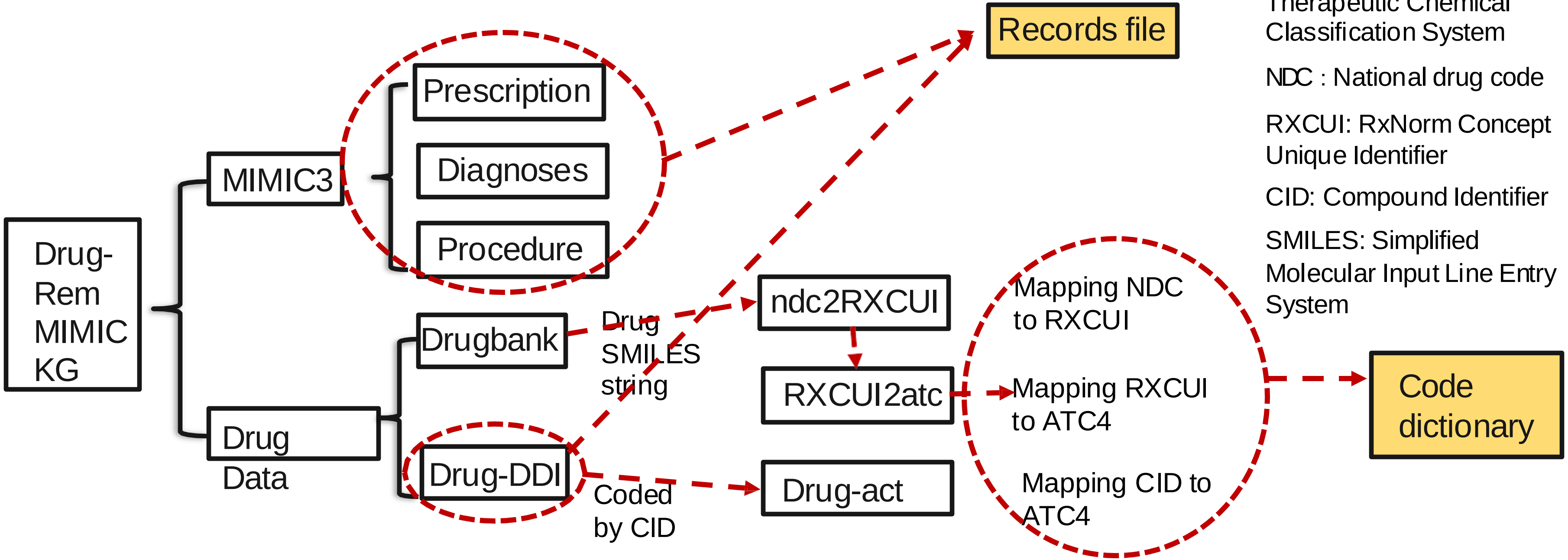
Procedure `ReconstructKG(`$\hat{x}_T$`):`
- Use $\hat{x}_T$ to build G'_k;
- **return** G'_k;

Procedure `AddTopKRelations(`G'_k, *items, entities,* k`):`
- **foreach** *item* i *in items* **do**
  - Select top-k entities $J = \{j_1, j_2, ..., j_k\}$ based on scores;
  - **foreach** *entity* $j \in J$ **do**
    - Add relation between item i and entity j in G'_k;
- **return** G'_k;

Function *MainPipeline*x_0, *items, entities,* k**:**
- $x_{T'} \leftarrow$ `AddNoiseToKG(`x_0`);`
- $\hat{x}_T \leftarrow$ `ReverseDenoise(`$x_{T'}$`);`
- $G'_k \leftarrow$ `ReconstructKG(`$\hat{x}_T$`);`
- $G'_k \leftarrow$ `AddTopKRelations(`G'_k, *items, entities,* k`);`
- **return** G'_k;

MIMIC Knowledge Graph

- MIMIC dataset: https://mimic.physionet.org/; (prescription, diagnose, procedure) NDC-RXCUI-ATC4 mapping, CID-ATC, NDC-RXCUI mapping, drugbank, drug DDI
- Preprocess data: https://github.com/ycq091044/SafeDrug.git



ATC4: Anatomical Therapeutic Chemical Classification System

NDC : National drug code

RXCUI: RxNorm Concept Unique Identifier

CID: Compound Identifier

SMILES: Simplified Molecular Input Line Entry System

MIMIC Knowledge Graph

- Construct KG on patient, diagnosis, medication, procedure:
  - ✓ Records_final.pkl

```
User_3 <has_procedure> Procedure_[1, 0, 13, 22, 27, 31, 14, 39, 28, 41, 26, 32, 17, 29]
User_4 <has_diagnosis> Diagnosis_[63, 64, 65, 66, 67, 68, 20, 47, 46, 52, 69, 70]
User_4 <has_diagnosis> Diagnosis_[28, 29, 30, 31, 32, 33, 2, 15]
User_4 <has_diagnosis> Diagnosis_[0, 1, 3, 4, 5, 6, 7, 8, 9, 11, 12, 14, 13, 2, 22, 26, 40, 41, 28, 42, 43, 44, 37, 45, 46, 47, 18, 20, 48, 49, 50, 17, 51, 52]
User_4 <has_procedure> Procedure_[71, 65, 72, 73, 74, 75, 76, 68, 46, 52, 77, 70]
User_4 <has_procedure> Procedure_[34, 28, 35, 36, 37, 38, 39]
User_4 <has_procedure> Procedure_[1, 5, 6, 0, 13, 22, 2, 29, 26, 38, 28, 41, 32, 45, 18, 17, 52, 44, 12, 39, 11, 42, 3, 47, 53]
User_4 <has_medication> Medication_[20, 50, 68, 46, 78, 69, 52, 70, 79]
User_4 <has_medication> Medication_[21, 40, 41, 14, 42, 43, 4, 1]
User_4 <has_medication> Medication_[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 22, 26, 40, 29, 47,
User_5 <has_diagnosis> Diagnosis_[80, 81, 82, 12, 83, 84, 85, 86, 87]
```

  - ✓ Get the actual ID of each patient, diagnosis, medication, procedure through voc_final dictionary
  - ✓ Drug-drug actions:
    - ➢ Map drug ID to drug SMILEs dictionary through act3toSMILES.pkl
    - ➢ connect as the value of adjacency matrix is 1 (ddi_A_final.pkl)
  - ✓ Flatten the updated KG and only keep the unique entities and relations.

[0, 2999] for diagnoses, [3000, 5999] for procedure, [6000, 8999] for med

Table 1: Dataset Statistics for MIMIC3-Drug

Statistics	MIMIC3-Drug
# Patients	6,350
# Diagnoses	1,917
# Procedures	1,898
# Medication	1,593
# Interacion	40,326
# Density	1.28×10^{-5}
Knowledge Graph	
# Entities	3,612
# Relations	4
# Triplets	41,000

```
(0, 0, 1)    (0, 3000, 1)    (0, 6043, 1)    (1, 3007, 1)
(0, 1, 1)    (0, 3001, 1)    (0, 6044, 1)    (1, 3009, 1)
(0, 2, 1)    (0, 3002, 1)    (0, 6046, 1)    (1, 3010, 1)
(0, 3, 1)    (0, 3003, 1)    (0, 6048, 1)    (1, 3012, 1)
(0, 6, 1)    (0, 3006, 1)    (0, 6049, 1)    (1, 3013, 1)
(0, 7, 1)    (0, 3007, 1)
(0, 8, 1)    (0, 3012, 1)
             (0, 3017, 1)
```

Experiment

- Experiment Setting

Train/test split: 0.8/0.2 diffusion steps: 60 epochs: 50 dropout rate: 0.2

Evaluation: Recall@N, NDCG@N as top-N recommendation metrics, N=20

$$\text{Recall@N} = \frac{\text{Number of relevant items in Top-N recommendations}}{\text{Total number of relevant items}}$$

$$\text{DCG@N} = \sum_{i=1}^{N} \frac{\text{rel}_i}{\log_2(i+1)}, \quad \text{NDCG@N} = \frac{\text{DCG@N}}{\text{IDCG@N}}$$

- rel_i: Relevance score of the item at position i (e.g., binary or graded relevance).
- $\log_2(i+1)$: A discount factor that penalizes lower-ranked items.

Baseline: GNN-based KG-enhanced: KGCN, KGAT, KGIN

Other generative models: multiVAE, CDAE, DiffRec

Result

- Q1: How does the performance of our proposed model compare to a diverse range of state-of-the-art models?

Model	Recall@20	Recall@10	NDCG@20	NDCG@10
DiffKG (ours)	**0.0716**		**0.2515**	
MultiVAE	0.0693		0.1988	
CDAE	0.0701		0.1917	
DiffRec	0.0706		0.2471	
KGIN	0.0615		0.2234	
KGCN	0.0621		0.1999	
KGAT	0.0572		0.1764	

Diagnosis	**Methods**	**Medicine Recommendations**
Sepsis, Acute respiratory failure, Hypertension	MultiVAE	Metoprolol Tartrate, Vancomycin, Furosemide
	CDAE	Furosemide, Metoprolol, Insulin, Norepinephrine
	KGIN	Vancomycin, Metoprolol Tartrate, Corticosteroids
	DiffKG	Furosemide, Furosemide, Amlodipine, Norepinephrine, Acetaminophen, Corticosteroids
Type 2 diabetes, Rheumatoid arthritis, Hypertension, Hyperlipidemia	MultiVAE	Phenylbutazone, Insulin, Fenofibrate, empagliflozin, liraglutide
	CDAE	Metformin, Tolbutamide, Phenylbutazone, Insulin, Acetaminophen, empagliflozin and liraglutide
	KGIN	Metformin, Amethopterin, Amiloride/HCTZ, Fenofibrate, empagliflozin, liraglutide
	DiffKG	Metformin, Insulin, Acetaminophen, Nifedipine, Fenofibrate

Future Works

- Compare with more baseline models
- Ablation study on different collaborative filtering method: MFBPR, LightGCN, SGL

  https://arxiv.org/pdf/2401.06982
- Ablation study on different noise scheduler method
- Incorporate with LLM for actual user input and provide denoised recommendation
- Efficient denoising strategy
- Visualization and comparison

References

1. *Jiang, Y., Yang, Y., Xia, L., & Huang, C. (2024, March). Diffkg: Knowledge graph diffusion model for recommendation. In Proceedings of the 17th ACM International Conference on Web Search and Data Mining (pp. 313- 321).*

2. *Zhao, J., Wenjie, W., Xu, Y., Sun, T., Feng, F., & Chua, T. S. (2024, July). Denoising diffusion recommender model. In Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 1370-1379).*

3. *Yang, C., Xiao, C., Ma, F., Glass, L., & Sun, J. (2021). Safedrug: Dual molecular graph encoders for recommending effective and safe drug combinations. arXiv preprint arXiv:2105.02711.*

Thank you!