Contrastive knowledge graph attention network for request-based recipe recommendation

Paper ID: 1127

Xiyao Ma, Zheng Gao, Qian Hu, Mohamed AbdelHady

Amazon Alexa AI

Introduction

Kitchen assistant is one of the enabled services in intelligent voice assistants. Current solutions for recipe recommendation have two limitations:

- Neglect the diversity of user preferences
- User item interaction noise

We propose a Contrastive Knowledge Graph ATtention network (C-KGAT), which includes:

- A knowledge graph attention-based recommender.
- Profiling user diversified preferences from user sequential behaviors.
- A contrastive learning module with two auxiliary tasks to improve model robustness.

Problem Formulation

Given a target user and her utterance request, we aim to recommend top-K relevant and personalized recipes.

The information used in the proposed model including:

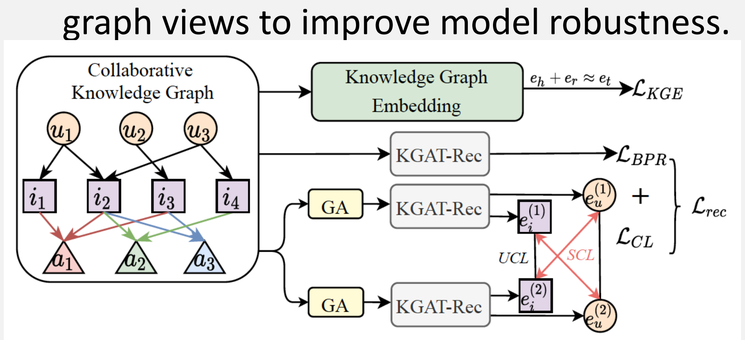
- Users
- Recipes with textual and categorial features
- User recipe interactions containing a sequence of behaviors.
- A knowledge graph with recipe entities and attribute entities (i.e. cuisine, ingredients, keywords).

In the end, we merge all information (user-recipe interactions and recipe knowledge graph) together to achieve a collaborative Knowledge Graph (CKG).

Method: C-KGAT

C-KGAT mainly consists of three components:

1. Knowledge graph embedding leverage the structure of CKG to learn entity embedding.
2. A KGAT-based recommender (KGAT-Rec) is proposed to learn collaborative user and recipe embeddings by modeling the diversified user preferences.
3. A contrastive learning module contrast user and recipe embeddings from different graph views to improve model robustness.



1. Knowledge Graph Embedding

Each user, recipe, and attribute entity is associated with an ID embedding, annotated by e_u, e_i, and e_a, respectively, which are used to initialize their entity embeddings in the collaborative knowledge graph. TransE is used to learn the entity embeddings with loss:

$$\mathcal{L}_{\mathrm{KGE}} = \sum_{(h,r,t,t')\in\mathcal{T}} -\ln\sigma\left(g\left(h,r,t'\right) - g(h,r,t)\right),$$
$$g(h,r,t) = \|e_h + e_r - e_t\|_2^2.$$

2. KGAT-Rec

In the collaborative knowledge graph, an entity embedding is updated by aggregating the rich semantic information from its neighbor triplets.

$$s(h,r,t) = (e_t)^\top \tanh(e_h + e_r),$$
$$\alpha_{h,r,t} = \frac{\exp(s(h,r,t))}{\sum_{(h,r',t')\in\mathcal{N}_h}\exp\left(s\left(h,r',t'\right)\right)},$$
$$e_{\mathcal{N}_h} = \sum_{(h,r,t)\in\mathcal{N}_h}\alpha_{h,r,t}e_t.$$

We learn different user preferences from their sequential behaviors towards each interacted recipe by a one layer bidirectional Gated Recurrent Unit (GRU)

$$h_{pref} = GRU([b_1, b_2, ..., b_k]),$$
$$e_r^* = W_p([e_r, h_{pref}]).$$

We update the target entity embedding by aggregating its entity embedding and embeddings of its neighbors with LeakyReLU.

$$\hat{e_h} = \mathrm{LeakyReLU}\left(W_a\left(e_h + e_{\mathcal{N}_h}\right)\right)$$

BERT is used to retrieve user request vector h_{req}. Final recipe vector is the concatenation of its entity embedding $e_{\hat{i}}$ and feature embedding e_f. Final loss is BPR loss:

$$\mathcal{L}_{\mathrm{BPR}} = \sum_{(u,i)\in\mathcal{Y}^+,(u,j)\notin\mathcal{Y}^+} -\log\sigma\left(\hat{y}_{ui} - \hat{y}_{uj}\right),$$
$$\hat{y}_{ui} = f_c(h_{req}, \hat{e_u}, f_{enc}(\hat{e_i}, e_f)).$$

3. Incorporating Contrastive Learning

In the *graph augmentation (GA)* stage, different views of the input graph are generated to expose novel patterns of representations to improve the model generalization. Operations including:

- Node Embedding Dropout
- Edge Dropout

we apply the operations on the input CKG graph to generate two different graph views.

Unsupervised Contrastive Learning (UCL)

InfoNCE loss is adopted to pull the different views of the same user entity close and push those of different user entities away:

$$\mathcal{L}_{UCL}(\mathcal{U}^{(1)},\mathcal{U}^{(2)}) = \sum_{u\in\mathcal{U}} -\log\frac{\exp\left(f\left(e_u^{(1)}, e_u^{(2)}\right)/\tau\right)}{\sum_{v\in\mathcal{U}}\exp\left(f\left(e_u^{(1)}, e_v^{(2)}\right)/\tau\right)}$$

Supervised Contrastive Learning (SCL)

Given an observed user-recipe interaction, we encourage the agreement between the user and recipe generated from different views. Meanwhile, we minimize the agreement between unobserved user-recipe pairs.

$$\mathcal{L}_{SCL}(\mathcal{U}^{(2)},\mathcal{I}^{(1)}) = \sum_{(u,i)\in\mathcal{Y}^+} -\log\frac{\exp\left(f\left(e_u^{(2)}, e_i^{(1)}\right)/\tau\right)}{\sum_{q\in\mathcal{Q}}\exp\left(f\left(e_u^{(2)}, e_q^{(1)}\right)/\tau\right)}$$

The total CL loss L_{CL} is the summation of UCL and SCL losses. The model is trained by alternatively minimizing recommendation loss $L_{rec} = L_{BPR} + \lambda L_{CL}$ and KGE loss L_{KGE} during each epoch.

Experiments

Datasets

We use Alexa data where customers can interact with devices equipped with screens by vocal request in Recipe-Voice dataset and touching the screen in Recipe-Touch dataset. To ensure the data quality, we take the 3(10)-core subset for the two datasets, where each user or recipe has at least 3 (10) interactions, respectively.

		Recipe-Voice	Recipe-Touch
Recipe	#Entity	50,070	18,740
Knowledge	#Relation Types	4	4
Graph	#Triplets	1,486,858	518,735

Baselines

YoutubeDNN, LightGCN, KGCN, KGAT

Model Performance Comparison

The table shows the relative performance improvement afforded by our method compared to all baseline models.

Datasets	Recipe-Voice		Recipe-Touch	
Models	Recall@5	NDCG@5	Recall@5	NDCG@5
YoutubeDNN	-5.5%	-3.0%	-3.9%	-4.1%
LightGCN	-4.8%	-2.1%	-1.5%	-2.4%
KGCN	-2.5%	-1.3%	-0.8%	-1.5%
KGAT	0.0%	0.0%	0.0%	0.0%
C-KGAT	**+5.2%**	**+7.4%**	**+4.9%**	**+5.8%**

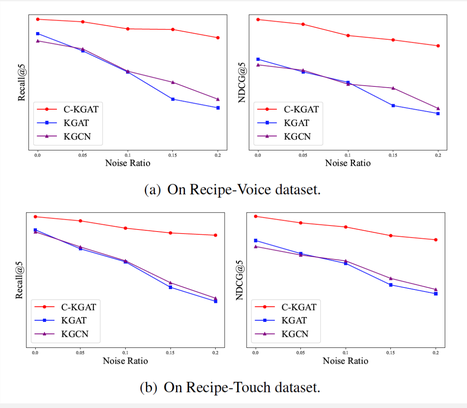
Ablation Study

We conduct ablation study to quantify the impact of the components in our proposed model and report the corresponding degradations, including the GRU for preference vector, contrastive learning (CL) that includes both UCL and SCL, supervised contrastive learning (SCL), node embedding dropout (NED), and edge dropout (ED).

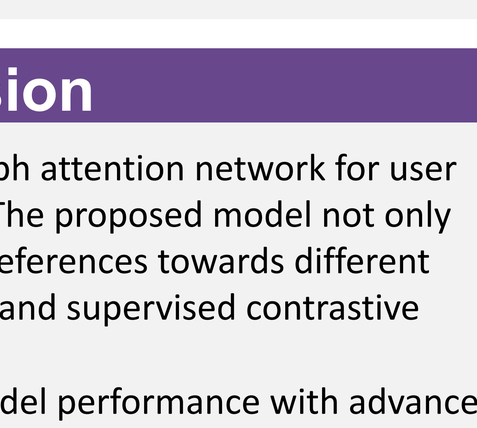
Datasets	Recipe-Voice		Recipe-Touch	
Models	Recall@5	NDCG@5	Recall@5	NDCG@5
C-KGAT	0.0%	0.0%	0.0%	0.0%
-GRU	-2.8%	-4.2%	-1.4%	-2.0%
-CL	-4.1%	-6.1%	-2.7%	-3.1%
-SCL	-2.1%	-3.2%	-1.3%	-1.8%
-NED	-1.4%	-2.6%	-0.9%	-1.4%
-ED	-1.0%	-2.1%	-0.6%	-0.9%

Model Robustness

We train models with different ratios of additional noise data sampled from unobserved interactions and compare their performances.



(a) On Recipe-Voice dataset.



(b) On Recipe-Touch dataset.

Conclusion

we propose a contrastive knowledge graph attention network for user request-based recipe recommendation. The proposed model not only boosts performance by modeling user preferences towards different recipes but also integrates unsupervised and supervised contrastive learning to improve model robustness.

In the future, we plan to improve the model performance with advanced negative sampling strategies and transfer learning for cross-domain recommendation.