Knowledge-aware Conversational Preference Elicitation with Bandit Feedback

Canzhe Zhao [1] Tong Yu [2]
Zhihui Xie [1] Shuai Li [1]

[1]Shanghai Jiao Tong University

[2]Carnegie Mellon University

April 28, 2022

Motivation - Conversational Recommender Systems

Traditional recommender systems:

- data sparsity
- cold-start problem

Conversational recommender systems:

- alleviate data sparsity by conversations
- mitigate the cold-start problem

Motivation - Conversational Recommender Systems

Traditional recommender systems:

- data sparsity
- cold-start problem

Conversational recommender systems:

- alleviate data sparsity by conversations
- mitigate the cold-start problem

Motivation - Limitations of existing CRSs

- Heavily depend on high-quality key-terms carefully labeled by humans
  - Incompletely-labeled key-terms $\Longrightarrow$ Performance degradation
- Leverage the feedback to different conversaional key-terms separately
  - Large candidate set of conversational key-terms $\Longrightarrow$ Sample-inefficiency
- This work: utilize the graph structure!

Motivation - Limitations of existing CRSs

- Heavily depend on high-quality key-terms carefully labeled by humans
  - Incompletely-labeled key-terms $\Longrightarrow$ Performance degradation
- Leverage the feedback to different conversaional key-terms separately
  - Large candidate set of conversational key-terms $\Longrightarrow$ Sample-inefficiency
- This work: utilize the graph structure!

Motivation - Limitations of existing CRSs

- Heavily depend on high-quality key-terms carefully labeled by humans
  - Incompletely-labeled key-terms $\Longrightarrow$ Performance degradation
- Leverage the feedback to different conversaional key-terms separately
  - Large candidate set of conversational key-terms $\Longrightarrow$ Sample-inefficiency
- This work: utilize the graph structure!

Problem Formulation

- Item set $\mathcal{A}$ with $|\mathcal{A}| = N$
- Item a's feature vector: $\boldsymbol{x}_a \in \mathbb{R}^d$
- Key-term set $\mathcal{K}$ with $|\mathcal{K}| = K$
- Key-term k's feature vector: $\tilde{\boldsymbol{x}}_k \in \mathbb{R}^d$
- Knowledge graph $\mathcal{G} = (\mathcal{E}, \mathcal{R})$ where $\mathcal{E} = \mathcal{A} \cup \mathcal{K}$ is the set of entities and $\mathcal{R}$ is the set of relations
- $\theta^* \in \mathbb{R}^d$ and $\tilde{\theta}^* \in \mathbb{R}^d$ are user preference vectors on items and key-terms respectively
- Receive rewards $r_{a_t,t} = \boldsymbol{x}_{a_t}^\top \boldsymbol{\theta}^* + \epsilon_t$ and $\tilde{r}_{k_t,t} = \tilde{\boldsymbol{x}}_{k_t}^\top \tilde{\boldsymbol{\theta}}^* + \tilde{\epsilon}_t$ after recommending item a_t and conducting one conversation on key-term k_t respectively
- The conversation frequency: $g(t)$[1]

Problem Formulation

- Item set $\mathcal{A}$ with $|\mathcal{A}| = N$
- Item a's feature vector: $\boldsymbol{x}_a \in \mathbb{R}^d$
- Key-term set $\mathcal{K}$ with $|\mathcal{K}| = K$
- Key-term k's feature vector: $\tilde{\boldsymbol{x}}_k \in \mathbb{R}^d$
- Knowledge graph $\mathcal{G} = (\mathcal{E}, \mathcal{R})$ where $\mathcal{E} = \mathcal{A} \cup \mathcal{K}$ is the set of entities and $\mathcal{R}$ is the set of relations
- $\boldsymbol{\theta}^* \in \mathbb{R}^d$ and $\tilde{\boldsymbol{\theta}}^* \in \mathbb{R}^d$ are user preference vectors on items and key-terms respectively
- Receive rewards $r_{a_t,t} = \boldsymbol{x}_{a_t}^\top \boldsymbol{\theta}^* + \epsilon_t$ and $\tilde{r}_{k_t,t} = \tilde{\boldsymbol{x}}_{k_t}^\top \tilde{\boldsymbol{\theta}}^* + \tilde{\epsilon}_t$ after recommending item a_t and conducting one conversation on key-term k_t respectively
- The conversation frequency: $g(t)$[1]

Problem Formulation

- Item set $\mathcal{A}$ with $|\mathcal{A}| = N$
- Item a's feature vector: $\boldsymbol{x}_a \in \mathbb{R}^d$
- Key-term set $\mathcal{K}$ with $|\mathcal{K}| = K$
- Key-term k's feature vector: $\tilde{\boldsymbol{x}}_k \in \mathbb{R}^d$
- Knowledge graph $\mathcal{G} = (\mathcal{E}, \mathcal{R})$ where $\mathcal{E} = \mathcal{A} \cup \mathcal{K}$ is the set of entities and $\mathcal{R}$ is the set of relations
- $\boldsymbol{\theta}^* \in \mathbb{R}^d$ and $\tilde{\boldsymbol{\theta}}^* \in \mathbb{R}^d$ are user preference vectors on items and key-terms respectively
- Receive rewards $r_{a_t,t} = \boldsymbol{x}_{a_t}^\top \boldsymbol{\theta}^* + \epsilon_t$ and $\tilde{r}_{k_t,t} = \tilde{\boldsymbol{x}}_{k_t}^\top \tilde{\boldsymbol{\theta}}^* + \tilde{\epsilon}_t$ after recommending item a_t and conducting one conversation on key-term k_t respectively
- The conversation frequency: $g(t)$[1]

Problem Formulation

Learning objective: minimizing the expected cumulative regret[2]

$$R(T) = \mathbb{E}\left[\sum_{t=1}^{T} \max_{a\in\mathcal{A}} \boldsymbol{x}_a^\top \boldsymbol{\theta}^* - \sum_{t=1}^{T} r_{a_t,t}\right].$$

Challenges:

- Key-terms are incompletely-labeled?
  - Propogate the user preference of key-terms on the graph
- Sample-inefficient when the candidate key-term set is large?
  - Leverage the graph to select the most informative key-terms

Problem Formulation

Learning objective: minimizing the expected cumulative regret[2]

$$R(T) = \mathbb{E}\left[\sum_{t=1}^{T} \max_{a \in \mathcal{A}} \boldsymbol{x}_a^\top \boldsymbol{\theta}^* - \sum_{t=1}^{T} r_{a_t,t}\right].$$

Challenges:

- Key-terms are incompletely-labeled?
  - Propogate the user preference of key-terms on the graph
- Sample-inefficient when the candidate key-term set is large?
  - Leverage the graph to select the most informative key-terms

Model

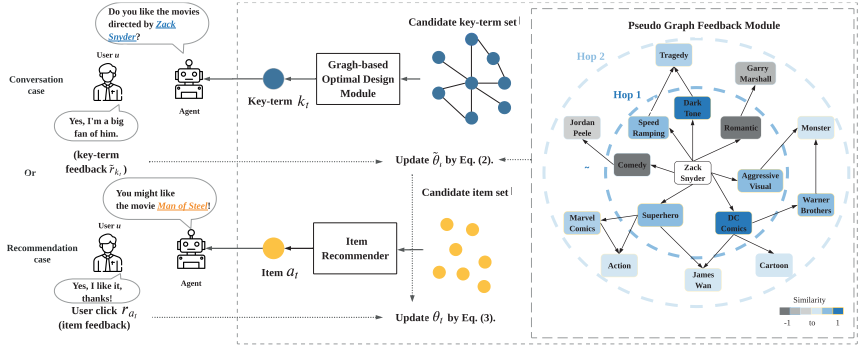


Figure 1. An overview of our knowledge-aware conversational preference elicitation framework.

Algorithm - Item Recommendation

- Learn $\tilde{\boldsymbol{\theta}}^*$ to accelerate the learning of $\boldsymbol{\theta}^*$

$$\begin{aligned}\tilde{\boldsymbol{\theta}}_t =& \arg\min_{\boldsymbol{\theta}} \sum_{\tau=1}^{t} \sum_{k \in \mathcal{K}_\tau} \left(\tilde{\boldsymbol{x}}_{k,\tau}^\top \boldsymbol{\theta} - \tilde{r}_{k,\tau} \right)^2 + \tilde{\lambda} \|\boldsymbol{\theta}\|_2^2 \\ =& \tilde{\boldsymbol{M}}_t^{-1} \tilde{b}_t \,, \end{aligned} \tag{1}$$

where $\mathcal{K}_\tau$ is the set of selected key-terms at iteration τ, $\tilde{\lambda}$ is the regularization parameter and

$$\tilde{\boldsymbol{M}}_t = \sum_{\tau=1}^{t} \sum_{k \in \mathcal{K}_\tau} \tilde{\boldsymbol{x}}_{k,\tau} \tilde{\boldsymbol{x}}_{k,\tau}^\top + \tilde{\lambda} \boldsymbol{I} \,,\ \tilde{\boldsymbol{b}}_t = \sum_{\tau=1}^{t} \sum_{k \in \mathcal{K}_\tau} \tilde{\boldsymbol{x}}_{k,\tau} \tilde{r}_{k,\tau} \,.$$

Algorithm - Item Recommendation

- Then $\boldsymbol{\theta}^*$ could be learned by

$$\begin{aligned}\boldsymbol{\theta}_t =& \arg\min_{\boldsymbol{\theta}} \lambda \sum_{\tau=1}^{t} (\boldsymbol{x}_{a_\tau}^\top \boldsymbol{\theta} - r_{a_\tau})^2 + (1-\lambda)\|\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}_t\|_2^2 \\ =& \boldsymbol{M}_t^{-1}(\boldsymbol{b}_t + (1-\lambda)\tilde{\boldsymbol{\theta}}_t)\,, \end{aligned} \tag{2}$$

where $\lambda \in [0, 1]$ balances the trade-off between the item-level and key-term-level information and

$$\boldsymbol{M}_t = \lambda \sum_{\tau=1}^{t} \boldsymbol{x}_{a_\tau} \boldsymbol{x}_{a_\tau}^\top + (1-\lambda)\boldsymbol{I}\,,\;\; \boldsymbol{b}_t = \lambda \sum_{\tau=1}^{t} \boldsymbol{x}_{a_\tau} r_{a_\tau}\,.$$

Algorithm - Item Recommendation

- Recommend item according to optimism principle in the face of uncertainty (OFU)

$$a_t = \arg\max_{a \in \mathcal{A}} \boldsymbol{x}_a^\top \boldsymbol{\theta}_t + \lambda\alpha \|\boldsymbol{x}_a\|_{\boldsymbol{M}_t^{-1}} + (1-\lambda)\tilde{\alpha} \|\boldsymbol{M}_t^{-1} \boldsymbol{x}_a\|_{\tilde{\boldsymbol{M}}_t^{-1}}, \qquad (3)$$

where α and $\tilde{\alpha}$ are the hyper-parameters representing the exploration level on items and key-terms respectively.

Algorithm - Conversational Preference Propagation

- Propogate user preference using graph structural and semantic information
- Graph structural information:

$$\mathrm{Sim}_{\mathcal{J}}(k, k') = |\mathcal{P}_k^h \cap \mathcal{P}_{k'}^h| / |\mathcal{P}_k^h \cup \mathcal{P}_{k'}^h| \, ,$$

  where $\mathcal{P}_k^h$ is the set of paths starting from key-term k with length no larger than h.
- Semantic information: cosine similarity $\mathrm{Sim}_{\mathcal{C}}(k, k')$
- Overall similarity metric:

$$\begin{aligned} \mathrm{Sim}(k, k') &= \gamma \, \mathrm{Sim}_{\mathcal{C}}(k, k') + (1 - \gamma) \, \mathrm{Sim}_{\mathcal{J}}(k, k') \\ &= \gamma \frac{\tilde{\boldsymbol{x}}_k^\top \tilde{\boldsymbol{x}}_{k'}}{\|\tilde{\boldsymbol{x}}_k\|_2 \|\tilde{\boldsymbol{x}}_{k'}\|_2} + (1 - \gamma) \frac{|\mathcal{P}_k^h \cap \mathcal{P}_{k'}^h|}{|\mathcal{P}_k^h \cup \mathcal{P}_{k'}^h|} \, . \end{aligned}$$

Algorithm - Conversational Preference Propagation

- Constructs a pseudo preference as $\mathrm{Sim}(k, k') * \tilde{r}_k$ to estimate $\tilde{r}_{k'}$ using true user key-term-level feedback $\tilde{r}_k$
- Updates $\tilde{\theta}_t$ utilizing the true feedback and all the pseudo graph feedback
- Referred to as pseudo graph feedback (PGF) module

Algorithm - Graph-based Conversation

- Assume the variance of key-term level noise $\tilde{\epsilon}_t$ is $\tilde{\sigma}^2$
- Gauss-Markov theorem shows that

$$\mathrm{Cov}(\tilde{\boldsymbol{\theta}}_t) = \tilde{\sigma}^2 \tilde{\boldsymbol{M}}_t^{-1} \, .$$

- Leverages optimal experimental design (OED) to select key-terms to make the determinant of $\tilde{\boldsymbol{M}}_t^{-1}$ diminish fast

Algorithm - Graph-based Conversation

- The optimal distribution π^* satisfies

$$\begin{aligned}\pi^* &= \arg\max_{\pi} \log\det \tilde{\boldsymbol{M}}_t(\pi) \\ &= \arg\max_{\pi} \log\det \left(g(t) \sum_{k\in\mathcal{K}} \pi(k) \boldsymbol{G}_k^h + \tilde{\lambda} \boldsymbol{I} \right) ,\end{aligned}$$

  where $\boldsymbol{G}_k^h = \sum_{k'\in\mathcal{N}_k^h} \tilde{\boldsymbol{x}}_{k'} \tilde{\boldsymbol{x}}_{k'}^\top$ is the Gramian matrix generated by the feature vectors of key-term k and k's h-hop neighbors
- To approximately solve this problem, compute the best rank-one approximation of $\boldsymbol{G}_k^h$ as

$$\tilde{\boldsymbol{x}}_k^h = \min_{\boldsymbol{x}\in\mathbb{R}^d} \| \boldsymbol{G}_k^h - \boldsymbol{x}\boldsymbol{x}^\top \|_F \, . \quad (4)$$

Algorithm - Graph-based Conversation

- Find π_h^* over $\{\tilde{\boldsymbol{x}}_k^h\}_{k\in\mathcal{K}}$ such that

$$\pi_h^* = \arg\max_{\pi} \log\det\left(\sum_{k\in\mathcal{K}} \pi(k)\tilde{\boldsymbol{x}}_k^h(\tilde{\boldsymbol{x}}_k^h)^\top + \tilde{\lambda}\boldsymbol{I}\right),$$

  which could be solved by canonical optimal design methods.

- Sample key-terms from π_h^* to conduct conversations
- Referred to as graph-based optimal design (GOD) module

Experiments

Research questions:

- **RQ1** Overall performance of GraphConUCB?
- **RQ2** Performance of GraphConUCB given the items with incompletely labeled key-terms?
- **RQ3** Ablation study of the PGF module and the GOD module?

Experiments - Overall Performance (RQ1)

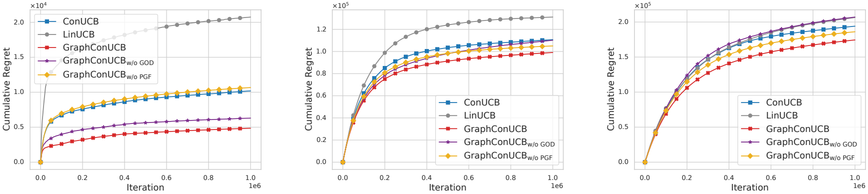
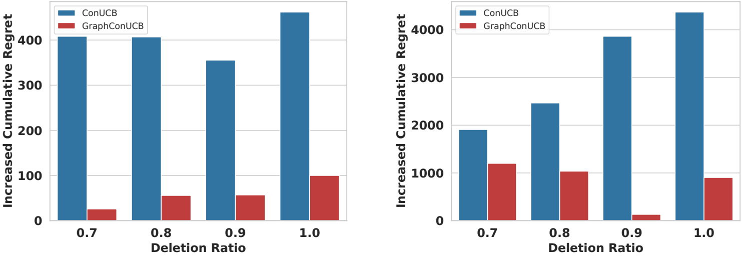


Figure 2. Overall performance comparison on MovieLens-20M, Last.FM and Amazon-Book datasets.

- GraphConUCB improve over ConUCB by 52.36%, 10.48% and 10.11% when $T = 1\text{M}$ on MovieLens-20M, Last.FM and Amazon-Book datasets respectively

Experiments - Learning with Incompletely Labeled Key-terms (RQ2)

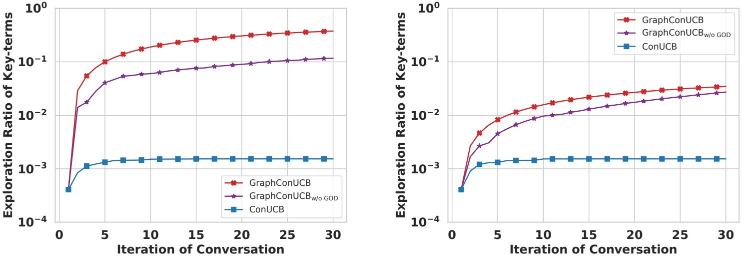


(a) MovieLens-20M dataset. (b) Last.FM dataset.

Figure 3. Increased cumulative regret under varying deletion ratio of key-terms.

- Compared to the baseline, our algorithm can handle the items with incompletely labeled key-terms more effectively

Experiments - Ablation Study (RQ3)



(a) MovieLens-20M dataset.

(b) Last.FM dataset.

Figure 4. Exploration ratio of key-terms in conversations.

- Exploration ratio of key-terms in GraphConUCB$_{w/o\ \text{GOD}}$ grows rapidly
- GraphConUCB achieves the fastest exploration ratio of key-terms

Conclusions

In this work:

- A pseudo graph feedback (PGF) module to effectively propagate the user preferences
- A graph-based optimal design (GOD) module which selects the most informative key-terms with the leverage of the graph structure

The End

- Q&A?

- Thank you!

References I

- Li, Lihong et al. "A contextual-bandit approach to personalized news article recommendation". In: *Proceedings of the 19th International Conference on World Wide Web, WWW 2010, Raleigh, North Carolina, USA, April 26-30, 2010*. ACM, 2010, pp. 661–670.
- Xie, Zhihui et al. "Comparison-Based Conversational Recommender System with Relative Bandit Feedback". In: *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '21. Virtual Event, Canada: Association for Computing Machinery, 2021, 1400–1409.
- Zhang, Xiaoying et al. "Conversational Contextual Bandit: Algorithm and Application". In: *WWW '20: The Web Conference 2020, Taipei, Taiwan, April 20-24, 2020*. ACM / IW3C2, 2020, pp. 662–672.