2024 IEEE International Conference on Multimedia and Expo (ICME) | 979-8-3503-9015-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICME57554.2024.10687762

RRE: A Relevance Relation Extraction Framework for Cross-domain Recommender System at Alipay

Jiayang Gu*†‡, Xovee Xu*†‡, Yulu Tian†, Yurun Hu†, Jiadong Huang†, Wenliang Zhong†, Fan Zhou*, Lianli Gao*§
*University of Electronic Science and Technology of China, Chengdu, Sichuan 610054, China
†Ant Group, Hangzhou, Zhejiang 310023, China
‡Equal Contribution, §Corresponding
Emails: jiayang.barrygu@gmail.com, xovee@std.uestc.edu.cn, vicky.tyl@antgroup.com, yurun.hyr@alibaba-inc.com, jiadong.hjd@antgroup.com, yice.zwl@antgroup.com, fan.zhou@uestc.edu.cn, lianli.gao@uestc.edu.cn

***Abstract*—Prevailing embedding-based cross-domain recommendation (CDR) techniques produce embeddings individually or transfer the overall feature distribution from one domain to another. However, in real-world applications, they might be ineffective due to semantic gap across domains, which arises from divergent purposes and descriptive styles. In this work, we aim to address this challenge between Mini Program and content channel in Alipay, the largest mobile payment platform in China. To bridge utility-oriented Mini Programs and advertisement-oriented contents, we utilize side information of entities to make the entity relevance scores trustworthy. Then we introduce a knowledge graph-based model to reduce the impact of embedding vibrating from contrastive learning and the biases from the pretrained language models. Extensive experiments conducted on a large-scale Alipay offline dataset as well as an online environment demonstrated the effectiveness of our proposed framework.**

***Index Terms*—Semantic relevance, relation extraction, knowledge graph, mini program**

I. Introduction

Mini Programs refer to lightweight applications integrated within a super-app ecosystem, encompassing a range of hierarchical services offered by third-party merchants. For instance, the *Ctrip* Mini Program on Alipay facilitates transportation ticket booking and short trip or hotel reservations. Within this ecosystem, the host of the Ctrip Mini Program may also seize the opportunity to generate additional revenue through advertising in the content channel. With respect to this gap, it is worth noting that the user's historical log of Mini Programs reflects their practical needs, while the content log can serve as a gauge of their expectations or intents. Specifically, it is important to recognize that an individual's booking history for occasional business trips does not necessarily imply that they are an avid traveler or enthusiast, and therefore recommending local traveling vlogs to them may not effective.

To bridge this gap that may occur in cross-domain recommendation (CDR) scenarios, we take the first step by formulating a Relevance Relation Extraction (RRE) problem that aims to predict the semantic relevance scores between Mini Program and content. From our empirical investigations, we have identified three notable obstacles that hinder the RRE performance on industrial-scale data. Accordingly, **(1)** Pretrained language models (PLMs) and text-based knowledge graphs (KGs) are prone to suffer from the isotropy problem [1] that makes the scores less comparable across the entire graph; **(2)** directly integrating the paradigm of traditional KG learning methods [2], [3] would cause extra computational cost, and additionally make the encoder model hard to converge; **(3)** there exists a wealth of potentially valuable side information linked to Mini Program or content in the super app, but they are not exploited in the current process of learning entity representations.

In this work, we propose a novel contrastive learning-based knowledge graph completion (KGC) model for addressing the aforementioned three obstacles in RRE problem. Specifically,

(1) With a dedicated bi-encoding KGC paradigm, our model is able to generate interpretable relevance scores between Mini Programs and content in Alipay's ecosystem; **(2)** By integrating topology side information and a more effective contrastive learning-based method, our model is able to fix the domain gap between entities; **(3)** We demonstrate the effectiveness of the proposed model on a synthetic RRE dataset and implement it in item-centric and user-centric online recommendation tasks. A reference source code is released at https://github.com/jkdxg8837/Relevance-Relation-Extraction.

II. Data and Problem Definition

Definition of Entities. In Alipay, *Mini Program* and *content* are two important components in its ecosystem, one providing services, and the other provides user- and merchant-generated content. To be specific, Mini Program can be accessed via links and cards in an article, user feed, QR code, search, live streaming, or through a Mini Program Hub. Common types of Mini Program include e-commerce, entertainment, news media, financial, travel, and health services. In our model, we represent Mini Programs by their titles and descriptions provided by their creators. On the other hand, content refers to articles and videos in Alipay's Platform. Third-party merchants can post articles to promote their products, while common users can share their product reviews by creating blogs or vlogs. We also represent content using the title and description provided by its creators. If these are missing, we extract

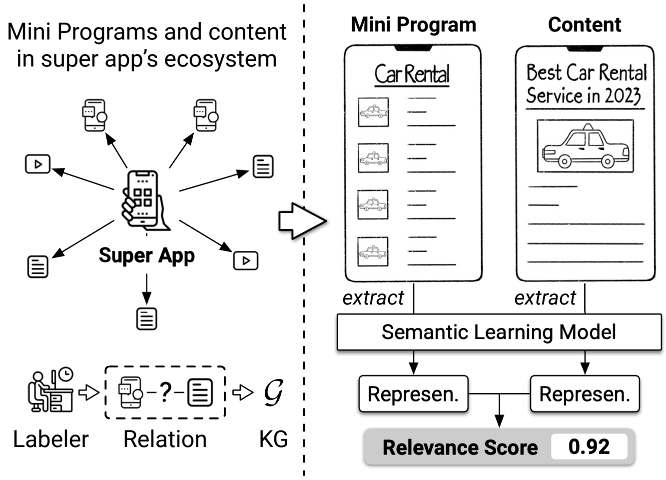

Fig. 1. An example of the relevance relation extraction problem between Mini Program and content.

descriptions from key pages/frames of the content, with the assistance of Document AI algorithms [4]. Furthermore, the generated descriptions undergo a quality supervision process.

Definition of RRE problem. Given a knowledge graph $\mathcal{G} = (m, r, c) \subseteq \mathcal{E}_m \times \mathcal{R} \times \mathcal{E}_c$, where each triplet (m, r, c) consists of a Mini Program entity $m \in \mathcal{E}_m$, a content entity $c \in \mathcal{E}_c$, and a symmetric relevance relation $r \in \mathcal{R}$ between them. The problem of RRE is to predict continuous relevance score $s(m, c)$ between Mini Programs and content in an incomplete $\mathcal{G}$. Fig. 1 shows an illustration of the RRE task.

Data Labeling. The Mini Program and its content possess multiple attributes, including brand, intention, main describing subject, localization and more. These attributes may be explicitly provided by the users, such as brand and localization, or can be implicitly extracted from their descriptions, such as intention and subject. We note that intention and subject play crucial roles in determining the correlation between two entities. For example, the *Weak correlation*, which serves as the boundary between related and non-related, requires either the intention, subject or localization to be the same or included. *Moderate correlation*, which is a slightly stronger relationship than weak correlation, necessitates all three components mentioned above to be connected, while the brands are generalized. *Strong correlation* demands not only the same subject or intention but also the same localization and related brand.

III. Relate works

Mini Program. The popularity of Mini Program in recent years has gradually attracted the attention of researchers and engineers in both academia and industry from different perspectives such as business value, user experience, program development & design, security and privacy [5]–[11]. For example, [12] studied the pros and cons of Mini Program compared to native apps and investigated the implications of Mini Program to mobile application platforms.

Knowledge Acquisition. It aims to obtain new knowledge via various KG representation learning, KG completion, and relation extraction techniques [2]. Traditional embedding-based methods such as TransE [13] and DistMult [14], focus on determining appropriate representation spaces and scoring functions for encoding entities and their relations. Text-based methods are able to utilize entity descriptions for KG reasoning via PLMs and can be used for inductive KG tasks [15].

IV. Methodology

A. Model Architecture

Given a Mini Program m, a content c, and their relevance relation type r, the RRE task can be seen as a knowledge graph completion (KGC) problem [2] for completing triplet $(m, ?, c)$. Fig. 2 presents an overview of the framework.

The scoring function in traditional KG typically integrates the embedding of relation r with m or c. However, it can incur additional computational overhead: to obtain the relevance score of a specific relation type, we need to combine every kind of relation token with the head or tail entity. We propose a new scoring function by removing the r in the input pair:

$$\phi(m, r, c) = 1 - |f(m, c) - g(r)| \tag{1}$$

$$f(m, c) = (\cos(\mathbf{h}_m, \mathbf{h}_c) + 1) / 2 \tag{2}$$

where $f(\cdot, \cdot) \in [0, 1]$ is a scaled cosine similarity between L_2 normalized representations $\mathbf{h}_m$ and $\mathbf{h}_c$ output by the BERT-based encoders, and $g(\cdot)$ is a mapping function from relation classes to the predefined relevance scores.

Given the distinctive characteristics of our RRE task, we adopt a novel approach by assigning an encoder to encode in-batch positive entities, while employing another to encode in-batch negative entities, with a momentum update technique. This new paradigm offers improved synergy with our newly devised score function. The advantage of this approach lies in its ability to not only generate high-quality negative samples but also disentangle the cross-effects arising from the involvement of the same entity playing both positive and negative roles within a batch. The momentum updating function is formally defined as follows:

$$\Theta_{\mathrm{BERT}_{neg}} \leftarrow \alpha\Theta_{\mathrm{BERT}_{neg}} + (1-\alpha)\Theta_{\mathrm{BERT}_{pos}} \tag{3}$$

where Θ is the parameters of the encoder and α is a momentum hyperparamter. At last, for a positive triplet (m_i, r_i, c_i), the contrastive loss $\mathcal{L}_{cts,i}$ [16] for optimizing the encoding network is defined as:

$$-\log \frac{e^{(\phi(m_i,r_i,c_i)/\tau)}}{e^{\phi(m_i,r_i,c_i)/\tau} + \sum_{j=1}^{\mathcal{N}} e^{\phi(m_i,r_i,c_j^{'})/\tau}} \tag{4}$$

B. Relevance Relation-Based Optimization

As we have mentioned before, descriptions from different channels vary in style and purpose. We utilize a self-supervised in-batch negative sampling strategy to fix the gap. Inspired by Dual Softmax Loss (DSL) [17], we revise the contrastive

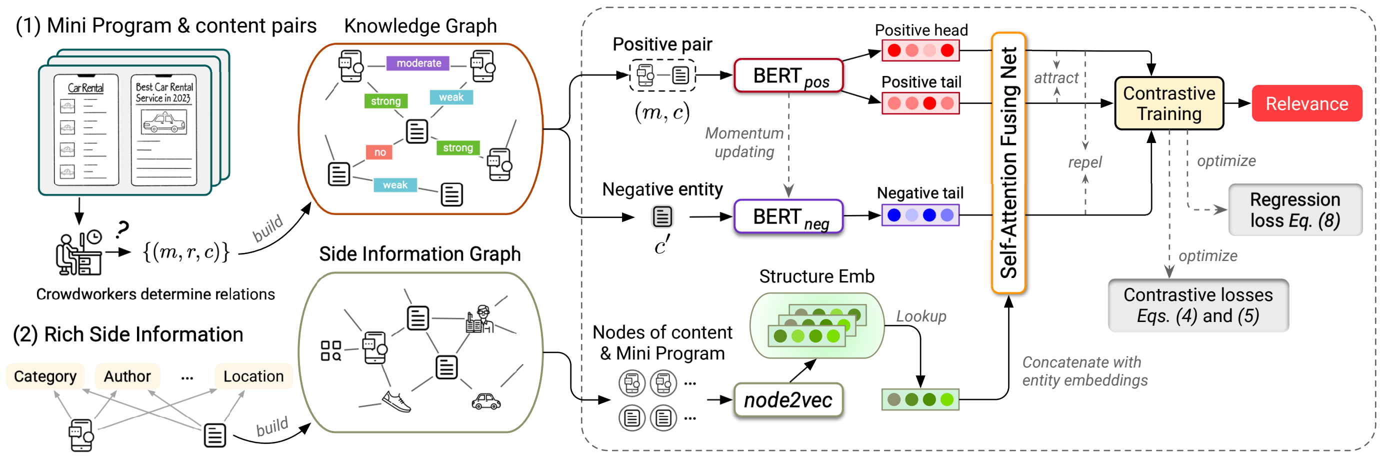

Fig. 2. Framework overview for extracting relevance relations between Mini Program and content in a super app.

loss in Eq. (4) by constructing a prior normalized matrix obtained by the Softmax operation on the columns of the entity similarity matrix. Then we multiply the prior normalized matrix with the original similarity matrix to highlight the scores of the ground-truth pair and reduce the impact of spurious correlations on our model. Revised contrastive loss is then defined as:

$$\mathcal{L}^{*}_{cts} = -\log \frac{e^{(\psi(i,i)/\tau)}}{e^{\psi(i,i)/\tau} + \sum_{j=1}^{\mathcal{N}} e^{\psi(i,j)/\tau}} \tag{5}$$

$$(i\ \ j) = \phi(m_i\ \ r_i\ \ c_i) \cdot P_{i,j} \tag{6}$$

$$P_{i,j} = \frac{e^{(\eta \cdot \phi(m_i, r_j, c_j))}}{\sum_{k=1}^{\mathcal{B}} e^{(\eta \cdot \phi(m_k, r_j, c_j))}} \tag{7}$$

where $P_{i,j}$ is entry of prior normalized matrix and η is a scale hyperparameter to smooth gradients.

Furthermore, in order to align the predicted relevance score with the labeled score from the mapping function, we adopt a regression loss based on mean squared error (MSE), which is defined as:

$$\mathcal{L}_{reg} = \frac{1}{\mathcal{B}} \sum_{i=1}^{\mathcal{B}} (f(m_i\ \ c_i) - g(r_i))^2 \tag{8}$$

C. Side Information Integration

Mini Programs and content in a super app are often associated with rich side information such as category, author, and location. Therefore, we propose to learn entity structure representations by constructing a large heterogeneous graph (more dense than the knowledge graph $\mathcal{G}$). The constructed graph not only contains nodes of Mini Program and content but also side information connected with them, which also exposes implicit node correlations beyond first-order neighborhoods. Specifically, we first utilize node2vec [18] to extract structure representations for each node in the graph. Then we create an MLP-based projection head to encode them. We fuse the structure representations with their corresponding semantic representations (output by the BERT-based encoders) via self-attention layers. The fused entity representations contain semantic-structural patterns that can be used for enhancing RRE performance.

Overall, our proposed model is optimized by the following loss:

$$\mathcal{L} = \mathcal{L}_{cts} + \beta \mathcal{L}^{*}_{cts} + \gamma \mathcal{L}_{reg} \tag{9}$$

where β and γ are loss-balance hyperparameters.

V. Experiments

Dataset. We construct a synthetic dataset for the RRE task, which comprises 13k Mini Programs and 22k contents. Regarding the labeled relation triplets, we have 19k for the "no correlation" class, 16k for "weak correlation", 8k for "moderate correlation", and 9k for "strong correlation".

The average sequence lengths of Mini Program and Content are 54 and 282, respectively.

Baselines. We adopt embedding-based methods including TransE [13], DisMult [14], ComplEx [19] and RotatE [20]; and text-based methods including KG-BERT [21], PrompKGC [22] and SimKGC [23]. As baselines are not designed to classify the relation between two entities, we also attach a regression head after these models. With this design, the confidence score of triplets can be granted the meaning of relevance strength.

Metrics. We use Precision, Recall, Accuracy, and F_1 score to measure the relation classification performance and use Pearson Correlation Coefficient (PCC) with MSE loss to measure the regression performance.

A. Main Result

The main results of the RRE performance comparison are shown in Table I, where we run each model five times and report the mean and standard deviation. We can observe that our proposed model significantly outperforms both embedding-based and text-based baselines, in comparison with the best-performed baseline SimKGC, we achieve an 11.76% improvement in terms of F_1 score. For other baselines, embedding-based methods generally surpass text-based methods, which

TABLE I
RELEVANCE RELATION EXTRACTION PERFORMANCE COMPARISON BETWEEN OUR MODEL AND BASELINES.

Method	Precision$_{\uparrow}$	Recall$_{\uparrow}$	F$_{1\uparrow}$	ACC$_{\uparrow}$	PCC$_{\uparrow}$	MSE$_{\downarrow}$
Embedding-based methods						
TransE [13]	$59.28_{\pm 0.134}$	$56.65_{\pm 0.110}$	$52.74_{\pm 0.170}$	$53.46_{\pm 0.299}$	$0.627_{\pm 0.002}$	$0.079_{\pm 0.001}$
DistMult [14]	$44.64_{\pm 0.208}$	$41.85_{\pm 0.104}$	$32.14_{\pm 0.024}$	$31.86_{\pm 0.031}$	$0.381_{\pm 0.011}$	$0.144_{\pm 0.001}$
ComplEx [19]	$37.38_{\pm 1.209}$	$35.39_{\pm 2.228}$	$28.04_{\pm 1.266}$	$27.60_{\pm 1.033}$	$0.281_{\pm 0.028}$	$0.160_{\pm 0.008}$
RotatE [20]	$42.86_{\pm 0.303}$	$45.68_{\pm 0.324}$	$41.47_{\pm 0.401}$	$43.76_{\pm 0.478}$	$0.329_{\pm 0.011}$	$0.160_{\pm 0.003}$
Text-based methods						
KG-BERT [21]	$24.24_{\pm 0.012}$	$26.64_{\pm 0.104}$	$23.66_{\pm 0.008}$	$37.86_{\pm 0.223}$	$0.105_{\pm 0.003}$	$0.169_{\pm 0.003}$
PromptKGC [22]	$25.07_{\pm 0.002}$	$25.20_{\pm 0.003}$	$22.15_{\pm 0.002}$	$23.43_{\pm 0.001}$	$-0.007_{\pm 0.001}$	$0.226_{\pm 0.001}$
SimKGC [23]	$55.34_{\pm 0.823}$	$58.97_{\pm 0.543}$	$55.42_{\pm 1.180}$	$59.36_{\pm 1.631}$	$0.559_{\pm 0.002}$	$0.099_{\pm 0.002}$
Ours	$\mathbf{68.57}_{\pm 0.029}$	$\mathbf{68.17}_{\pm 0.884}$	$\mathbf{67.18}_{\pm 0.186}$	$\mathbf{73.95}_{\pm 0.111}$	$\mathbf{0.774}_{\pm 0.010}$	$\mathbf{0.046}_{\pm 0.002}$

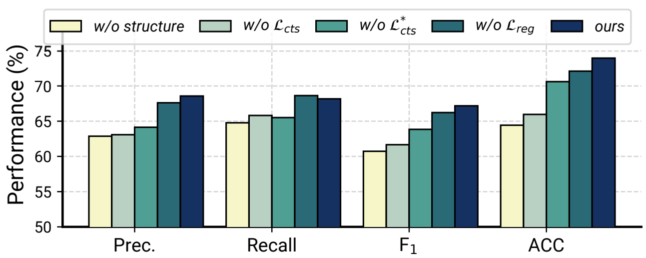

Fig. 3. Ablation on encoding network design.

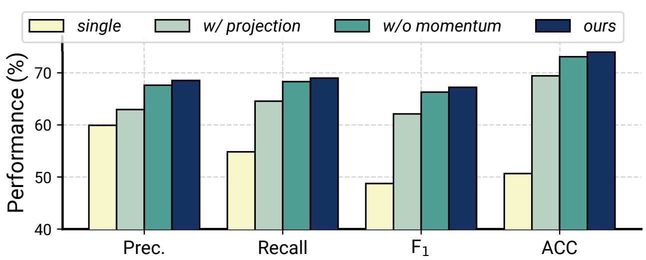

Fig. 4. Ablation on side information and loss.

is in line with the observations in [23], [24] that the KGC performance of text-based methods lags behind embedding-based methods. SimKGC expands the number and types of negatives in contrastive training and overtakes embedding-based methods on KGC. Our model is built on top of SimKGC with several new modules designed for the RRE task. In the next subsection, we conduct an ablation study to quantify the contribution of each module.

Ablation on Encoding Network. To demonstrate the effectiveness of our designed momentum encoding network, we test another three encoder architectures: (1) *single* means we only use one encoder without momentum updating; (2) *w/ projection* indicates the inclusion of a projection head on the *single*, demonstrating the impact of data augmentation; Similarly, (3) *w/o momentum* means we adopt bi-encoders but without momentum updating, allowing us to evaluate the effectiveness of disconnecting the connection between the encoders. The ablation result is shown in Fig. 3. It is evident that the

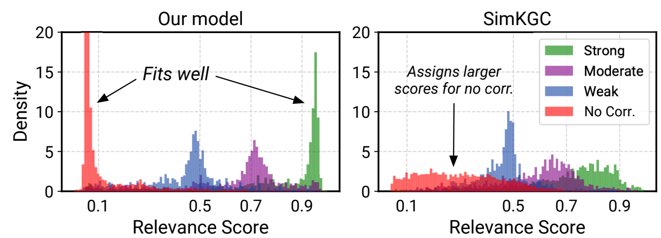

Fig. 5. The distribution of predicted relevance scores colored by their relevance classes of strong (0.9), moderate (0.7), weak (0.5), and no correlation (0.1).

single model exhibits the lowest performance. The inclusion of a projection head and the adoption of momentum updating both contribute to improvements in the RRE performance. This validates our underlying motivation to mitigate the detrimental effects arising from the interaction between in-batch positive and negative entities.

Ablation on side information and loss. We remove each of the following modules in our model to ablate their unique contributions to the performance, the result is shown in Fig. 4. Incorporating side information and modeling their structures can help us learn better entity representations. Both contrastive losses contribute to the RRE performance, and combining them yields even better results than using either individually. The using of regression loss also slightly increases the performance.

Analysis on Distribution of Relevance Score Fig. 5 plots the relevance score distributions (of 6,680 test triplets) predicted by our model and SimKGC. The scores are colored by their ground-truth classes. We can clearly observe that our model predicts much more accurate relevance scores (especially for pairs of strong and no correlations) in which their distributions fit well with the predefined scores.

Visualization. We use t-SNE to visualize the learned semantic-structural representations in Fig. 6. We can see that the entities with similar relevance scores or belonging to the same category are clustered together.

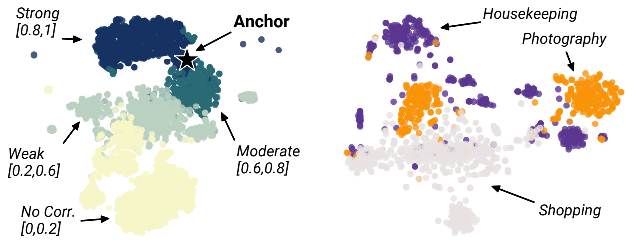

Fig. 6. Visualizations of our learned representations. Left: 500 randomly sampled entities, colored by their relevance scores to the **Anchor** entity (marked as black star ★). Right: 1,215 Mini Program entities from three categories.

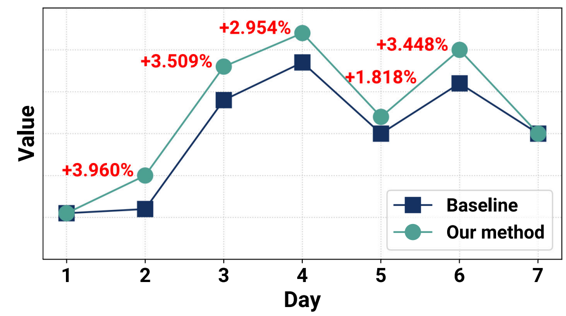

Fig. 7. Online CTRs performance during 7 days.

TABLE II
EXPERIMENT RESULTS IN DOWNSTREAM OFFLINE ITEM-CENTRIC RECOMMENDATIONS

Method	Recall	Hit	Precision	NDCG
LightGCN	0.0125	0.0837	0.0073	0.0100
UltraGCN	0.0243	0.1157	0.0086	0.0104
BC-Loss$_{raw}$	0.0297	0.1252	0.0109	0.0364
BC-Loss$_{ours}$	0.0328	0.1626	0.0260	0.0481

B. Offline and Online A/B Experiment

To demonstrate the effectiveness of our designed relevance score, we conducted offline item-centric recommendation and online user-centric recommendation experiments, which showed that the proposed framework could provide a semantic embedding extractor in a specific downstream scenario.

Offline item-centric recommendation experiment. To verify the effectiveness of our semantic relevance score, we compare it with other state-of-the-art recommendation methods [25], [26] and convert them from user-centric to item-centric by placing users with the Mini Programs and placing items with the contents. In the offline recommendation experiment, we used three months of user click behavior as training and validation sets, and used one-week future user click behavior as the test set. We constructed an offline item-centric recommendation dataset based on the same rules as the one used in BC-Loss [27]. After comparing with several commonly used recommendation algorithms, we chose BC-Loss as the baseline and integrated the relevance score we parsed into the margin score in a reasonable way. The experimental results showed that the integrated relevance score can alleviate the popularity bias in the recommendation task and improve the overall performance of the model.

Online recommendation experiment. We conducted the online A/B testing on the homepage of Alipay for 7 days. For policy reasons, we only publish improvements relative to the base model. Compared to traditional semantic feature extraction models that require extensive pre-training to adapt to special styles of downstream semantics [28], or rely on several models to work together to complete the semantic feature extraction and fusion, our model is trained via a semi-supervised learning approach with lower costs. The chosen metric is click-through rate (CTR). The results of the 7-day A/B test with 5% of user flow are shown in Fig. 7.

VI. Conclusion

Mini Programs have more than hundreds of millions of monthly active users. Understanding the semantics of Mini Program and content as well as their correlations in a super app is challenging and still under-explored. We take the first step to formulate relevance relation extraction as a KGC problem and propose a text-based contrastive learning framework to learn relevance-guided entity representations. We design a momentum encoding network with a new relevance scoring function and two optimization losses. We also incorporate rich side information associated with the Mini Program and content to enhance semantic representation learning. Extensive experiments showed the effectiveness of our framework in both relevance score extraction and cross-domain recommendation. In our future work, we aim to incorporate multimodal information and explore modeling techniques for interactions among various entities within the super app ecosystem. Specifically, we plan to investigate the relationships between Mini Programs and coupons, along with other entities.

Acknowledgment

This work was supported in part by the Ant Group, in part by the National Natural Science Foundation of China under Grants No. 62020106008, No. U22A2097 and No. U23A20315, in part by the Kuaishou, and in part by the SongShan Laboratory under Grant No. YYJC012022019.

References

[1] S. Rajaee and M. T. Pilehvar, "An isotropy analysis in the multilingual BERT embedding space," in *Findings of ACL*, 2022, pp. 1309–1316.

[2] S. Ji, S. Pan, E. Cambria, P. Marttinen, and S. Y. Philip, "A survey on knowledge graphs: Representation, acquisition, and applications," *IEEE TNNLS*, vol. 33, no. 2, pp. 494–514, 2021.

[3] I. Ferrari, G. Frisoni, P. Italiani, G. Moro, and C. Sartori, "Comprehensive analysis of knowledge graph embedding techniques benchmarked on link prediction," *Electronics*, vol. 11, p. 3866, 2022.

[4] L. Cui, Y. Xu, T. Lv, and F. Wei, "Document AI: Benchmarks, models and applications," in *DIL Workshop in ICDAR*, 2021.

[5] H. Pei-jun, C. Peng, W. Xuan, Z. Jian-Le, and C. Fang-Fang, "Research on key technology of wechat mini program based on secure QR code," *Annual International Conference on Network and Information Systems for Computers*, pp. 283–287, 2018.

[6] L. Hao, F. Wan, N. Ma, and Y. Wang, "Analysis of the development of wechat mini program," *Journal of Physics: Conference Series*, vol. 1087, 2018.

[7] F. Zhou, X. Xu, G. Trajcevski, and K. Zhang, "A survey of information cascade analysis: Models, predictions, and recent advances," *ACM Computing Surveys*, vol. 54, no. 2, pp. 1–36, 2021.

[8] M. Schreieck, ang Shu Ou, and H. Krcmar, "Mini-app ecosystems," *Business & Information Systems Engineering*, vol. 65, pp. 85–93, 2022.

[9] G. Wang, T. Zhong, X. Xu, K. Zhang, F. Zhou, and Y. Wang, "Vector-quantized autoencoder with copula for collaborative filtering," in *CIKM*, 2021, pp. 3458–3462.

[10] J. Zhang, L. Yang, Y. Han, Z. Sun, and Z. Xiang, "A small leak will sink many ships: Vulnerabilities related to Mini Programs permissions," *arXiv:2205.15202*, 2022.

[11] F. Zhou, X. Qi, X. Xu, J. Wang, T. Zhong, and G. Trajcevski, "Meta-learned user preference for topic participation prediction," in *GLOBECOM*, 2020, pp. 1–6.

[12] K. Cheng, M. Schreieck, M. Wiesche, and H. Krcmar, "Emergence of a post-app era – an exploratory case study of the WeChat mini-program ecosystem," in *International Conference on Wirtschaftsinformatik*, 2020.

[13] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, "Translating embeddings for modeling multi-relational data," in *NIPS*, 2013.

[14] B. Yang, W. Yih, X. He, J. Gao, and L. Deng, "Embedding entities and relations for learning and inference in knowledge bases," in *ICLR*, 2015.

[15] K. Teru, E. Denis, and W. Hamilton, "Inductive relation prediction by subgraph reasoning," in *ICML*, 2020, pp. 9448–9457.

[16] X. Xu, F. Zhou, K. Zhang, and S. Liu, "CCGL: Contrastive cascade graph learning," *IEEE TKDE*, vol. 35, no. 5, pp. 4539–4554, 2023.

[17] X. Cheng, H. Lin, X. Wu, F. Yang, and D. Shen, "Improving video-text retrieval by multi-stream corpus alignment and dual Softmax loss," *arXiv:2109.04290*, 2021.

[18] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," *SIGKDD*, 2016.

[19] T. Trouillon, J. Welbl, S. Riedel, É. Gaussier, and G. Bouchard, "Complex embeddings for simple link prediction," in *ICML*, vol. 48, 2016, pp. 2071–2080.

[20] Z. Sun, Z. Deng, J. Nie, and J. Tang, "RotatE: Knowledge graph embedding by relational rotation in complex space," in *ICLR*, 2019.

[21] L. Yao, C. Mao, and Y. Luo, "KG-BERT: BERT for knowledge graph completion," *arXiv:1909.03193*, 2019.

[22] N. Zhang, X. Xie, X. Chen, D. Shumin, Y. Hongbin, and C. Huajun, "Knowledge collaborative fine-tuning for low-resource knowledge graph completion," *Journal of Software*, vol. 33, no. 10, 2022.

[23] L. Wang, W. Zhao, Z. Wei, and J. Liu, "SimKGC: Simple contrastive knowledge graph completion with pre-trained language models," in *ACL*, 2022, pp. 4281–4294.

[24] X. Wang, T. Gao, Z. Zhu, Z. Liu, J.-Z. Li, and J. Tang, "KEPLER: A unified model for knowledge embedding and pre-trained language representation," *Transactions of the ACL*, vol. 9, pp. 176–194, 2019.

[25] X. He, K. Deng, X. Wang, Y. Li, Y. Zhang, and M. Wang, "LightGCN: Simplifying and powering graph convolution network for recommendation," in *SIGIR*, 2020, pp. 639–648.

[26] K. Mao, J. Zhu, X. Xiao, B. Lu, Z. Wang, and X. He, "UltraGCN: Ultra simplification of graph convolutional networks for recommendation," in *CIKM*, 2021, pp. 1253–1262.

[27] A. Zhang, W. Ma, X. Wang, and T. Chua, "Incorporating bias-aware margins into contrastive loss for collaborative filtering," in *NeurIPS*, 2022.

[28] Y. Huang, T. Lv, L. Cui, Y. Lu, and F. Wei, "Layoutlmv3: Pre-training for document AI with unified text and image masking," in *ACM MM*, 2022, pp. 4083–4091.