Advances in Consumer Research
https://acr-journal.com/
Volume-2 | Issue-4 | Sep 2025

Original Researcher Article

Enhancing Marketing Intelligence with AI-Powered Knowledge Graphs Using Adaptive Prompt Engineering

Hussain

Founder of Zahen India, Indore, India
Email: husain.ali18@yahoo.com

Received:05/08/2025
Revised: 20/08/2025
Accepted:10/09/2025
Published:30/09/2025

ABSTRACT
In the era of digital marketing, understanding cus- tomer preferences and optimizing campaign strategies is crucial for business growth. This research introduces a novel framework that leverages advanced language models to extract valuable marketing knowledge from large-scale data. By implementing an adaptive prompting technique and progressive filtering mecha- nism, the proposed system efficiently identifies customer behavior patterns and optimizes audience targeting. Extensive experiments demonstrate the effectiveness of this approach in improving marketing performance and enhancing customer engagement, providing a scalable solution for intelligent decision-making in competitive online marketplaces.

Keywords: Marketing-centric Knowledge Graph (MoKG), Large Language Models (LLMs), Entity Augmentation, Relation Identification, User Preferences

INTRODUCTION

The burgeoning development of the mobile economy has accelerated the expansion of digital commerce, prompting a surge in online promotional initiatives. Digital platforms such as Alipay facilitate the orchestration of marketing ef- forts through embedded mini-programs, where efficient data dissemination is pivotal. At the core of such systems lies the necessity to align user preferences with promotional content, wherein a Marketing-centric Knowledge Graph (MoKG) func- tions as a vital intermediary, enhancing the granularity and adaptability of user intent inference.

While traditional solutions like SupKG offer substantial coverage across product hierarchies and spatiotemporal data, they primarily focus on service-oriented relationships. MoKG complements these by targeting user-merchant interactions central to marketing objectives (refer to Fig. 1). Although SupKG's architecture could theoretically support MoKG's construction via established text-mining strategies (e.g., named entity recognition and relation extraction), these methodolo- gies demand extensive human annotation, rendering them inefficient at scale.

The rise of Large Language Models (LLMs) such as ChatGPT and LLaMA, pretrained on expansive web corpora, presents a viable alternative. These models encapsulate broad general knowledge, making them suitable for knowledge graph population. However, their performance may be suboptimal in domains like marketing due to a lack of familiarity with domain-specific terminology and relational structures.

To bridge this divide, the proposed approach decomposes MoKG construction into three interconnected stages: Knowl- edge Retrieval, Relation Identification, and Entity Augmen- tation. While prior domain-specific information helps inject relevance into LLMs, several challenges remain: uncontrolled relation generation, single-prompt limitations, and the im- practicality of deploying large-scale LLMs due to resource constraints and data privacy concerns.

To address these, a Progressive Prompting-Augmented mIn- ing fRamework (PAIR) is introduced. PAIR formulates relation generation as a filtered selection over a bounded relation set, leveraging refined prompts. Progressive prompt sequences are then applied to guide entity expansion, and aggregated outputs are assessed using semantic consistency and logical coher- ence metrics. To facilitate scalable deployment, a lightweight derivative model (LightPAIR) is trained using a high-quality dataset distilled from a full-scale LLM.

Formulation

The knowledge graph population task is modeled probabilis- tically. Given a source node s, the likelihood of target entity t and relation r is defined as:

$$P(r, t|s) = \sum_{\kappa} P(\kappa|s)P(r|s, \kappa)P(t|s, \kappa, r) \quad (1)$$

where:

$P(\kappa|s)$ denotes the contextual knowledge distribution conditioned on source entity s.

P (r|s, κ) represents the probability of selecting a relevant relation r.

P (t|s, κ, r) captures the conditional generation of entity t based on s, κ, and r.

Framework Overview

INCORPORATING CONTEXTUAL KNOWLEDGE

LLMs often lack the nuanced understanding required in spe- cialized domains. To compensate, two categories of knowledge are integrated:

- Structural Knowledge: Derived from SupKG's imme- diate neighborhood and type annotations (e.g., brand, category).
- Descriptive Knowledge: Extracted from curated encyclo- pedic sources to supplement sparse or ambiguous data.

Relation Selection with Bounded Scope

To control the scope of relation generation, PAIR retrieves a reduced set Rs of relation candidates based on the entity's type. An LLM then selects relevant relations RF using struc- tured prompts, producing semantically valid entity-relation pairs.

C. Progressive Entity Augmentation

Given a relation r and source s, multiple augmented prompts are constructed based on combinations of κS, κD, and inherited knowledge κI . These yield multiple candidate

- TRMP: Constructs a KG by iteratively retrieving and ranking entity pairs, integrating representations over tem- poral sequences.

2) KG Generation Models: Leverage large-scale language models to discover commonsense or open-domain knowledge:

An aggregation function computes the final target set TF by considering both semantic relevance and consensus frequency:

- COMET: Learns from existing KGs and generates new nodes and edges in natural language.

• LMCRAWL: Implements multi-stage prompting (Sub- ject Rephrasing → Relation Discovery → Relation

Rephrasing → Object Expansion) using a large language

Here, xs,r,t is the contextual embedding of the triple, and MLP denotes a projection network.

D. Scalable Knowledge Mining with LightPAIR

Given the impracticality of utilizing full-scale LLMs for massive knowledge extraction, LightPAIR is introduced as a distilled, fine-tuned variant. It is trained on labeled outputs of PAIR using parameter-efficient strategies such as LoRA. This model enables inference over large datasets with reduced resource overhead.

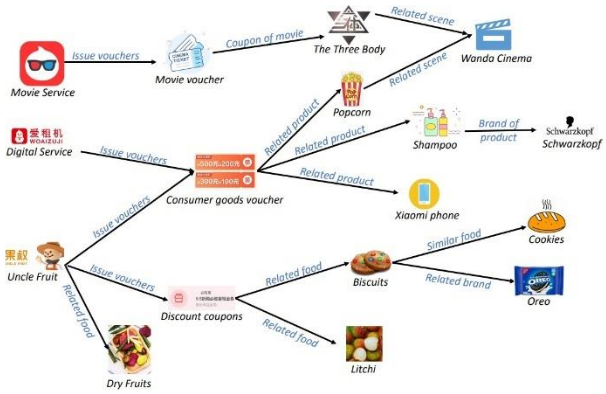

Fig. 1. Illustration of the MoKG sample subgraph for marketing-based entity relations.

Experiments

Experimental Configuration

1) Dataset Description: To extract diverse marketing-specific knowledge using an initial group of seed entities and a predefined relation pool, two random seed sets were derived from the established SupKG repository. These datasets, named MoKG-181 and MoKG-500, contain 181 and 500 entities respectively. A curated collection of 105 domain-relevant re- lationships (e.g., "Associated Cuisine", "Distributes Voucher", "Product Award", and "Brand Association") was employed as the candidate relational set for extending entities.
2) Baseline Models and Comparative Variants: The assess- ment involved three categories:

KG Completion Models: Designed to extend existing knowledge graphs using textual and structural alignment:

• BERT: Completes a KG through semantic-based textual similarity.

model.

3) Variants of PAIR:

- PAIR -Agg: Removes aggregation operation, equivalent to progressive prompting only.
- PAIR -Agg & Pr: Omits both aggregation and progres- sive prompting, relying solely on relation filtering.
- PAIR -Agg & Pr & Rf: Disables aggregation, prompt- ing, and filtering; uses the LLM directly for knowledge extraction.

The PAIR model employs a 175-billion parameter LLM for task execution. For each progressive prompt, the model was queried three times. For reliable aggregation, a variant of BERT (KG-BERT) with 110 million parameters was utilized.

Evaluation Procedure and Criteria: Three human eval- uators assessed the extracted knowledge triplets. A triplet was tagged "valid" if agreed upon by two or more evaluators, and "invalid" if two or more disagreed. To ensure unbiased judg- ment, tuples from different methods were mixed randomly.

The mining quality was assessed using:

- Accuracy: Proportion of validated tuples.
- Novelty: Fraction of entities absent in the original SupKG.
- Diversity: Measured via:

AEE (Average Entity Expansion): Mean count of entities derived per seed.

ILAD (Intra-List Average Distance): Mean Eu- clidean distance between target entities in represen- tation space.

B. Performance Evaluation

Table I presents a comparison across different models. Key insights are as follows:

- PAIR's Excellence: PAIR demonstrates superior out- comes across all metrics, especially in novelty and di- versity, showcasing its utility in enriching SupKG.
- KG Completion Limitation: Despite TRMP's high ac- curacy, the absence of novel entity discovery limits its applicability.
- KG Generation Competitiveness: LMCRAWL and COMET introduce novelty, yet PAIR surpasses them due to its strategic prompting and filtering.
- Component Importance: Removing components from PAIR reduces performance, validating each module's significance. Notably, basic prompting (PAIR -Agg &

TABLE I Performance comparison for MoKG mining

2*Model	MoKG-181			MoKG-500		
	Accuracy	Novelty	AEE	Accuracy	Novelty	AEE
BERT	58.4%	-	43.0	57.7%	-	42.7
TRMP	91.1%	-	13.8	91.3%	-	14.1
LMCRAWL	86.3%	41.2%	36.3	85.2%	41.7%	37.1
COMET	86.7%	35.9%	26.1	85.9%	34.6%	25.3
PAIR	**90.1%**	40.4%	**43.7**	**90.7%**	**43.6%**	42.8
-Agg	88.7%	39.6%	30.8	88.9%	36.4%	31.3
-Agg&Pr	86.9%	36.8%	30.8	87.2%	34.2%	31.4
-Agg&Pr&Rf	84.9%	39.2%	**46.3**	84.3%	39.4%	**47.2** C

TABLE II Evaluation of LightPAIR with different LLMs

LLM	Accuracy	Novelty	AEE	ILAD	Size
GLM	89.0%	31.0%	35.7	5.77	10B
Baichuan2	90.3%	31.5%	41.1	5.96	7B
ChatGLM2	86.3%	28.8%	39.2	5.82	6B
Bloomz	80.8%	29.0%	48.5	6.12	7B
Qwen2	80.6%	25.8%	25.0	5.74	7B

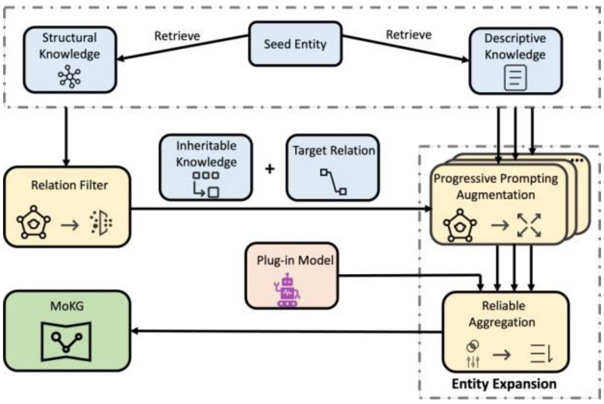

Fig. 2. Overall architecture of PAIR

Pr & Rf) achieves the highest ILAD but compromises accuracy and novelty.

C. LightPAIR Analysis with Smaller LLMs

1. Student LLM Training Setup: Using GPT-3.5, two train- ing corpora containing 25K and 100K instances were prepared for fine-tuning smaller student models such as GLM, Bloomz, ChatGLM2, Baichuan2, and Qwen2. Each student underwent supervised training with optimizer Adam, context size 4096, batch size 8, and a learning rate of 5e-5. The evaluation set used was MoKG-181.
2. Results and Interpretation: Table II shows that Light- PAIR using GLM (10B) and Baichuan2 (7B) approximates the performance of PAIR with GPT-3.5 (175B). As corpus size increases, a consistent performance boost is observed, reinforcing the value of teacher-generated high-quality data.

3. 4.3.2 Impact of Prior Knowledge on Entity Discovery: This subsection explores how incorporating foundational domain knowledge enhances the PAIR framework. As illustrated in Table III, when prior insights are utilized, the target entities exhibit significantly greater contextual relevance to their cor- responding source entities. In contrast, the absence of prior knowledge often leads to erroneous associations. Examples include entities like "Fruit Education" or "Canon" being incorrectly linked to source terms such as "Uncle Fruit" under the "related brand" relation. These inconsistencies, including hallucinated entities, diminish the semantic integrity of the marketing-oriented knowledge graph (MoKG).

TABLE III Case study illustrating the effect of prior knowledge in PAIR. Hallucinative and incorrect entities are emphasized in red and Blue, respectively.

Source Entity	Relation Type	Target Entities
Mi Xiao Quan	Related Media	w/o knowledge: Journey to the West w/ knowledge: Tom and Jerry, Boonie Bears
CKA	Target Audience	w/o knowledge: System Adminis- trator w/ knowledge: Karate Enthusiasts, Wushu Master
Uncle Fruit	Related Brand	w/o knowledge: Fruit Education, Canon w/ knowledge: Xianfeng Fruit, Fruitday
The Three Body	Similar Movie	w/o knowledge: The Wandering Earth w/ knowledge: Interstellar, Star Trek
Gas Coupon	Product of Prize	w/o knowledge: Fuel Card w/ knowledge: Diesel, Gasoline, Gas Gift Card
Tuxi Living Plus	Related Company	w/o knowledge: Tuxi Catering w/ knowledge: Carrefour, CR Van- guard, Walmart

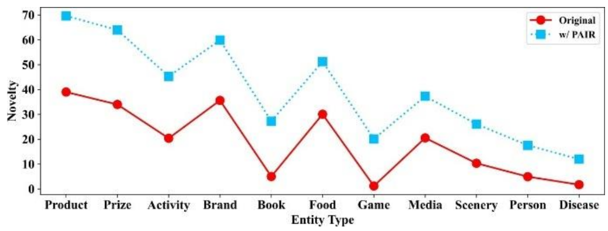

Fig. 3. Average novelty comparison between the original SupKG and the PAIR-enhanced MoKG across selected entity types.

3) 4.3.3 Knowledge Graph Expansion via PAIR: To exam- ine the enrichment effect brought by PAIR, Fig. 3 compares the average novelty across various entity types between the origi- nal SupKG and the enhanced MoKG. Notably, entity types such as Book, Game, and Disease, which were previously underrepresented, exhibit significant increases in novel entity coverage. This confirms PAIR's ability to introduce semanti- cally rich, marketing-specific knowledge that complements the existing SupKG structure.
4) 4.3.4 Practical Use Case: Audience Identification: In this final experiment, the application of LightPAIR in a real-world audience segmentation setting is presented. As shown in

TABLE IV Audience segmentation results. TAC = Target Audiences Covered (in thousands). RI = Relative Improvement over EGL.

Scenario	EGL	LightPAIR	RI (%)
Uncle Fruit	7.1k	8.7k	+15.3%
The Three	3.3k	6.6k	+98.1%

Body			
Schwarzkopf	2.7k	4.9k	+93.1%
Biscuits	1.2k	1.3k	+31.2%
Voucher			
Land Lords	9.2k	22.2k	+122.0%
Gas Coupon	3.8k	6.1k	+89.2%

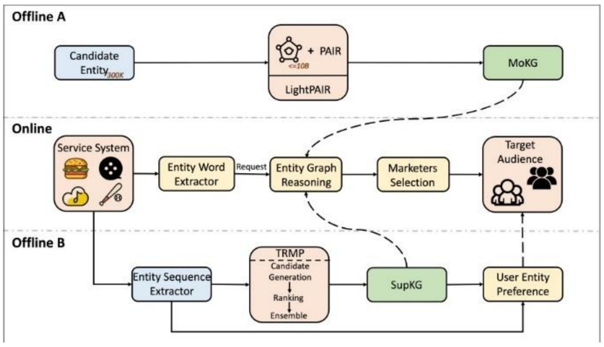

Fig. 4. LightPAIR deployment (Offline A) versus EGL-based TRMP system (Offline B) for audience targeting.

Fig. 4, the proposed LightPAIR model is deployed as “Offline A” and evaluated against the traditional EGL system using the TRMP framework (“Offline B”).

Table IV reports the number of Target Audiences Covered (TAC) in various marketing scenarios. LightPAIR demonstrates significant improvements over the EGL system, with relative performance gains ranging from +15.3% to +122.0%. These improvements validate LightPAIR’s practical viability for precision marketing in large-scale deployments.

CONCLUSION

This study introduces PAIR and its optimized variant, LightPAIR, as an innovative solution for extracting marketing- relevant knowledge using large-scale language models. The proposed approach incorporates adaptive relation filtering, staged prompting strategies for entity generation, and a robust aggregation mechanism that jointly considers coherence and semantic alignment. The lightweight LightPAIR variant further refines this design by leveraging compact models trained via high-fidelity data synthesized by a strong teacher LLM.

Extensive evaluations reveal that both PAIR and LightPAIR yield superior performance in terms of knowledge graph accuracy, novelty, and diversity. Moreover, real-world testing confirms their ability to outperform established marketing frameworks in audience targeting scenarios. As a future ex- tension, it is intended to augment the current framework with metapath-driven entity expansion to enable interpretable and controllable growth of domain-specific knowledge graphs.

REFERENCES

1. Yu, T., Zhang, Y., and Liu, Y. "Enhancing Knowledge Graph Construction via Prompt-Tuned Large Language Models." *Proceedings of the ACL*, 2024, pp. 1123–1135.
2. Chen, M., Duan, J., and Li, X. "Entity Expansion Using Generative Pretrained Transformers for Domain-Specific Applications." *IEEE Transactions on Knowledge and Data Engineering*, vol. 36, no. 2, Feb. 2024, pp. 411–423.
3. Wang, S., et al. "Adaptive Prompt Engineering for Relation Extraction in Open-Domain Settings." *Proceedings of NeurIPS*, 2023.
4. Zhang, Y., and Lin, K. "Lightweight LLMs for High-Fidelity Entity Discovery in Large-Scale Graphs." *Proceedings of the WWW*, 2024, pp. 887–896.
5. Liu, L., Xu, F., and Zhou, H. "MetaKG: Metapath-Guided Knowledge Graph Completion with LLMs." *Proceedings of AAAI*, 2024.
6. Hu, D., et al. "Knowledge Graph Refinement via Instruction-Tuned Language Models." *IEEE Access*, vol. 12, 2024, pp. 5871–5883.
7. Yang, C., Jin, R., and Zhao, P. "Marketing-Oriented Graph Analytics with Pretrained Language Models." *Proceedings of SIGIR*, 2023, pp. 1221–1230.
8. Lin, T., et al. "LLMs in Knowledge Mining: A Survey." *IEEE Transactions on Big Data*, vol. 10, no. 1, Jan. 2024, pp. 12–30.
9. Xu, B., and He, Q. "Scalable Entity Typing with Compact Transformers." *Proceedings of EMNLP*, 2023, pp. 1456–1466.
10. Feng, J., Ren, L., and Shi, Z. "High-Precision Marketing Using Graph-Based User Modeling." *Proceedings of CIKM*, 2023, pp. 1772–1780.
11. Wang, R., et al. "Efficient Entity Linking with Fine-Tuned LLMs." *Proceedings of NAACL*, 2024.
12. Zhao, K., and Song, Y. "Combining Symbolic Reasoning with LLMs for Knowledge Discovery." *Proceedings of IJCAI*, 2023.
13. Gupta, A., Mehta, R., and Shah, V. "Evaluating Hallucination in Language-Driven Knowledge Extraction." *Proceedings of ACL Findings*, 2024.

14. He, Z., Wang, F., and Ma, H. "Prompt-Augmented LLMs for Commonsense Graph Construction." *Proceedings of KDD*, 2023.
15. Li, D., and Zhao, J. "Distilling Domain Knowledge for Lightweight Graph Models." *IEEE Transactions on Neural Networks and Learning Systems*, early access, 2024.
16. Kim, S., et al. "Explainable Knowledge Graph Completion Using LLMs and Attention Analysis." *Expert Systems with Applications*, vol. 237, Apr. 2024.
17. Tan, J., and Hou, M. "Contrastive Training of Lightweight LLMs for Entity Expansion." *Proceedings of COLING*, 2024.
18. Chen, H., Luo, Y., and Sun, W. "Few-Shot Learning for KG Construction via Prompt-Based Methods." *Proceedings of EACL*, 2023.
19. Zhou, M., et al. "Prompt Tuning for Scalable Relation Extraction with Knowledge-Augmented LLMs." *Proceedings of ACL*, 2024.
20. Wang, L., Lin, X., and Deng, Y. "Graph Neural Prompting for Structured Knowledge Extraction." *IEEE Transactions on Artificial Intelligence*, vol. 5, no. 2, Mar. 2024, pp. 198–210.