Fusion of Structural-Aware Prompts and Contrastive Representation Learning for Commonsense Knowledge Graph Completion

Yu Liu, Jun Ma, Xuan Xie, Xun Liu, and Yanmei Kang*
University of International Relations, Beijing, China
{liuyu, majun, xiexuan, xun}@uir.edu.cn, kymei@163.com
*Corresponding author

***Abstract*—Commonsense Knowledge Graph Completion (CKGC) aims to infer missing relationships between entities in a commonsense knowledge graph (CKG). However, existing methods face two significant challenges: the loss of topological information when mapping structured triples into linear text for pre-trained language models (PLMs), and the semantic ambiguity of similar entities that can lead to representation confusion. To address these challenges, we propose a novel inductive commonsense knowledge graph completion method, Fusion of Structural-Aware Prompts and Contrastive Representation Learning for Commonsense Knowledge Graph Completion. Our method introduces two key innovations: (1) structure-aware prompt generation, which encodes local subgraph structures to guide PLMs in generating structure-sensitive representations; (2) contrastive learning, which uses an adversarial loss to better distinguish semantically similar but structurally different entities. Extensive experiments on benchmark datasets demonstrate that our method significantly improves performance, verifying the effective synergy of structural prompts and contrastive representation learning.**

***Index Terms*—Knowledge Graph Completion, Structure-aware Prompting, Contrastive Representation Learning**

I. Introduction

Commonsense Knowledge Graphs (CKGs) aim to model a vast array of everyday concepts in human cognition, such as "*water quenches thirst*" or "*birds can typically fly.*" Commonsense knowledge plays a crucial role in various artificial intelligence applications [1], [2]. For instance, CKGs enhance the reasoning capabilities of AI systems, particularly in ambiguous and context-dependent scenarios. Furthermore, CKGs can help mitigate the "hallucination" phenomenon in large language models by providing structured knowledge anchors.

By simulating the human cognitive framework, the vast array of implicit relational facts in CKGs significantly enhances the adaptability of AI systems in real-world environments. As the real world is open and dynamic, the inherent incompleteness of CKGs limits their practical applicability. To address this, Commonsense Knowledge Graph Completion (CKGC) has become a core task within the knowledge graph community, aiming to infer missing triples through reasoning mechanisms. For instance, an incomplete triple (e.g., "*eating too much sugar, causes, ?*") can be completed as (e.g., "*eating too much sugar, causes, health problems*"). Current approaches primarily rely on graph-structure-based reasoning mechanisms [3]–[5].

In recent years, Pre-trained Language Models (PLMs), with their implicitly encoded commonsense knowledge base and powerful sequential reasoning capabilities, have introduced a new paradigm for CKGC. Compared to traditional knowledge graph completion methods, PLMs leverage commonsense knowledge from large-scale corpora to infer missing facts without relying on explicit rule designs. However, as the complexity of tasks increases, the limitations of PLMs in commonsense reasoning gradually become evident, particularly when it comes to structural information modeling and prompt mechanism design, where key bottlenecks exist. Existing approaches [6], [7] primarily transform knowledge graph triples into natural language text, i.e., a serialized representation of "*head entity - relation - tail entity,*" to adapt to the input format of PLMs.

However, this transformation introduces a modality gap, where the topological structure of the knowledge graph is weakened or even completely lost when converted into linear text. In knowledge graphs, the way entities are connected often carries significant semantic information. For example, in the ATOMIC dataset, the triple *(X, xNeed, Y)* can be transformed into the sentence "*X needs Y,*" but this text sequence fails to explicitly express the specific relationship type between X and Y and their network structure in the graph. In a subgraph containing multiple related entities, X may simultaneously have several different "$xNeed$" relations pointing to different objects, yet PLMs, which input a single text sequence, struggle to capture this structured information. The loss of structural information leads to PLMs being unable to accurately model the complex interactions between entities, thereby affecting the completion results. Moreover, existing methods, in the absence of graph structures, often rely on the commonsense knowledge encoded internally within PLMs for reasoning. This makes the model's predictions more susceptible to the distribution of the training corpus and limits its ability to generalize to more complex knowledge graph completion tasks.

Another significant challenge is the semantic ambiguity of PLMs and their high sensitivity to prompt mechanisms. Many entities in commonsense knowledge graphs have highly similar

Supported by the Student Academic Research Training Project of the University of International Relations (No. 3262025SYJ07).

textual descriptions, yet in actual semantic and reasoning tasks, they may play entirely different structural roles. For instance, "*sugar*" and "*carbohydrate*" are semantically close, but their connections in a knowledge graph could differ significantly. The former may be more related to concepts such as "*sweetness*" or "*food additives*," while the latter is more closely associated with "*nutrients*" or "*energy sources*." In the absence of structural information, PLMs are prone to erroneously treat these two entities as interchangeable based solely on textual similarity, leading to representation confusion and ultimately impacting the accuracy of reasoning.

To address the aforementioned challenges, we propose a novel commonsense knowledge graph completion method that fuses structural-aware prompts with contrastive representation learning. Our method introduces a graph convolutional network to encode the k-hop neighborhood subgraph of the target triple, generating structure-aware soft prompt vectors. These soft prompts are concatenated with manually defined hard prompt templates to form composite prompts, allowing the PLM to explicitly perceive the structural roles of entities in the graph while retaining language priors. To overcome the representation ambiguity when handling semantically similar entities, we also construct positive and negative sample pairs based on structural similarity. By minimizing the contrastive loss, we force the PLM to distinguish between semantically and structurally decoupled entities. Notably, during training, we freeze the PLM parameters and only fine-tune the prompt vectors and contrastive modules, ensuring that the PLM's linguistic knowledge is preserved while avoiding catastrophic forgetting, thus further enhancing the model's generalization ability. We conducted experiments on the public datasets ConceptNet and ATOMIC. The experimental results demonstrate significant performance improvements in the CKGC task, validating the tremendous potential of our method in this domain. Our main contributions can be summarized as follows:

- We propose a novel method for CKGC that combines structure-aware prompting with contrastive learning to enhance PLM performance.
- We introduce a contrastive learning mechanism is introduced to resolve semantic ambiguity, improving reasoning accuracy for similar entities.
- We conducted extensive experiments on ConceptNet and ATOMIC datasets show significant performance improvements, validating the method's effectiveness.

II. Related Work

KGC as a core technique for enhancing KG quality, has evolved along two main methodological paradigms: transductive and inductive approaches. Transductive methods have achieved significant progress through embedding-based and GNN-based models. Embedding models like TransE [8] and its variants (TransH [9], TransR [10], and TransD [11]) model relations as translations between entity embeddings. RotatE [12] captures symmetry via rotations in complex space, while bilinear models like RESCAL [13] and DistMult [14] balance expressiveness and efficiency. GNN-based models, such as R-GCN [15], assign relation-specific weights, and R-GAT [15] adds attention to dynamically weigh neighbors. Methods like SGBC [16] and RGAT [15] densify graph structures, while Dense-ATOMIC [17] uses event templates to constrain reasoning. However, these models struggle with short texts or complex causal chains in CKGs due to node sparsity and semantic ambiguity.

Inductive methods, aimed at generalizing to unseen entities, use subgraph reasoning (e.g., Grail [3], RMPI [4]) to extract patterns from local subgraphs, while relation-topology models (e.g., TACT [5], InGram [18]) capture dependencies among relations. To address the challenges of abstraction and dynamism in CKGs, inductive methods integrate textual semantics and graph structure, as seen in InductivE [19] and MICO [20], which combine entity descriptions and causal reasoning. Despite their success, efficiently combining structure and semantics in large, sparse, and multimodal CKGs remains a key challenge.

III. Methodology

The framework of our method is shown in Figure 1, with a three-stage progressive architecture design: (a) The Subgraph Structure Encoder aggregates local neighborhood information of the target triple(h, r, t) through hierarchical graph convolution, generating topology-aware entity-relation embeddings; (b) The Structure-Aware Prompt Fusion Module jointly encodes the structure embeddings with predefined hard prompt templates (e.g., "*[X] motivated by goal [Y]*"), generating hybrid prompt vectors that are both semantically aligned and structurally sensitive; (c) The Structure-Constrained Contrastive Optimizer constructs positive and negative sample pairs based on structural similarity, optimizing the PLM's ability to distinguish between semantically similar but structurally different entities, reducing its over-reliance on superficial text features, and thus improving reasoning accuracy.

A. Subgraph Structure Encoder

This module aims to encode the local subgraph $G_k(h)$ of the target triple (h, r, t) into structure-sensitive vector representations using Knowledge Graph Embedding (KGE) techniques. By centering on the head entity h, we sample its k-hop neighborhood to construct the subgraph $G_k(h)$, which includes the associated entity set V_h and relation edge set ε_h. The initial embeddings of entities and relations $E_h^{(0)} \in \mathbb{R}^d$, $E_r^{(0)} \in \mathbb{R}^d$ are randomly initialized using Xavier normal distribution [21], where d is the dimension of the latent space. Through multiple layers of graph convolution operations, the node representations are iteratively updated. The mathematical expression for the update at the l-th layer is:

$$E_h^{(l+1)} = \sigma\left(W_s^{(l)} E_h^{(l)} + \sum_{v \in N(h)} W_n^{(l)} E_v^{(k)} \odot W_r^{(l)} E_r^{(l)}\right) \tag{1}$$

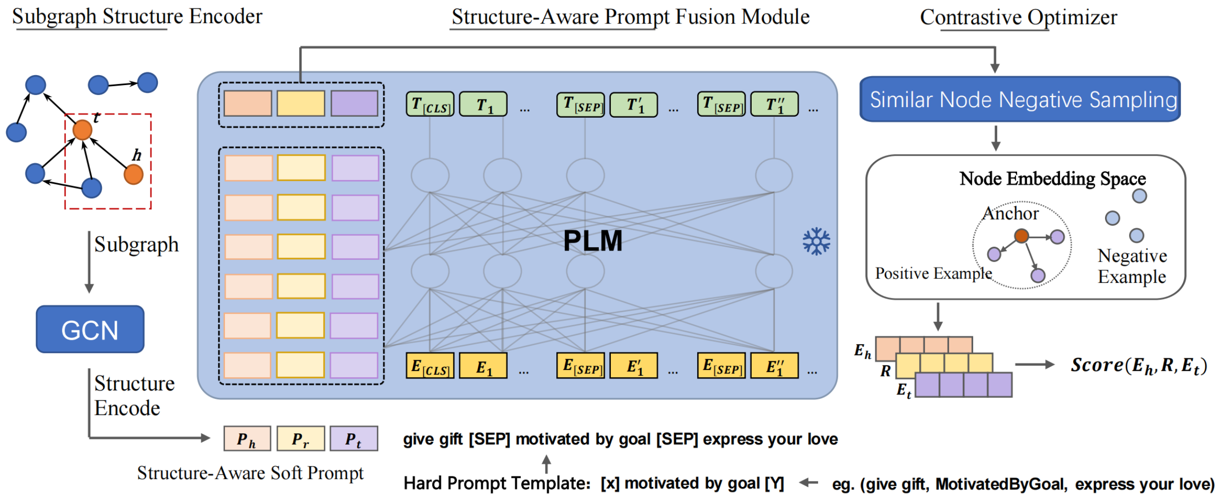

Fig. 1. Framework for Commonsense Knowledge Graph Completion via Structure-Aware Prompting and Contrastive Representation Learning.

where $N(h)$ represents the set of direct neighbors of h, and $W_s^{(l)}, W_n^{(l)}, W_r^{(l)}$ are trainable parameter matrices, with $\odot$ denoting element-wise multiplication. This design captures the evolving structural roles of entities at different graph depths $(l = 1, ..., L)$ through relation-specific weight transformations and neighborhood feature aggregation. After L layers of convolution, we obtain the structured embeddings of the target triple (E_h^g, E_r^g, E_t^g).

B. Structure-Aware Prompt Fusion Module

This module integrates graph structure information with PLM semantics via a soft-hard prompt collaboration mechanism, consisting of three stages:

Sequence Construction and Input Representation. The target triple (h, r, t) is mapped from discrete symbols to a continuous semantic space, forming a computable textual representation. Special identifiers such as [SEP] are used to structure the concatenation of entities and relations, aligning with Transformer input format. Following LAMA's prompt engineering methodology [22], the relation set R is restructured into natural language predicates to enhance PLM's semantic adaptability.

Structure-Aware Prompt Generation. The subgraph structure encoder generates entity embeddings E_h, E_t and relation embeddings E_r. To transfer this graph structure information effectively to the PLM, a soft prompt strategy is employed. Assuming the pre-trained language model has L Transformer layers, each with hidden size h, the soft prompts consist of a trainable sequence of vectors, forming preset content for the input. Inspired by Liang et al.'s [23] layer-wise prompt approach, shorter prompt sequences (e.g., 5-10 vectors) are inserted at each layer to enhance PLM's ability to capture entity and relation information. Additionally, conditional soft prompts are introduced, with each soft prompt encoding graph structure embeddings of entities and relations via trainable vectors. To prevent overfitting on textual information, PLM parameters are frozen during training, ensuring that only the generated prompts are processed.

The generation process of structure-aware prompt P_s is as follows:

$$P_s = [F(E_h^g) \parallel F(E_r^g) \parallel F(E_t^g)] \tag{2}$$

where $F(\cdot) = W_2\,\mathrm{GELU}(W_1 \cdot)$ is a two-layer feedforward network with shared weights. $W_1 \in \mathbb{R}^{d_h \times d_g}$ and $W_2 \in \mathbb{R}^{d_p \times d_h}$ are projection matrices, where d_g is the graph embedding dimension and d_p is the PLM hidden dimension. This process adapts the graph structure information to the PLM's representation space.

Layer-Wise Prompt Injection and Joint Encoding. In the PLM, structure-aware prompts are injected not only into the input layer but also distributed across each layer to enable multi-level interactions. Specifically, the vector $H^{(j)}$ input to the j-th layer is processed based on the soft prompt P_s and hard prompt $T(h, r, t)$:

Input Layer Concatenation: At the input embedding layer of the PLM, the soft prompt vector is concatenated at the beginning of the text sequence:

$$H^{(0)} = [P_s; \mathrm{Embed}(T(h, r, t))] \tag{3}$$

Hidden Layer Enhancement: At the j-th layer $(1 \leq j \leq p)$, interactive soft prompt vectors are inserted:

$$H^{(j)} = \mathrm{PLM}_j([P_s^{(j)}; H^{(j-1)}]) \tag{4}$$

where $P_s^{(j)}$ is the soft prompt specific to the j-th layer. This method enables structure-aware prompts to interact with textual information at multiple levels within the PLM, ensuring that graph structure information is fully utilized across the model.

C. Structure-Constrained Contrastive Optimizer

Although frozen PLMs generate semantically rich and efficient text representations, they lack the flexibility to distinguish between similar entity nodes due to the inability to fine-tune. This is particularly problematic when handling semantically similar but structurally different entities. To address this issue, we propose a structure-constrained contrastive optimizer, which jointly models the similarity between semantic and structural spaces to improve fine-grained reasoning, especially in distinguishing similar entity nodes. The design of the optimizer is as follows: **Contrastive Sample Construction.** The goal of the contrastive sample construction module is to generate representative positive and negative samples to enhance the model's discriminative ability. As shown in Figure 2, we first use the joint representation of the head entity and relation from the structure-aware prompt module as the anchor a, calculated as:

$$a = \mathrm{MLP}([E_h^p \oplus E_r^p]) \tag{5}$$

where E_h^p, E_r^p are the PLM-encoded representations of the head entity and relation. For positive samples, we select the PLM representation E_t^p associated with the true tail entity t. To further enhance contrastive learning, we introduce an adversarial negative sample generation method. Negative samples are not randomly selected but are filtered based on both semantic and structural similarity. Specifically, entities semantically similar to t but structurally different are chosen as negative samples. The construction process is:

$$N_{\text{adv}}(t) = \{e \in \mathcal{E} \mid \mathrm{sim}_{\text{PLM}}(t, e) > \tau_s \wedge D_{\text{JS}}(G_k(t), G_k(e)) > \tau\} \tag{6}$$

where $\mathrm{sim}_{\text{PLM}}$ is the cosine similarity based on semantic representations, and D_{JS} is the Jensen-Shannon divergence measuring subgraph structural similarity.

Contrastive Loss Function. To maximize the model's ability to distinguish between positive and negative samples, the standard InfoNCE contrastive loss function [24] is used. InfoNCE loss calculates the cosine similarity between samples, encouraging the model to maximize the similarity between positive samples and the anchor, while minimizing the similarity between negative samples and the anchor. The loss function is:

$$\mathcal{L}_{\text{cl}} = -\log \frac{\exp(\cos(a, E_t^p)/\tau)}{\sum_{e \in N_{\text{adv}}(t)} \exp(\cos(a, E_e^p))/\tau} \tag{7}$$

where the temperature coefficient τ controls the smoothness of the distribution.

Through this design, the contrastive optimizer not only generates effective positive and negative samples but also improves the model's discriminative ability in handling similar entity nodes by guiding with structural constraints and contrastive loss, thus enhancing fine-grained reasoning capability.

IV. Experiments

A. Experimental Settings

We selected two widely used datasets, ConceptNet and ATOMIC, for performance evaluation of the model. All experimental settings were conducted on these two datasets to ensure the results' generalizability and comparability. In the experimental setup, the pre-trained language model used was RoBERTa-Large [25]. The Adam optimizer was employed during optimization, with related hyperparameters adjusted accordingly. The learning rate for the language model was set to 2×10^{-5}, the batch size was set to 64, and the adversarial sample ratio was set to 40%. In the InfoNCE loss function, the temperature coefficient τ was selected from the set $\{0.05, 0.1, 0.2\}$. The initial value of α in the joint loss function was set to 0.5. During training, the model's Mean Reciprocal Rank (MRR) was evaluated on the validation set every 10 epochs. Training stopped once the MRR on the validation set no longer showed improvement, and the model checkpoint with the highest MRR on the validation set was chosen for final testing.

Several existing baseline models were selected for comparison, including graph-based models, PLM-based models, and graph-text fusion methods. These models are as follows:

- ConvE [26]: Employs graph convolution to capture complex interactions among head entities, relations, and tail entities. By applying 2D convolution over their embeddings and a nonlinear projection, ConvE generates efficient KG representations that improve both accuracy and efficiency in KGC tasks.
- RotatE [12]: Embeds entities and relations in the complex space and models each relation as a rotation from head to tail. This approach naturally captures symmetry and periodicity, making it well suited for link prediction involving cyclic patterns.
- SGBC [16]: Combines BERT and GCN by fine-tuning BERT to obtain rich semantic embeddings and leveraging GCN to encode graph structure. SGBC enhances CKG reasoning by jointly modeling textual and structural information, thus improving the understanding of entities and relations.
- InductivE [19]: Uses an inductive learning strategy that captures commonsense patterns to complete missing triples, excelling at handling previously unseen entities. Its strong inductive bias yields superior generalization in KGC and reasoning tasks.
- MICO [20]: Frames KGC as a contrastive learning problem by converting triples into sequence pairs. Through contrastive optimization of node representations, MICO enhances commonsense reasoning accuracy by strengthening the distinction between positive and negative samples.

We used **Hits@N** and **MRR** as evaluation metrics. Specifically, Hits@N measures the model's accuracy in the ranking of candidate results; a higher Hits@N value indicates better identification of correct entities during the inference process.

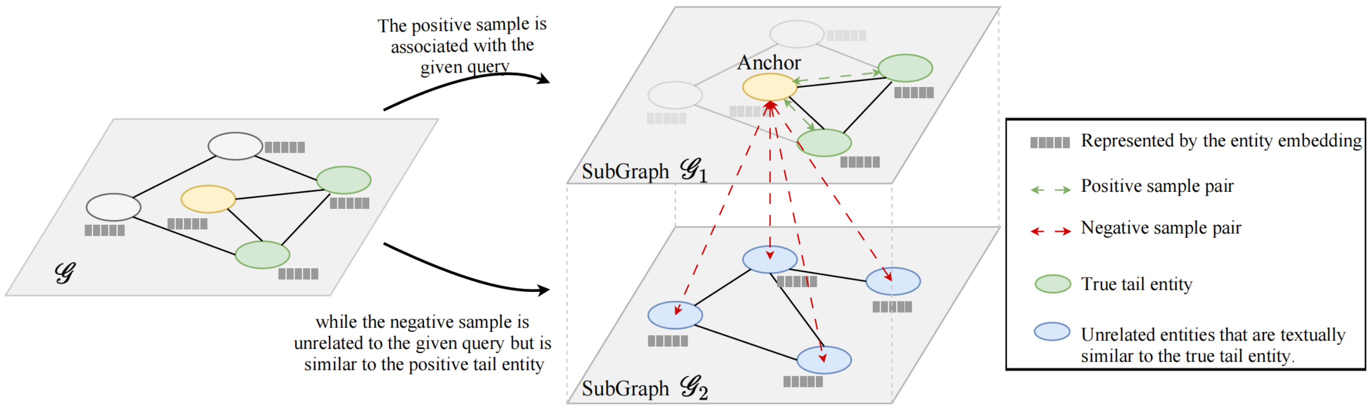

Fig. 2. Illustration of Positive and Negative Sample Construction for Contrastive Optimizer.

MRR measures the model's ranking ability for the correct answer; a higher MRR value signifies the model's precision in predicting missing entities during ranking.

B. Experimental Results

As shown in Table I, our method demonstrates significant advantages in the inductive CKGC task. Compared to three baseline models—graph structure-based methods, pre-trained language model-based methods, and graph-text fusion methods—our method achieves optimal performance on both commonsense knowledge graph benchmarks. Specifically, with respect to the MRR metric, our method outperforms the best baseline models by 10.28 and 16.78 on the ConceptNet and ATOMIC datasets, respectively. For the Hits@10 metric, the improvements are 14.53 and 32.19, respectively. Notably, on the event-driven ATOMIC dataset, even though MICO improves the compatibility of the PLM by converting event triples into semantically coherent text sequences, our method further boosts Hits@10 from 15.69 to 47.88 with the integration of a structural-aware prompt learning mechanism, demonstrating the importance of explicitly modeling graph topological features in complex event reasoning. Early graph-text fusion methods, such as SGBC and InductivE, primarily use PLM as a semantic feature extractor and fail to fully exploit PLM's powerful capabilities in semantic understanding and reasoning tasks. In contrast, our method freezes the PLM and incorporates prompt engineering strategies, retaining the powerful reasoning ability of PLM while injecting structural inductive biases, thereby significantly enhancing the reasoning and completion capabilities of PLM in CKGC tasks.

C. Experimental Analysis

To comprehensively evaluate the performance of the proposed method in this paper, we conducted an ablation study aimed at analyzing the contribution of each module to the overall model performance. The effectiveness of key components, such as structural-aware prompts, contrastive representation learning, and the frozen language model design, was assessed. These experiments aim to provide a deeper understanding of the model's performance in commonsense knowledge graph completion tasks and to offer insights for its further optimization.

TABLE I
EXPERIMENTAL RESULTS OF OUR METHOD ON CKGC TASKS IN INDUCTIVE SCENARIOS

Type	Model	CN-82K-Ind		ATOMIC-Ind	
		MRR	Hits@10	MRR	Hits@10
Graph-Based	ConvE	0.21	0.40	0.08	0.09
	RotatE	0.32	0.50	0.10	0.12
PLM-Based	MICO (BERT-large)	9.00	19.06	7.07	13.52
	MICO (RoBERTa-base)	9.08	18.73	7.52	14.46
	MICO (RoBERTa-large)	10.92	22.07	8.13	15.69
Fusion-Based	SGBC	12.29	19.36	0.02	0.07
	InductivE	18.15	29.37	2.51	5.45
	Ours	**28.43**	**43.90**	**24.91**	**47.88**

In the ablation experiments, we designed several variants targeting key modules to evaluate the impact of each component on the model's performance:

- **Ours w/o struc_prompt**: This variant removes the structural-aware prompt and uses randomly initialized soft prompts to verify the importance of structural-aware prompts on the model's performance.
- **Ours w/o contrastive**: This variant removes the contrastive optimizer and directly uses the representations output by the frozen PLM for triple scoring, to explore the role of contrastive representation learning in the model.
- **Ours w/o frozen**: This variant allows the model to fine-tune all parameters to test whether freezing the PLM has a positive impact on performance under different training settings.

The experimental results, as shown in Table II, indicate that removing any key module leads to lower performance on the CN-82K-Ind dataset, suggesting that each component of the model contributes positively to the overall performance. Furthermore, as illustrated in Figure 3, freezing the PLM parameters leads to more stable and superior performance

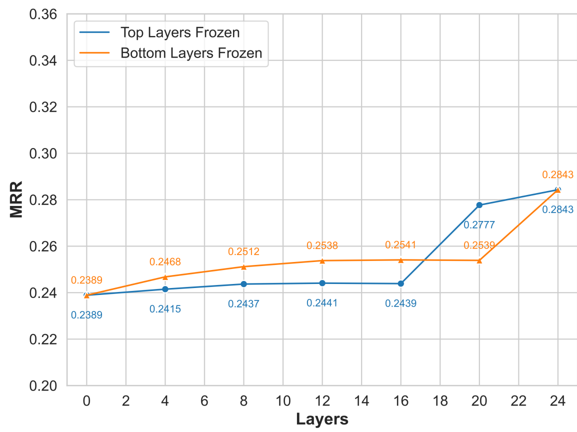

Fig. 3. Impact of Freezing PLM Parameters on CN-82K-Ind.

compared to full fine-tuning, highlighting the effectiveness of preserving the pre-trained knowledge while introducing task-specific structural prompts.

TABLE II
ABLATION EXPERIMENT RESULTS

Model	MRR	$Hits@1$	$Hits@3$	$Hits@10$
Ours	28.43	16.08	29.51	43.90
- w/o struc_prompt	21.73	11.50	23.82	36.09
- w/o contrastive	25.57	12.74	25.88	40.13
- w/o frozen	23.89	11.96	25.52	38.92

- (a) **w/o struc_prompt**: After removing the structural-aware prompt, the model's performance noticeably declines. Although a combination of soft and hard prompts is retained, the lack of structural awareness makes it difficult for the model to accurately predict the tail entity. The MRR metric drops by 16%, highlighting the importance of structural awareness in enhancing the PLM's semantic understanding and reasoning ability.
- (b) **w/o contrastive**: After removing the contrastive optimizer, the model's performance also declines to varying degrees. Notably, the performance drops by 21% on the Hits@1 metric, indicating that without contrastive representation learning, the model's ability to distinguish between similar entities is significantly weakened, making it harder to predict the most accurate entity as the top candidate. This further confirms the critical role of adversarial negative samples in optimizing model representations and improving prediction accuracy.
- (c) **w/o frozen**: Allowing the model to fine-tune all PLM parameters results in a 4.54 drop in MRR, indicating that more trainable parameters can degrade performance. This suggests that freezing certain PLM layers helps maintain generalization and performance. Further experiments on freezing strategies, including bottom and top layers, show that freezing more layers leads to significant performance improvements, with freezing bottom layers slightly outperforming top layers. This may be because the lower layers of BERT capture more fundamental semantic information, which is more beneficial for knowledge graph completion tasks.

V. CONCLUSION

We propose an inductive CKGCs method that integrates structural-aware prompts with contrastive representation learning. The core innovation of this method lies in using hierarchical prompt engineering to guide the PLM to effectively leverage its implicitly encoded commonsense knowledge for reasoning. Specifically, the subgraph structure of the target entity's neighborhood is encoded as dynamic soft prompts, providing explicit topological information to the PLM, thereby assisting it in understanding the potential relationships between entities. In addition, we introduce contrastive representation learning based on text similarity, which jointly optimizes the similarity between the semantic space and the graph structure space, enhancing the PLM's ability to distinguish between textually similar but logically conflicting nodes. To validate the effectiveness of our method, experiments were conducted on the public datasets ConceptNet and ATOMIC. The experimental results show that our method achieves significant performance improvements in the commonsense knowledge graph completion task, demonstrating the great potential of introducing structural-aware PLM in this task.

REFERENCES

[1] Sap, M., Le Bras, R., Allaway, E., Bhagavatula, C., Lourie, N., Rashkin, H., Roof, B., Smith, N.A., Choi, Y.: Atomic: An atlas of machine commonsense for if-then reasoning. In: Proceedings of the AAAI conference on artificial intelligence, vol. 33, no. 01, pp. 3027–3035. AAAI (2019).
[2] Speer, R., Chin, J., Havasi, C.: Conceptnet 5.5: An open multilingual graph of general knowledge. In: Proceedings of the AAAI conference on artificial intelligence, vol. 31, no. 1, pp. 4444–4451. AAAI (2017).
[3] Teru, K., Denis, E., Hamilton, W.: Inductive relation prediction by subgraph reasoning. In: International conference on machine learning, pp. 9448–9457. PMLR (2020).
[4] Geng, Y., Chen, J., Pan, J.Z., Chen, M., Jiang, S., Zhang, W., Chen, H.: Relational message passing for fully inductive knowledge graph completion. In: 2023 IEEE 39th international conference on data engineering (ICDE), pp. 1221–1233. IEEE (2023).
[5] Chen, J., He, H., Wu, F., Wang, J.: Topology-aware correlations between relations for inductive link prediction in knowledge graphs. In: Proceedings of the AAAI conference on artificial intelligence, vol. 35, no. 7, pp. 6271–6278. AAAI (2021).
[6] Yao, L., Mao, C., Luo, Y.: KG-BERT: BERT for knowledge graph completion. In: arXiv preprint arXiv:1909.03193 (2019).
[7] Zha, H., Chen, Z., Yan, X.: Inductive relation prediction by BERT. In: Proceedings of the AAAI conference on artificial intelligence, vol. 36, no. 5, pp. 5923–5931. AAAI (2022).
[8] Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., Yakhnenko, O.: Translating embeddings for modeling multi-relational data. In: Burges, C.J., Bottou, L., Welling, M., Ghahramani, Z., Weinberger, K.Q. (eds.) Advances in Neural Information Processing Systems, vol. 26, Curran Associates, Inc. (2013).
[9] Wang, Z., Zhang, J., Feng, J., Chen, Z.: Knowledge graph embedding by translating on hyperplanes. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 28, no. 1. AAAI (2014).
[10] Lin, Y., Liu, Z., Sun, M., Liu, Y., Zhu, X.: Learning entity and relation embeddings for knowledge graph completion. In: Proceedings of the AAAI conference on artificial intelligence, vol. 29, no. 1. AAAI (2015).

[11] Ji, G., He, S., Xu, L., Liu, K., Zhao, J.: Knowledge graph embedding via dynamic mapping matrix. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pp. 687–696. ACL (2015).
[12] Sun, Z., Deng, Z.-H., Nie, J.-Y., Tang, J.: Rotate: Knowledge graph embedding by relational rotation in complex space. In: arXiv preprint arXiv:1902.10197 (2019).
[13] Nickel, M., Tresp, V., Kriegel, H.-P., et al.: A three-way model for collective learning on multi-relational data. In: Proceedings of the International Conference on Machine Learning (ICML), vol. 11, pp. 3104482–3104584. ACM (2011).
[14] Yang, B., Yih, W.-t., He, X., Gao, J., Deng, L.: Embedding entities and relations for learning and inference in knowledge bases. In: arXiv preprint arXiv:1412.6575 (2014).
[15] Ju, J., Yang, D., Liu, J.: Commonsense knowledge base completion with relational graph attention network and pre-trained language model. In: Proceedings of the 31st ACM International Conference on Information & Knowledge Management, pp. 4104–4108. ACM (2022).
[16] Malaviya, C., Bhagavatula, C., Bosselut, A., Choi, Y.: Commonsense knowledge base completion with structural and semantic context. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 03, pp. 2925–2933. AAAI (2020).
[17] Shen, X., Wu, S., Xia, R.: Dense-ATOMIC: Towards densely-connected ATOMIC with high knowledge coverage and massive multi-hop paths. In: arXiv preprint arXiv:2210.07621 (2022).
[18] Lee, J., Chung, C., Whang, J.J.: InGram: Inductive knowledge graph embedding via relation graphs. In: Proceedings of the International Conference on Machine Learning, pp. 18796–18809. PMLR (2023).
[19] Wang, B., Wang, G., Huang, J., You, J., Leskovec, J., Kuo, C.-C.J.: Inductive learning on commonsense knowledge graph completion. In: Proceedings of the 2021 International Joint Conference on Neural Networks (IJCNN), pp. 1–8. IEEE (2021).
[20] Su, Y., Wang, Z., Fang, T., Zhang, H., Song, Y., Zhang, T.: MICO: A multi-alternative contrastive learning framework for commonsense knowledge representation. In: arXiv preprint arXiv:2210.07570 (2022).
[21] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. Int. Conf. Artificial Intelligence and Statistics (AISTATS)*, 2010, pp. 249–256.
[22] Petroni, F., Rocktäschel, T., Lewis, P., Bakhtin, A., Wu, Y., Miller, A.H., Riedel, S.: Language models as knowledge bases? In: CoRR, vol. abs/1909.01066 (2019).
[23] Li, X.L., Liang, P.: Prefix-tuning: Optimizing continuous prompts for generation. In: CoRR, vol. abs/2101.00190 (2021).
[24] Van den Oord, A., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding. In: CoRR, vol. abs/1807.03748 (2018).
[25] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V.: RoBERTa: A robustly optimized BERT pretraining approach. In: CoRR, vol. abs/1907.11692 (2019).
[26] Dettmers, T., Minervini, P., Stenetorp, P., Riedel, S.: Convolutional 2d knowledge graph embeddings. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32, no. 1. (2018)