HILL PUBLISHING

Future Trends in AI Research, 2025, 2(1), 1-5
https://www.hillpublisher.com/Journals/ftair/
ISSN Online: 3066-3083

Semantic Enhancement Strategy for Knowledge Graph Completion

Jian Sun[1,*], Yizheng Xu[2], Yansong Li[3]

[1]Iowa State University, Ames, Iowa 50011, USA.
[2]University of Malaya, Kuala Lumpur 50603, Malaysia.
[3]Zhengzhou Police College, Zhengzhou 450000, Henan, China.

How to cite this paper: Jian Sun, Yizheng Xu, Yansong Li. (2025) Semantic Enhancement Strategy for Knowledge Graph Completion. *Future Trends in AI Research*, 2(1), 1-5.
DOI: 10.26855/ftair.2025.06.001

Received: June 21, 2025
Accepted: July 11, 2025
Published: August 11, 2025

[*]**Corresponding author:** Jian Sun, Iowa State University, Ames, Iowa 50011, USA.

Abstract

Knowledge graph completion (KGC) is a critical task aimed at enhancing the integrity and completeness of knowledge bases by predicting missing triples. Traditional methods often struggle with complex relations due to their limited semantic modeling capabilities, which restrict their ability to accurately infer missing information in diverse and intricate knowledge structures. To address these limitations, this paper systematically examines advanced semantic enhancement strategies for KGC. These strategies include external knowledge integration, deep representation learning, and logical reasoning reinforcement. By incorporating multi-source data, these methods enrich the semantic context of the knowledge graph, enabling more accurate predictions. Hierarchical semantic encoding further refines the representation of entities and relations, capturing nuanced meanings and dependencies. Additionally, causal reasoning mechanisms are introduced to handle complex relational patterns, ensuring that the inferred triples align with real-world causal structures. These enhanced methods demonstrate significant improvements in completion tasks, particularly in predicting long-tail entities that are often underrepresented in traditional approaches. Experiments on benchmark datasets such as FB15k show substantial performance gains, highlighting the effectiveness of the proposed strategies.

Keywords

Knowledge graph completion; Semantic enhancement; Representation learning; Logical reasoning

1. Introduction

Knowledge graphs, as crucial carriers of structured knowledge, play a pivotal role in intelligent search, recommendation systems, and other domains. However, existing knowledge graphs commonly suffer from incompleteness, severely limiting their practical utility. To address this issue, knowledge graph completion techniques have emerged, with the core objective of algorithmically predicting and filling missing triples in knowledge graphs. Traditional completion methods primarily rely on statistical co-occurrence patterns or simplistic vector-space representations. While these approaches perform adequately for common entities and simple relations, they often fall short when handling semantically complex long-tail entities. With advancements in artificial intelligence, researchers have gradually recognized the necessity of incorporating richer semantic information to significantly improve completion

performance, giving rise to semantic enhancement research in knowledge graph completion [1]. This paper systematically examines current mainstream semantic enhancement strategies from three perspectives: multi-source knowledge fusion, deep representation learning, and logical reasoning reinforcement. Through comparative experiments, we evaluate the strengths and weaknesses of different methods and thoroughly discuss the challenges and future directions in this field, providing comprehensive technical references for researchers.

2. Relevant Basis

Knowledge graph completion technology serves as a pivotal approach to enhancing the integrity of knowledge bases, with its core objective being the prediction of missing relationships between entities to refine existing knowledge networks. The foundational theories in this field are built upon knowledge representation learning frameworks. Early models such as TransE laid critical groundwork by embedding entities and relationships into low-dimensional vector spaces, using simple translation operations to simulate relational patterns. With advancements in deep learning, graph neural network-based methods have achieved more sophisticated semantic capture by aggregating neighborhood information, while the introduction of attention mechanisms has further improved models' ability to identify key features [2]. These technological evolutions reflect the trajectory of knowledge representation—progressing from shallow vectorization to deep semantic understanding. Innovations such as relational path modeling and entity-type constraints have significantly enhanced models' accuracy in parsing the topological structure of knowledge graphs.

At the theoretical level of semantic enhancement, multimodal knowledge fusion has emerged as a crucial direction for overcoming the limitations of traditional methods. By integrating heterogeneous data sources such as textual descriptions, ontological constraints, and visual features, models can establish richer semantic representations of entities, particularly in scenarios involving low-frequency entities and complex relationships. Representation learning guided by external knowledge—such as contextual semantic priors provided by pre-trained language models—effectively mitigates data sparsity issues. Concurrently, the integration of symbolic logic and neural networks has pioneered a new paradigm of neuro-symbolic reasoning. By incorporating first-order logical rules into the representation learning process, models retain the generalization capabilities of neural networks while adhering to explicit logical constraints. This hybrid architecture significantly improves the interpretability of prediction results, offering technical support for application scenarios requiring traceable decision-making, such as medical diagnosis.

The unique structural characteristics of knowledge graphs also present specific challenges for algorithm design. Among these, relational asymmetry (e.g., the irreversible nature of relationships like "father" and "child") requires models to distinguish relational directions, while multi-hop reasoning demands have spurred the development of reinforcement learning-based path-search strategies. Furthermore, the dynamic evolutionary nature of knowledge graphs necessitates the use of temporally aware embedding methods—such as timestamp encoding or memory networks—to capture time-varying features of entity relationships. Recent research trends indicate that meta-learning frameworks show promise in few-shot relationship prediction tasks, while contrastive learning-based negative sampling strategies enhance model robustness by constructing more challenging training samples. These foundational breakthroughs collectively form the theoretical underpinnings of modern knowledge graph completion systems, providing multi-layered technical reserves for subsequent innovations in semantic enhancement methods.

3. Semantic Enhancement Strategy Classification and Analysis

Semantic enhancement strategies in knowledge graph completion (KGC) exhibit a multidimensional technical spectrum, with their core idea centered on augmenting models' reasoning capabilities by incorporating additional semantic information beyond traditional triplet structures. From a methodological perspective, three mainstream enhancement approaches can be observed:

External knowledge injection primarily leverages contextual semantics from pre-trained language models or domain ontologies. For instance, fusing BERT-generated entity description vectors with graph neural network representations significantly improves the representation quality of rare entities. This strategy proves particularly effective in open-domain knowledge graphs, where semantic similarity computations can uncover potential cross-graph alignment relationships [3]. Structural enhancement methods, on the other hand, focus on mining implicit semantic features within the knowledge graph itself. Notable examples include relation-path attention mechanisms and subgraph reasoning techniques. By analyzing semantic propagation patterns in multi-hop relation paths, these methods effectively identify logical consistency in complex relational chains. For example, in medical knowledge graphs reasoning

over "drug-target-disease" indirect associations, path-weight learning modules can automatically filter out noisy paths.

Cross-modal semantic fusion represents the most cutting-edge direction in current research, with its technical essence lying in establishing semantic mapping bridges between heterogeneous data. Vision-text joint embedding models, which align entity image features with structured knowledge, enable the capture of implicit correlations that pure symbolic systems struggle to express [4]. In geographic knowledge graphs, for instance, the co-learning of satellite image texture features and entity spatial relationships significantly improves the prediction accuracy of topological relations like "adjacent regions." Temporal semantic enhancement, specifically designed for dynamic knowledge graphs, employs time-aware graph attention networks to model the evolution trajectories of entity relationships. This temporal encoding mechanism can distinguish between different semantic patterns—such as periodic events versus sudden incidents—as seen in financial knowledge graphs when judging the timeliness of corporate merger relationships.

In the comparative analysis of these three strategy types, external knowledge injection excels in data-sparse scenarios but incurs domain adaptation costs. Structural enhancement methods demonstrate strong performance in complex reasoning tasks on dense knowledge graphs yet rely on high-quality path sampling. Cross-modal approaches, while offering high extensibility, face challenges in multimodal alignment [5]. The latest hybrid enhancement frameworks attempt to dynamically balance the contributions of different semantic sources through gating mechanisms. For example, in academic knowledge graph construction, simultaneously leveraging paper text semantics, citation network structures, and author collaboration timelines, this multidimensional semantic fusion approach outperforms single-strategy methods in predicting the evolution trends of emerging research topics.

It is worth noting that the choice of semantic enhancement strategies essentially reflects an art of trade-off between knowledge completeness and computational complexity. When enhancement dimensions align with the semantic characteristics of the target knowledge graph, even simple relational rotation embedding combined with moderate type constraints can yield significant performance improvements—a phenomenon repeatedly validated in the construction of domain-specific knowledge graphs.

4. Semantic Enhancement Methods for Knowledge Graph Completion

In the field of knowledge graph completion (KGC), typical semantic enhancement methods overcome the limitations of traditional knowledge representation through multi-level semantic fusion, demonstrating unique advantages in both theory and practice. Approaches based on external knowledge typically integrate structured knowledge with unstructured text—for instance, by transforming descriptive texts from Wikipedia summaries into vector representations via entity linking techniques, followed by joint training with symbolic entities in the knowledge graph. This hybrid representation effectively resolves entity ambiguity issues, significantly improving the recognition accuracy of rare medical conditions in tasks such as clinical terminology normalization by comparing semantic similarity between medical literature descriptions and standardized terms.

Deep mining of relational semantics extends traditional relation embedding by incorporating relation path encoding and contextual modelling [6]. Particularly in multi-hop reasoning tasks, attention mechanisms dynamically aggregate information from paths of varying lengths, enabling the model to differentiate between core and auxiliary relations with human-like precision. For example, in legal knowledge graphs, when determining the logical correlation strength between "contract breach" and "force majeure clauses," path-sensitive relation encoders exhibit superior discriminative ability.

Cross-modal semantic fusion introduces a new dimension in knowledge representation, with its core technology lying in the construction of a unified multimodal semantic space. Vision-language pre-training models align visual region features with entity descriptions through contrastive learning, allowing purely symbolic systems to perceive visual semantics. In product knowledge graph construction, for instance, analyzing the consistency between product design and functional descriptions can automatically detect potential cases of false advertising [7]. Temporal enhancement methods, on the other hand, focus on capturing dynamic evolutionary patterns in knowledge, employing neural ordinary differential equations (NODEs) to model the continuous change processes of entity attributes. This differential equation-driven representation learning is particularly effective in financial risk prediction, precisely capturing gradual trends and abrupt changes in corporate credit ratings.

These methods exhibit distinct yet complementary characteristics in application. External knowledge injection excels in sparse data scenarios but depends on high-quality corpora; relational semantic mining performs well on dense graphs but incurs high computational costs; cross-modal approaches offer strong extensibility but require

carefully designed alignment loss functions. Recent research trends indicate that meta-learning frameworks dynamically allocating contribution weights to different enhancement strategies can achieve performance close to that of combined methods while maintaining simplicity. This adaptive enhancement paradigm has demonstrated unique advantages in open-domain question-answering systems, automatically selecting the most relevant semantic pathways based on question types.

Notably, successful semantic enhancement often requires deep alignment with domain-specific characteristics. In biomedical knowledge graphs, for example, a simple injection of gene ontology constraints may prove more effective than complex multimodal fusion, revealing the principle that "the most suitable is the best" in the selection of semantic enhancement techniques.

5. Experimental Evaluation and Analysis

In the systematic evaluation of semantic enhancement methods, experimental design must balance benchmark performance validation with domain adaptability analysis. Comparative experiments on standard datasets (e.g., FB15k-237 and WN18RR) typically employ link prediction metrics (Mean Rank and Hits@K). However, it is crucial to account for the differential impact of various semantic enhancement strategies on sparse entities (e.g., rare word senses in WN18RR) and dense relations (e.g., multi-hop associations in FB15k-237). For instance, entity enhancement methods that inject descriptive text under the TransE framework can improve Hits@10 by 5-8%, with gains exceeding 15% for long-tail entities, highlighting the mitigating effect of external knowledge on data sparsity. Further error case analysis reveals that cross-modal approaches excel with visually related entities (e.g., "musical instruments" or "architectural styles") but may underperform purely symbolic methods for abstract concepts (e.g., "legal clauses") due to modal alignment noise [8]. Such performance divergence necessitates fine-grained entity-type evaluation rather than reliance solely on macro-level metrics.

Ablation experiments are critical for understanding the contributions of individual semantic components. By incrementally removing modules such as relation path encoding, temporal modeling, or external knowledge injection, synergistic or antagonistic interactions between components can be observed. In financial knowledge graph tasks like default prediction, using only the temporal differential equation module yields an AUC of 0.82, while joint training with relational context boosts performance to 0.89, indicating strong coupling between dynamic attributes and structural dependencies. However, forcibly integrating mismatched modalities (e.g., irrelevant product images) can degrade prediction performance by 3-5%. These nonlinear interactions demonstrate that semantic enhancement is not merely a modular assembly but requires domain-informed fusion strategies.

Scalability evaluation must address the trade-off between computational efficiency and resource consumption. Although cross-modal methods based on graph neural networks lead in accuracy, their training time may be 7-10 times longer than traditional embedding approaches, with GPU memory usage growing exponentially with the number of modalities. For real-time applications like medical Q&A systems, lightweight knowledge-distilled versions can be deployed, trading a 2% accuracy loss for a 5x inference speedup. Domain adaptation experiments further reveal the generalization boundaries of semantic enhancement: when vision-language models pre-trained on general domains are transferred to remote sensing knowledge graphs, adversarial training becomes necessary to maintain modal alignment stability due to distributional disparities between satellite and natural images. This finding offers universal insights for cross-domain applications.

Ultimately, experimental results should feed back into iterative methodological refinements. For example, adaptive step-size adjustments in temporal differential equations and sparsity-pruned relation path encoding both stem from addressing bottlenecks identified during evaluation. This "evaluation-analysis-improvement" closed loop drives semantic enhancement technologies toward greater robustness and interpretability while providing empirical guidance for technical selection across diverse application scenarios.

6. Conclusion

Research on semantic enhancement for knowledge graph completion has effectively improved model performance by incorporating richer semantic information. This paper systematically reviews and evaluates the main approaches in this domain. While significant progress has been made, challenges persist in handling dynamic knowledge updates, achieving cross-domain adaptation, and enhancing model interpretability. Looking ahead, integrating the semantic understanding capabilities of large language models, introducing causal reasoning mechanisms, and incorporating

human feedback learning may represent key pathways to overcoming current bottlenecks. As semantic enhancement technologies continue to evolve, knowledge graph completion will undoubtedly play a central role in more intelligent applications, driving AI systems toward greater efficiency and reliability. Research in this field holds not only substantial theoretical value but also provides robust technical support for knowledge-driven intelligent applications.

References

[1] Cai Q, XU B, Dong X. A knowledge graph completion model using semantically enhanced prompts and structural information[J]. Computer Science, 2024: 1-17.

[2] Zhao F, Wang S. Conceptual model construction of international Chinese digital-intelligent textbooks based on knowledge graphs[J]. Language Teaching and Research, 2025(4): 13-23.

[3] Ji X, Ying T, WANG X, et al. Research on programming knowledge retrieval and intelligent Q&A system based on knowledge graph[J]. Intelligent Computers and Applications, 2025, 15(7): 186-193.

[4] Cao J, Guo J. Three dimensions of applying knowledge graphs in ideological and political education[J]. Ideological and Political Education Research, 2025, 41(3): 130-136.

[5] Zhang L. A knowledge graph completion method based on reinforcement learning and graph attention mechanism[J]. Informatization Research, 2025, 51(3): 60-66.

[6] Liu M, Zhao Y, QIN B. Rethinking knowledge engineering courses in the era of large models[J]. Heilongjiang Education (Higher Education Research & Evaluation), 2025(6): 43-48.

[7] Peng C, Li Z. Construction of oil-gas exploration and development case database based on large language models and knowledge graphs[J]. Computer Programming Skills & Maintenance, 2025(5): 14-16, 47.

[8] Chen X, Li S, ZHOU C, et al. Review on fault knowledge graph technology system[J]. Computer Measurement & Control, 2025, 33(5): 1-12.