A Brief Review of Multi-modal Knowledge Graph Completion

Haoyue Zhao

School of Computer Science & Technology, Beijing Institute of Technology, Liangxiang East Road, Fangshan District, Beijing, China

228501259@qq.com

Abstract. Knowledge graphs have become the infrastructure of artificial intelligence. However, most current knowledge graphs are incomplete. Consequently, knowledge graph completion (KGC) has become a hot research topic. Researchers primarily focus on unimodal knowledge graph completion, which consists solely of textual information. With the rapid progress of AI, the demand for multi-modal knowledge graphs (MMKGs) is increasing. However, research on multi-modal knowledge graph completion (MMKGC) is still in its initial stages. There is no clear recognition of its status and trends. First, we summarize the multi-modal knowledge graph and its significance. Second, we classify the comparisons between unimodal knowledge graph completion and MMKGC. Finally, we discuss different methods of MMKGC. This paper may provide guidance for future research.

Keywords: Knowledge Graph Completion, Multi-modal Knowledge Graph, Unimodal Knowledge Graph.

1. Introduction

The Knowledge Graph (KG) was first introduced by Google in 2012. It is defined as a large-scale knowledge base composed of numerous entities and the relationships between them[1]. Knowledge graphs (KGs) provide support for search engines, helping to make informational services more intelligent and convenient. Since 2013, various fields, including biomedicine, have focused on KGs to accelerate their progress. Entities are typically learned through human effort or representation learning. Human learning is more comprehensive but requires significant resources, while representation learning is more efficient but offers less interpretability[2]. There is a lack of complete relationships within these graphs. Therefore, knowledge graph completion is necessary. Currently, single text-based knowledge graph completion has made rapid progress. For instance, S. Guan et al. proposed a Shared Embedding-based Neural Network (SENN) model[3]. R. Zhang et al. presented a neural network-based literature discovery (LBD) approach to identify drug candidates from PubMed and other COVID-19-focused research literature[4]. In recent years, with the development of computer vision and multi-modal learning, researchers have discovered that multi-modal information has advantages over text alone. It can enrich the representation of entities and concepts, enhancing reasoning ability and narrowing information gaps, such as in the identification of entities. A multi-modal knowledge graph (MMKG) includes text, audio, video, images, and more. It constructs entities in various forms and the relationships between them. Multi-modal knowledge graph completion (MMKGC) has emerged because many facts are still missing, and many implicit relationships between entities have not been fully explored.

MMKGC plays a vital role in mining missing triples from existing KGs. This process involves three sub-tasks: Entity Prediction, Relation Prediction, and Triple Classification[5]. However, current research on the differences between single-modal and multi-modal approaches is insufficient, and the development of MMKGC is still in its initial stages. Research efforts are relatively scattered, lacking a clear progression trend and prospects.

To address this problem, we organized the basic information on multi-modal KGs, including their progression from unimodal KGs and the differences between them. We then introduce the necessity of MMKGC and discuss the reasons for this completion as well as its research methods. Section III compares KGC and MMKGC, highlighting their different methods and application scenarios. Finally, detailed methods of MMKGC are analyzed.

2. Literature Review and Analysis

2.1. Multi-modal Knowledge Graph

A knowledge graph is represented as a set of triplets, consisting of two entities and their relationships. A knowledge graph acquires and integrates information into an ontology and applies a reasoner to derive new knowledge[6]. The original data of KGs is divided into three types: structured data, semi-structured data, and unstructured data. Unstructured data primarily comes from MMKGs. Currently, the storage technologies for KGs include Resource Description Framework (RDF) and graph databases. The construction of an MMKG is referred to as knowledge acquisition and is divided into three categories: entity recognition, relation extraction, and event extraction[7].

Unlike traditional KGs, a MMKG is not limited to text alone; it extends its information resources to various forms. A KG qualifies as multi-modal (MMKG) when it contains knowledge symbols expressed in multiple modalities, which can include, but are not limited to, text, images, sound, and video[5].

In practical research, entities and relationships are often missing or incomplete. Therefore, MMKGC is introduced, which is essential for several reasons:

1. Graphs often contain missing and ambiguous information, such as entities and relations.
2. To fully incorporate all kinds of information, it is necessary to collect multiple modalities comprehensively and representatively.
3. Incomplete information may lead to incorrect or inaccurate predictions.

2.2. Multi-modal Knowledge Graph Completion Research Method

The main focus of MMKGC is to complete the structure of a knowledge graph by predicting missing entities or relationships and mining unknown facts[7]. Currently, research on MMKG [7]completion lacks depth and abundance. The primary goal in this field is to explore how to integrate different forms of entities and whether the model can handle large-scale complex relational data, including its computational efficiency and complexity as well as the model's completion accuracy[1]. Another significant challenge is how to continuously integrate new information into the overall model, given that the information landscape is constantly changing. Therefore, there is a demand for new models that are both efficient and accurate.

MMKGC plays a crucial role in the advancement of artificial intelligence. With a completed knowledge graph, AI can predict our needs more rapidly and accurately. Search engines and internet-based domains can also benefit from better development. Existing KGs tend to be relatively static, making it difficult to meet evolving demands. Dynamic KGs represent a key trend for future development, as most fields rely on ever-changing information[1].

The research methods for MMKGC can be divided into three main categories:

1. Different Modal Information Fusion and Alignment: This involves linking two different pieces of information that have the same equivalent meaning[8].
2. Using Rule-Based Reasoning and Generative Models to Generate New Entities: For example, employing generative adversarial networks (GANs) to create new entities and relationships to complete the graph[9].

3. Multi-modal Inference to Better Integrate Various Models: Unlike traditional methods for single-modal KGs, MMKGC requires the capability to capture all types of modalities while simultaneously creating new entities to fill in the missing parts.

3. Discussion

3.1. Comparison Between KG and MMKG Completion

Table 1. Comparison between KG and MMKG completion.

	KGC	MMKGC
Type	Entity Relationship	Entity Relationship Different modal
Modal	Text	Text, image, audio, video
Method	Text embedding, rule-based, graph-based	Rule-based, graph-based, modals alignment and fusion
Technical complexity	Simple	Complex
Application	Knowledge Retrieval, search engine	Computer vision, autonomous driving, video clarification

Knowledge graph completion (KGC) is typically restricted to mere textual information. Z. Chen et al. classified its methods into traditional and representation learning methods. The former category includes rule-based methods derived from machine learning algorithms, such as rule-based learning and path ranking algorithms. The latter includes translation models (e.g., TransE, TransH) and other neural network models, including graph neural networks (GNNs) and attention-based techniques[1]. The primary difference between KGC and multi-modal knowledge graph completion (MMKGC) is the extension of modalities involved. In addition to entity and relationship reasoning, MMKGC requires the integration of different modalities. The introduction of additional modalities presents a greater number of challenges and choices in how to represent knowledge. For example, during the construction and completion of an MMKG, a new modality representing the same entity can be added as another entity or as an attribute of the entity [10].

As Table I shows, traditional KGC can be achieved through text embedding, rule-based methods, and graph-based methods. For instance, researchers focus on fact embedding [11], where facts are embedded into vector spaces to facilitate the prediction of missing relationships or entities. However, when expanding into the multi-modal domain, we must further consider modality alignment and fusion to correspond images and videos with their respective texts. This adds layers of complexity to the task, making MMKGC significantly more intricate than simple KGC. One important aspect is the increase in evaluation criteria; MMKGC must account for additional factors like modality consistency and coherence compared to KGC[12]. In other words, MMKGC needs to ensure that knowledge extracted from different modalities aligns properly and forms a coherent representation of the same entity.

MMKGC shares a broader range of application scenarios. Beyond simple tasks like knowledge retrieval and text-based search engines, MMKGC significantly enhances the reasoning abilities of systems. As a result, search engines can handle images, audio, video, and other types of information more fluently[13]. Furthermore, when integrated with advanced computer vision techniques, MMKGC can assist with autonomous driving and video clarification. In this context, MMKGC enables the real-time recognition of objects and relationships across different environments, a crucial capability for the automotive and robotics industries. In conclusion, MMKGC plays a necessary role in the advancement of the artificial intelligence field [14]. It expands the scope of AI applications by integrating various models into reasoning systems.

3.2. Analysis of MMKGC Methods

Table 2. Comparison between MMKGC methods.

Method	Type	Model	Purpose
Integration	Fusion	Early, late	Comprehensive understanding
	Alignment	Fine-grained, coarse-grained	Unify different modal
	Representation	Joint, coordinated	Vector mapping
Inference	Prediction	Translate, GNN	Discover missing entity and relations
	Generation	GAN	Generate new information

The process of multimodal knowledge graph completion (MMKGC) is generally divided into two key aspects: integration and inference. The former aspect focuses on combining multiple data sources, including information prediction and generation. These two aspects are not separate; rather, they should be considered together when dealing with a multimodal knowledge graph. It is important to note that integration and inference are interconnected. Only by fully integrating downstream data can inference achieve precise and complete results.

As shown in Table II, integration encompasses fusion, alignment, and representation. Fusion and alignment focus on methods for linking and correlating different modalities. For example, a text labeled “car” can be linked with an image of a car. Specifically, early fusion requires the integration of attributes before output, while late fusion operates in reverse. Similarly, modal alignment can be performed at different levels of granularity. Coarse-grained alignment involves mapping entire pieces of information across modalities, such as connecting an entire image with its corresponding textual description. In contrast, fine-grained alignment involves more detailed mapping, such as linking specific visual features to corresponding text attributes, allowing for more nuanced connections between modalities[15].

Representation learning in the integration phase leverages various advanced techniques, such as translation-based models, neural networks, and attention mechanisms. It aims to capture the internal relationships between entities and their multimodal attributes. The goal of representation learning is to embed these relationships into a unified vector space, allowing for a more precise and robust representation of knowledge. By incorporating data from different modalities into the graph structure, researchers can ensure that the knowledge graph reflects a comprehensive understanding of entities, relationships, and their attributes[16].

As for the inference of MMKGC, prediction and generation are two major aspects. Prediction aims to anticipate possible entities and relationships across different modalities. For example, when given an image of a car, the system must be capable of predicting and linking it to the appropriate text label, such as the word "car." This process is often achieved using models like translation-based methods and GNNs [17]. A common issue encountered in the prediction process is the presence of incomplete or sparse entities. This results from insufficient data for certain information and limits the system’s ability to make accurate predictions. To address this problem, C. Zhang et al. proposed a solution[18], which ensures robustness in MMKGC tasks and helps maintain the accuracy of the graph, even when data is scarce.

In addition to prediction, generation is another critical component of MMKGC inference. Generation models, such as generative adversarial networks (GANs), have been adopted to create new entities and relationships. These techniques help fill in the gaps in the knowledge graph by generating new, plausible knowledge from existing data. For example, if certain relationships or entities are missing from the graph, GANs can generate synthetic entities and relationships based on learned patterns from the available data. This not only reduces uncertainty but also improves the overall completeness of the knowledge graph[19].

4. Conclusion

This paper briefly reviews the origins of MMKGC, analyzing the key differences between traditional unimodal KGC and MMKGC. It delves into the mainstream research directions in the field of MMKG completion, highlighting both the theoretical foundations and practical applications. Specifically, we categorize the completion process into two primary aspects: integration and inference. The integration approach focuses on identifying and leveraging correlations between different entities across modalities, enhancing the existing connections within the knowledge graph. In contrast, the inference approach emphasizes the incorporation of new factors, such as introducing novel entities, relationships, or attributes that were previously absent. These two approaches work in tandem to enhance the overall completeness of the MMKG. However, despite the progress made, current developments still face various challenges. Many existing models struggle to efficiently handle the growing complexity and diversity of information, which influences the accuracy and completeness of the entire system. Nevertheless, the field presents promising prospects in the following areas:

More models with high efficiency and accuracy are needed to complete tasks in shorter time frames. As data in the real world continues to grow in size and complexity, there is a pressing need for fast and precise methods.

Automation of the completion process is essential. The high density of information demands efficient interactions between structured data and the knowledge graph. In this regard, we need to develop systems that can autonomously update and maintain KGs with minimal human intervention.

Expansion of application scenarios is anticipated, including human-computer interaction, intelligent search engines, and recommendation systems. An important development direction for MMKGC is to achieve breakthroughs in more diverse application contexts.

References

[1] Z. Chen, Y. Wang, B. Zhao, J. Cheng, X. Zhao, and Z. Duan, “Knowledge graph completion: A review, ” IEEE Access, vol. 8, pp. 192435–192456, 2020, doi: 10.1109/ACCESS.2020. 3030076.

[2] B. Wang, T. Shen, G. Long, T. Zhou, Y. Wang, and Y. Chang, “Structure-augmented text representation learning for efficient knowledge graph completion, ” Web Conf. 2021 - Proc. World Wide Web Conf. WWW 2021, no. 1, pp. 1737–1748, 2021, doi: 10.1145/3442381. 3450043.

[3] S. Guan, X. Jin, Y. Wang, and X. Cheng, “Shared embedding based neural networks for knowledge graph completion, ” Int. Conf. Inf. Knowl. Manag. Proc., pp. 247–256, 2018, doi: 10.1145/3269206.3271704.

[4] R. Zhang, D. Hristovski, D. Schutte, A. Kastrin, M. Fiszman, and H. Kilicoglu, “Drug repurposing for COVID-19 via knowledge graph completion, ” J. Biomed. Inform., vol. 115, no. October 2020, p. 103696, 2021, doi: 10.1016/j.jbi.2021.103696.

[5] Z. Chen et al., “Knowledge Graphs Meet Multi-Modal Learning: A Comprehensive Survey, ” vol. 14, no. 8, pp. 1–54, 2024, [Online]. Available: http://arxiv.org/abs/2402.05391

[6] Y. Lu, W. Zhao, N. Sun, and J. Wang, “Enhancing Multimodal Knowledge Graph Representation Learning through Triple Contrastive Learning, ” pp. 5963–5971, 2021.

[7] Y. Chen, X. Ge, S. Yang, L. Hu, J. Li, and J. Zhang, “A Survey on Multimodal Knowledge Graphs: Construction, Completion and Applications, ” Mathematics, vol. 11, no. 8, pp. 1–27, 2023, doi: 10.3390/math11081815.

[8] B. Cheng, J. Zhu, and M. Guo, “MultiJAF: Multi-modal joint entity alignment framework for multi-modal knowledge graph, ” Neurocomputing, vol. 500, pp. 581–591, 2022, doi: 10.1016/ j.neucom.2022.05.058.

[9] Q. Wang, Y. Ji, Y. Hao, and J. Cao, “GRL: Knowledge graph completion with GAN-based reinforcement learning, ” Knowledge-Based Syst., vol. 209, p. 106421, 2020, doi: 10.1016/j.knosys.2020.106421.

[10] X. Zhu et al., “Multi-Modal Knowledge Graph Construction and Application : A Survey, ” IEEE Trans. Knowl. Data Eng., vol. 36, no. 2, pp. 715–735, 2024, doi: 10.1109/TKDE.2022.3224228.

[11] X. Long, L. Zhuang, A. Li, H. Li, and S. Wang, Fact Embedding through Diffusion Model for Knowledge Graph Completion, vol. 1, no. 1. Association for Computing Machinery, 2024. doi: 10.1145/3589334.3645451.

[12] T. Shen, F. Zhang, and J. Cheng, “A comprehensive overview of knowledge graph completion, ” Knowledge-Based Syst., vol. 255, p. 109597, 2022, doi: 10.1016/j.knosys.2022.109597.

[13] R. Sun et al., “Multi-modal Knowledge Graphs for Recommender Systems, ” pp. 1405–1414, 2020, doi: 10.1145/3340531.3411947.

[14] W. Liang, P. De Meo, Y. Tang, and J. Zhu, “A Survey of Multi-modal Knowledge Graphs: Technologies and Trends, ” ACM Comput. Surv., vol. 56, no. 11, pp. 1–41, 2024, doi: 10.1145/3656579.

[15] X. Chen et al., “Hybrid Transformer with Multi-level Fusion for Multimodal Knowledge Graph Completion, ” SIGIR 2022 - Proc. 45th Int. ACM SIGIR Conf. Res. Dev. Inf. Retr., pp. 904–915, 2022, doi: 10.1145/3477495.3531992.

[16] S. Liang, A. Zhu, J. Zhang, and J. Shao, “Hyper-node Relational Graph Attention Network for Multi-modal Knowledge Graph Completion, ” ACM Trans. Multimed. Comput. Commun. Appl., vol. 19, no. 2, 2023, doi: 10.1145/3545573.

[17] K. Liang et al., “A Survey of Knowledge Graph Reasoning on Graph Types: Static, Dynamic, and Multi-Modal, ” IEEE Trans. Pattern Anal. Mach. Intell., vol. PP, pp. 1–20, 2024, doi: 10.1109/TPAMI.2024.3417451.

[18] L. Wang, W. Zhao, Z. Wei, and J. Liu, “SimKGC: Simple Contrastive Knowledge Graph Completion with Pre-trained Language Models, ” 2019.

[19] D. Chen and R. Zhang, “Building Multimodal Knowledge Bases With Multimodal Computational Sequences and Generative Adversarial Networks, ” IEEE Trans. Multimed., vol. 26, pp. 2027–2040, 2024, doi: 10.1109/TMM.2023.3291503.