Angus Academic Press CCAM 2025
ISSN:3007-9640 DOI: https://doi.org/10.70114/acmsr.2025.2.1.P370

Contrastive Knowledge Graph Error Detection Model of TransR Based on Spatial Transformation

Jiawei Fan [1, a] *, Yifeng Yue [1, b], and Xianghui Ren [1]

[1] The 15th Research Institute of China Electronics Technology Group Corporation, China.

[a] fanjw_lucky@qq.com, [b] yueyf1230@qq.com

Abstract. Contrastive Knowledge Graph Error Detection (CAGED) is designed based on the contrastive learning of knowledge graphs. CAGED employs an error-aware knowledge graph neural network (EaGNN) that utilizes a gated attention mechanism to inhibit the spread of information from incorrect triplets. However, EaGNN's architecture does not account for the transformation of entities and relations across different spaces, thereby limiting the expressive capacity of the model. Furthermore, CAGED uses a static balance parameter to modulate the impact of contrastive loss and embedding loss, which relies heavily on human expertise and increases the difficulty of model training. The paper introduces contrastive knowledge graph error detection model of TransR (TransR-CAGED), an enhanced EaGNN approach that maps entities into relation-specific spaces, thereby capturing intricate interactions among triplets. It features a dynamic balance parameter informed by the variances in triplet representations across different perspectives, enabling the dynamic adjustment of weights assigned to contrastive loss and embedding loss. This innovation renders model training more reliable. Compared to the conventional CAGED method, the proposed strategy exhibits distinct advantages, especially in scenarios where high recall rates are critical.

Keywords: Knowledge Graph, Error Detection, Space Transformation, Contrastive Learning.

1. Introduction

The knowledge graph(KG) stands as a pivotal technology for the representation and management of knowledge, finding extensive application across various domains including medicine [1,2], military [3], electric power [4,5], and education [6,7]. The assessment of a KG's quality is a critical process in evaluating its data, serving not only to bolster the dependability of knowledge graph-based applications but also to guarantee the optimal performance of these applications.

The domain of knowledge graph quality evaluation has evolved from rule-based methods to more advanced embedding-based and contrastive learning-based approaches [8]. Specifically, contrastive learning methods assess the reliability of KGs by comparing different graphs or distinct parts within a single graph, thus identifying inconsistencies and inaccuracies. Additionally, contrastive learning techniques enhance the interpretability of the evaluation process.

Zhang [8] conducted research on KG quality assessment, which is grounded in contrastive learning, and introduced an innovative framework known as CAGED. This framework adeptly identifies errors by harnessing the synergy between contrastive learning and KG embedding techniques. Central to this approach is an EaGNN designed to mitigate the spread of misinformation emanating from incorrect triplets. The framework concatenates the vectors of graph triplet in a straightforward manner before applying subsequent gated attention mechanisms. However, this direct concatenation method does not account for the fact that entities and relations are situated in distinct vector spaces, potentially overlooking their unique characteristics.

TransR [9] is a KG embedding method that regards an entity as a complex of various attributes, and believes that different relations have different semantic spaces and focus on different attributes of entities. Therefore, this paper applies the idea of TransR space transformation to the improvement of the EaGNN in CAGED, proposing a TransR-CAGED method for KG quality assessment. This method makes adjustments to the directly concatenated triple parts in CAGED, that is, it first performs TransR space transformation operations, and then applies gated attention mechanisms to filter the concatenated triples. The TransR-CAGED method mainly has the following innovative points:

ISSN:3007-9640 DOI: https://doi.org/10.70114/acmsr.2025.2.1.P370

(1) The proposed method ensures that entities are mapped to spaces corresponding to each relationship, thereby facilitating the rationalization of subsequent triplet concatenation operations.

(2) Our method utilizes the space transformation matrix, which enhances the model expressiveness for triplet and effectively captures the intricate interactions between entities and relations.

(3) The TransR-CAGED method introduces dynamic balance parameters into the framework's loss function, allowing for a flexible adjustment of the weights assigned to contrastive loss and embedding loss based on variations in triple representations from different perspectives. This dynamic adjustment significantly increases the reliability of model training.

2. Related Work

2.1 Contrastive Learning

Contrastive learning emerged as a potent technique for managing unlabeled data, finding extensive application in the fields of computer vision [10,11] and natural language processing [12,13]. Its utility in detecting errors within KGs has also captured the interest of the academic community [8]. Ma et al. [14] introduced the Graph-Level and Local Semantic Enhancement Contrastive Framework (GLSEC), which supplants the traditional random graph structure contrast method, to delve into the underlying global semantics connecting entities within a KG. Employing contrastive learning for the assessment of KG quality yielded insights that are both intuitive and precise.

To harness the strengths of various techniques, some researchers have integrated contrastive learning with embedding methodologies. Zhang et al. [8] developed a framework known as CAGED, which combines contrastive learning with KG embedding techniques for error detection within KGs. This innovative framework aimed to alleviate the adverse effects of incorrect triplets on model training to a certain extent. They proved that the CAGED framework outperforms existing error detection methods in terms of effectiveness. Gao et al. [15] addressed the necessity of accommodating the graph's structure and intricate relationships by introducing a KG embedding technique called TracKGE. This method converted the structural information of the KG into a sequential format, which streamlines the processing by the transformer model and bolsters the capacity to extract latent information from rich semantics and intricate relationships.

2.2 KG Embedding Method

Embedding-based methods assess the validity of triples by mapping entities and relations into a continuous vector space. It can be primarily categorized into two key directions: semantic matching methods and translation-based methods. Scholars have extensively studied semantic matching methods [16-18]. Zhang et al. [18] proposed an integrated framework named NativE, which includes a module called CoMAT that enhances multimodal embeddings through adversarial training, thereby facilitating multimodal knowledge reasoning in complex environments. TransE [19], a highly influential translation-based method, posits that the sum of the head entity vector and the relation vector should approximately equal the tail entity vector, a principle used to identify errors and assess the quality of KGs. Inspired by TransE, numerous other translation-based methods for KG quality assessment have been developed, such as TransD [20], TransH [21], and TransR [9].

TransR [9], a technique for KG embedding, is founded on the concept of embedding entities and relations into separate vector spaces, utilizing a relation transformation matrix to map entities from the entity space to the relation space. This approach has earned significant acclaim among scholars [22,23]. Zhang et al. [22] proposed a representation learning model that employs soft translation and relation matrix projection, effectively distinguishing between entities and relations within their own vector spaces. Cai et al. [23] crafted a method for constructing KGs using TransR, streamlining the complexity and curtailing the training time required for KGs. Zhao et al. [24] pioneered an

ISSN:3007-9640 DOI: https://doi.org/10.70114/acmsr.2025.2.1.P370

innovative integration of TransR with Graph Neural Networks (GNNs), revealing the implicit semantics embedded within COVID-19 data.

3. Methodology

High-quality KGs can more accurately reflect relevant textual information, ensuring the reliability of decision-making assistance. CAGED, as a method for detecting errors in KGs, can cover multiple dimensions of graph quality, safeguarding the effectiveness and completeness of the graph.

This paper presents an enhanced version of the CAGED method, dubbed TransR-CAGED, which refines the original CAGED framework in two key dimensions. Firstly, it refines the EaGNN method incorporated within the framework, as illustrated in Fig. 1. The method begins by spatially transforming the entities and then concatenates them with their respective relationships to feed into the subsequent gated attention mechanism processes. Secondly, the paper introduces modifications to the loss function design. By substituting the static trade-off parameter with a dynamic cosine distance metric that adapts to the differences between various views, the significance of the two loss components can be dynamically adjusted in accordance with the sample variations.

In this section, we analyze the TransR-CAGED from the following two parts. The first part is about the process of entity space transformation, and the second part is the design of the loss function.

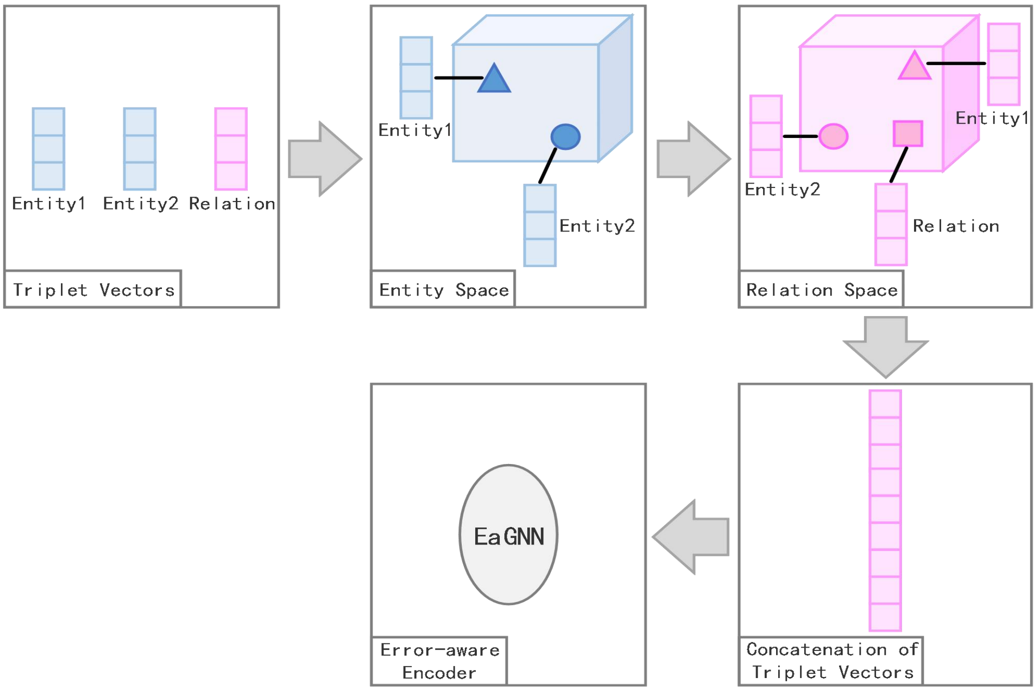

Fig. 1 Improved EaGNN module of the TransR-CAGED framework

3.1 Entity Space Transformation

TransR spatial transformation is integrated into the EaGNN method within the CAGED framework. Different entity space transformation matrices are constructed according to various relationships, and the entity vector is mapped to the relationship space through the spatial transformation matrix. The specific process is shown in Fig. 1. First, the spatial transformation matrix is utilized to project the entities in the triplets into their respective relation spaces, resulting in entity-relationship triplets that occupy the same spatial context. Subsequently, these spatially transformed triplets are used as input for the following operations of the EaGNN method.

The traditional EaGNN method within the CAGED framework overlooks the distinction between entity and relational spaces, which can impair the precision of model analysis. The method outlined

ISSN:3007-9640 DOI: https://doi.org/10.70114/acmsr.2025.2.1.P370
in this paper employs TransR for spatial transformation, mapping entities from the entity space into a specific relational space. This approach ensures that entities with an actual relationship are in close proximity within the relational space, while those without a relationship are distant from each other. Furthermore, by concatenating entities and relations within the same space, this method can more accurately reflect the semantic relationships among triplets, thereby ensuring the reliability of processing complex relational triples.

3.2 Loss Function

The loss within the CAGED framework is composed of two principal elements. The first element is the KG embedding loss, which addresses the inconsistencies within individual triplets; the second is the contrastive learning loss, capturing the representational differences of the same triplet across various views. Traditional CAGED employed a fixed trade-off parameter to balance the contributions of these two loss components to error detection, with the flexibility to adjust this parameter according to different scenarios. However, since this parameter is static, it cannot be dynamically modified during the model training process in response to the varying importance of the sample's embedding loss and contrast loss. Moreover, the adjustment of this parameter is subject to human factors, which can influence the difficulty and reliability of the model training process.

Consequently, we refine the framework's loss function by substituting the original trade-off parameter with the cosine similarity distance between disparate views. The formula is as follows:

$$\mathrm{C(h,r,t)} = \sigma(\frac{1}{2} \times \cos\_\mathrm{distan}(x_i, z_i) \times \mathrm{sim}(x_i, z_i) - \mathrm{E(h,r,t)}), \tag{1}$$

where $\mathrm{E(h,r,t)} = \|e_h + e_r - e_t\|_2$ denotes the degree of internal inconsistency within the triplets, whereas $\mathrm{sim}(x_i, z_i)$ represents the consistency of triple representations from the same sample in View I and View II [8]. Cosine distance is formulated as $\cos\_\mathrm{distan}(x_i, z_i) = 1 - \frac{x_i \cdot z_i}{\|x_i\|\|z_i\|}$, and σ is sigmoid function.

When the directions of the representation vectors under different views of the same triplet are very different, it means that the prediction error of the triplet vector is large. The importance of this part of the loss should be strengthened, that is, the impact of the contrastive learning loss should be increased. On the contrary, the impact of the contrastive learning loss on the overall loss can be reduced. This design concept meets the actual loss change requirements of KG error detection, adjusts the impact of the two parts of the error in real time, and improves the accuracy of model training.

4. Experiment

In this section, we conducted a validation of the TransR-CAGED method using the WN18RR dataset, a process crucial for establishing the method's effectiveness. To bolster the method's credibility, we assessed its performance of different anomaly ratios. The WN18RR dataset, derived from authentic real-world data, stands as a pivotal benchmark within the realm of knowledge graph embedding and associated domains. It is crafted from the WordNet lexical database, encompassing a comprehensive structured knowledge base that encompasses a diverse array of entities and relationships. Our experimental dataset encompasses 93,003 triples, spanning across 40,943 entities and 11 distinct relationship types. Considering the dataset's dimensions, we set the batch size to 64 for optimal processing.

To mitigate the discrepancies arising from randomness in experimental results, we conducted five trials for each method and selected the optimal outcomes. We assess the precision and recall metrics for knowledge graph error detection associated with both methods. Precision is defined as the percentage of erroneous triples identified among the K triples exhibiting the lowest confidence scores [8]; Recall denotes the percentage of erroneous triples detected out of all actual erroneous triples [8]. The experiments varied the anomaly ratio to investigate each method's sensitivity to

ISSN:3007-9640 DOI: https://doi.org/10.70114/acmsr.2025.2.1.P370

noisy data while also evaluating how different parameter values for selecting the top K% lowest confidence scores influenced model performance.

The comparative results for the CAGED and TransR-CAGED methods are presented in Table 1, Table 2, and Table 3. Firstly, the data presented in the three tables clearly illustrate the complementary nature of precision and recall for the two methods under examination. An observable trend emerges where an increase in recall corresponds to a decrease in precision, consistent with findings from other studies [8]. Secondly, the TransR-CAGED method demonstrates its strengths as the K value increases. Despite starting at a disadvantage when K is small, it progressively narrows the performance gap with CAGED and ultimately surpasses it. This indicates that our method possesses enhanced comprehensive analytical capabilities and superior recall rates, which are particularly advantageous in scenarios requiring high recall, such as medical diagnostics and fraud detection. Moreover, the sustained relative advantage of our method at higher K values across various anomaly ratios highlights its stability and reliability.

Table 1. Error detection results for different K on the WN18RR dataset (anomaly ratio=0.05)

	K	0.001	0.01	0.03	0.05	0.07	0.09	0.1
Precision	CAGED	1.0000	0.9118	0.6832	0.5262	0.4272	0.3648	0.3406
	TransR-CAGED	1.0000	0.8581	0.6627	0.5359	0.4473	0.3834	0.3568
Recall	CAGED	0.0200	0.1824	0.4099	0.5262	0.1000	0.6566	0.6813
	TransR-CAGED	0.0200	0.1716	0.3976	0.5359	0.6262	0.6901	0.7135

Table 2. Error detection results for different K on the WN18RR dataset (anomaly ratio=0.1)

	K	0.001	0.01	0.05	0.1	0.11	0.15	0.2
Precision	CAGED	1.0000	0.9763	0.8163	0.6099	0.5774	0.4741	0.3900
	TransR-CAGED	1.0000	0.9333	0.7578	0.5904	0.5625	0.4753	0.3970
Recall	CAGED	0.0100	0.0976	0.4082	0.6099	0.6352	0.7112	0.7800
	TransR-CAGED	0.0100	0.0933	0.3789	0.5904	0.6187	0.7130	0.7940

Table 3. Error detection results for different K on the WN18RR dataset (anomaly ratio=0.15)

	K	0.001	0.01	0.05	0.1	0.11	0.15	0.2	0.3
Precision	CAGED	1.0000	0.9892	0.8791	0.7438	0.7152	0.6229	0.5293	0.4080
	TransR-CAGED	0.9892	0.9624	0.8501	0.7234	0.6999	0.6244	0.5446	0.4221
Recall	CAGED	0.0067	0.0660	0.2930	0.4958	0.5244	0.6229	0.7057	0.8160
	TransR-CAGED	0.0066	0.0642	0.2834	0.4823	0.5133	0.6244	0.7262	0.8442

5. Conclusion

The TransR-CAGED method introduced in this paper extends CAGED by integrating the concept of entity-to-relation space transformation and dynamically adjusting the loss weights based on the significance of the two components of the model's loss function. Our comparative experiments show that the method is not only feasible but also highly effective, particularly for applications requiring a high recall rate.

Although our approach demonstrates practical utility, we recognize that there is room for improvement, especially at lower K values where its performance trails behind conventional methods. This area remains a focus of our ongoing research efforts to enhance the method's overall performance.

Acknowledgements

Thanks to the National Key R&D Program of China under Grant No.2022YFB3103603 for supporting this work.

ISSN:3007-9640 DOI: https://doi.org/10.70114/acmsr.2025.2.1.P370

References

[1] Liu Wensu, Tang Tianyu, Feng Jianwei, et al. Knowledge graph construction based on granulosa cells transcriptome from polycystic ovary syndrome with normoandrogen and hyperandrogen. Journal of Ovarian Research, 2024, 17(1).

[2] Xinyan Wang, Kuo Yang, Ting Jia, et al. KDGene: knowledge graph completion for disease gene prediction using interactional tensor decomposition. Briefings Bioinform, 2024, 25(3).

[3] Cao Bo, Xing Qinghua, Li, Longyue, et al. KGTLIR: An Air Target Intention Recognition Model Based on Knowledge Graph and Deep Learning. Computers, Materials & Continua, 2024, 80(7): 1251-1275.

[4] Ye Zhenhao, Qi Donglian, Liu Hanlin, et al. RoUIE: A Method for Constructing Knowledge Graph of Power Equipment Based on Improved Universal Information Extraction. Energies (19961073), 2024, 17(10).

[5] Gao Bo, Yan Zhenhua, Li Xuefeng, et al. Fine-grained partial discharge pattern recognition method based on substation equipment knowledge graph. IOP Publishing Ltd, 2024.

[6] Chiara Alzetta, Ilaria Torre, Frosina Koceva. Annotation Protocol for Textbook Enrichment with Prerequisite Knowledge Graph. Technology, knowledge and learning: Learning mathematics,science and the art in the context of figital technologies, 2024, 29(1): 197-228.

[7] Abu-Salih Bilal, Alotaibi Salihah. A systematic literature review of knowledge graph construction and application in education. Heliyon, 2024, 10(3).

[8] Qinggang Zhang, Junnan Dong, Keyu Duan, et al. Contrastive Knowledge Graph Error Detection. Proceedings of the 31st {ACM} International Conference on Information & Knowledge Management, 2022, 2590-2599.

[9] Yankai Lin, Zhiyuan Liu, Maosong Sun, et al. Learning entity and relation embeddings for knowledge graph completion. Proceedings of the Twenty-Ninth {AAAI} Conference on Artificial Intelligence, 2015, 2181-2187.

[10] Jihoon Tack, Sangwoo Mo, Jongheon Jeong, et al. Csi: Novelty detection via contrastive learning on distributionally shifted instances. Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, 2020, 11839–11852.

[11] Chengwei Chen, Yuan Xie, Shaohui Lin, et al. Novelty detection via contrastive learning with negative data augmentation. Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, 2021, 606-614.

[12] Yingheng Wang, Yaosen Min, Xin Chen, et al. Multi-view graph contrastive representation learning for drug-drug interaction prediction. The Web Conference 2021, 2021, 2921-2933.

[13] Xin Xia, Hongzhi Yin, Junliang Yu, et al. Self-supervised hypergraph convolutional networks for session-based recommendation. Thirty-Fifth {AAAI} Conference on Artificial Intelligence, 2021, 4503–4511.

[14] Ruixin Ma, Xiaoru Wang, Cunxi Cao, et al. GLSEC: Global and local semantic-enhanced contrastive framework for knowledge graph completion. Expert Systems with Application(Sep.), 2024, 250: 123793.

[15] Mandeng Gao, Shengwei Tian, Long Yu. CONHyperKGE: Using Contrastive Learning in Hyperbolic Space for Knowledge Graph Embedding. International Journal of Pattern Recognition & Artificial Intelligence, 2024, 38(4): 2451005:1-2451005:23.

[16] Bishan Yang, Wen-tau Yih, Xiaodong He, et al. Embedding entities and relations for learning and inference in knowledge bases. 3rd International Conference on Learning Representations, 2015.

[17] Théo Trouillon, Johannes Welbl, Sebastian Riedel, et al. Complex embeddings for simple link prediction. Proceedings of the 33nd International Conference on Machine Learning, 2016, 48: 2071–2080.

[18] Yichi Zhang, Zhuo Chen, Lingbing Guo, et al. NativE: Multi-modal Knowledge Graph Completion in the Wild. Proceedings of the 47th International {ACM} {SIGIR} Conference on Research and Development in Information Retrieval, 2024, 91-101.

[19] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, et al. Translating embeddings for modeling multi-relational data. Advances in Neural Information Processing Systems 26, 2013, 2787–2795.

ISSN:3007-9640 DOI: https://doi.org/10.70114/acmsr.2025.2.1.P370
[20] Guoliang Ji, Shizhu He, Liheng Xu, et al. Knowledge Graph Embedding via Dynamic Mapping Matrix. Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, 2015,687–696.

[21] Zhen Wang, Jianwen Zhang, Jianlin Feng, et al. Knowledge Graph Embedding by Translating on Hyperplanes. Proceedings of the Twenty-Eighth {AAAI} Conference on Artificial Intelligence: 2014, 1112-1119.

[22] Zhenghang Zhang, Jinlu Jia, Yalin Wan, et al. TransR*: Representation learning model by flexible translation and relation matrix projection. Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology, 2021, 40(5):10251-10259.

[23] Yingkai Cai, Chu Wang, Zhongfeng Wang, et al. An improved Knowledge Graph Model Based on Fuzzy Theory and TransR. IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), 2020.

[24] Zhao Hongxi, Li Hongfei, Liu Qiaoming, et al. Using TransR to enhance drug repurposing knowledge graph for COVID-19 and its complications. Methods, 2024, 221: 82-90.