Doctoral Dissertation

# Exploration of Large Language Models for Noise Detection and Refinement in Knowledge Graphs

Na DONG

Supervisor: Kiyoaki SHIRAI

Graduate School of Advanced Science and Technology
Japan Advanced Institute of Science and Technology
(Information Science)

March 2026

**Abstract**

Knowledge graphs (KGs) have emerged as a powerful paradigm for representing structured knowledge about the real world. KGs encode information as a collection of triples, each consisting of a head entity, a relation, and a tail entity. This kind of structure enables not only the integration of heterogeneous data sources but also the support of downstream tasks such as question answering, recommender systems, and semantic search. However, the practical application of KGs is often hindered by several common challenges. KGs automatically constructed from large text corpora are prone to errors due to natural language ambiguity and extraction processes; while manually constructed KGs, while more reliable, are still susceptible to human error, have limited scalability, and are incomplete. As a result, KGs frequently contain misinformation, redundancy, and noisy triples, all of which can undermine their reliability and reduce their effectiveness in real-world applications.

This dissertation addresses the critical problem of improving the quality of noisy KGs through a two-step framework: (1) detecting noisy triples, and (2) refining the noisy triples by correcting the entities involved. Noisy KGs not only distort factual knowledge but also propagate errors to downstream reasoning and knowledge completion tasks. Unlike conventional approaches that rely heavily on hand-crafted rules, statistical heuristics, or embedding-based anomaly detection, our work leverages recent advances in large language models (LLMs), which possess strong semantic understanding and contextual reasoning capabilities, making them well-suited for identifying and repairing noisy knowledge structures at scale.

Within this framework, we propose two methods: *LLM_sim* and *LLM_rule*. The first method, *LLM_sim*, evaluates the plausibility of a candidate triple by comparing it with existing triples in the KG, thereby capturing subtle semantic inconsistencies that traditional models often miss. In addition to detection, we designed *LLM_sim* to repair detected noisy triples by grouping and calculating similarity. The second method, *LLM_rule*, induces logical rules directly from the KG to capture semantic constraints between entities and relations. We generate relevant rules through *LLM_rule*, which can propose effective corrections, providing interpretability and transparency that are often missing in similarity-based or embedding-based methods.

To validate the effectiveness of the proposed methods, we conducted extensive experiments on both synthetic and real-world noisy KGs. Synthetic datasets allow controlled injection of noise. These datasets enable systematic evaluation of detection and refinement capabilities under varying noise

ratios. Real-world datasets reflect the complex and heterogeneous nature of naturally occurring noise. Across both settings, our results demonstrate that *LLM_sim* and *LLM_rule* achieve strong performance in identifying and refining noisy triples.

Crucially, our comparative analysis reveals that **LLM_rule** generally emerges as the superior model, outperforming LLM_sim in both detection accuracy and refinement quality across fact-oriented datasets (e.g., FB15k-237, s-NELL). While LLM_sim remains competitive in linguistically rich domains like WN18RR, LLM_rule's integration of explicit logical constraints offers greater robustness against hallucination and structural inconsistencies.

We further investigated the subsequent impact of noise refinement on knowledge graph completion (KGC). KGC aims to infer missing triples in a KG by learning patterns from observed triples. However, noise in the training data severely impairs the generalization performance of KGC models. By incorporating refined KGC generated by *LLM_sim* and *LLM_rule*, we observed significant improvements in standard KGC metrics, including mean reciprocal rank (MRR) and hits@k. These improvements confirm that high-quality input data is a crucial factor in improving the performance of completion models and that noise refinement can serve as a key preprocessing step in the KGC pipeline.

The contributions of this work can be summarized as follows. First, we formulate a general framework for addressing noise in KGs that combines detection and refinement in a unified process. Second, we propose two novel methods, *LLM_sim* and *LLM_rule*, that leverage the semantic reasoning and rule induction capabilities of LLMs to achieve robust noise handling. Third, we provide extensive empirical evidence from both synthetic and real-world datasets, demonstrating that our methods consistently outperform baseline approaches in both detection accuracy and refinement quality. Finally, we show that our noise refinement framework directly enhances the performance of downstream KGC models, thereby underscoring its practical significance for KG applications. This study advances the state of the art in KG refinement by integrating the semantic power of LLMs with interpretable rule-based reasoning. The proposed methods offer a scalable, effective, and interpretable solution to the longstanding problem of noise in KGs. By improving KG quality, our work not only enhances their reliability but also paves the way for more accurate and trustworthy knowledge-driven AI systems. The findings of this dissertation hold broader implications for the development of robust intelligent systems that depend on high-quality structured knowledge.

**Keywords:** Knowledge Graph, Knowledge Acquisition, Error Correction, Large Language Model, Rule Induction, Knowledge Graph Completion