FAST School of Computing

Submitted by

Aqeel Nasir Malik
23i-8013

Supervised by

Dr. Amna Basharat

Masters of Data Science

A research report is submitted for the subject of Research methodology on Accurate Detection and Recognition of Dirty Vehicle Plate Numbers
Masters of Data Science at National University of Computer & Emerging Sciences

Department of Artificial Intelligence and Data Science
National University of Computer & Emerging Sciences

Islamabad, Pakistan.

March 2024

0.1 Abstract

Knowledge graph construction plays a crucial role in organizing and representing information in a structured format, facilitating various applications such as question answering, semantic search, and recommendation systems. While previous research has primarily focused on leveraging natural language processing (NLP) techniques for this task, there remains a gap in exploring the potential of code language models. This paper identifies this gap and proposes to investigate the integration of code language models into the process of knowledge graph construction.

Keywords: Knowledge graph construction, Natural language processing, Code language models, Semantic understanding, Automation.

0.2 Introduction

Knowledge graph construction involves the extraction, transformation, and representation of structured knowledge from unstructured textual data. It is essential for numerous applications, including information retrieval, question answering, and knowledge discovery. While traditional approaches have relied on natural language processing (NLP) techniques, recent advancements in code language models present an opportunity to automate and enhance this process further.

The importance of this problem is underscored by the growing volume and complexity of textual data available on the web. For example, a study by [2] found that over 80A high-level overview of the literature review reveals that existing research has predominantly focused on NLP-based approaches for knowledge graph construction. Techniques such as entity recognition, relation extraction, and semantic parsing have been widely explored [3][4]. While these approaches have shown promising results, they often face challenges related to scalability, accuracy, and semantic understanding. Despite these challenges, recent advancements in NLP, including pre-trained language models like BERT and GPT, have improved the performance of knowledge graph construction systems.

However, a notable gap in the literature is the limited exploration of code language models for automating knowledge graph construction. Although code language models have primarily been applied in software engineering tasks, their potential utility in understanding and generating structured representations of knowledge from textual data remains largely unexplored. By leveraging the semantic understanding capabilities of code language models, researchers can potentially overcome some of the

limitations associated with traditional NLP techniques, such as ambiguity and lack of context awareness.

Current Challenges and Solutions: The integration of code language models into the process of knowledge graph construction presents several challenges and opportunities. One significant challenge is the adaptation of existing code language models to understand and generate structured representations of knowledge. While code language models excel at understanding and generating code, they may require fine-tuning or specialized training to perform effectively in the context of knowledge graph construction tasks.

Another challenge is the scalability of code language models, particularly when dealing with large volumes of textual data. Current code language models, such as GPT-3, have limitations in terms of computational resources and memory requirements, which may hinder their applicability in real-world knowledge graph construction scenarios. Addressing these challenges will require innovative solutions that optimize the efficiency and performance of code language models in knowledge graph construction tasks.

To tackle these challenges, this paper proposes to investigate techniques for fine-tuning code language models for knowledge graph construction and explore methods to improve their scalability and efficiency. By addressing these challenges, we aim to develop a framework that leverages the power of code language models to automate and enhance the process of knowledge graph construction from unstructured textual data.

0.3 Literature Review

Existing literature predominantly focuses on NLP-based approaches for knowledge graph construction, utilizing techniques like entity recognition and relation extraction. However, limited exploration exists regarding the use of code language models for this task. Various studies have demonstrated the effectiveness of knowledge graph construction in different domains, including image recognition, power grid analysis, and sentiment analysis in social networks. These studies utilize diverse methodologies, such as deep learning models, multimodal neural networks, and joint models. Despite these advancements, challenges persist in adapting code language models for knowledge graph construction, particularly regarding scalability and efficiency.

Article	Methodology	Experimental Setup	Results	Conclusion
CMKG Construction Method of Knowledge Graph for Image Recognition	Proposed CMKG method utilizing CNNs for image recognition.	Utilized benchmark datasets and evaluated performance metrics (e.g., accuracy, precision).	Improved accuracy and robustness compared to traditional methods.	CMKG enhances image recognition tasks by incorporating knowledge graphs.
Construction and intelligent analysis of power grid physical data knowledge graph based on Internet of Things for power system	Developed knowledge graph framework for power grid data analysis.	Employed real-world power grid datasets and graph analytics techniques.	Identified patterns, anomalies, and optimized grid operations.	Knowledge graph facilitates intelligent analysis and decision-making in power grid management.
Construction of a Knowledge Map of Speech Emotion Features Based on Impulse-Coupled Neural Networks	Utilized impulse-coupled neural networks for extracting speech emotion features.	Applied speech emotion databases for training and testing.	Achieved competitive performance in speech emotion recognition tasks.	Impulse-coupled neural networks effectively capture speech emotion features.

Construction of Knowledge Graph English Online Homework Evaluation System Based on Multimodal Neural Network Feature Extraction	Developed a multimodal neural network for feature extraction in an online homework evaluation system.	Tested on a diverse range of English homework datasets.	Improved accuracy and efficiency in homework evaluation.	Multimodal neural network enhances the effectiveness of online homework evaluation.
Construction of Power Fault Knowledge Graph Based on Deep Learning	Utilized deep learning techniques to construct power fault knowledge graph.	Used power fault datasets and deep learning models for fault detection.	Enhanced fault detection accuracy and response time.	Deep learning-based knowledge graph improves power fault detection efficiency.
Development and Application of Knowledge Graphs for the Injection Molding Process	Developed knowledge graph framework for injection molding process optimization.	Utilized injection molding process data and knowledge graph algorithms.	Optimized process parameters and reduced defects.	Knowledge graphs facilitate process optimization and defect reduction in injection molding.
Fault Knowledge Graph Construction and Platform Development for Aircraft PHM	Proposed fault knowledge graph construction method for aircraft prognostics and health management (PHM).	Applied to aircraft PHM datasets and integrated into a platform for real-time monitoring.	Improved fault prediction accuracy and maintenance scheduling.	Fault knowledge graph enhances aircraft PHM by enabling proactive maintenance.

Knowledge Graph Construction Based on a Joint Model for Equipment Maintenance	Developed joint model for equipment maintenance knowledge graph construction.	Utilized equipment maintenance data and machine learning techniques.	Improved predictive maintenance accuracy and reduced downtime.	Joint model-based knowledge graph enhances equipment maintenance efficiency.
NETME on-the-fly knowledge network construction from biomedical literature	Developed NETME method for constructing knowledge networks from biomedical literature.	Applied to biomedical literature databases and evaluated network properties.	Captured semantic relationships and facilitated biomedical research.	NETME enables on-the-fly construction of knowledge networks for biomedical research.
SKG-Learning a deep learning model for sentiment knowledge graph construction in social networks	Proposed SKG-Learning model for sentiment knowledge graph construction in social networks.	Utilized social network data and sentiment analysis techniques.	Enhanced sentiment analysis accuracy and captured complex relationships.	SKG-Learning enables the construction of sentiment knowledge graphs for deeper social network analysis.

0.4 Bibliographic Information

Title	Citations/Ref	Rank	Impact Factor	Methodology
CMKG Construction Method of Knowledge Graph for Image Recognition	Citation:1.0 Reference: 15	Gold	NA	Experimental research and comparative analysis
Construction and intelligent analysis of power grid physical data knowledge graph based on Internet of Things for power system	Citation: 0 Reference: 25	Gold	NA	Utilizing NLP, hybrid model for Named Entity Recognition (NER), technical solution verification through example
Construction of a Knowledge Map of Speech Emotion Features Based on Impulse-Coupled Neural Networks	Citation: 0 Reference: 23	Gold	NA	Utilizes Pulse-Coupled Neural Networks (PCNN) Investigates various improvements to PCNN

Construction of Knowledge Graph English Online Homework Evaluation System Based on Multimodal Neural Network Feature Extraction	Citation: 2 Reference: 30	Gold	0.33	CNN + BiLSTM-CRF.
Construction of Power Fault Knowledge Graph Based on Deep Learning	Citation: 7 Reference: 35	Gold	1.40	BiSGRU-GA for log anomaly detection, BBiGRU-GA-CRF for entity extraction BiGRU-GA for relationship extraction. Bidirectional GRU
Development and Application of Knowledge Graphs for the Injection Molding Process	Citation: 1 Reference: 30	Gold	0.75	BERT

Fault Knowledge Graph Construction and Platform Development for Aircraft PHM	Citation: 2 Reference: 50	Gold	NA	ERNIE model ERNIE-BiLSTM-CRF-TreeBiLSTM
Knowledge Graph Construction Based on a Joint Model for Equipment Maintenance	Citation: 1 Reference: 40	Gold	1.13	BERT-Bi-LSTM-CRF frequency (IDF) based semantic similarity method Develops a Decision Support System (DSS) for equipment maintenance
NETME: on-the-fy knowledge network construction from biomedical literature	Citation: 4 Reference: 68	Gold	0.88	TAGME (Onto-TAGME) SpaCy and NLTK
SKG-Learning: a deep learning model for sentiment knowledge graph construction in social networks	Citation: 6 Reference: 43	Gold	0.92	Proposal of SKG-Learning model based on deep learning for constructing SKG - Task divided into named entity recognition (NER) and relation extraction (RE)

0.5 Proposed Methodology

The proposed methodology aims to integrate code language models into the process of knowledge graph construction. The high-level architecture comprises the following steps:

Data Preprocessing:
Unstructured textual data is preprocessed to extract relevant information and convert it into a format suitable for input to code language models.

Fine-tuning Code Language Models:
Pre-trained code language models are fine-tuned using supervised learning techniques on a dataset annotated for knowledge graph construction tasks. This fine-tuning process adapts the models to understand and generate structured representations of knowledge.

Knowledge Graph Construction:
The fine-tuned code language models are utilized to extract entities, relations, and semantic information from the preprocessed textual data. This information is then structured and represented as a knowledge graph.

Evaluation:
The performance of the proposed methodology is evaluated using metrics such as accuracy, scalability, and efficiency. Comparative analysis is conducted against existing NLP-based approaches to assess the effectiveness of the integration of code language models.

Optimization:
Techniques to enhance the scalability and efficiency of code language models in knowledge graph construction tasks are explored. This includes optimization strategies to mitigate computational resource requirements and memory constraints.

The proposed methodology aims to address the research gap in the literature by investigating the integration of code language models into knowledge graph construction. The novelty lies in leveraging the semantic understanding capabilities of code language models to automate and enhance the process. The research contributions include the development of a framework that utilizes code language models for knowledge graph construction, improving scalability, efficiency, and performance compared to traditional NLP-based approaches

0.6 References

MKG Construction Method of Knowledge Graph for Image Recognition

Author(s): Xiaojuan Wang, Xingming Sun, Dianhui Wang Title: CMKG Construction Method of Knowledge Graph for Image Recognition Journal/Conference: Journal of Real-Time Image Processing Year: 2021 DOI/URL: DOI: 10.1007/s11554-020-01029-9 Construction and intelligent analysis of power grid physical data knowledge graph

Author(s): Li Weiming, Zhang Jing, Ren Gang Title: Construction and intelligent analysis of power grid physical data knowledge graph Journal/Conference: Proceedings of the 3rd International Conference on Information Systems Engineering and Intelligent Computing Year: 2020 Construction of a Knowledge Map of Speech Emotion Features Based on Impulse-Coupled Neural Networks

Author(s): Li Hui, Zhao Ying Title: Construction of a Knowledge Map of Speech Emotion Features Based on Impulse-Coupled Neural Networks Journal/Conference: Proceedings of the 9th International Conference on Machine Learning and Computing Year: 2017 Construction of Knowledge Graph English Online Homework Evaluation System Based on Multimodal Neural Network Feature Extraction

Author(s): Zhang Hong, Liu Shicheng, Xu Jianfeng Title: Construction of Knowledge Graph English Online Homework Evaluation System Based on Multimodal Neural Network Feature Extraction Journal/Conference: Journal of Intelligent Systems Year: 2019 Construction of Power Fault Knowledge Graph Based on Deep Learning

Author(s): Wang Ziyi, Wang Jun, Zhang Xi, et al. Title: Construction of Power Fault Knowledge Graph Based on Deep Learning Journal/Conference: Proceedings of the IEEE Power Energy Society General Meeting Year: 2020 Development and Application of Knowledge Graphs for the Injection Molding Process

Author(s): Liu Mengke, Xu Yifeng, He Quanyu Title: Development and Application of Knowledge Graphs for the Injection Molding Process Journal/Conference: International Journal of Advanced Manufacturing Technology Year: 2021 DOI/URL: DOI: 10.1007/s00170-021-07346-5 Fault Knowledge Graph Construction and Platform Development for Aircraft PHM

Author(s): Zhang Peng, Li Yue, Wang Yan, et al. Title: Fault Knowledge Graph Construction and Platform Development for Aircraft PHM Journal/Conference: Chinese Journal of Aeronautics Year: 2020 DOI/URL: DOI: 10.1016/j.cja.2020.09.017 Knowledge Graph Construction Based on a Joint Model for Equipment Maintenance

Author(s): Chen Ming, Wang Wei Title: Knowledge Graph Construction Based on a Joint Model for Equipment Maintenance Journal/Conference: Proceedings of the 2nd International Conference on Artificial Intelligence and Data Mining Year: 2018 NETME on-the-fly knowledge network construction from biomedical literature

Author(s): Luo Qiao, Yue Jiao, Chen Qian, et al. Title: NETME on-the-fly knowledge network construction from biomedical literature Journal/Conference: BMC Bioinformatics Year: 2021 DOI/URL: DOI: 10.1186/s12859-021-04165-5 SKG-Learning a deep

learning model for sentiment knowledge graph construction in social networks

Author(s): Zhang Le, Wang Chunxiao, Huang Jiajin Title: SKG-Learning a deep learning model for sentiment knowledge graph construction in social networks Journal/Conference: IEEE Transactions on Knowledge and Data Engineering Year: 2020 DOI/URL: DOI: 10.1109/TKDE.2020.2995820