A Review of Knowledge Graph Embedding Methods of TransE, TransH and TransR for Missing Links

Salwana Mohamad @ Asmara*
Faculty of Computing
Universiti Malaysia Pahang Al-Sultan Abdullah
Pekan, Malaysia
salwanamohamad@ump.edu.my

Noor Azida Sahabudin
Faculty of Computing
Universiti Malaysia Pahang Al-Sultan Abdullah
Pekan, Malaysia
azida@ump.edu.my

Nor Syahidatul Nadiah Ismail
Faculty of Computing
Universiti Malaysia Pahang Al-Sultan Abdullah
Pekan, Malaysia
nadiahismail@ump.edu.my

Ily Amalina Ahmad Sabri
Faculty of Ocean Engineering Technology and Informatics
Universiti Malaysia Terengganu,
Terengganu, Malaysia
ilylina@umt.edu.my

2023 IEEE 8th International Conference On Software Engineering and Computer Systems (ICSECS) | 979-8-3503-1093-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICSECS58457.2023.10256354

***Abstract*— Knowledge representation and reasoning require knowledge graph embedding as it is crucial in the area. It involves mapping entities and relationships from a knowledge graph into vectors of lower dimensions that are continuous in nature. This encoding enables machine learning algorithms to effectively reason and make predictions on graph-structured data. This review article offers an overview and critical analysis specifically about the methods of knowledge graph embedding which are TransE, TransH, and TransR. The key concepts, methodologies, strengths, and limitations of these methods, along with examining their applications and experiments conducted by existing researchers have been studied. The motivation to conduct this study is to review the well-known and most applied knowledge embedding methods and compare the features of those methods so that a comprehensive resource for researchers and practitioners interested in delving into knowledge graph embedding techniques is delivered.**

Keywords—knowledge graph, knowledge graph embedding, TransE, TransH, TransR

I. INTRODUCTION

Knowledge graphs (KG) are intelligent systems that integrate knowledge and data on a large scale, with their origins dating back to the expert systems developed in the late 1960s [1]. A knowledge graph comprises three fundamental components which are a collection of entities, *E*, a set of relations, *R*, and the interactions between entities and relations, which are expressed through triples. These triples follow a specific structure, defined as (*h, r, t*), where *h* represents the head entity, *t* represents the tail entity, and *r* signifies the relation that connects them [2]. KGs are widely utilized in various areas such as biomedical, search engine, and e-commerce to represent structured data and knowledge. KG embedding involves the conversion of entities and relationships from a knowledge graph into low-dimensional continuous vectors within a vector space. This transformation allows for efficient representation and analysis of the KG, enabling the performance of diverse reasoning and machine learning tasks [3], and the semantic and structural characteristics of the knowledge graph can be captured efficiently and concisely. KG embedding allows for the representation of entities and relationships as continuous vectors, which in turn facilitates a wide range of reasoning tasks. By mapping entities and relationships to vector representations, machine learning algorithms can effectively operate on graph-structured data. Task such as link prediction, entity categorization, and suggestion can be supported by entities and relations that is presented within a KG.

Entities and relations which are the embedding elements of a knowledge graph can be utilized to predict missing links by optimizing their probabilities based on a specified scoring function. This process allows for the generation of new facts that do not currently exist in the graph [2].

Knowledge graph embedding offers an adaptable approach for representing knowledge, which can be applied in various domains. These techniques find utility in a diverse range of applications, encompassing the prediction of missing information, the construction of recommender systems, the facilitation of question answering capabilities, the support for query expansion, and numerous other use cases. By employing knowledge graph embedding, these applications can benefit from the efficient representation and reasoning capabilities provided by embedding techniques. [4]. However, triplets of (*h, r, t*) alone offer limited information. It acquiring a relevant text corpus within the domain is not always viable, leading to embedding outcomes that diverge from the intended meaning [5].

Within the context of knowledge graphs (KGs), there exists a task similar to inference in knowledge bases (KBs) known as graph completion. This task involves various subtasks such as prediction of entity, relation, and triple classification, all of which rely on knowledge graph embeddings. TransE, TransH, and TransR are three popular methods for knowledge graph embedding, each with its own strengths and limitations. TransE is a modest and effective

Authorized licensed use limited to: Universiti Malaysia Pahang Al Sultan Abdullah (UMPSA).. Downloaded on March 06,2024 at 08:01:28 UTC from IEEE Xplore. Restrictions apply.