IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 10 October 2023, accepted 1 November 2023, date of publication 8 November 2023, date of current version 17 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3331585

APPLIED RESEARCH

Improving FMEA Comprehensibility via Common-Sense Knowledge Graph Completion Techniques

HOUSSAM RAZOUK[1,2], XING LAN LIU[3], AND ROMAN KERN[2,3]
[1]Infineon Technologies Austria AG, 9500 Villach, Austria
[2]Institute of Interactive Systems and Data Science, Graz University of Technology, 8010 Graz, Austria
[3]Area of Knowledge Discovery, Know-Center GmbH, 8010 Graz, Austria

Corresponding author: Houssam Razouk (Houssam.razouk@infineon.com)

This work is conducted under the framework of the Electronic Components and Systems for European Leadership (ECSEL) AI4DI "Artificial Intelligence for Digitising Industry" project. The project has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 826060. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Germany, Austria, Czech Republic, Italy, Latvia, Belgium, Lithuania, France, Greece, Finland, Norway. The publication costs are funded by the "Edge AI Technologies for Optimised Performance Embedded Processing" project, received funding from Key Digital Technologies (KDT) JU under grant agreement No 101097300. The KDT JU receives support from the European Union's Horizon Europe research and innovation program and Austria, Belgium, France, Greece, Italy, Latvia, Luxembourg, Netherlands, Norwayg.

ABSTRACT The Failure Mode Effect Analysis process (FMEA) is widely used in industry for risk assessment, as it effectively captures and documents domain-specific knowledge. This process is mainly concerned with causal domain knowledge. In practical applications, FMEAs encounter challenges in terms of comprehensibility, particularly related to inadequate coverage of listed failure modes and their corresponding effects and causes. This can be attributed to the limitations of traditional brainstorming approaches typically employed in the FMEA process. Depending on the size and diversity in terms of disciplines of the team conducting the analysis, these approaches may not adequately capture a comprehensive range of failure modes, leading to gaps in coverage. To this end, methods for improving FMEA knowledge comprehensibility are highly needed. A potential approach to address this gap is rooted in recent advances in common-sense knowledge graph completion, which have demonstrated the effectiveness of text-aware graph embedding techniques. However, the applicability of such methods in an industrial setting is limited. This paper addresses this issue on FMEA documents in an industrial environment. Here, the application of common-sense knowledge graph completion methods on FMEA documents from semiconductor manufacturing is studied. These methods achieve over 20% MRR on the test set and 70% of the top 10 predictions were manually assessed to be plausible by domain experts. Based on the evaluation, this paper confirms that text-aware knowledge graph embedding for common-sense knowledge graph completion are more effective than structure-only knowledge graph embedding for improving FMEA knowledge comprehensibility. Additionally we found that language model in domain fine-tuning is beneficial for extracting more meaningful embedding, thus improving the overall model performance.

INDEX TERMS Natural language processing, common-sense knowledge, failure mode effect analysis, FMEA, semiconductor manufacturing, knowledge graph completion.

I. INTRODUCTION

Risk assessment and root cause analysis are two practices that capture and document causal domain knowledge. This causal domain knowledge is crucial for downstream tasks such as decision making, data analysis, and more. In industry, Failure Mode Effect Analysis (FMEA) process is widely used for risk assessment. On a high level, FMEA process can be split into three main sub-types, which are system FMEA, design FMEA, and process FMEA [1], [2], [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Arianna D'Ulizia.

Process FMEA is applied to enhance and optimise the manufacturing processes and workflow [4]. In the FMEA process, a multidisciplinary team is often involved in brainstorming approaches to identify lists of failure modes, their causes and effects. Failure modes along with their causes and effects are documented in FMEA documents. These FMEA documents typically use a standardised tabular format. This tabular format columns contain the failure mode, effect, cause, and other columns such as risk priority number, detection, etc. Each row in FMEA documents tabular format represents a particular failure mode of a process or product, along with its corresponding effect and cause. As such, the tabular structure of the FMEA documents indicates implicit relations between designated FMEA cells belonging to the same row. Mainly, cell text belonging to the failure mode column describes the cause of cell text belonging to the effect column. At the same time, cell text belonging to the failure mode column describes the effect of the cell text belonging to the cause column. In addition, the FMEA tabular format incorporates columns such as Element ID and Characteristic. These columns provide context information of the causal relations, specifying process step type in process FMEA. Finally, a document header is employed to outline the scope of the FMEA. This header specifies the product or technology, for which the FMEA is conducted, provides context information for the entire FMEA. To facilitate comprehension, Table 1 offers an illustrative example with two rows extracted from actual process FMEA document from semiconductor manufacturing.

Prior researchers have emphasized a notable issue in FMEA documents regarding the non-standardized descriptions of failure modes [5]. This results in instances where similar or even identical failure modes are articulated and documented in varying manners. This challenge extends to the effect, cause, and characteristic columns as well. Consequently, methods relying on exact matching for processing this information are significantly impacted.

Also, Bluvband et al. have raised concerns regarding the clarity and comprehensibility of brainstorming approaches employed in FMEA, potentially resulting in instances of missing information, such as the omission of possible failure modes [6]. Moreover, it is not uncommon for causes or effects to be overlooked. As a consequence, FMEA documents may exhibit significant variation, dependent on factors like the team size, diversity in disciplines, and prior experience in conducting the analysis. These disparities in FMEA documents may imply limitations in terms of assignability, precision, and plausibility of the documented information.

In summary, the variability in FMEA documents may raise concerns regarding the comprehensibility of the information they contain [7]. While increasing the size of the FMEA team holds promise for improving the clarity of information in these documents, practical constraints such as time and resource limitations may present hurdles in implementing this approach. Consequently, the absence of standardized and comprehensible descriptions in FMEA documents can significantly hinder the practical applicability and transferability of causal knowledge for downstream tasks.

In response to the challenges of non-standardized descriptions of failure modes in FMEA, scholars turned to text mining techniques. The effectiveness of these techniques has been demonstrated in various ways, such as extracting knowledge from FMEA documents for verification and validation planning [8], standardizing components failure modes [5], and assessing the consistency of cells and relations in the FMEA tabular format [9].

At the same time, noteworthy advances in the common-sense knowledge completion techniques include new publicly available large scale data sets such as Atomic [10] and ConceptNet [11], modelling the common-sense knowledge bases as knowledge graphs and the use of knowledge graph embedding algorithms for downstream tasks such as common-sense knowledge graph completion [12], [13]. These advances have made it possible to integrate knowledge from various data sources. For instance, in [13] and [12] the authors leverage language models pre-trained on a large scale text data set to enrich the embedding of the nodes in common-sense knowledge bases. This technique can be especially important for root cause analysis and risk assessment in an industrial setting where a significant amount of knowledge is documented in a semi-structured way, as seen in FMEA, and in a unstructured textual format, as evident in failure analysis reports.

Complementary to existing research, that has explored ontology-based solutions for improving the transferability of causal domain knowledge contained in FMEA documents as outlined in [14], [15], and [16] and has showcased notable progress in common-sense knowledge completion techniques [12], [13], our study aims to examine the potential impact of these advancements on the FMEA process effectiveness. Specifically, we explore whether leveraging common-sense knowledge completion techniques can enhance the comprehensibility of brainstorming sessions by treating FMEA documents as the initial knowledge base. As such, in this work, the following intriguing research questions are raised:

- *(RQ1)* What are the similarities and key differences between the knowledge documented within FMEA and common-sense knowledge bases?
- *(RQ2)* How effective are common-sense knowledge completion techniques to enhance the comprehensibility of information articulated in FMEA documents?

To the best of our knowledge, no previous studies have explored the use FMEA documents as a domain-specific common-sense knowledge base. To address this gap, we propose to leverage novel common-sense techniques, in particular graph embedding, to perform knowledge graph completion tasks on actual FMEA documents from the semiconductor manufacturing industry. By answering the research questions we have posed, we make concrete contributions to the extension of common-sense reasoning techniques to industrial applications. Specifically, our contributions include:

TABLE 1. Example of an actual process FMEA document from semiconductor manufacturing. The scope header (first row) provides context for the entire FMEA document. The tabular structure indicates implicit causal relations between designated FMEA cells belonging to the same row. Two causal relations are identified in each FMEA row: the first between the failure mode and its effect, and the second between the failure mode and its cause. The Element ID and Characteristic columns describe the context information for each individual causal relation.

Scope	Technology T	Version V	Voltage Class C	Wafer Diameter D
Element ID	**Characteristic**	**Failure Mode**	**Effect**	**Cause**
123	START RAW WAFER	use of wrong material	malfunction of device -> high yield loss	mismatching
123	START RAW WAFER	use of material with low/high substrate resistivity	reduced breakdown voltage BVDSS/Rdson -> high yield loss	substrate doping not in spec (auto doping tail)

- Highlighting the FMEA comprehensibility issue with respect to possible causes and effects.
- Assessing the similarities between common-sense knowledge bases and industrial risk assessment documents.
- Highlighting the unique difficulties associated with FMEA data to enable the successful implementation of said techniques.
- Be the first to apply techniques from common-sense completion to domain-specific knowledge extracted from risk assessment documents.
- Empirical evaluation on actual FMEA documents in a case study from semiconductor manufacturing.

II. RELATED WORK

In [17], Zhongyi et al. highlighted that the success of FMEA depends on two key factors: (1) whether the FMEA process completely and accurately identifies the failure modes contained in the system and (2) whether it can scientifically evaluate the risk level of these failure modes. The former is also applicable to the causes and effects of the identified failure modes. Typically, the FMEA failure modes identification process along with its causes and its effects is carried out by a multidisciplinary team of domain experts using brainstorming approaches. Although the brainstorming approaches are the most common method, Bluvband et al. in [6] believe that they could easily lead to the omission of failure modes. Therefore, Bluvband et al. in [6] propose a checklist consisting of 10 types of failure modes as a supplement to the brainstorming approaches enabling the FMEA team to establish a personalised list of failures that are associated with a specific activity or item. Other approaches for improving the FMEA comprehensibility are summarised in section II-A.

Unlike conventional knowledge bases, common-sense knowledge bases, similar to FMEA, are characterized by entities that are represented by non-standardized free-form text. These knowledge bases attempt to model general common-sense knowledge including causal relationships. In context of common-sense knowledge completions, which involves predicting missing links in a knowledge base, text-aware knowledge graph embedding methods have shown to be effective [12], [13]. These methods leverage natural language processing techniques to extract information from the free-form text descriptions of entities in the knowledge base, which can then be used to improve the accuracy of the knowledge graph embedding. There are several approaches for common-sense knowledge completions based on text-aware knowledge graph embedding, including those summarized in section II-B.

A. IMPROVING FMEA COMPREHENSIBILITY

Several approaches have been proposed to support failure modes, effects, and causes identification process including:

- Theoretical modelling: considers the system structure and its function in failure mode identification process. Fault Tree Analysis (FTA) [18] is a commonly used theoretical modelling approach. Additional examples on theoretical modelling approach include the extended Go model [19], the module-based fault propagation (MFP) model [20] and the functional fault identification and propagation (FFIP) model [21].
- Clustering approaches: try to identify similarities between products/processes and failure modes. Such approaches are based on analysing historical data and applying statistical clustering methods (e.g. K-means). As a practical example, in [22] the authors present methods for the extraction of component failure modes from maintenance text data set using Ward agglomeration and similarity-based histogram clustering. An ontology-based text mining system to process millions of unstructured information and cluster them with different techniques is presented in [23].
- Knowledge reasoning approaches: intend to obtain new knowledge or conclusions. The most common methods of knowledge reasoning approaches are expert systems and case-based reasoning (CBR). Such approaches are characterised by leveraging experiences accumulated over long-term practice to build knowledge base. As a welcome consequence, these approaches avoid complex quantitative models. In practice, CBR can be integrated with rule-based reasoning (RBR). For example, CBR is leveraged as RBR for decisions making imitating the experts decisions in fault analysis [24]. CBR is also used to provide a recommendation list for cross validation [25].
- Natural Language Processing (NLP) methods: are typically leveraged to automatically classify and extract

information from textual documents. In industry, NLP methods are mainly applied to process instructions, recorded text or warning information. For example, NLP methods are used to analyse online user comments, which describe faults that end users have experienced [26].

More recent works on FMEA explore integration with other techniques such as decision trees and fuzzy groups [27], bayesian networks [28] and deep learning [9]. However, to the best of our knowledge, no previous scholars considered the use of the FMEA data set as domain specific common-sense knowledge base. Thus, exploring the applicability of techniques from common-sense completion to FMEA documents is limited.

B. TEXT AWARE KNOWLEDGE GRAPH EMBEDDING FOR COMMON-SENSE KNOWLEDGE BASE COMPLETION

A knowledge graph is a structured representation of knowledge, denoted by $G = (N, T, RT)$, where N is a set of nodes (entities), T is a set of triples, and RT is a set of relation types. In a knowledge graph, where a triplet consists of a subject node, a relation type, and an object node, given the incomplete triplet "use of material with low/high substrate resistivity- Causes - ??", the objective of knowledge graph completion methods is to predict the missing object node, which is likely to be "reduced breakdown voltage BVDSS/Rdson". However, this task is challenging as it requires modeling complex patterns and dependencies among nodes and relation types in the knowledge graph.

Knowledge graph embedding is the process of representing nodes and relations types of a knowledge graph as low-dimensional vectors in a continuous vector space. Structure only knowledge graph embedding methods, such as TransE [29] DistMult [30] ConvE [31], learn the nodes and relation types embeddings only from the knowledge graph structure treating the nodes and relation types as categorical entries. For example, Convolution knowledge graph embedding models, such as ConvE [31], use embedding layers to learn the mapping function from categorical representation of nodes and relations types in a knowledge graph to a denser, continuous representation. The ConvE method also uses a 2D convolutional layer to capture the interactions between the head node and the relation in a triple. Other approaches, such as Structure Aware Convolutional network (SACN) [32], combines different structure only knowledge graph embedding methods, e.g. Graph Convolutional Network (GCN) and ConvE, to provide solution for learning graph node embedding utilizing graph connectivity structure. While structure only knowledge method for knowledge graph embedding effectively encodes the structure information of the knowledge graph, it does not incorporate any text information that might be available in a knowledge graph.

In common-sense knowledge graphs and FMEA knowledge graphs, conceptually related nodes but not equivalent are often represented as distinct nodes. It can be difficult to capture the nuances of these relationships without access to text information as part of the knowledge graph embedding process. This is because the semantics of the text contained in the nodes can have a significant impact on the relationships between them.

Recent research have been focused on developing methods to more effectively integrate text information into knowledge graph embedding. For example, BERT KG [33] leverages knowledge graph to directly adapt language model wights. Other approaches leverages language models to extract the text attributes embeddings. These embeddings are concatenated to nodes embeddings, which are learned for the knowledge graph structure [12] or even replaced them completely taking BERT-ConvE as a prime example [13].

The BERT-ConvE method uses transformer based language models to encode the text of the knowledge graph nodes and ConvE to encode the structure information of the knowledge graph. This approach has been shown to achieve state-of-the-art performance on common-sense benchmark knowledge graph completion tasks [13]. Both ConvE and BERT-ConvE use the same pre-processing pipeline, where all nodes and relation types are categorically encoded. The knowledge graph is represented as triples and split into a training and testing set. The training set contains the known knowledge graph, while the testing set contains triples to be predicted, which are not limited to the nodes in the known knowledge graph. The primary difference between BERT-ConvE and ConvE is that in ConvE, the node and relation type from the training example are passed to embedding layers, while in BERT-ConvE, a pre-trained language model (e.g. BERT) is leveraged to extract the nodes embeddings based on the nodes text and replaces the nodes embedding layer. This allows for encoding of nodes unknown to the training set and ensures text-awareness of the embedding. These nodes are referred to as unknown nodes. Furthermore, a variation of BERT-ConvE named Triples-BERT-ConvE extracts the context-aware embeddings of known nodes based on their neighbourhood in the known knowledge graph. Meanwhile, the embeddings of the relation types are learned during the training of the graph embedding model. It should be noted that while training the BERT-ConvE knowledge graph embedding model, the weights of the language model (i.e. BERT) are frozen. Thus, BERT-ConvE performance depends on the embeddings quality of the texts contained in the node. The quality of these text embeddings can be significantly impacted by the relevance of the model original pre-training data and the fine-tuning approach used to adapt the model to a domain specific text.

III. METHOD

FMEA documents contain an incomplete set of failure modes, causes, and effects, which can limit their usefulness in developing appropriate root cause analysis and risk mitigation strategies. The objective of this study is to develop a method that can increase the usability of the causal domain knowledge contained in a set of FMEA documents by transforming the information into a more structured format, while also completing the lists of causes, and effects. To achieve this objective, we propose a method that involves the following steps:

1) Structured Data Representation: This step involves representing the information in a structured format (i.e. knowledge graph) that can be easily analyzed and used to develop knowledge completion method. This step is elaborated in section III-A.
2) Knowledge Completion: This step involves completing missing causes, and effects using knowledge graph embeddings by considering the FMEA documents as a commons sens knowledge base. This step is elaborated in section III-B.

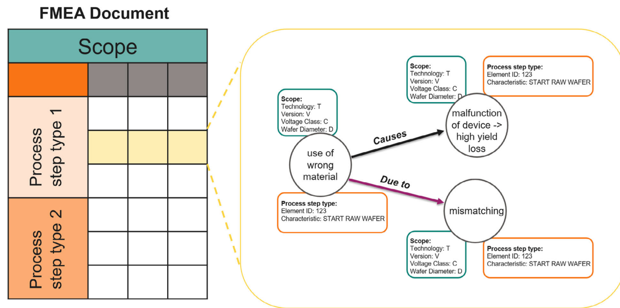

FIGURE 1. Illustrating a knowledge graph extracted from an actual FMEA document. The nodes depict entries in the failure mode, cause, and effects columns. Relationships, denoted as *Causes* and *Due_to*, link failure modes to causes and effects, respectively. Supplementary attributes, such as the scope of the FMEA and process step type, are incorporated into the nodes to uphold consistent information granularity with the original FMEA document.

A. KNOWLEDGE GRAPH CONSTRUCTION FROM FMEA DOCUMENTS: MAPPING INFORMATION TO A GRAPH STRUCTURE

Given a knowledge graph the definitions of nodes and relation types within a knowledge graph are heavily dependent on the downstream task at hand. For example, if the objective is to enhance the comprehensibility of FMEA with regard to the listed failure modes, it is critical to model context information in a structured manner. This includes modeling context information, such as process step type, and their relationship to corresponding failure modes as separate nodes, which can be linked together to represent the semantics of how a particular failure mode might occur in a given context.

In contrast, if the downstream objective is to generate hypotheses about the causes or effects of events (i.e. failure modes), a different approach is necessary. The proposed approach involves representing each cell from the failure mode, effect, or cause columns in the FMEA document as a node in a knowledge graph. Next, two edges are created to represent the causal relations documented in each FMEA row. Namely, one edge from the failure mode to the effect with the type *Causes*, and another edge from the failure mode to the cause with the type *Due_to*. This approach is aligned with the data generation process, i.e. the FMEA process, where failure modes are first identified and then their causes and effects are determined, and has the potential to improve model predictions as suggested in previous research [34]. To maintain the same level of information granularity, the nodes in the knowledge graph also include additional attributes representing the context information. These additional attribute include the scope of the FMEA and process step type, which is represented by the Element ID and Characteristics. Figure 1 illustrates an example of extracted nodes and relations from a row contained in an actual Process FMEA also described in Table 1.

Duplicated nodes and edges are removed to ensure that the resulting FMEA knowledge graph is concise and informative Transferring the FMEA documents into a knowledge graph allows for a more streamlined and manageable representation of the information contained in the FMEA documents.

In order to compare the knowledge graph extracted from FMEA documents (FMEA knowledge graph) with common-sense knowledge bases, several metrics are used to analyze the structure and content of the graphs. These metrics include the number of nodes, the number of triples, and the number of relation types in the graphs. Additionally, the overall density for each of the knowledge graphs is also computed.

Another important aspect of comparing the FMEA knowledge graph with common-sense knowledge graphs is to assess and compare the semantic relations represented by relation types. Also, language model tokenizers are used to assess the relevance of the language used in the knowledge graph nodes text to the language used in the data set, on which the language models are initially pre-trained. This can help to identify biases induced by domain-specific language typically used in the FMEA documents.

B. KNOWLEDGE GRAPH COMPLETION: PREDICTING MISSING CAUSES AND EFFECTS

FMEA documents often contain short text with domain-specific language and abbreviations [9]. In these documents, conceptually related nodes are often represented as distinct nodes. To address these challenges, we are exploring and improving methods for integrating text information into knowledge graph embedding. In particular, we test the effectiveness of BERT-ConvE in integrating text information into the embeddings for the FMEA knowledge graph case. The proposed method for FMEA knowledge graph completion is depicted in Figure 2. This method leverages the use of high-quality node text embeddings, including pre-trained language models, in-domain fine-tuning approaches, and context-aware node embeddings, to improve the accuracy and completeness of knowledge graph embeddings.

The implementation of BERT-ConvE is based on Dettmers et al. ConvE model [31], but there are some key differences. Specifically, inverse relation types are added to the training set, and training examples are generated by randomly sampling nodes and relation types from the known knowledge graph. The resulting model output is a vector of dimensions 1xK_N, where K_N is the number of known nodes in the graph. To generate the expected output vector for each training example, the known object nodes are retrieved from the training set and encoded in a one-hot vector format, where the values at the index of the known targets are set to 1, and the rest of the values are set to 0. A 10% label smoothing factor is applied to the expected output vector, and binary cross-entropy loss is used to calculate the loss based on the model prediction and the expected output vector.

The main difference between ConvE and BERT-ConvE lies in the initialization of the node embedding layer. In the ConvE model, the node embedding layer is initialized randomly and adapted during training. However, in BERT-ConvE, the node embedding layer is initialized with the text embedding of the nodes, leveraging the power of BERT to embed the text information into the node embeddings. Importantly, the node embedding layer is not trained during the training of BERT-ConvE.

Also, in this research, we explore the impact of the quality of node text embeddings on knowledge graph embeddings. Specifically, we compare the effectiveness of various approaches, including the selection of pre-trained language models, in-domain fine-tuning of language models, and the use of context-aware node embeddings.

When selecting the pre-trained language model, we considered several criteria, such as the training data set used, the number of unknown words in the new data set, and the number of tokens per unknown word, as suggested by Tosone et al. [35].

The impact of in-domain fine-tuning of language models on downstream tasks, such as knowledge graph embedding methods for completing the FMEA knowledge graph, is investigated. Specifically, we build up on Tosone's work [35], who conducted a series of experiments by fine-tuning different language models on a data set from semiconductor manufacturing using a masked language modelling approach with both uniform masking (UM) and point-wise mutual information (PMI) masking objectives [36]. The PMI masking objective maximizes the point-wise mutual information while selecting the span of tokens to mask, while the UM masking objective masks tokens randomly with uniform masking probability. We also explore the relationship between the masking objective and the relevance of the language used in the knowledge graph node text to the language used in the data set, on which the language models were initially pre-trained.

Finally, in this research, we also investigate the effectiveness of context-aware node embeddings on the performance of the method. To achieve this, we follow the approach of Liu et al. [13] and retrieve the triples, in which the node is a subject or object node. Next, these triples are transformed into text sequences and feed to a language model to extract context-aware embeddings. As each node appears in multiple contexts, the context-aware embeddings of the node is averaged over these multiple contexts.

Empirical analysis is used to compare different language models and identify the best-performing embeddings for completing the FMEA knowledge graph. The models are evaluated using the mean reciprocal rank (MRR) as the score function, which is computed in a filtered setting. This means that candidates from known triples are not ranked together with their inverse triples. MRR measures the rank of the correct answer in the list of all possible answers, and it is particularly useful when working with large knowledge graphs. By computing MRR in a filtered setting, the model ability to predict missing causes and effects is evaluated more accurately. Higher MRR indicates improved model performance. However, it is important to note that there are limitations to the automatic metrics used to evaluate the performance of knowledge graph embedding models, as mentioned by Malaviya et al. [12]. As such, In addition to the automatic metrics, human evaluations have been conducted to provide a more comprehensive evaluation of the models.

IV. EXPERIMENTS AND RESULTS

For the purpose of this research, a total of 1,059 risk assessment documents exemplified by process FMEA documents from semiconductor manufacturing industry are collected. In total, these documents cover 438 scopes and results in a data set of 492,507 FMEA rows. However, after removing duplicate content, a data set of 188,227 FMEA rows remains. The data set also includes 28,685 distinct process step types. The data set is compared to common-sense knowledge bases benchmark data sets from three perspectives.

The first perspective focuses on the structure information present in the FMEA documents. Specifically, the proposed transformation of the information contained in the FMEA documents to a knowledge graph form is employed to facilitate this assessment. This involves representing the FMEA data set as a knowledge graph and comparing its structure to the benchmark knowledge graph. Various metrics such as density, number of connected components, number of nodes, and number of relation types are used to assess the similarities and differences between the FMEA data set as a knowledge graph and common-sense knowledge bases benchmark data sets.

Next we study the impact of context information, such as process step type and scope, on the FMEA knowledge graph in terms of graph density. It should be noted that incorporating context information in the FMEA knowledge graph significantly impacts its density. Here, the FMEA knowledge graph includes both process step type and scope as context information and is much less dense compared to common-sense knowledge graphs such as ConceptNet 100k and Atomic. Specifically, the FMEA knowledge graph density is almost nine times lower than that of ConceptNet 100k and five times lower than that of Atomic. Excluding context information represented by scope and process step type from the FMEA knowledge graph leads to an increase in density values. Specifically, when only the process step

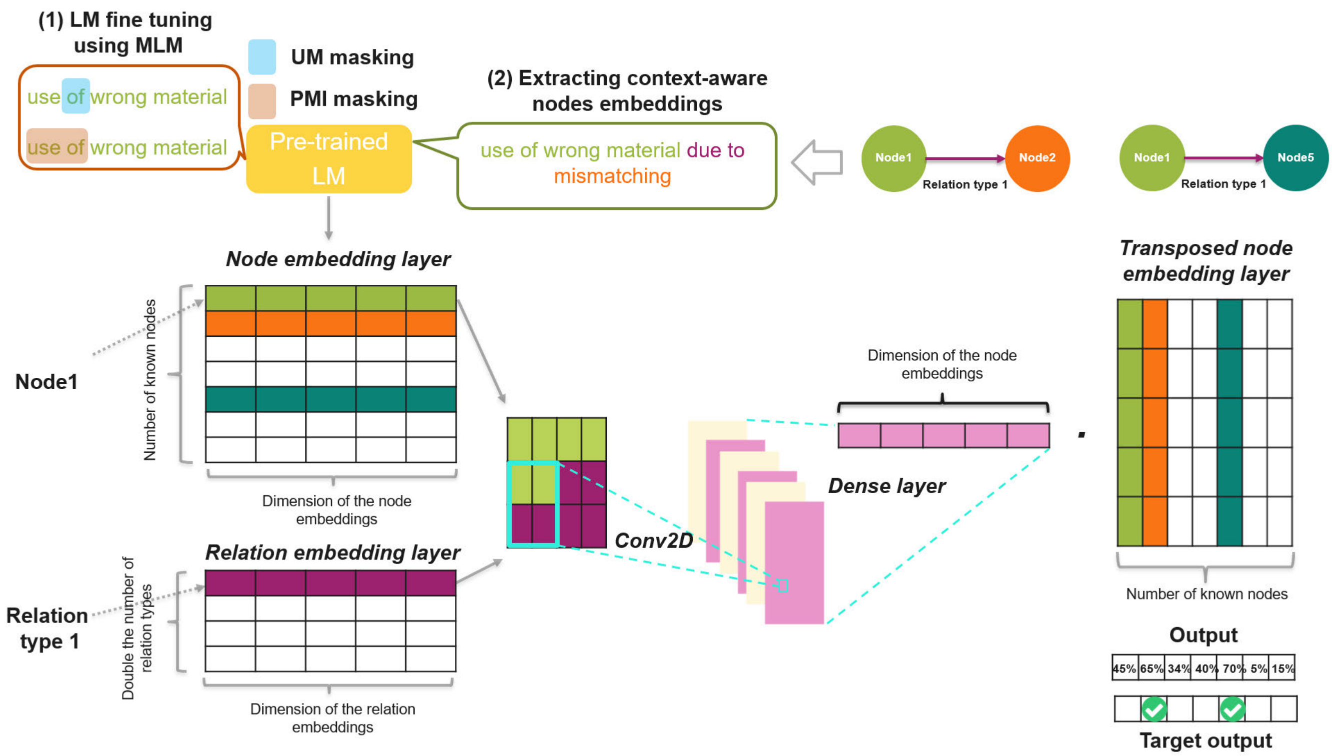

FIGURE 2. Training pipeline overview: Triple-BERT-ConvE. FMEA documents are converted into a knowledge graph. This graph undergoes a series of transformations using the Triple-BERT-ConvE workflow: (1) Fine-tuning S2ORC-SciBERT/BERT language models using masked language modeling with UM/PMI masking objectives on nodes texts. (2) Converting the knowledge graph into text sequences and extracting context-aware embeddings of the nodes.Training examples are generated by selecting a node and a relation type from the graph, retrieving its first neighbors based on the relation type, and constructing a target output vector. All values in the target vector are set to 0 except for those corresponding to neighboring nodes, which are set to 1. Binary cross-entropy loss is employed to train the ConvE model layers, comprising relation embedding, Conv2D, and dense layers. At the same time, the node embedding layer is populated with context-aware embeddings based on their graph neighborhood, rather than being trained directly.

type is included, the density value increases to more than double its original value. When only scope is included, the density value increases to three times its original value. Finally, when all context information is are excluded, the density value increases even further to more than eleven times its original value. A summary of the analysis results is presented in Table 2. The results show significant differences between the FMEA knowledge graph and the common-sense knowledge graphs such as Atomic and ConceptNet 100k. The table provides a comprehensive overview of the structural characteristics of the different knowledge graphs used in this analysis. Based on the analysis of knowledge graph density, it has been found that the FMEA knowledge graph without context information (i.e. scope and process step type) is the most comparable to the ConceptNet 100k knowledge graph in terms of density. This suggests that the removal of context information leads to a more connected structure in the FMEA knowledge graph, which is similar to the structure of the ConceptNet 100k graph.

The second perspective of this comparison is focused on the content information that is present in the FMEA documents. This involves assessing the semantics relations in the FMEA knowledge graph and the text attributes of the failure mode, effect, and cause columns, and comparing them with benchmark data sets represented by Atomic and ConceptNet. To assess the text attributes, we compare the text present in the three data sets with text found in the general domain. To assist in this assessment, we utilize language models trained on the general domain. Specifically, we utilize the tokenizers of these language models to identify unknown words, which are typically split into multiple tokens. The rationale behind this is that the higher the number of unknown words for a language model, the less aligned its original data set (which it was trained on) is with the data that is being applied to it. In the first phase, we primarily focus on evaluating the well-known BERT language model [37]. However, based on the results presented by Tosone [35], we also explore the use of other language models such as S2ORC-SciBERT [38], which have a higher degree of overlap with the text in the FMEA data set [35]. As such, the number of unknown words, along with the ratio of unknown words per node and the average number of splits for unknown words in the three data sets are reported in Table 3. We can see that the FMEA data set contains a higher ratio of unknown words per node compared to common-sense benchmark data sets. However, by utilizing S2ORC-SciBERT for the FMEA case, we observe a decrease in the ratio of unknown words per node.

In the third and final perspective of this comparison, we examine the semantics relations modelled by the three

TABLE 2. Comparison of FMEA knowledge graphs to benchmark common-sense knowledge graphs based on their node and edge counts, number of relation types, density, and connected components. The table includes benchmarks such as Atomic and ConceptNet, as well as variations of FMEA knowledge graphs with different context information included or excluded. The results show the varying sizes and structures of the different knowledge graphs. FMEA knowledge graph without context information comparable is the most to the ConceptNet 100k knowledge graph in terms of density.

Data set	Nodes	Edges	Relation Types	Density	Connected Components
Atomic	256,609	610,536	9	9.27×10^{-6}	256,609
ConceptNet	78,093	100,000	34	16.40×10^{-6}	75,083
FMEA including process step type and scope as context information	413,827	32,1837	2	1.88×10^{-6}	413,815
FMEA including only process step type as context information	208,151	171,300	2	3.95×10^{-6}	208,126
FMEA including only scope as context information	171,963	166,960	2	5.65×10^{-6}	171,814
FMEA excluding context information	51,358	56,932	2	21.58×10^{-6}	51,215

TABLE 3. Text statistics analysis for three data sets (ConceptNet, Atomic, and FMEA) using two language models (BERT and S2ORC-SciBERT). The table reports the number of unknown words, the ratio of unknown words per node, and the average number of splits per unknown word. Higher numbers of unknown words indicate less alignment between the original language model data set and the data being applied to. FMEA data set has a higher ratio of unknown words per node compared to common-sense benchmark data sets. By utilizing S2ORC-SciBERT for the FMEA case, a decrease in the ratio of unknown words per node is observe.

Data set	Language Model	Unknown Words	The Ratio of Unknown Words Per Node	Average Splits Per Unknown Word
ConceptNet	BERT	12,657	14%	2.77
ConceptNet	S2ORC-SciBERT	15,484	23%	2.66
Atomic	BERT	14,513	6%	2.66
Atomic	S2ORC-SciBERT	18,121	14%	2.50
FMEA	BERT	23,454	32%	3.75
FMEA	S2ORC-SciBERT	23,123	27%	3.56

data sets. Our approach focuses on causal relationships in FMEA documents and proposes two types of relations. ConceptNet and Atomic include relation types similar to the first type found in the FMEA documents (i.e *Causes*), which are the *Causes* type in ConceptNet and *Effects* and *Reacts* in Atomic. These benchmarks also contain more detailed relations of the same type, such as relations of the type *CausesDesire* in ConceptNet and relations of type *Intents* in Atomic. Furthermore, ConceptNet and Atomic include relation types that share some semantic similarity with the second type of relations found in the FMEA documents (i.e *Due_to*), which are *HasPrerequisite* relations in ConceptNet and *Needs* in Atomic. However, ConceptNet and Atomic also include non-causal relation types, such as *RelatedTo* and *LocatedNear* relations in ConceptNet and *Attributes* in Atomic. Finally, it is worth noting that neither ConceptNet nor Atomic contain information about the context in separate attributes, such as the scope and process step type, as seen in the FMEA documents. Hence, various techniques for graph embeddings to complete the knowledge graph extracted from FMEA documents excluding context information are evaluated.

To assess the effectiveness of common-sense knowledge completion techniques in improving the comprehensibility of existing FMEA documents, we propose utilising knowledge graph embedding for knowledge graph completion. As such, a total of 11 models are evaluated on the FMEA data set, and their performance is used to compare different criteria when selecting the most suitable approach for FMEA knowledge graph completion.

The first criterion is to evaluate the effectiveness of models that incorporate the FMEA textual attribute compared to models that only incorporate the structure of the FMEA data set. ConvE model have been chosen as baseline representing models that only incorporate the structure of the FMEA data set. For the former, It is important to choose an appropriate language model for the textual attribute, as this can affect the quality of the embeddings and the overall performance of the method. Therefore, two language models, BERT and S2ORC-SciBERT, are tested resulting in BERT-ConvE and S2ORC-SciBERT-ConvE models.

The second criterion is to evaluate the effectiveness of models that incorporate in-domain fine-tuned language models. Here, we make use of BERT and S2ORC-SciBERT, which are fine-tuned by Tosone [35] using Masked Language Modelling on semiconductor manufacturing related data set, which includes the same FMEA data set. Each FMEA cell text is considered as an independent text snippet and used for fine-tuning the language models. Specifically, this study compares the effectiveness of two different Masked Language Modelling techniques on two language models, BERT and S2ORC-SciBERT. One technique is based on the standard masking objective using uniformed masking (UM), while the other technique leverages point-wise mutual information to select masked tokens (PMI). The comparison is based on the resulting models, which include BERT-UM-ConvE, BERT-PMI-ConvE, S2ORC-SciBERT-UM-ConvE, and S2ORC-SciBERT-PMI-ConvE.

The third criterion is to examine the effectiveness of incorporating the neighbourhood of a specific node in graph embeddings, as it has been shown to improve the quality of embeddings in common-sense knowledge graphs [27]. This is evaluated by incorporating the neighbourhood information of the FMEA data set nodes into the models resulting in Triples-BERT-PMI-ConvE and Triples-S2ORC-SciBERT-PMI-ConvE.

Moreover, we evaluate the effect of increasing the number of trainable parameters of the graph embedding models by increasing the size of ConvE kernel and the number of channels resulting in Triples-BERT-PMI-ConvE-Large and Triples-S2ORC-SciBERT-PMI-ConvE-Large.

All the models developed for these experiments are trained for 1,000 epochs and evaluated using train-test splits. Specifically, the average MRR values over the last 200 epochs are reported. To ensure a fair comparison, we evaluated relations from known node to known node and from unknown node to known node separately, as ConvE initialises unknown nodes randomly. The experimental results are presented in Table 4. The results demonstrate that all models that incorporate the textual attribute of FMEA outperform models that only consider the structure of FMEA data. Surprisingly, models that use S2ORC-SciBERT without fine-tuning perform worse than those that use BERT without fine-tuning. Also, fine-tuning the language model using UM reduces the performance of models that use BERT but improves the performance of models that use S2ORC-SciBERT. As anticipated, fine-tuning the language model using PMI improves the performance of both models. Furthermore, extracting context-aware embeddings improves the performance of both models, and increasing the number of trainable parameters enhances the performance of both models.

Finally, a panel of domain experts conducted a manual evaluation of the outcomes. The experts are provided with examples from the testing set, as well as the top 10 filtered predictions recommended by Triples-S2ORC-SciBERT-PMI-ConvE. They are asked to identify which of the top 10 predictions are plausible based on their extensive domain knowledge. Experts are also given the option to mark ambiguous examples. 260 examples are validated in accordance with these guidelines. Based on the feedback received by the experts, 8% of the annotated data set is reported to contained ambiguous information and required disambiguation using context information such as the scope and process step type. Interestingly, according to the experts' assessment of the remaining examples, 70% of the top 10 predictions are plausible answers. Namely, in 99% of cases, at least one possible answer is present in the top 10 filtered predictions recommended by Triples-S2ORC-SciBERT-PMI-ConvE. The 99% figure is significantly higher than the hit@10 of 26% figure, which can be acquired by comparing if the target of an example is part of the top 10 predictions.

V. DISCUSSIONS

The results of this study demonstrate that there are several similarities between the knowledge documented during FMEA and common-sense knowledge bases. Firstly, both FMEA and common-sense knowledge bases aim to represent knowledge about the world in a structured format. FMEA captures knowledge about failure modes and their causes and effects in a specific domain within a specific context (e.g. scope and process step type), while common-sense knowledge bases capture knowledge about the everyday world, including common objects, events, and relationships between them. Secondly, both FMEA and common-sense knowledge bases rely on the notion of causality to represent knowledge. In FMEA, causal relationships are used to capture the relationships between failure modes and their causes and effects. In common-sense knowledge bases, causal relationships are used to capture the relationships between events and their causes and effects.

However, there are several differences between the knowledge documented during FMEA and common-sense knowledge bases.

First, FMEA data is highly specialised and domain-specific, whereas common-sense knowledge data sets are more general and applicable to many different domains. As such, the language used in the FMEA data set is domain specific language, which differ from the language used in the general domain. This is highlighted in Table 3 where the number of unknown words to a language model, which is trained on data sets from the general domain (i.e. BERT) contained in the FMEA data set is significantly larger compared to number of unknown words to the same language model contained in common-sense knowledge bases. Concretely, in the FMEA data set case, the number of unknown words to BERT is 1.85 and 1.62 times the number of unknown words to BERT in the cases of ConceptNet and Atomic respectively. Also, this is translated to a higher ratio of unknown words per node in the FMEA case, which is 2.3 and 5.3 times the ratio of unknown words per node in ConceptNet and Atomic respectively.

Second, the proposed approach in modelling FMEA knowledge is primarily aimed at enhancing the comprehensibility of FMEA by identifying the causes and effects of a given failure mode. Therefore, the proposed approach is limited to modelling causal relationships only, in contrast to other common-sense knowledge bases, which also include non-causal relationship types. However, it is possible to extend the proposed approach by incorporating other types of relationships, such as the relationship between the context information and the failure mode or by integrating FMEA data with other data sources to improve the coverage. This can be achieved by modifying the approach based on its intended use.

TABLE 4. Performance comparison of FMEA knowledge graph completion using different model architectures and configurations. The study evaluates the impact of feature extraction, language model fine-tuning, and context-aware embeddings on the achieved mean reciprocal rank (MRR) metric. Text-aware knowledge graph embeddings based models that use PMI masking objective for in-domain fine-tuning of the language models and context-aware embeddings based on the node neighbourhood achieve the highest MRRs. Increasing the number of model trainable parameters also improves the achieved MRR. Textual attribute of FMEA improves model performance, while different language models and fine-tuning have varying effects. Context-aware embeddings and increased trainable parameters enhance model performance.

Model	Overall avg MRR	Known to known avg MRR	Unknown to known avg MRR
ConvE	-	6.4%	-
BERT-ConvE	13.9%	13.6%	16.3%
S2ORC-SciBERT-ConvE	12.8%	12.7%	14.4%
BERT-UM-ConvE	13.3%	13.2%	14.9%
S2ORC-SciBERT-UM-ConvE	14.0%	13.9%	15.2%
BERT-PMI-ConvE	15.2%	15.1%	15.8%
S2ORC-SciBERT-PMI-ConvE	14.7%	14.6%	16.3%
Triples-BERT-PMI-ConvE	15.9%	15.8%	17.7%
Triples-S2ORC-SciBERT-PMI-ConvE	16.6%	16.5%	18.2%
Triples-BERT-PMI-ConvE-Large	18.1%	18.0%	19.6%
Triples-S2ORC-SciBERT-PMI-ConvE-Large	20.3%	20.0%	24.0%

Third, contextual information such as the scope and process step type is explicitly expressed in FMEA documents, whereas in some common-sense knowledge bases like Atomic, it is implicitly included in the relation type and in text attributes, and in others like ConceptNet, it is absent. In the FMEA case, the inclusion or exclusion of this context greatly affects the structure of the resulting knowledge graph. The FMEA knowledge graph, when constructed to include process step type and scope as context information, is much more sparsely connected compared to benchmark graphs. This may be because the nature of FMEA data and its specific context may not have as many connections between different nodes as more general common-sense knowledge graphs. Excluding context information may result in a higher density of the graph due to fewer nodes and relations, but this could also lead to important information loss and decreased analysis accuracy. Therefore, further research is necessary to develop methods that explicitly consider context information as separate input.

Despite their differences, the results of this study demonstrate the effectiveness of common-sense knowledge completion techniques applied on FMEA documents in enhancing the comprehensibility of information in existing FMEA documents. Specifically, incorporating text-aware embeddings into knowledge graph embedding models has shown significant improvement in the performance of the BERT-ConvE model, compared to the ConvE model that uses structure-only embeddings, as shown in Table 4.

While the FMEA domain-specific language was considered, the study also examined the impact of using different language models that are more aligned with the FMEA data set, such as S2ORC-SciBERT, on the performance of the knowledge graph embedding model. Interestingly, the study found that using a language model that shares more similar tokens with the FMEA knowledge graph text may not necessarily lead to improved model performance. In fact, the use of S2ORC-SciBERT resulted in lower performing model (S2ORC-SciBERT-ConvE) compared to models that uses language models trained on general data sets, such as BERT-ConvE. This drop in performance might be attributed to the assumption that while the tokens may be more aligned with the FMEA text, their semantics may differ significantly from the original data set that the language model is trained on, compared to the FMEA texts.

To address this challenge, the study also examined the impact of using in-domain fine-tuning techniques of the considered language models such as masked language modelling using uniformed masking (UM) on the performance of the knowledge graph embedding model. However, the effectiveness of this approach is highly dependent on the similarity of the language models tokens to the text in the FMEA documents. The experiments show that fine-tuning with UM on a language model with less similar tokens to the FMEA text (such as BERT-UM-ConvE) resulted in a decrease in performance compared to BERT-ConvE. Conversely, fine-tuning with UM on a language model with more similar tokens to the FMEA text (such as S2ORC-SciBERT-UM-ConvE) leads to improved performance compared to S2ORC-SciBERT-ConvE. In the case of BERT-UM-ConvE, the drop in performance is attributed to be in line with the authors findings in [36] where the authors argued that the use of uniform masking in a MLM limits the language model learning capabilities by reducing its training objective to superficial local cues, resulting in an ineffective in-domain fine-tuning process and inferior performance in downstream tasks. To further assert this claim, the results show that further improvement of the model performance is achieved by fine-tuning the model using a modified masking objective (i.e. PMI). Here, both models (i.e. S2ORC-SciBERT-PMI-ConvE and BERT-PMI-ConvE) achieve performance improvements over models, which use these language models with out in-domain fine tuning regardless of the language models token similarity to the text in the FMEA.

Due to the similarity between the knowledge graph extracted from FMEA documents and common-sense knowledge graph, we investigate the effectiveness of the approach proposed by Liu et al. [13], which involves extracting context-aware embeddings of nodes based on their neighbourhoods in the graph. The results demonstrated that models incorporating context-aware embeddings (i.e. Triples-S2ORC-SciBERT-PMI-ConvE and Triples-BERT-PMI-ConvE) outperformed models that only considered node embeddings (i.e. S2ORC-SciBERT-PMI-ConvE and BERT-PMI-ConvE). These findings are consistent with Liu et al.'s argument that by exploiting the context dependency of transformer-based language models, the quality of node embeddings can be improved, resulting in an overall improvement in model performance.

The results show that increasing the number of trainable parameters in a model can improve its performance by allowing the model to capture more complex patterns and relationships in the data. However, larger models require more computational resources to train and evaluate, which can be a limiting factor for practical applications. This is particularly relevant for applications like FMEA case, where the model needs to be retrained on a regular basis to incorporate new data and to maintain its accuracy.

It is worth noting that the reported filtered MRR results for all models show a higher unknown-to-known MRR than the known-to-known MRR. However, it is important to highlight that even if a node describes the same semantic meaning as an already known node, it can be expressed in a different way and thus still be considered unknown. This limits the effectiveness of the filter for the unknown-to-known scenario because the head node is unknown while the target node is known, making it easier to predict the target node based on captured semantics. This explains why the unknown-to-known MRR is generally higher. In contrast, for the known-to-known scenario, we assume that all easy targets are already known and filtered out, resulting in a lower MRR. Therefore, the model needs to rely more on its ability to accurately predict the next node, which can be challenging due to the complex and varied nature of information documented in the FMEA. Despite this, the known-to-known scenario is still an important measure of the model effectiveness since it is more representative of real-world use cases where both the head and target nodes are known, and the goal is to predict a hypothesis about possible relations between them.

Finally, as FMEA knowledge graphs tend to contain nodes with non-standard text description and arguably incomplete, the top of the ranking could often includes several false negative target entities, which are not captured by automated evaluations. As such, the human evaluation indicates significantly better performance compared to the automatic evaluation, revealing the limitations of automated evaluations for increasing the comprehensibility of FMEA documents through knowledge graph embedding models. The substantial discrepancy between human and machine assessments can be attributed to the incompleteness and non-standard notations of the FMEA graph.

VI. CONCLUSION

FMEA documents suffer from limitations related to the comprehensibility of the brainstorming approaches used to identify failure modes, causes, and effects. This study proposes leveraging recent advancements in common-sense knowledge completion techniques to address this issue. These techniques involve using language models pre-trained on large-scale data sets, modelling the knowledge contained in FMEA documents as knowledge graphs, and using graph embedding algorithms to complete the knowledge contained in FMEA document and enrich future FMEAs.

The study is the first to apply techniques from common-sense completion to domain-specific knowledge extracted from risk assessment documents and provides an empirical evaluation on actual FMEA documents in a case study from semiconductor manufacturing. The study findings reveal that there are several noteworthy similarities between the knowledge documented in FMEA and common-sense knowledge bases. Both FMEA and common-sense knowledge bases aim to structure knowledge about the world. While FMEA documents information on failure modes and their causes and effects within a specific domain and context, common-sense knowledge bases encompass knowledge about everyday objects, events, and their relationships. Furthermore, both rely on causality to represent knowledge. FMEA utilizes causal relationships to capture the connections between failure modes and their causes and effects, while common-sense knowledge bases utilize causal relationships to represent the links between events and their causes and effects.

However, in contrast to the similarities between FMEA and common-sense knowledge bases, there are also some differences. Firstly, FMEA data is domain-specific and uses specialised language, making it different from the general language used in common-sense knowledge bases. Secondly, the proposed approach only models causal relationships, which is different from other knowledge bases that also include non-causal relationships. However, it is possible to modify the approach to include other relationships or integrate it with other data sources. Thirdly, contextual information is explicitly expressed in FMEA documents, but it is implicit or absent in some common-sense knowledge bases, which affects the resulting knowledge graph structure and density. Therefore, further research is needed to develop methods that explicitly consider context information as a separate input.

Overall, the study highlights the suitability of incorporating common-sense knowledge completion techniques and domain-specific language models into knowledge graph embedding models for enhancing the comprehensibility of information in FMEA documents. The use of text-aware embeddings, in-domain fine-tuning with modified masking objectives, and context-aware embeddings based on the nodes neighbourhood are effective approaches for improving model performance also for the FMEA case. The findings also suggest that using language models with more similar tokens to the FMEA text may not necessarily lead to better performance. Furthermore, the study revealed that FMEA knowledge graphs often contain nodes with non-standard text descriptions that may be incomplete. This can result in the top-ranking including several false negative target entities, which are not captured by automated evaluations. Therefore, human evaluation may be a better approach to increase the comprehensibility of FMEA documents through knowledge graph embedding models. Based on the human evaluation 70% of the top ten prediction of the model are plausible answers, which illustrates the effectiveness of the proposed techniques for improving the comprehensibility of the knowledge documented in the FMEA documents.

REFERENCES

[1] D. H. Stamatis, *Failure Mode and Effect Analysis: FMEA From Theory to Execution*. Atlanta, GA, USA: Quality Press, 2003.

[2] *Failure Mode and Effects Analysis-FMEA Handbook: Design FMEA, Process FMEA, Supplement FMEA for Monitoring and System Response*, Automot. Ind. Action Group, Southfield, MI, USA, Verband der Automobilindustrie e. V., Berlin, Germany, 2019.

[3] S. Gray. (2019). *AIAG & VDA FMEA Handbook*. [Online]. Available: https://myscma.com/wp-content/uploads/2019/09/SCAC-2019-FMEA-Handbook-SG-Final-13-Sep.-2019.pdf

[4] B. Cabanes, S. Hubac, P. Le Masson, and B. Weil, ''Improving reliability engineering in product development based on design theory: The case of FMEA in the semiconductor industry,'' *Res. Eng. Des.*, vol. 32, no. 3, pp. 309–329, Jul. 2021.

[5] Z. Xu, Y. Dang, P. Munro, and Y. Wang, ''A data-driven approach for constructing the component-failure mode matrix for FMEA,'' *J. Intell. Manuf.*, vol. 31, no. 1, pp. 249–265, Jan. 2020.

[6] Z. Bluvband and P. Grabov, ''Failure analysis of FMEA,'' in *Proc. Annu. Rel. Maintainability Symp.*, Jun. 2009, pp. 344–347.

[7] B. Gladysz and A. Albers, ''How do C&C^2-models improve efficiency, comprehensibility and scope in failure analysis—An empirical study based on two live-labs,'' in *Proc. 15th Int. Design Conf.*, 2018, pp. 1127–1138.

[8] Q. Huang, G. Wu, and Z. S. Li, ''Design for reliability through text mining and optimal product verification and validation planning,'' *IEEE Trans. Rel.*, vol. 70, no. 1, pp. 231–247, Mar. 2021.

[9] H. Razouk and R. Kern, ''Improving the consistency of the failure mode effect analysis (FMEA) documents in semiconductor manufacturing,'' *Appl. Sci.*, vol. 12, no. 4, pp. 1–12, 2022. [Online]. Available: https://www.mdpi.com/2076-3417/12/4/1840

[10] M. Sap, R. Le Bras, E. Allaway, C. Bhagavatula, N. Lourie, H. Rashkin, B. Roof, N. A. Smith, and Y. Choi, ''Atomic: An atlas of machine commonsense for if-then reasoning,'' in *Proc. AAAI Conf. Artif. Intell.*, 2019, vol. 33, no. 1, pp. 3027–3035.

[11] R. Speer, J. Chin, and C. Havasi, ''ConceptNet 5.5: An open multilingual graph of general knowledge,'' in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 4444–4451.

[12] C. Malaviya, C. Bhagavatula, A. Bosselut, and Y. Choi, ''Commonsense knowledge base completion with structural and semantic context,'' in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 3, pp. 2925–2933.

[13] X. Liu, H. Hussain, H. Razouk, and R. Kern, ''Effective use of BERT in graph embeddings for sparse knowledge graph completion,'' in *Proc. 37th ACM/SIGAPP Symp. Appl. Comput.*, Apr. 2022, pp. 799–802.

[14] X. Zhao and Y. Zhu, ''Application research of ontology-enabled process FMEA knowledge management method,'' *Int. J. Intell. Syst. Appl.*, vol. 4, no. 3, pp. 34–40, Apr. 2012.

[15] A. T. James, O. P. Gandhi, and S. G. Deshmukh, ''Knowledge management of automobile system failures through development of failure knowledge ontology from maintenance experience,'' *J. Adv. Manage. Res.*, vol. 14, no. 4, pp. 425–445, Oct. 2017.

[16] W. L. Mikos, J. C. E. Ferreira, P. E. A. Botura, and L. S. Freitas, ''A system for distributed sharing and reuse of design and manufacturing knowledge in the PFMEA domain using a description logics-based ontology,'' *J. Manuf. Syst.*, vol. 30, no. 3, pp. 133–143, Aug. 2011.

[17] Z. Wu, W. Liu, and W. Nie, ''Literature review and prospect of the development and application of FMEA in manufacturing industry,'' *Int. J. Adv. Manuf. Technol.*, vol. 112, nos. 5–6, pp. 1409–1436, Jan. 2021.

[18] X. Zhang, Y. Li, Y. Ran, and G. Zhang, ''A hybrid multilevel FTA-FMEA method for a flexible manufacturing cell based on meta-action and TOPSIS,'' *IEEE Access*, vol. 7, pp. 110306–110315, 2019.

[19] L. Liu, D. Fan, Z. Wang, D. Yang, J. Cui, X. Ma, and Y. Ren, ''Enhanced GO methodology to support failure mode, effects and criticality analysis,'' *J. Intell. Manuf.*, vol. 30, no. 3, pp. 1451–1468, Mar. 2019.

[20] K.-W. Noh, H.-B. Jun, J.-H. Lee, G.-B. Lee, and H.-W. Suh, ''Module-based failure propagation (MFP) model for FMEA,'' *Int. J. Adv. Manuf. Technol.*, vol. 55, nos. 5–8, pp. 581–600, Jul. 2011.

[21] S. Sierla, I. Tumer, N. Papakonstantinou, K. Koskinen, and D. Jensen, ''Early integration of safety to the mechatronic system design process by the functional failure identification and propagation framework,'' *Mechatronics*, vol. 22, no. 2, pp. 137–151, Mar. 2012.

[22] L. Chen and R. Nayak, ''A case study of failure mode analysis with text mining methods,'' in *Proc. 2nd Int. Workshop Integrating Artif. Intell. Data Mining*, 2007, pp. 49–60.

[23] D. G. Rajpathak, ''An ontology based text mining system for knowledge discovery from the diagnosis data in the automotive domain,'' *Comput. Ind.*, vol. 64, no. 5, pp. 565–580, Jun. 2013.

[24] V. H. Jacobo, A. Ortiz, Y. Cerrud, and R. Schouwenaars, ''Hybrid expert system for the failure analysis of mechanical elements,'' *Eng. Failure Anal.*, vol. 14, no. 8, pp. 1435–1443, Dec. 2007.

[25] H. Wang and H. Wang, ''A hybrid expert system for equipment failure analysis,'' *Expert Syst. Appl.*, vol. 28, no. 4, pp. 615–622, May 2005.

[26] D. Law, R. Gruss, and A. S. Abrahams, ''Automated defect discovery for dishwasher appliances from online consumer reviews,'' *Expert Syst. Appl.*, vol. 67, pp. 84–94, Jan. 2017.

[27] S. C. Liu and S. Y. Liu, ''An efficient expert system for machine fault diagnosis,'' *Int. J. Adv. Manuf. Technol.*, vol. 21, no. 9, pp. 691–698, Jun. 2003.

[28] B. Li, T. Han, and F. Kang, ''Fault diagnosis expert system of semiconductor manufacturing equipment using a Bayesian network,'' *Int. J. Comput. Integr. Manuf.*, vol. 26, no. 12, pp. 1161–1171, Dec. 2013.

[29] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, ''Translating embeddings for modeling multi-relational data,'' in *Proc. Neural Inf. Process. Syst. (NIPS)*, 2013, pp. 1–9.

[30] B. Yang, W.-T. Yih, X. He, J. Gao, and L. Deng, ''Embedding entities and relations for learning and inference in knowledge bases,'' 2014, *arXiv:1412.6575*.

[31] T. Dettmers, P. Minervini, P. Stenetorp, and S. Riedel, ''Convolutional 2D knowledge graph embeddings,'' in *Proc. AAAI Conf. Artif. Intell.*, 2018, vol. 32, no. 1, pp. 1811–1818.

[32] C. Shang, Y. Tang, J. Huang, J. Bi, X. He, and B. Zhou, ''End-to-end structure-aware convolutional networks for knowledge base completion,'' in *Proc. AAAI*, vol. 33, 2019, pp. 3060–3067.

[33] L. Yao, C. Mao, and Y. Luo, ''KG-BERT: BERT for knowledge graph completion,'' 2019, *arXiv:1909.03193*.

[34] Z. Jin, J. von Kügelgen, J. Ni, T. Vaidhya, A. Kaushal, M. Sachan, and B. Schoelkopf, ''Causal direction of data collection matters: Implications of causal and anticausal learning for NLP,'' in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2021, pp. 9499–9513.

[35] D. Tosone, ''Improving FMEA consistency in semiconductor manufacturing through text classification,'' M.S. thesis, Dept. Math., Comput. Sci. Phys., Univ. Udine, Dept. Inform. Syst., Alpen-Adria-Universität Klagenfurt, Klagenfurt, Austria, 2022.

[36] Y. Levine, B. Lenz, O. Lieber, O. Abend, K. Leyton-Brown, M. Tennenholtz, and Y. Shoham, ''PMI-masking: Principled masking of correlated spans,'' 2020, *arXiv:2010.01825*.

[37] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, ''BERT: Pre-training of deep bidirectional transformers for language understanding,'' 2018, *arXiv:1810.04805*.

[38] K. Lo, L. Lu Wang, M. Neumann, R. Kinney, and D. S. Weld, ''S2ORC: The semantic scholar open research corpus,'' 2019, *arXiv:1911.02782*.

HOUSSAM RAZOUK received the B.Sc. degree in computer engineering from Damascus University, Syria, and the M.Sc. degree in systems design from the Carinthia University of Applied Sciences, Austria. He is currently pursuing the Ph.D. degree in computer sciences with the Institute of Interactive Systems and Data Science, Graz University of Technology, Austria.

In addition to his academic pursuits, he is a Senior Specialist in quality with Infineon Technologies Austria. His research interests include machine learning, natural language processing, and causal data science. Notably, he has researched urban blight analysis and failure mode effect analysis, utilizing advanced machine learning techniques to gain valuable insights and propose effective solutions.

XING LAN LIU received the B.S. degree in mechanical engineering from Tsinghua University, China, in 2002, the M.S. degree in nano-science from the University of Groningen, The Netherlands, in 2006, and the Ph.D. degree in applied physics from the Delft University of Technology, The Netherlands, in 2010. After having worked on several industrial jobs in data analytics, she moved to Know-Center GmbH, Austria, in 2018. She is currently a Senior Researcher. Her main research interest includes natural language processing (NLP), specifically in knowledge graphs and language models.

ROMAN KERN received the Ph.D. degree from the Graz University of Technology.

He is currently a Computer Scientist and an Assistant Professor with the Institute for Interactive Systems and Data Science, Graz University of Technology. He is also the Chief Scientific Officer with the Know-Center Research Centre (Competence Centre for Trustworthy AI). His research interests include natural language processing and machine learning, with a focus on causal data science. He applies these methods to achieve trustworthy AI in fields like digital libraries, intelligent transportation systems, and smart production.