**Constructing a Knowledge Graph for Automated HAZOP Analysis**

Zhiyuan Li,a Jinsong Zhao a,b\*

aState Key Laboratory of Chemical Engineering, Department of Chemical Engineering, Tsinghua University, Beijing 100084, China

bBeijing Key Laboratory of Industrial Big Data System and Application, Tsinghua University, Beijing 100084, China

jinsongzhao@tsinghua.edu.cn

Abstract

Hazard and Operability (HAZOP) analysis is widely acknowledged as a prominent approach for conducting process hazard analysis. In order to enhance the efficiency of the HAZOP analysis process, several approaches have been proposed in the past few decades to automate the HAZOP analysis. Nevertheless, only a limited number of them have been successfully incorporated into practical chemical processes. This paper presents a novel automated approach for conducting HAZOP analysis, utilizing a comprehensive knowledge graph. The knowledge graph integrates relevant data on processes, equipment, streams, chemicals, and the outcomes of HAZOP analysis. A reasoning model named COMPGCN is introduced to predict missing links within a knowledge graph, ultimately aiming to facilitate the automation of HAZOP analysis. In the experimental section, the deep desulfurization process was selected as a case study, and a knowledge graph consisting of 2988 entities, 72 relationships, and 7867 facts was constructed. The reasoning result over the knowledge graph shows the validity of this approach, and helps to discover the hidden relationships between possible risk causes and consequences.

**Keywords**: HAZOP analysis, Knowledge graph, Graph neural network, Deep desulfurization process

* 1. Introduction

The Hazard and Operability Study (HAZOP) is a systematic and essential approach used to analyze potential process hazards during the planning or design stage. The heavy dependence on experiential knowledge in the manual process hazard analysis leads to a lack of consistency in the quality of HAZOP reports. In recent decades, many researchers have proposed various automated HAZOP systems. The automated HAZOP system mainly consists of three key components, namely system modeling, knowledge representation, and reasoning engine. The categorization of systems can be based on their reasoning engine, which includes rule-based, model-based, case-based, and process-history-based systems (Single et al., 2019). The advanced HAZOP systems integrates models with case-based reasoning, such as PHASuite (Zhang et al., 2005), PetroHAZOP (Zhao et al., 2009), and KROSA (Daramola et al., 2013). These systems reuse the knowledge derived from previous HAZOP studies and exhibit promising potential for the analysis of hazardous scenarios. However, the progress of these systems is being hindered by the lack of integration between system modeling and knowledge representation.

In this paper, a HAZOP knowledge graph is proposed to effectively integrate system modeling with safety knowledge. Furthermore, automated HAZOP analysis is performed by reasoning over the knowledge graph. The schema of the knowledge graph, and methodology for knowledge graph construction and reasoning were introduced in Section 2. In Section 3, the deep desulfurization process was chosen as a case study. We validated the effectiveness of our reasoning method and derived new knowledge from the existing knowledge graph. The conclusions drawn from our findings are presented in Section 4.

* 1. Methodology

Figure 1 provides an overview of the process involved in developing a knowledge graph. Firstly, the extraction of triples that represent the relation between two entities from process safety information was conducted. Subsequently, an integration framework was utilized to align entities across various knowledge sources. Based on the existing graph, we further discovered unmined relationships using knowledge graph reasoning techniques. By predicting the relations between deviations and analysis results, the ultimate goal of automating HAZOP analysis could be achieved.



Figure 1. An overview of knowledge graph construction, reasoning, and application in HAZOP analysis.

* + 1. Ontology Design

To define the types of entities and relationships in the HAZOP knowledge graph, an ontology was developed in a top-down manner, as illustrated in Figure 2. By defining the ontology, we aim to enhance data quality and combine process system information and HAZOP analysis results into a unified knowledge graph. The process system consists of discrete nodes, each of which contains information about the equipment, streams, and their connections. The safety knowledge relates to structured results of the HAZOP analysis, which include parameter, deviation, reason, consequence and preventions.

* + 1. Methodology for Knowledge Graph Construction

As previously stated, the HAZOP knowledge graph serves as a representation of process system and safety knowledge. Therefore, the selection and integration of diverse knowledge sources is highly important. To acquire knowledge about process systems, the interconnections among important equipment components were derived from Process Flow Diagrams (PFDs). Subsequently, the tables related to equipment, chemicals, or streams were transformed into a graph structure. To obtain knowledge regarding safety, we utilized named entity recognition (NER) to extract crucial information about chemicals or equipment from HAZOP reports, the results of the HAZOP analysis were then imported into a graph database. Finally, we integrated the information into a comprehensive knowledge graph for HAZOP analysis. Figure 3 presents an integration framework for the knowledge graph. We started by dividing the process system into nodes and then used stream or equipment ID to align entities from various sources, such as the PFDs, material balance sheets, and Material Safety Data Sheets (MSDSs).



Figure 2. An ontology model of the HAZOP knowledge graph.



Figure 3. An integration framework for the HAZOP knowledge graph.

* + 1. Methodology for Knowledge Graph Reasoning

Knowledge graph reasoning aims to infer unknown facts based on existing facts in the graph. This method can be applied to estimate the causes and consequences of deviations in process parameters. Knowledge graph reasoning can be categorized into four distinct categories: rule-based reasoning, embedding model-based reasoning, Convolutional Neural Network (CNN)-based reasoning, and Graph Convolutional Network (GCN)-based reasoning. In contrast, GCN-based reasoning aggregates the message from neighboring nodes to obtain an updated embedding for the central node. This enables a more comprehensive description of the central node features. By combining scoring functions derived from embedding models or CNN-based models, GCN-based reasoning has shown superior performance in link prediction tasks.

In this paper, the COMPGCN model (Vashishth et al., 2015) is used to represent relations and entities as vectors and combine their representations using vector subtraction, multiplication, or circular-correlation. Besides conventional relationship types, this model also has the capability to represent reverse and self-loop relationship types.

* 1. Case study

In this section, deep desulfurization, an essential component of petroleum processing technologies, is chosen as a case to develop a knowledge graph for automated HAZOP analysis. The process of deep desulfurization is crucial for maintaining the quality of oil products. Hydrogen sulfide is also widely acknowledged as the main occupational hazard in the petrochemical industry. Moreover, the process of burning this substance results in the formation of sulfur oxides, which are the main cause of acid rain. Therefore, HAZOP analysis is essential for ensuring the reliability of the deep desulfurization process design.

* + 1. Implementation of Knowledge Graph

We have collected a total of 19 design works focused on deep desulfurization processes, each containing process flow diagrams, material balance tables, energy balance tables, equipment selection tables and HAZOP tables. Additionally, we provided MSDS for the chemicals used in the process. Then, a comprehensive knowledge graph was constructed by the integration framework. The graph contains a total of 2988 entities, 72 types of relationships, and 7867 facts, as described in Table 1. Since HAZOP reports only analyze production or storage devices with significant hazards, we selected 13 types of equipment for analysis. These include pumps, compressors, heat exchangers, reactors, distillation towers, absorbers, furnaces, gas-liquid separators, electrolytic tanks, reflux drums, buffer tanks, storage tanks, and molecular sieve dehydrators.

Table 1. A description of the HAZOP knowledge graph.

|  |  |  |
| --- | --- | --- |
| **Entities** | **Relations** | **Facts** |
| Chemical | Equipment | Stream | Parameter | Cause | Effect | Prevention |
| Name | Type |
| 36 | 133 | 13 | 165 | 37 | 909 | 700 | 995 | 72 | 7867 |

For sentences in HAZOP reports, the BERT-CRF (Bidirectional Encoder Representations from Transformers-Conditional Random Field) framework is chosen as a solution for NER to extract entities such as equipment names, chemicals, and equipment IDs. We also manually annotated 599 sentences that describe the chemical process to validate the feasibility of this method. The experimental results are shown in Table 2. The performance in identifying the entity type 'Equipment ID' is superior. The reason is that the equipment ID follows a standardized format and typically consists of no more than 6 characters. However, the identification of entities such as chemicals and equipment names can be challenging due to the use of abbreviations and aliases.

Table 2. Performance of BERT-CRF in the named entity recognition task evaluated on descriptive texts of chemical processes. F1 is defined as the harmonic mean of precision and recall.

|  |  |  |  |
| --- | --- | --- | --- |
| Entity type | Precision | Recall | F1 |
| Chemical | 0.842 | 0.865 | 0.853 |
| Equipment | 0.866 | 0.882 | 0.874 |
| Equipment ID | **0.988** | **0.977** | **0.982** |

In the case of deep desulfurization process, a total of 199 chemicals, 33 equipment ID and 684 equipment names were extracted from HAZOP reports. Figure 4 presents an example of the results. After identifying the equipment ID mentioned in the sentence, we can determine that E0302 has been identified as the equipment related to the root cause of a low feed liquid level in E0303. The included link in the figure has the potential to enhance machine understanding of the intricacies involved in human reasoning.



Figure 4. A demonstration of the effect of named entity recognition.

* + 1. Results of Knowledge Graph Reasoning

The COMPGCN model was used in this section to predict links on the HAZOP knowledge graph. To evaluate the algorithm's reliability, we partitioned the facts in the original knowledge graph into a training set and a test set, and we attempted to reconstruct the facts in the test set using the facts in the training set. To adapt to the automated HAZOP analysis task, we specifically curated a test set consisting of 715 facts that include the relations 'causeby', 'resultsin', and 'preventedby'. The training set comprised a total of 7152 facts. We compared the performance differences of the COMPGCN model by employing three different scoring functions (TransE, Distmult, ConvE) and three different approaches for combining entity and relationship representations (vector subtraction, multiplication, circular-correlation). The results are shown in Table 3, where 'MR' denotes the mean rank of the triples in prediction results, 'MRR' denotes the mean reciprocal ranking of the triples, and 'Hits@K' denotes the average percentage of the positive triples ranked in the top-K positions.

Table 3. Performance of COMPGCN in link prediction task evaluated on HAZOP knowledge graph.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Score function | Compositionoperator | MR | MRR | Hits@1 | Hits@3 | Hits@10 |
| TransE | Sub | 232.39 | 0.030 | 0.021 | 0.022 | 0.027 |
| Mult | 232.49 | 0.029 | 0.021 | 0.021 | 0.025 |
| Corr | 232.48 | 0.029 | 0.021 | 0.021 | 0.025 |
| DistMult | Sub | 49.07 | 0.479 | 0.375 | **0.547** | 0.657 |
| Mult | 49.97 | 0.446 | 0.350 | 0.499 | 0.634 |
| Corr | 53.62 | 0.455 | 0.345 | 0.523 | 0.646 |
| ConvE | Sub | 44.82 | **0.494** | **0.407** | 0.543 | 0.656 |
| Mult | **35.36** | 0.464 | 0.354 | 0.529 | **0.666** |
| Corr | 38.81 | 0.474 | 0.379 | 0.527 | 0.634 |

It can be observed that the performance of TransE's scoring function is significantly worse. This is because the TransE model is limited to modeling one-to-one relationships, while in HAZOP analysis, it is common to encounter situations where there are multiple causes for a single deviation, in which the TransE model fails to capture. The scoring functions derived from DistMult and ConvE exhibit comparable reasoning effects, and the impact of composition operators on the reasoning effect is not significant. When considering only the top-ranked prediction from the model, approximately 40% of the entities can be correctly predicted. When considering the top ten ranked predictions, it is observed that around 67% of the answer entities can be accurately predicted. We have also noticed that some expected facts have not been included in the knowledge graph. An illustration of the outcomes can be observed in Figure 5, we predicted the reason for reduced or absence of inlet flowrate in reactor R101 based on COMPGCN model. The true label is ranked second in predictions, while the first one is also a valid reason. This fact can be used to complete the original knowledge graph, thereby improving its overall comprehensiveness.



Figure 5. An illustration depicting the outcomes of knowledge graph reasoning.

* 1. Conclusions

In this paper, we proposed an automated method for constructing and reasoning with a knowledge graph for HAZOP analysis. We first defined the schema of the knowledge graph and outlined the integration framework that facilitated the connection between various knowledge sources. Additionally, a BERT-CRF framework was used to recognize named entities in HAZOP reports. Based on the initial construction of the HAZOP knowledge graph, we utilized knowledge graph reasoning technology, specifically the COMPGCN model, to automate the analysis of causes and consequences related to deviations in process parameters. Our experimental findings indicate that around 67% of the correct responses can be found among the top ten predicted results. This suggests the possibility of using a knowledge graph to implement the automated HAZOP analysis. However, it is important to note that the construction process of the HAZOP knowledge graph did not consider the instrumentation and control information. Therefore, we will further explore the automated extraction of process control information from the P&ID.

References

Single, J. I., Schmidt, J., and Denecke, J. 2019. “State of research on the automation of HAZOP studies”. Journal of Loss Prevention in the Process Industries 62 (November): 103952. https://doi.org/10.1016/j.jlp.2019.103952.

Zhao, C., Bhushan, M., and Venkatasubramanian, V. 2005. “PHASuite: An automated HAZOP analysis tool for chemical processes: Part II: Implementation and Case Study”. Process Safety and Environmental Protection 83(6): 533-548. https://doi.org/10.1205/psep.04056.

Jinsong Zhao, Lin Cui, Lihua Zhao, Tong Qiu, and Bingzhen Chen. 2009. “Learning HAZOP expert system by case-based reasoning and ontology”. Computers & Chemical Engineering 33(1): 371-378. https://doi.org/10.1016/j.compchemeng.2008.10.006.

Daramola, O., Stålhane, T., Omoronyia, I., and Sindre, G. 2013. “Using Ontologies and Machine Learning for Hazard Identification and Safety Analysis”. In Managing requirements knowledge: 117-141. https://doi.org/10.1007/978-3-642-34419-0\_6.

Vashishth, S., Sanyal, S., Nitin, V., & Talukdar, P. 2019. “Composition-based Multi-Relational Graph Convolutional Networks”. In International Conference on Learning Representations (September). https://doi.org/10.48550/arXiv.1911.03082.