Literal-Aware Knowledge Graph Embedding for Welding Quality Monitoring

Zhipeng Tan[1,2,*], Zhuoxun Zheng[1,3], Antonis Klironomos[1,4], Muhammed Gad[1],
Guohui Xiao[5], Ahmet Soylu[6], Evgeny Kharlamov[1,3,*] and Baifan Zhou[6,3,*]

[1]*Bosch Center for AI, Germany*

[2]*RWTH Aachen, Germany*

[3]*Department of Informatics, University of Oslo, Norway*

[4]*University of Mannheim, Germany*

[5]*University of Bergen, Norway*

[6]*Department of Computer Science, Oslo Metropolitan University, Norway*

Abstract

Recently there has been a series of studies in knowledge graph embedding (KGE), which attempts to learn the embeddings of the entities and relations as numerical vectors and mathematical mappings via machine learning (ML). However, there has been limited research that applies KGE for industrial problems in manufacturing. This paper investigates whether and to what extent KGE can be used for an important problem, that is quality monitoring for welding in manufacturing industry. It is an important process accounting for production of millions of cars annually. The work is in line with our research of data-driven solutions that intends to replace the traditional costly quality monitoring. The paper tackles two challenging questions simultaneously: how large the welding spot diameter is; and to which car body the welded spot belongs to. The problem setting is difficult for traditional ML because there exist a high number of car bodies that should be assigned as class labels. We formulate the problem as link prediction, and experimented popular KGE methods with literals on real industry data, with consideration of literals. This paper accompanies the full paper in in-use track and provides additional discussion on problem formulation, literal handling strategies, and included information in industrial KG construction.

1. Introduction

Background and Challenge. Research in knowledge graphs and their industrial applications has attracted increasing attention [1]. Recently there has been a series of studies in knowledge graph embedding (KGE), but there has been limited research that applies KGE for industrial problems in manufacturing. This paper investigates whether and to what extent KGE can be used for an important problem, that is quality monitoring for welding in manufacturing industry. We discuss automated welding, which plays a critical role in the automotive industry for manufacturing high-quality car bodies, with millions of cars produced annually. The welding process generates a vast amount of data, considering the number of welding machines in car production lines and the thousands of spots on each carbody (Fig. 1a). This large amount of data increases the demand of data-driven solutions, which aim to reduce and eventually replace
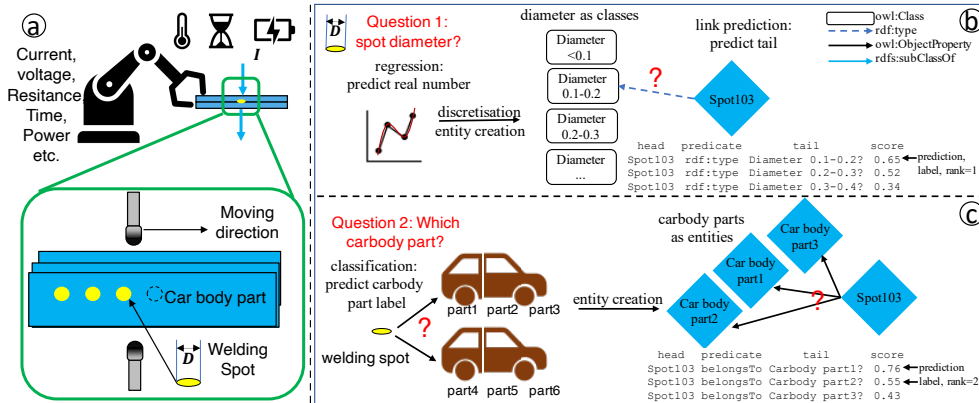
Figure 1: Two core questions in welding quality monitoring: *Question 1* (Q1), how large is the spot diameter? *Question 2* (Q2), which car body part does this spot diameter belongs to?

conventional destructive, yet extremely expensive and inefficient, methods. Addressing this challenges, two core questions need to be answer here as shown in Fig. 1b and Fig. 1c. First, spot diameter (Q1) is the key quality indicator for evaluating welding quality. Since the diameter must be neither too large nor too small. Second, the carbody part of the spot is important because it is essential to know the percentage of good spots for each car body, but the carbody-spot correspondence information does not exist in a large amount of historical data silos from past welding systems. Currently there are research applying ML and Digital Twin in industry [2], but limitedly solve this challenge especially the two core questions.

Our Approach. We investigate KGE for answering the two questions in the automotive industry with production data, and compare with representative classic ML. We first construct KGs from tabular data with special handling on literals; then, we conduct experiments and compare mainstream KGE methods such as TransE, RotatE, AttH with multilayer perceptron (MLP), and compare with variant KGE applied in [3]. This poster paper accompanies our full paper in ISWC 2023 [4]. This paper provides additional discussions on problem formulation, literal handling, and industrial KG construction, which were not present in the full paper.

2. Approach

Welding KG construction. A welding KG is constructed from tabular data collected from a German factory. We have used welding-related information, such as time of welding processes, welding machines, welding programs, and welding parameters (e.g., voltage, current, resistance). The constructions are conducted on welding spots and the car body and diameters. We transform the values of the welding data table into entities and the relationships between these entities as edges in the KG. Fig.2b shows the construction of literal entities, which are entities generated from numeric values. The literals are handled as described in the later part.

Problem formulation. Fig. 1b and Fig. 1c shows in more details the two research questions of the quality monitoring in the use case and our approach to reformulate them: given the information of the welding spot, we want to predict the carbody of the this welding spot and the diameter of the welding spot. As shown in the Fig. 1, both of the problems are reformulated. The Spot diameter prediction was a regression problem based on the welding data to predict the real values for the diameters size. Due to resolution when measuring the spot diameter,
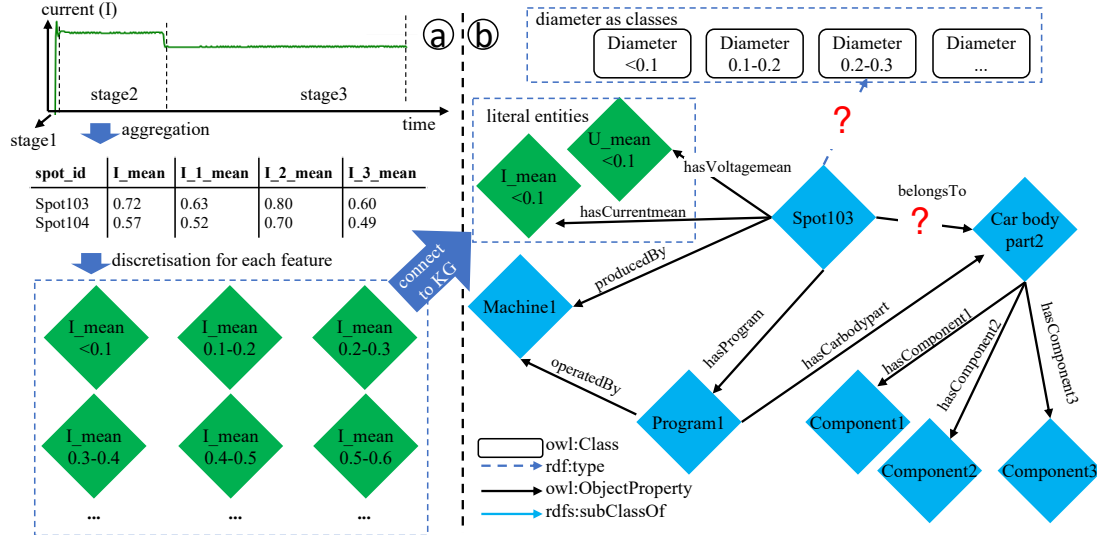
Figure 2: (a) Procedure of literal embedding (b) Partial illustration of the welding KG

we discretise the diameters into different diameter classes as classification problem. We then constructed the entities based on the diameter classes and carbody class. The links between welding spots and the diameter classes or the carbody are predicted.

Literal handling. We did the following steps for literal embeddings inspired by [5]. The numeric literals of the knowledge graphs are embedded following aggregation, discretisation of features, entity creation, and linking. As shown in the Fig. 2, in the aggregation step, the sensor measured values are aggregated into the mean values of the three stages and the overall mean values in real numbers. Then in the discretisation step, we discretise the real values into different ranges. And then we create entities based on the discretised ranges and link them.

3. Evaluation and Discussion

We discuss the experiments and provide additional discussion on the problem formulation, the literal handling and discretisation techniques, and the industrial KG construction.

Discussion on problem formulation. We discuss three promising ways: (1) classic ML with MLP; (2) classic KGE with link prediction; (3) binary triple classification inspired by [3].

MLP. In manufacturing, quality monitoring aims at predicting diameters and carbody. These questions are formulated as classification problems, where carbodies and diameters are formulated as the predicted classes. This formulation is verified

Table 1: Model performance comparison on answering Q1 and Q2. Bold: best results. Underlined: second best.

		MLP	**TransE**	**RotatE**	**AttH**
Q1	Acc(Hits@1)	<u>0.39</u>	**0.42**	0.25	0.31
	MRR	-	**0.65**	0.49	<u>0.57</u>
	nrmse	**0.05**	<u>0.06</u>	0.08	<u>0.06</u>
Q2	Acc(Hits@1)	<u>0.61</u>	**0.64**	0.52	0.53
	MRR	-	**0.77**	0.69	<u>0.70</u>
	Hits@Groupby3	-	**0.85**	0.81	<u>0.79</u>

by the MLP model. This model is proven to be most efficient and have best performance only in *nrmse* in Q1, but provides inadequate performance for Q2, see Tab. 1.

KGE. Since there can be over hundreds and thousands of carbodies, Q2 is not very suitable as a classic ML problem. We adapt the problem and reformulate it as link prediction (Fig. 2), where the correct link between a welding spot and the correct carbody should have the highest score among all the other carbodies. Similar principles also hold for diameter predictions. This formulation is verified by the KGE model. We use metrics, including Acc, MRR, and Hits@Groupby3. We introduce a new metric Hits@Groupby3, because no KGE model deliver satisfactory Acc. We thus adopt the adaptation to relax the metric, to test the accuracy based on the group of 3 carbodies. The results, see Tab. 1, show TransE delivers the best results for both Q1 and Q2 on Acc, MRR, and Hits@Groupby3.

KGE-MLP. The third possible problem formulation is inspired by [3], where the link is formulated as binary classification with the output score between 0 and 1, where a value closer to 1 indicates the link exists. The score is compared with all other potential diameters or carbodies, and the predicted link the highest score is selected as the prediction. This formulation is verified by the KGE-MLP model. This model proved to be not good on Acc, MRR, and other metrics compared with the MLP and KGE models in our welding quality monitoring use case, but is State-of-the-Art (SotA) model on other use case [3]. So we still choose this model in our paper.

Discussion on literal landling. In SotA research, there exist mainly two ways to handle literals: discretisation (KGA) [5] and literal as embedding vector (LiteralE) [6]. According to the experiments in [5], the discretisation with bins methods yield SotA results with large improvements on the traditional KGE. We also consider [6] not suitable, because it requires fixed embedding size for fixed number of literals, while in real application the number of literals can vary a lot. This paper chooses the discretisation with bins as in [5] method to encode the literal information. Other discretisation methods are compared and discussed in the next part.

Table 2: Best KGE model compared with KGE-MLP models. The KGE-MLP models are marked with*. Bold: best.

	Metric	TransE	TransE*	DistMult*	HolE*
Q1	Acc(Hits@1)	**0.42**	0.17	<u>0.22</u>	0.21
	MRR	**0.65**	0.45	<u>0.48</u>	<u>0.48</u>
	nrmse	**0.06**	0.11	<u>0.09</u>	0.10
Q2	Hits@1	**0.64**	0.34	0.34	<u>0.37</u>
	MRR	**0.77**	0.48	<u>0.52</u>	0.41
	Hits@GroupBy3	**0.85**	0.45	0.46	<u>0.52</u>

Discussion on discretisation strategy. There are different discretisation approaches discussed in the paper, including the single setting without overlapping, overlapping and hierarchical settings. There are also two different bins creation methods based on frequencey or the fixed value. According to the experiment results the single bin with fixed value is simple and the performance differences between different discretisation strateties are insignificant, so we choose this method on the welding quality monitoring.

Discussion on included information in industrial KG construction. In the KG construction step, we also notice it is important to consider the impact of the tabular columns on the KGE performance. The number of columns in production data are very large (over 200), but only few information are important based on domain knowledge, and most of them are meta-settings, or overlapping information that do not contribute to the KGE performance. This is a common problem for industrial KGs, especially for industries such as manufacturing, mining, chemistry, where large amount of information is collected, but only few are essential for the quality

monitoring.. Thus, we need to choose the least amount of columns that represent the most important information for welding. This selection process is done by iteratively updating the features and evaluating the performance. Our observations are, the most important information are (1) those that are crucial to the graph structure of the KG, such as welding machine, welding program, the materials and thickness of carbody, that have impact define the graph structure; (2) sensor values (literals in KG).

4. Conclusion and Outlook

This poster paper presents an extended abstract of our full paper [4] and provides additional discussions. The research is under the under the umbrella of Neuro-Symbolic AI for Industry 4.0 at Bosch. We aim at enhancing manufacturing technology with both symbolic AI [7] (such as semantic technologies) for improving transparency [1], and ML for prediction power. We will further improve the performance of the KG embedding method and develop other complementary technologies, such as ontologies [8, 9], ontology-based data access, etc.

References

[1] Z. Zheng, B. Zhou, D. Zhou, A. Soylu, E. Kharlamov, Executable knowledge graph for transparent machine learning in welding monitoring at bosch, in: CIKM, 2022, pp. 5102–5103.

[2] Z. Huang, M. Fey, C. Liu, E. Beysel, X. Xu, C. Brecher, Hybrid learning-based digital twin for manufacturing process: Modeling framework and implementation, Robotics and Computer-Integrated Manufacturing 82 (2023) 102545.

[3] E. B. Myklebust, E. Jimenez-Ruiz, J. Chen, R. Wolf, K. E. Tollefsen, Knowledge graph embedding for ecotoxicological effect prediction, in: ISWC, 2019.

[4] Z. Tan, B. Zhou, Z. Zheng, O. Savkovic, Z. Huang, I. G. Gonzalez, A. Soylu, E. Kharlamov, Literal-aware KGE for welding quality monitoring, in: ISWC, 2023.

[5] J. Wang, F. Ilievski, P. A. Szekely, K.-T. Yao, Augmenting knowledge graphs for better link prediction, in: IJCAI, 2022.

[6] A. Kristiadi, M. A. Khan, D. Lukovnikov, J. Lehmann, A. Fischer, Incorporating literals into knowledge graph embeddings, in: ISWC, 2019.

[7] D. Rincon-Yanez, M. H. Gad-Elrab, D. Stepanova, K. T. Tran, C. C. Xuan, B. Zhou, E. Karlamov, Addressing the scalability bottleneck of semantic technologies at bosch, ESWC Industry (2023).

[8] B. Zhou, Z. Zheng, D. Zhou, Z. Tan, O. Savković, H. Yang, Y. Zhang, E. Kharlamov, Knowledge graph-based semantic system for visual analytics in automatic manufacturing, ISWC, 2022.

[9] Z. Zheng, B. Zhou, D. Zhou, A. Q. Khan, A. Soylu, E. Kharlamov, Towards a statistic ontology for data analysis in smart manufacturing, in: ISWC Posters, volume 3254, 2022.