Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2021) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Industry 4.0 and Smart Manufacturing

Exploiting Knowledge Graph for Multi-faceted Conceptual Modelling using GCN

Yuwei Wan[a]*, Zhenyuan Chen[a], Fu Hu[a], Ying Liu[a], Michael Packianather[a], Rui Wang[b]

[a] *Institute of Mechanical and Manufacturing Engineering, School of Engineering, Cardiff University, Cardiff, CF24 3AA, UK*
[b]*UNICMICRO, Science Town Commercial Plaza, Kexue Road, Guangzhou, 510000, China*

Abstract

The relevant information obtained from multiple sources usually contributes to one intricate phenomenon in the industrial processes. Data fusion of different sources usually leads to more expressive and informative information than that of each single data source. Integrated information has been widely used to model a multi-faceted conceptual phenomenon, which provides a comprehensive and versatile view of understanding of the process. However, the conventional approaches concatenate feature vectors to integrate different facets, not considering the semantic gaps between them. Meanwhile, knowledge graph (KG) receives considerable attention in recent years as it comprises rich relational information among elements. Thus, KG provides a promising way to fuse multiple data sources by bridging the semantic gaps, which can be exploited in the modelling of a multi-faceted phenomenon. Inspired by the advancement of KG, we proposed an approach based on KG and a machine learning algorithm for multi-faceted modelling. Firstly, a domain-specified ontology was built to eliminate the varying distance metrics across facet boundaries, and KGs were generated by populating the data surrounding a multi-faceted phenomenon into this ontology. Secondly, the KGs were fed into a graph convolutional neural network (GCN) to learn the node features and the graph structure for graph embedding simultaneously with the shared parameters. Lastly, with the aim of multi-faceted conceptual modelling, the features obtained from the GCN model were used as inputs for machine learning algorithms to learn the hidden patterns of KGs. An experimental study using real-world data from the cold rolling process was conducted to demonstrate the feasibility of the proposed model.

Keywords: knowledge graph; graph embeddings; graph convolutional neural network; multi-faceted data

* Corresponding author. Tel.: +44-07410584532.
E-mail address: wany8@cardiff.ac.uk

1877-0509

1. Introduction

Enormous volumes of information or data are created, collected, and stored, which started as a result of the continuous development of complex systems and sensor devices. This type of data is diverse, heterogeneous, and fragmented [1]. In most cases, the relevant information of different modalities or sources is related to one intricate phenomenon or system in real complex industrial applications. In this context, integrating data from multiple sources is of great significance to obtain more comprehensive and informative information than that of each single data source. Meanwhile, owing to the paradigm shift of data-driven approaches, the fusion information has been widely used to model a multi-faceted conceptual system or phenomenon, which provides a united and versatile view of understanding of the system or phenomenon [2]. Hence, data fusion (DF) is one of the hot research topics currently. However, the conventional data fusion approaches concatenate feature vectors to fuse different facets, not considering varying distance metrics across facet boundaries. In this case, one fundamental challenge in fusing disparate facets lies in bridging the semantic gaps among them [3]. Meanwhile, graph-structured data receives considerable attention in recent years since it is the natural target to cover the domain information in many applications, such as social networks and molecular structures. As a typical type of graph-structured data, KG encodes knowledge in a specific domain, that comprises rich relational information among elements. KG is composed of entities (nodes) and edges. Edges are regarded as a fact, which shows a specific relational connection between two entities. Specifically, an edge is demonstrated in the form of a triplet, including head entity, relation, tail entity. In this context, KG captures each source domain's intrinsic ontology and achieves cross-links among facets. Therefore, KG provides a promising way to fuse disparate data for bridging the semantic gaps of disparate data sources.

Based on the integration of multi-faceted information, it is significant to exploit KGs for obtaining the collective and comprehensive information of multi-source data. For the task of multi-faceted phenomenon modelling, machine learning is a prevailing technical choice. However, KG, which represents structured data effectively, is challenging to manipulate by conventional machine learning models due to such triple's underlying symbolic nature [4]. When it comes to the inputs of conventional machine learning models, they usually take feature vectors representing objects in terms of tabular attributes (such as numeric attributes and categorical attributes) [5]. In this context, conventional machine learning approaches cannot be directly applied in graph-structured data.

Consequently, with the aim of representation of KG to bridge these semantic gaps, graph embedding algorithms have been proposed to learn the expressive and informative representation of graph-structured data [6]. Considered that the graph information is conserved, graph embedding maps a graph into a low dimensional space. In other words, given that the data lies in a low manifold, graph embedding approaches aim at reducing the dimensionality of the non-relational data. Specifically, a similarity graph is built on the similarity of the pairwise features obtained from a block of non-relational high-dimensional features. Each node is embedded into a low-dimensional space where connected nodes are closer to each other. According to the different types of graph-structured data, the inputs of graph embedding models are diverse in different cases. However, the outputs of graph embedding models are usually low-dimensional vectors that contain the graph information. In this context, achieving the low-dimensional vectors enables the feasibility of inputs of multi-faceted conceptual modelling using machine learning.

Conventional graph embedding methods have been utilised to analyse the graph-structured data in an effective yet efficient way, such as DeepWalk, Node2vec, Graph2vec [7, 8]. Although graph embedding methods provide an available way to represent graph information in low-dimensional vectors, these methods suffer from two major limitations [9]. Firstly, the implementations of embedding methods are accompanied by a high computational cost, because that there are no shared parameters between nodes in the encoder. It also means that the number of parameters increases dramatically with the number of nodes. Secondly, it is difficult to apply the same graph embedding models in different graphs. Especially, the embedding methods are unable to handle the dynamic graphs. In other words, conventional graph embedding methods lack the ability of generalisation.

To address these limitations, graph embedding based on GCN has been designed to learn both structure information and node features for graph embedding in an end-to-end manner [10]. Recent works have applied GCN for graph embedding successfully in different scenarios [11-13]. Firstly, compared with conventional graph embedding methods that learn the features of the node in a specific order, GCN propagates each node respectively, which leads to the ignorance of the input order of nodes. In this context, GCN avoids traversing all possible orders for presenting a graph completely, which leads to high computational cost. Secondly, each node of the graph is propagated with the shared parameters between nodes, which also avoids high computational costs. Thirdly, unlike conventional graph embedding models that the state of an edge is represented as states of the connected nodes, GCN learns the hidden information to update the information of nodes by a weighted sum of the information of their neighbourhoods.

In this study, with the aim of multi-source data integration by KGs, an approach based on GCN is proposed to learn representations of KGs for the multi-faceted phenomenon modelling. Firstly, a domain-specified ontology was utilised to generate knowledge graphs (KGs) by aggregating the knowledge of a multi-faceted conceptual phenomenon. Secondly, a GCN model was applied to take KGs as inputs for connected feature extraction and graph embedding. Thirdly, by using graph structure and node content information simultaneously, the results of graph embedding by GCN were used to model the multi-faceted conceptual phenomenon. Lastly, an experimental study using real-world data for multi-faceted conceptual modelling in the cold rolling process was conducted to demonstrate the feasibility of the proposed model.

The main contributions of this study can be summarised as follows:

- We have proposed that KG provides a promising way to fuse multiple data for bridging the semantic gaps of disparate data sources.
- We have prepared a graph embedding model based GCN to covert the KGs into a low-dimensional space.
- We have provided a comprehensive evaluation with a detailed analysis of experimental results and comparisons with different prevailing machine learning models.
- We found that the feasibility of the proposed model was validated in a real-world case application.

The rest of this study is structured as follows. Section 2 provides literature reviews on multi-faceted information modelling in machine learning, graph embeddings, and applications of KG in machine learning. In section 3, three frameworks for KG construction, KG embedding, and KGs for machine learning are outlined. An experimental study using real-world data is presented to demonstrate the feasibility of the proposed model in section 4. The conclusions of this study are presented in section 5.

2. Literature review

2.1. Multi-faceted information modelling in machine learning

DF is defined as a joint analysis framework of multiple-source data to contain the information that the single source does not recover [2]. Unlike the data collected from a single source, the data fused from multiple sources usually assembles versatile information about one phenomenon [14]. Depended on the different objectives of DF, three levels of DF have been defined and have formed the most prevailing taxonomy, which are low-level, mid-level, and high-level respectively [2].

For low-level, DF methods apply integration algorithms that concatenate observational data directly, which is known as the raw sensory data. In low-level fusion, the major characteristic is to build a data-driven model after achieving fusion [15]. In this context, the major superiority of this level is to offer possibilities of result interpretation directly in terms of the original variables obtained from each data block. Meanwhile, the major limitation is that many multivariate data analysis approaches cannot be used in the data directly, since the number of observations is much smaller than the number of variables. Hence, the preliminary operation of regularisations is necessary to handle the dimensionality problem [16]. According to the characteristic of low-level DF, the integration algorithms of this level are usually used in some similar scenarios, such as signal fusion and image fusion [17].

Mid-level DF can be regarded as feature level, where features drawn from each data block are fused to accomplish the most informative feature vector through feature selection algorithms. Specifically, the feature sets extracted from multiple sources are fused into a new high-dimension feature vector [4]. In this type of fusion, the obtained feature vector is fed into data-driven models to produce the desired outcome. In this context, the major advantage of mid-level DF lies in reducing the dimensionality of each data block separately before achieving to couple them [16]. However, when it comes to outcome interpretation of models, mid-level DF bridge the connections between features selected in the final model and the original variables from the raw data blocks, thus resulting in the interpretation predicament [16]. According to the characteristic of the mid-level DF, the approaches at this level are usually applied in different scenarios, such as different sensor nodes. A framework of feature fusion was used to combine passive and active remote sensing imagery, and to generate a more expressive and informative feature set. Based on that, a support vector machine (SVM) model took the feature set as inputs to recognise the American Bramble (a special kind of plant) from surrounding vegetation [18]. An approach of mid-level DF was adopted to couple the medical data acquired from the daily activity for the generation of a high-quality activity feature space. The integrated features were fed into an improved random forest (RF) model to predict early heart disease [19].

At high-level DF, the decisions of different analytical outputs from the processing of each data block are fused to accomplish the final result, which is known as decision level or information level. For this type of fusion, the variables from each raw data platform are not considered, because the outcomes of models are only integrated. In this context, the major advantage lies in greater confidence of the final performance by information integration from the different model outcomes, thus leading to diminish the holistic uncertainty among the single model uncertainties. Meanwhile, another obvious merit of high-level DF is to offer the bond possibility of the heterogeneous sensors. In this context, the information from the heterogeneous sensors can be processed with different algorithms. However, there is a corresponding limitation of high-level DF is that does not offer insight on relevant variables or features, thus giving rise to the interpretation difficulty of links among analytical sources [16]. In the context of high-level data fusion, a decision-level data fusion method was proposed to improve the accuracy of soil moisture content estimation [20]. A decision-level data approach was applied to transform low-dimensional decisions based on individual sensor data into high-dimensional decisions. Two case studies of quality control in additive manufacturing and predictive maintenance in aircraft engines showed that this approach can decrease the prediction uncertainty and improve the prediction accuracy respectively [21].

2.2. Knowledge graph embeddings

KGs represented by a collection of interlinked descriptions of entities are exploited to integrate the rich relational information among elements. However, as a type of non-Euclidean data structure, conventional machine learning models cannot be directly deployed in the graph domain. Specifically, conventional machine learning models are applied in tackling data based on tabular format and the split-apply-combine paradigm [5]. With the aim of fitting the input format for conventional machine learning models, KG embedding algorithms have been proposed to convert the KG into a low dimensional space in which the graph structural information and graph properties are maximumly preserved. Depended on the different theories, three types of KG embedding methods were defined and have formed the most prevailing taxonomy, which are factorisation based, random walk based and deep learning based [7].

Factorisation based algorithms represent the edges in KGs by a matrix, and further factorise this matrix to achieve the KG embedding. The matrices are applied to describe the connections between two nodes, which include adjacency matrix, Katz similarity matrix, node transition probability matrix and Laplacian matrix. Various approaches based on factorisation have been proposed, such as LLE [22], Laplacian Eigenmaps [23], Graph Factorisation [24]. For factorisation-based approaches, they are not allowed to learn an arbitrary function. In this context, unless explicitly included in their objective function, they cannot learn structural equivalence.

Random walks have been proposed to estimate many properties (such as node centrality and similarity) of the KGs. They are especially useful when one can either only partially observe the graph, or the graph is too large to measure in its entirety. Various embedding techniques using random walks on graphs to obtain the representations of KGs have been proposed, such as DeepWalk [25], node2vec [26], HARP [27]. In random walk based methods, the mixture of equivalences can be controlled to a certain extent by varying the random walk parameters.

Deep neural network-based methods have been designed to take the overall neighbours of each node as input, for example, SDNE [28], VGAE [29] and LINE [30], thus leading to high computational cost and inappropriate for large sparse graphs. To tackle these two problems, GCN has been proposed to define a convolutional operator on a graph [10]. The embeddings of each node's neighbours are aggregated to this node iteratively using an embedding function. In this context, the embedding aggregation of local neighbours makes this process scalable, and the representations of global neighbours are achieved to each node by multiple iterations.

2.3. Applications of knowledge graph in machine learning

Depended on different task objectives, four types of applications of KG in machine learning were defined and have formed the most prevailing taxonomy, which are node classification, link prediction, community detection and graph classification [31]. In terms of node classification, the missing types of nodes have been predicted using the given information of the edges and nodes, which is regarded as the machine learning task. In link prediction, the task objective is to obtain the outcome that if there is a link between two nodes. For community detection, the final goal is to achieve subgraphs from a single large KG, which can be regarded as a clustering of the nodes. When it comes to graph classification, the results of machine learning models are usually related to graphs, such as the label of a single intact KGs.

As mentioned previously, various embedding algorithms already have been designed and implemented to extract information from graphs in many industrial applications. For instance, regarding the node classification, researchers proposed a promising technique (namely MINDWAL) to mine interpretable and informative walks from KGs [32]. The walks are a type of format to generate input features that can be fed into a classification algorithm for node classification. Compared with other baseline techniques (such as decision tree, random forest), a competitive performance of the accuracy was demonstrated by applying in the biomedical domain. For link prediction, a competitive model based on deep learning was proposed to predict the interaction between two different drugs using the features learned from KGs [33]. In terms of graph classification, a GMADL model was designed to learn the subgraph features for mining graph data. Specifically, dictionary learning approaches were deployed to learn the hidden patterns of graph data for generating sparse code. The feature matrices of the graph were represented by the sparse codes. By building the graph classifier of SVM, the decent performances of accuracy and efficiency were realised with eight graph benchmarks [34].

3. Methodology

3.1. Knowledge graph construction

Ontology is defined as a form of knowledge conceptualisation, which is a model to describe structured and unstructured information through the general entities or concepts within a domain, properties, and the relationship between these concepts [35]. As formal semantic models, ontologies have been used to display the related knowledge representation of a specific domain rather than the description of a specific individual in a specified domain. Compared with the KG, the ontology is regarded as a model to express both the structure and information of the knowledge of the semantic environment and capture the relevant data from multiple sources within one phenomenon [36]. With the aim of the KG construction, the data within a specific domain is imported into the classes and the related network of ontologies to achieve multi-source data fusion. Hence, the KGs constructed are collected and stored to build a graph database.

Fig. 1 shows the flowchart of the construction of KG. In this study, a KG construction approach using ontology and data is used, which includes two stages.

For stage one, the relations among data blocks from multiple sources are identified that they contribute to one phenomenon. According to different application scenarios, various types of multiple sources exist in industrial applications, such as different sensors, different modalities, different features. Subsequently, the data that contributes to one phenomenon need to be collected and prepared for multi-faceted conceptual modelling.

In stage two, a KG construction approach using ontology and data is used. Firstly, an ontology is built in a specific domain, including a set of relevant concepts and their relations. Then, the data collecting from multiple sources is fed into the ontology to generate KGs. As shown in Fig. 1, the KG contains information of entities and relations (is represented by V and E) after importing multi-source data.

An undirected KG is defined as $G = (V,\ E)$ with N nodes $v_i \in V$ and edges $\left(v_i, v_j\right)$. An adjacency matrix $A \in \mathbb{R}^{N\times N}$ is defined as a square matrix used to represent a finite graph. In the adjacency matrix, elements present whether pairs of vertices are adjacency or not in the graph. A degree matrix is defined as $D_{ii} = \sum_i A_{ij}$.

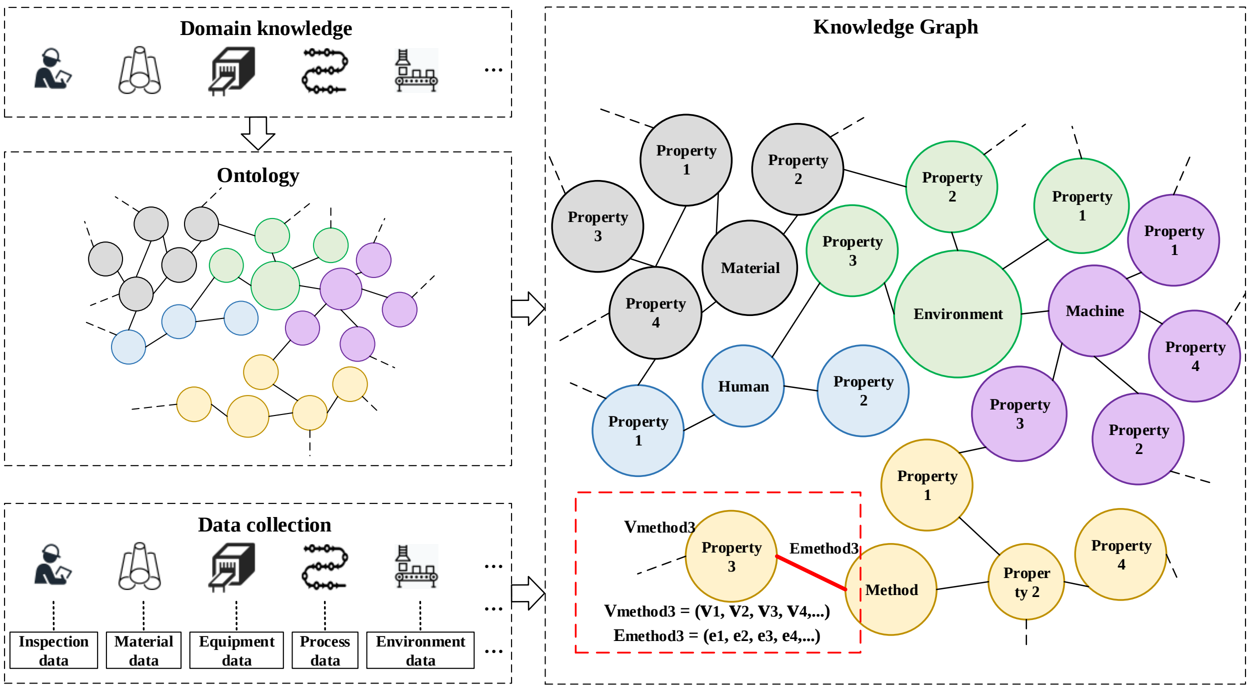

Fig. 1. The schematic diagram for the construction of KG.

3.2. RF-GCN model based on graph embedding

Through constructing KGs for populating multi-source data, it is necessary to use an embedding algorithm that transforms the KGs into a low-dimensional space for multi-faceted conceptual modelling. In this study, GCN is used to extract the connected features and learn the representation of KGs for graph embedding in an end-to-end manner. A convolutional operator defined on a localised first-order approximation has been proposed to handle the graph-structured data directly, such as KG. Specifically, GCN propagates the information of each node by its neighbours iteratively. In this context, the convolutional architecture learns hidden layer representations of KG, where both local graph structure and node features are preserved simultaneously [37].

Given a specific undirected KG $G = (V,\ E)$ with N nodes $v_i \in V$ and edges (v_i, v_j), a GCN model is defined as $f(X, A)$ with the adjacency matrix $A \in \mathbb{R}^{N\times N}$ and the node feature matrix X. That also means that the GCN model takes the adjacency matrix (which is represented as the graph structure) and the node feature matrix (which is represented as node content information) as inputs. GCN model is stacked by multiple convolutional layers. A single convolutional layer of GCN is demonstrated with the following rule:

$$H^{(l+1)} = \boldsymbol{ReLU}\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right) \tag{1}$$

Here, $\tilde{A} = A + I_N$ is defined as the adjacency matrix of the undirected graph G with added self-connections. I_N is defined as the identity matrix. Moreover, $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$ and $W^{(l)}$ describes a layer-specific trainable weight matrix, and $ReLU(\cdot)$ is an activation function. $H^{(l)} \in \mathbb{R}^{N\times D}$ is the matrix of activations in the l^{th} layer. $H^{(0)} = X$, and X is the original node feature matrix.

Fig. 2 shows the workflow of the RF-GCN model based on graph embedding. Specifically, the KG embedding process includes three stages, which are KG expression, representation learning, and multi-faceted conceptual modelling. Firstly, the node feature matrices and adjacency matrices are obtained and computed from KGs respectively. In other words, the entities' attributes (is represented as node features) and the relation features (is represented as adjacency matrix) are used to express the KGs. Secondly, expressing KG by the entity and the graph structure, these two types of matrices are imported into a two-layer GCN model. A two-layer GCN model is built to take the two matrices as inputs for the representation learning of KGs automatically. For each node, the feature updates itself by its neighbours within the shared parameters. Thirdly, based on the KG embedding, the information contained by KGs is converted into embedding features, which are appropriate to be inputted into machine learning models. As the outcomes of the GCN model, the embedding features are accomplished at the node level, which can be regarded as feature extraction from KGs. With the aim of multi-faceted conceptual modelling, the embedding matrices are imported into the RF model subsequently for machine learning tasks.

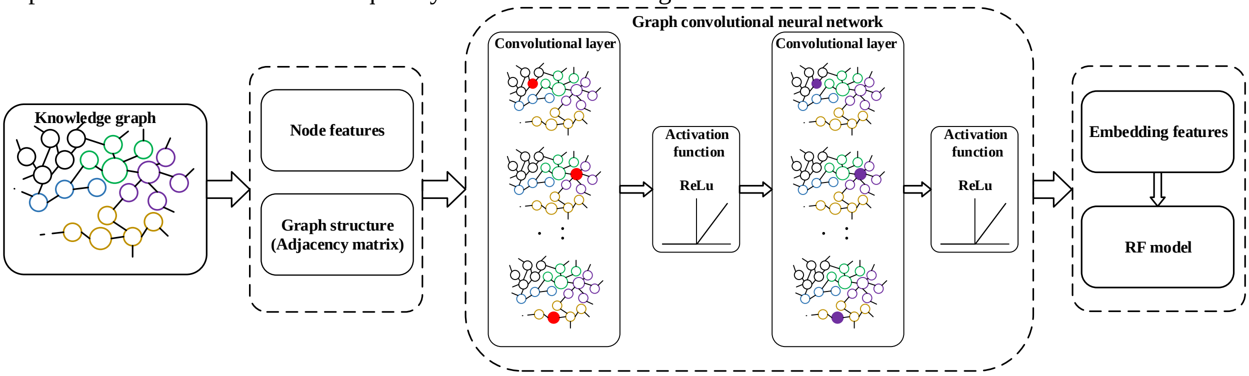

Fig. 2. The RF-GCN model based on graph embedding.

4. Case study

4.1. Data description and preparation

In this section, an experimental study using real-world data for multi-faceted conceptual modelling was conducted to illustrate the feasibility of the proposed model. We collected data from an electrical steel manufacturer where a reversing mill is equipped for the cold rolling process. Electrical steel is a high-silicon iron alloy. High silicon content in alloys enhances the electrical resistivity of the material, which results in reducing magnetic losses. A higher concentration of silicon leads to the brittleness of the strip, which results in breakages occurred in the process of cold rolling.

Based on our previous study [38], the relevant data that contribute to breakage are collected from hot-rolled coils, annealing and pickling process, emulsion record, cold rolling process and quality inspections. Fig. 3 shows how data are collected across the manufacturing process of the cold-rolled strip. The data sources for strip breakage modelling are shown in Table 1, while the quality inspections data are used as the label.

The dataset is collected and stored from these five processes of the steel industry in this study, which covers the production period of six months. For this historical dataset, 956 of them were labelled as "normal", while 368 of them were labelled as "breakage". In this case, the dataset is regarded as an imbalanced dataset.

Table 1. Details of relevant features for four processes.

Processes	Number of features	Examples of features
Hot-rolled coil	15	Chemical content, gauge, crown, quench temperature etc.
Annealing and Pickling	8	Annealing temperature, Jetflow speed etc.
Emulsion	9	Dirt result, pH, conductivity, chloride index etc.
Cold rolling	17	Rolling, tension, measured slip etc.

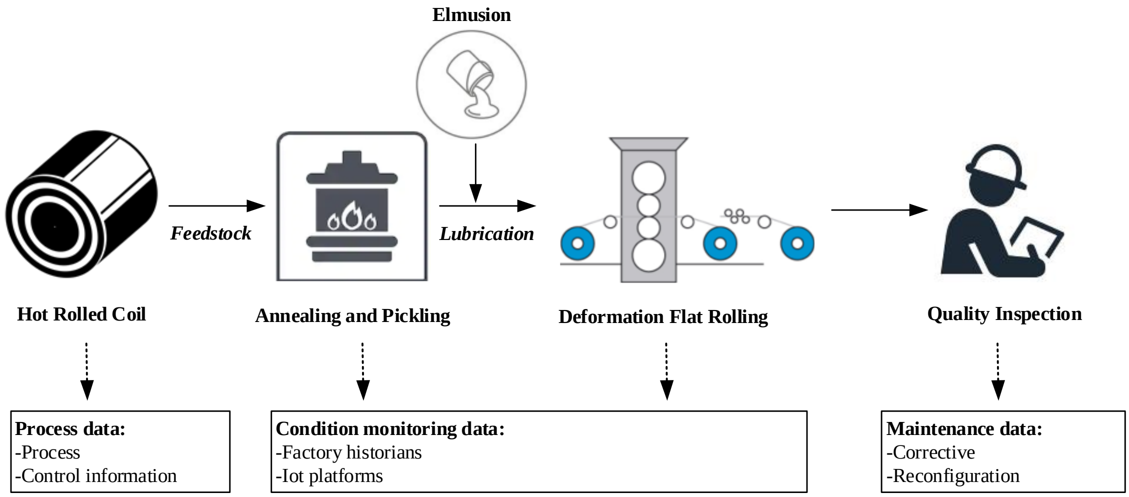

Fig. 3. The workflow of five processes of the steel industry.

4.2. Experiments

The experiments were conducted using our proposed KG-based fusion scheme. To evaluate the performance, four popular machine learning algorithms are built using the open-source package Scikit-learn with the default hyperparameters, including complement Naïve Bayes (CNB), multi-Layer perceptron (MLP), SVM and RF. For the SVM, the kernel type was radial basis function with reciprocal of feature number set as the kernel coefficient. In terms of MLP, two hidden layers were designed. Within each layer, the number of neurons was set as the same as the number of features. Lastly, the number of the estimators of the RF model was set as 100, and the square root of the number of input features was selected for the best split.

To be specific, each coil was fed to the domain ontology for KG generation. As shown in Fig. 2, each graph representing one coil was fed into the proposed GCN pipeline for graph embedding. Subsequently, the embedding feature matrices were imported into four machine learning models for breakage prediction.

In addition, the 5-fold cross-validation tests were conducted to assess the performances, and the mean values of the five folds were recorded. Although accuracy (ACC) is typically used as the performance metric for binary classification tasks, the ACC is not sufficient for the class imbalance problem. In this case, the area under the receiver operating curve (AUC) was also implemented along with ACC to evaluate the performance of our models. To be specific, the configuration of the processing system used for these experiments is a CPU Core i7-9700 K 3.8 GHz.

4.3. Results and discussions

The feasibility of the proposed model for multi-faceted conceptual modelling was verified in four different models. The performance in terms of ACC and AUC for the designed KG-based fusion scheme using different machine learning algorithms are shown in Fig. 4. The comparisons among the four machine learning algorithms reveal that the RF model show advantages over the other three machine learning models.

To be specific, it is obvious that RF surpassed the other machine learning baselines in terms of ACC (0.78). Moreover, the experimental results show that a significant reduction in terms of AUC was found using the CNB, MLP and SVM models, while the RF model performs best on this metric. Specifically, the RF achieved the best AUC (0.75) comparing with other models. In this context, the RF model presents the potential for imbalanced datasets.

Compared with the MLP and SVM, CNB achieved the best result in terms of AUC, while it achieved the lowest ACC. In this context, CNB is more suitable for imbalanced data sets in these three algorithms. Moreover, CNB is a very simple classifier, which is embodied in implementing easily.

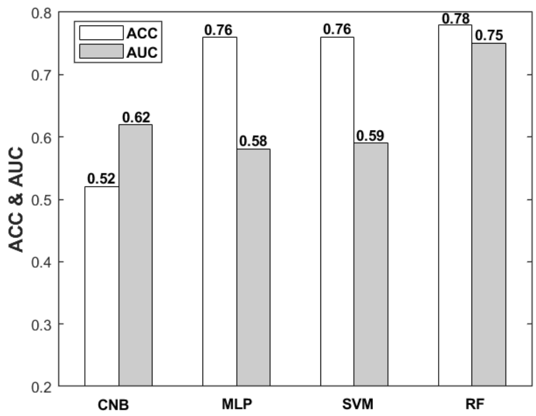

Fig. 4. Results across different machine learning algorithms.

Additionally, the MLP model and the SVM model achieved much better performances on ACC than the CNB model. However, it can be observed that there is a significant reduction in terms of AUC for these two models. For the MLP model and the SVM model, the AUCs are 0.58 and 0.59, respectively. Typically, conventional SVM and MLP algorithms perform barely satisfactory for the imbalanced dataset, which was revealed in the experiments that these two models obtained low values in terms of AUC.

5. Conclusion

In this paper, we proposed a fusion scheme to bridge the semantic gaps among disparate data sources, which is the application of KGs to integrate data from these sources. Since the KGs cannot be fed into machine learning models directly, we proposed the graph embedding method based on the GCN to covert KGs into a low-dimensional space. Specifically, a two-layer GCN model was built to learn the graph structure and node features of KGs simultaneously. Based on that, a dense vector representation of KGs was created and imported into various machine learning models for multi-faceted phenomenon modelling. A preliminary case study using the real-world data from the steel-making process was conducted. The experimental results were evaluated, and the performances of different prevailing machine learning models were assessed. The feasibility of the proposed model was validated in the steel industrial application. In the future, a more representative and informative way of extracting the representations of KGs will be explored.

References

[1] K. Zhang and J. Liu, "Review on the Application of Knowledge Graph in Cyber Security Assessment," in IOP Conference Series: Materials Science and Engineering, 2020, vol. 768, no. 5, p. 052103: IOP Publishing.

[2] M. Cocchi, "Introduction: Ways and means to deal with data from multiple sources," in Data Handling in Science and Technology, vol. 31: Elsevier, 2019, pp. 1-26.

[3] W. Zheng et al., "Pay attention to doctor-patient dialogues: Multi-modal knowledge graph attention image-text embedding for COVID-19 diagnosis," Information Fusion, 2021.

[4] C. Chen, R. Jafari, and N. Kehtarnavaz, "A survey of depth and inertial sensor fusion for human action recognition," Multimedia Tools and Applications, vol. 76, no. 3, pp. 4405-4425, 2017.

[5] M. Nickel, K. Murphy, V. Tresp, and E. Gabrilovich, "A review of relational machine learning for knowledge graphs," Proceedings of the IEEE, vol. 104, no. 1, pp. 11-33, 2015.

[6] W. L. Hamilton, R. Ying, and J. Leskovec, "Representation learning on graphs: Methods and applications," arXiv preprint arXiv:1709.05584, 2017.

[7] P. Goyal and E. Ferrara, "Graph embedding techniques, applications, and performance: A survey," Knowledge-Based Systems, vol. 151, pp. 78-94, 2018.

[8] P. Cui, X. Wang, J. Pei, and W. Zhu, "A survey on network embedding," IEEE Transactions on Knowledge and Data Engineering, vol. 31, no. 5, pp. 833-852, 2018.

[9] Z. Liu and J. Zhou, "Introduction to Graph Neural Networks," Synthesis Lectures on Artificial Intelligence and Machine Learning, vol. 14, no. 2, pp. 1-127, 2020.

[10] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," arXiv preprint arXiv:1609.02907, 2016.

[11] Y. Xie, C. Yao, M. Gong, C. Chen, and A. K. Qin, "Graph convolutional networks with multi-level coarsening for graph classification," Knowledge-Based Systems, vol. 194, p. 105578, 2020.

[12] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and S. Y. Philip, "A comprehensive survey on graph neural networks," IEEE transactions on neural networks and learning systems, 2020.

[13] H. Jiang, P. Cao, M. Xu, J. Yang, and O. Zaiane, "Hi-GCN: A hierarchical graph convolution network for graph embedding learning of brain network and brain disorders prediction," Computers in Biology and Medicine, vol. 127, p. 104096, 2020.

[14] L. Kong, X. Peng, Y. Chen, P. Wang, and M. Xu, "Multi-sensor measurement and data fusion technology for manufacturing process monitoring: a literature review," International Journal of Extreme Manufacturing, vol. 2, no. 2, p. 022001, 2020.

[15] H. F. Nweke, Y. W. Teh, G. Mujtaba, U. R. Alo, and M. A. Al-garadi, "Multi-sensor fusion based on multiple classifier systems for human activity identification," Human-centric Computing and Information Sciences, vol. 9, no. 1, pp. 1-44, 2019.

[16] M. Cocchi, Data Fusion Methodology and Applications. Elsevier, 2019.

[17] T. Meng, X. Jing, Z. Yan, and W. Pedrycz, "A survey on machine learning for data fusion," Information Fusion, vol. 57, pp. 115-129, 2020.

[18] P. Rajah, J. Odindi, and O. Mutanga, "Feature level image fusion of optical imagery and Synthetic Aperture Radar (SAR) for invasive alien plant species detection and mapping," Remote Sensing Applications: Society and Environment, vol. 10, pp. 198-208, 2018.

[19] M. Muzammal, R. Talat, A. H. Sodhro, and S. Pirbhulal, "A multi-sensor data fusion enabled ensemble approach for medical data from body sensor networks," Information Fusion, vol. 53, pp. 155-164, 2020.

[20] O. Yahia, R. Guida, and P. Iervolino, "Weights based decision level data fusion of landsat-8 and sentinel-L for soil moisture content estimation," in IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, 2018, pp. 8078-8081: IEEE.

[21] Y. Wei, D. Wu, and J. Terpenny, "Decision-Level Data Fusion in Quality Control and Predictive Maintenance," IEEE Transactions on Automation Science and Engineering, vol. 18, no. 1, pp. 184-194, 2020.

[22] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," science, vol. 290, no. 5500, pp. 2323-2326, 2000.

[23] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in Nips, 2001, vol. 14, no. 14, pp. 585-591.

[24] A. Ahmed, N. Shervashidze, S. Narayanamurthy, V. Josifovski, and A. J. Smola, "Distributed large-scale natural graph factorization," in Proceedings of the 22nd international conference on World Wide Web, 2013, pp. 37-48.

[25] B. Perozzi, R. Al-Rfou, and S. Skiena, "Deepwalk: Online learning of social representations," in Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, 2014, pp. 701-710.

[26] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, 2016, pp. 855-864.

[27] H. Chen, B. Perozzi, Y. Hu, and S. Skiena, "Harp: Hierarchical representation learning for networks," in Proceedings of the AAAI Conference on Artificial Intelligence, 2018, vol. 32, no. 1.

[28] D. Wang, P. Cui, and W. Zhu, "Structural deep network embedding," in Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, 2016, pp. 1225-1234.

[29] T. N. Kipf and M. Welling, "Variational graph auto-encoders," arXiv preprint arXiv:1611.07308, 2016.

[30] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei, "Line: Large-scale information network embedding," in Proceedings of the 24th international conference on world wide web, 2015, pp. 1067-1077.

[31] W. L. Hamilton, "Graph representation learning," Synthesis Lectures on Artifical Intelligence and Machine Learning, vol. 14, no. 3, pp. 1-159, 2020.

[32] G. Vandewiele, B. Steenwinckel, F. De Turck, and F. Ongenae, "MINDWALC: mining interpretable, discriminative walks for classification of nodes in a knowledge graph," BMC Medical Informatics and Decision Making, vol. 20, no. 4, pp. 1-15, 2020.

[33] M. R. Karim, M. Cochez, J. B. Jares, M. Uddin, O. Beyan, and S. Decker, "Drug-drug interaction prediction based on knowledge graph embeddings and convolutional-LSTM network," in Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, 2019, pp. 113-123.

[34] X. Zheng, S. Liang, B. Liu, X. Xiong, X. Hu, and Y. Liu, "Subgraph feature extraction based on multi-view dictionary learning for graph classification," Knowledge-Based Systems, vol. 214, p. 106716, 2021.

[35] L. Otero-Cerdeira, F. J. Rodríguez-Martínez, and A. Gómez-Rodríguez, "Ontology matching: A literature review," Expert Systems with Applications, vol. 42, no. 2, pp. 949-971, 2015.

[36] A. Konys, "An ontology-based knowledge modelling for a sustainability assessment domain," Sustainability, vol. 10, no. 2, p. 300, 2018.

[37] Z. Yan, J. Ge, Y. Wu, L. Li, and T. Li, "Automatic virtual network embedding: A deep reinforcement learning approach with graph convolutional networks," IEEE Journal on Selected Areas in Communications, vol. 38, no. 6, pp. 1040-1057, 2020.

[38] Z. Chen, Y. Liu, A. Valera Medina, and F. Robinson, "A multi-source feature-level fusion approach for predicting strip breakage in cold rolling," presented at the 2020 IEEE 16th International Conference on Automation Science and Engineering (CASE), Virtual, 20-24 August, 2020.