

RESEARCH ARTICLE

Relationship Extraction and Processing for Knowledge Graph of Welding Manufacturing

KAINAN GUAN¹, LIANG DU, AND XINHUA YANG

Liaoning Key Laboratory of Welding and Reliability of Rail Transportation Equipment, School of Materials Science and Engineering, Dalian Jiaotong University, Dalian 116028, China

Corresponding author: Xinhua Yang (yangxh@djtu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 51875072 and Grant 52005071, and in part by the Foundation for Overseas Talents Training Project in Liaoning Colleges and Universities under Grant 2018LNGXGJWPY-YB012.

ABSTRACT Acquiring welding domain relationships and forming a knowledge graph can positively impact complex engineering problem solving and intelligent manufacturing applications. However, relationships are lacking in the welding domain. The relationship extraction and processing solution are designed to handle data with different characteristics in welding fabrication. The BiLSTM+Attention and CR-CNN models are employed to extract relations in unstructured documents. The neighborhood rough set-based association rule model is proposed for project-specific documents to accomplish relationship acquisition, in which invalid attributes are removed via neighborhood rough sets and attribute values are related via association rules. In addition, the knowledge graph is built based on extracted relationships, and unique empirical relationships are handled by introducing relational nodes and databases. The results show that BiLSTM+Attention gets a good score with Macro-average metrics (0.788 for Precision, 0.846 for Recall, and 0.816 for F1-score). The relational rules obtained via the proposed model are consistent with the production experience. The constructed knowledge graph effectively handles empirical relationships while positively impacting knowledge retrieval, intelligent question and answer, and decision-making for complex engineering problems.

INDEX TERMS Welding manufacturing, relationship extraction, neighborhood rough sets, association rule, knowledge graph.

I. INTRODUCTION

Data is a critical factor in domain operation and development, and data-driven has gradually become an essential means to solve complex domain problems [1], [2]. In recent decades, data-based welding manufacturing has focused on meeting the development requirements for high quality and efficiency. Methods such as case-based [3] retrieval, rule-based reasoning [4], and fuzzy expert systems [5] are applied to welding process design, welding material selection, and production process planning. Although these intelligent data-based methods [6], [7] have become an effective means to solve engineering problems, the challenge of complex knowledge representation and relationship construction remains

The associate editor coordinating the review of this manuscript and approving it for publication was Sathish Kumar¹.

unresolved. The knowledge graph is considered to address the proposed limitations due to its compatibility in the knowledge representation. Relationship extraction is an indispensable step in knowledge graph construction, and it is an essential medium for the logical composition of knowledge and linking of domain entities. However, the specialized and complex nature of the domain relationships makes relationship extraction challenging in manufacturing, especially in welding fabrication.

Relational extraction methods are generally classified as template-based [8], supervised-based [9], and weakly supervised-based relationship extraction [10]. The template-based approach uses pre-defined relationship templates by domain experts and then matches relationships from the text. This method has high applicability in a small range of texts but relies on extensive manual work making it less

portable. Supervised learning models are used for relationship extraction based on pre-labeled data to acquire many textual relationships. The method is also divided into pipeline operations [11], [12] and joint operations [13], considering the training process of the entity and relation. Entities and relationships are trained separately in a pipeline approach, such as the CR-CNN [14] and Attention CNNs [15] models. Joint extraction methods consider entity and relationship model training to combine entity and relationship extraction. Miwa and Bansal [16] propose an end-to-end neural model to extract entities and their relationships. Word sequences and dependency tree substructure information are captured via LSTM-RNN. Such methods require a large amount of annotated data before training. Weakly supervised approaches have been proposed to address the limitation of insufficient corpus data. Their model training is achieved with a small number of labeled samples, such as the APCNNs [17] model and the NELL system [18]. However, most research focuses on enhancing models and techniques on non-domain data or biomedical [19] domains for relationship extraction. Relationships in the manufacturing domain are neglected, especially in welding manufacturing. Several existential challenges: (i) Almost no relevant studies concentrate on extracting weld manufacturing relationships, resulting in a lack of research data. (ii) Relation extraction in weld-specific files is challenging because the association discovery between attributes is hard. (iii) Relationship handling for engineering applications needs to be explored. Automatic acquisition of basic relations, refinement of specific relations, and processing of empirical relations are the significant challenges of relation extraction in the welding domain.

In this study, we aim to support the solution of complex problems in welding manufacturing via extracting data relationships, reducing the limitations of knowledge representation, and building a relational network. However, the challenges of missing data, domain specificity, and practical engineering applications must be solved. Therefore, we design a scheme that integrates neural networks, attribute reduction, and association rules and construct a knowledge graph for engineering applications. Our approach has the following features: (i) Professionalism and reference. Oriented to practical welding production, we dig deeper into the domain data characteristics by extracting unique relations and processing empirical relations. In addition, the method we describe has a substantial reference value for domain relation extraction. (ii) Covering multiple types of data. The data includes the underlying unstructured files, structured result files, and empirical data. (iii) High engineering application value. Industry-specific method design with relationships embedded in the knowledge graph forms a knowledge system in the welding field. Data support is provided for developing intelligent technologies and engineering problem-solving in the industry.

We divided the relationship extraction task into three subtasks: general relationship extraction, engineered document relationship extraction, and empirical relationship

processing, based on several challenges in welding relationship extraction. Models BiLSTM+Attention [20] and CR-CNN are employed to extract relationships from unstructured documents such as welding standards, production guidance documents, design guidelines, etc. The actual production knowledge text is trained in advance as word vectors to support the training of the relational extraction model through the word2vec [21] method. We propose a neighborhood rough set-based association model for engineered documents to extract special relations from the normalized result file. Valid attribute links are determined based on neighborhood rough sets [22], and relationships between attribute values are extracted based on association rules [23]. In addition, empirical formulas and models are collected in a relational database, and the knowledge graph embedding is implemented using key-value pairs. Actual data from the production of welded bogies for high-speed trains is collated to support model validation. The results show that model BiLSTM+Attention has good scores (Precision: 0.788, Recall: 0.846, F1-score: 0.816) in our data with Macro-average evaluation metrics. Special relations in real production are extracted from the 110 result file data via the proposed special relation extraction model and have high consistency with experience. Empirical data and extracted relationships are embedded in a domain knowledge graph to support complex problem-solving and important decision-making in welding manufacturing. The main contributions of this paper are as follows:

(i) We implemented the relationship extraction of unstructured documents oriented to the lack of knowledge in welding manufacturing. The baseline for welding relationship extraction was listed on our data to support the relevant studies.

(ii) A neighborhood rough set-based association rule model for welding data structure characteristics is proposed to extract the relationships in the attributes and attribute values from the actual welded structured documents, respectively.

(iii) Empirical relational databases and extracted relationships are embedded in domain knowledge graphs to support complex problem-solving and critical decision-making in welding manufacturing.

The aim is to extract relations and build a knowledge graph for welding manufacturing data. For this reason, we designed three kinds of relationship extraction and processing methods for different data forms. In addition, the experimental results were obtained via data processing, attribute restoration, and relationship model construction. We predict that our work can support the digital construction of welding manufacturing and the data requirements of intelligent systems.

II. METHODS

A. RELATIONSHIP EXTRACTION

Unstructured data such as welding standards, production requirements, and guidance documents contain essential knowledge to ensure weld fabrication. Automatic acquisition of relationships benefits welding knowledge system

construction and automated production. Relationships are carried out in sentences expressed with natural language. Extracting dependency features between word sequences is the key to extracting relationships. BiLSTM+Attention is employed for sentence-level relation extraction due to its good bidirectional semantic feature acquisition performance. Long short-term memory (LSTM) [24] is an essential element of BiLSTM+Attention construction which contains three vital structures: forget gate, input gate, and output gate. This model calculates the current output and cell state via the output and cell state at the previous time. The behavior of this model is shown in FIGURE 1.

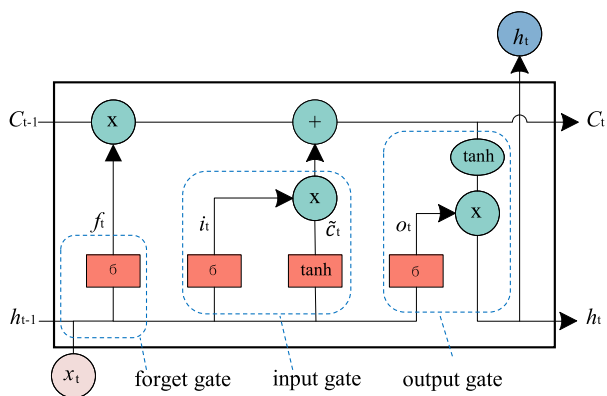


FIGURE 1. LSTM cell structure diagram.

The function of the forget gate is to discard negative information in the cell state. The sigmoid activation function (σ) defines whether the information is retained based on the output at the last moment and the input at the moment. The calculation formula is as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

where σ is the sigmoid activation function, W_f and b_f represent weights and biases, h_{t-1} represents the output of the last time, and x_t is the input at the moment.

The input gate is employed to update the state at the current moment. The calculation process is expressed in (2), (3), and (4). Where “tanh” is the hyperbolic tangent activation function, it represents the current moment of the reserved information, \tilde{C}_t and C_t are the current cell input and output status.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (3)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (4)$$

The positive state of the cell is assigned to the current output via the output gate. The output is calculated via (5) and (6). Where o_t is the current output cell status, and h_t is the current output value.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t * \tanh(C_t) \quad (6)$$

BiLSTM+Attention obtains bidirectional semantic information in a sentence via fusing a forward LSTM, a reverse LSTM, and an attention mechanism, including an input layer, an encoding layer, an LSTM layer, attention, and an output layer. The input layer splits the sentence into several words to complete the data input. The embedding layer implements vocabulary mapping to low latitude space LSTM layers acquire high-level features from the forward and reverses directions. The Attention layer is weighted by the position weights to obtain the sentence vector. The output layer completes relational classification based on sentence feature vectors. The schematic diagram is described as shown in FIGURE 2.

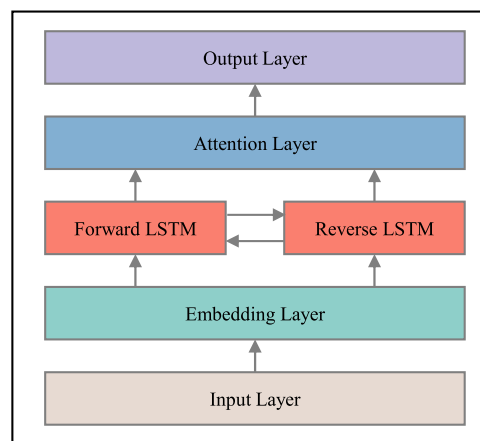


FIGURE 2. BiLSTM+Attention model principle structure.

B. SPECIAL RELATIONSHIP COMPLEMENT

Companies usually keep many standardized result files in welding production, guiding production and ensuring quality. These files contain multiple attributes, and it is not easily defined whether the attributes are related. Extracting relationships oriented towards valid attributes is positive for our work. The neighborhood rough set-based association rule model is proposed, which takes the field attributes in the data as decision attributes and the remaining attributes as conditional attributes, respectively. And complete attribute simplification and relationship extraction for welding production data. The reduced attributes and the decision attributes form an inference relationship. The specific workflow of the model can be divided into four steps as follows:

(1) The information decision system $S = (U, A, V, f)$ is built separately for the field attributes in the data. Where U is a non-empty finite set of objects called the theoretical domain; A is a non-empty set of attributes that is the concatenation of conditional attributes C and decision attributes D ; V is the value domain; f is an information function that meets $\forall x \in U, a \in A, f(x, a) \in V_a$. Defining $x_i \in U$, the neighborhood of x needs to satisfy (7).

$$\delta(x_i) = \{x_j | x_i \in U, \Delta(x_i, x_j) \leq \delta\} \quad (7)$$

(2) Lower approximation of computational information decision system. The decision attribute D divides the argument domain U into N equivalence classes $(X_1, X_2 \dots X_N)$, $\forall B \subseteq A$. The lower approximation of the decision attribute D concerning subset B is calculated by (8) and (9).

$$\underline{N}_B D = \bigcup_{i=1}^N \underline{N}_B X_i \tag{8}$$

$$\underline{N}_B X = \{x_i | \delta_B(x) \subseteq X, x_i \in U\} \tag{9}$$

(3) Reduction of attribute sets based on importance. The dependence of the decision attribute D on the conditional attribute B can be expressed as the lower approximation to the upper theoretical domain of B , as in (10). Importance can be understood as the degree of influence of a conditional attribute on a decision attribute and is a vital reference for attribute simplification. It is calculated as in (11). The reduced set of attributes is represented in (12).

$$k_D = \gamma_B(D) = \frac{|\text{Pos}_B(D)|}{|U|}, \text{Pos}_B(D) = \underline{N}_B D \tag{10}$$

$$\text{Sig}(a, B, D) = \gamma_B(D) - \gamma_{B-\{b\}}(D) \tag{11}$$

$$R = \{b | \text{Sig}(a, B, D) \geq \Delta, b \in A\} \tag{12}$$

(4) Define the dataset $P = R \cup D$, and calculate the support of the item set p in the dataset. The number of occurrences of item set p in the dataset P as a proportion of the total dataset is the support degree, as in (13). Items greater than a threshold in the set of items are called frequent itemsets. For the set that satisfies the condition of the frequent itemset, define $r \in R, d \in D$, and $r \cup d = p$, then the confidence of r and d is as in (14). Relationships can be extracted from rules that exceed the confidence threshold t , as in (15).

$$\text{Support}(p) = \frac{\text{num}(p)}{\text{num}(P)} \tag{13}$$

$$\text{Confidence}(d \leftarrow r) = \frac{\text{Support}(p)}{\text{Support}(r)} \tag{14}$$

$$\text{Rules}(d \leftarrow r) = \{d \leftarrow r | \text{Confidence}(d \leftarrow r) > t\} \tag{15}$$

C. EXPERIENCE RELATIONSHIP PROCESSING

The knowledge graph is a collection of entities and relations; generally, the entities are linked one-to-one via the corresponding relations. However, many relationships in welding manufacturing require many-to-many, many-to-one, and linear mappings between entities. For example, “welding method and groove form corresponding to the assembly gap,” “development of welding process specification,” and “the number of weld passes have a linear correlation with the weld depth.” Many nodes need to be built to satisfy the knowledge expression. The numerous scoped nodes reduce the knowledge’s accuracy and increase the knowledge graph’s complexity. Therefore, we construct the particular relationship database and introduce relationship nodes to overcome the proposed limitations.

The relational database contains fundamental logical relations (greater than, less than, starts at, ends at, not equal to, etc.), empirical relations (logic rules, empirical formulas, etc.), and model relations (classification or regression models based on actual engineering data, etc.). In addition, the relationship is called according to the unique key. The physical form of a relational library can be seen as a collection of multiple interfaces. When knowledge search involves relational nodes, we can obtain relational content based on key-value pairs. The schematic structure is shown in FIGURE 3.

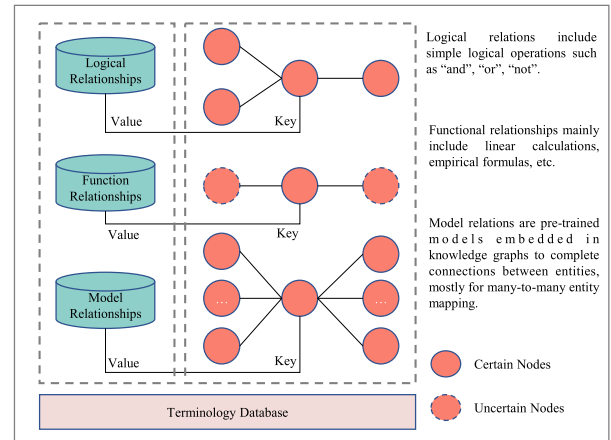


FIGURE 3. Relationship processing schematic.

III. EXPERIMENTS

A. RELATIONSHIP SYSTEM CONSTRUCTION

Relationships are essential ties between properties of things and connect entities to form domain knowledge networks in the welding manufacturing knowledge system. In actual production, relationships are often contained in unstructured documents, standardized production data, empirical formulas, etc. Extracting domain relationships from different structures is significant for domain knowledge system construction and welding engineering applications. The relationship extraction for practical welding production is divided into three sub-tasks: unstructured data extraction, relationship complementation, and experience relationships processing. Relationship extraction and knowledge graph construction are expressed in FIGURE 4.

Unstructured data extraction is the process of converting unstructured data into standard relational triples. The relationships are obtained through data processing, word vector, and the relational model. The engineering data characteristics are considered in the relationship extraction of standardization result documents. The attribute dependency relationship is established through attribute reduction, and the standard triplet relationship between attribute values is obtained through association rules. Empirical data and formulas are collected into relational databases and associated with knowledge maps through key-value pairs. In addition, a knowledge map is constructed based on the acquired standard triples.

TABLE 1. Word vector training parameters.

Parameter	Value	Parameter Interpretation
vector_size	100	The dimensionality of the word vector. Select vector_size 100 according to corpus size.
window	10	Word vector context maximum distance. The selection window is 10 due to a long-distance association of data.
min_count	3	The minimum frequency of the word vector. Minor frequency specialized features affect subsequent work reduction default value 5 to 3.
workers	4	Select 4 threads to train the model based on hardware conditions.
sg	1	The 1 indicates the selection of the Skip-Gram model.
alpha	0.025	The initial step of the iteration.

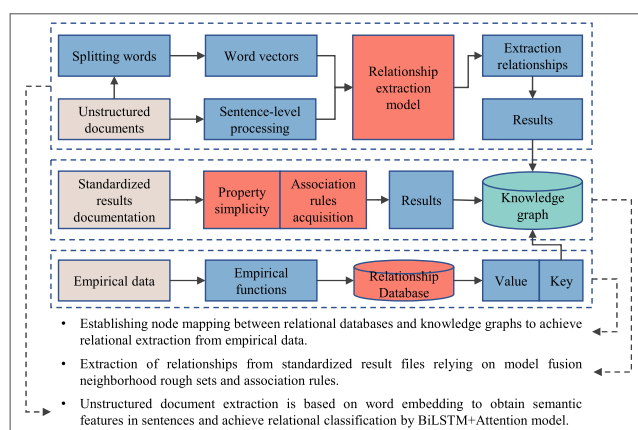


FIGURE 4. The process of relationship extraction and knowledge graph construction.

B. DATA PROCESSING

In welding manufacturing, data exists mainly in unstructured, semi-structured empirical data and standardized production documents. In unstructured data relationship extraction, these confusing and disordered data are divided into design, process, production, inspection, and maintenance according to manufacturing stages. Punctuation classification data is at the sentence level and is manually checked to form the initial standardized data.

In unstructured relation extraction tasks, relation determination requires reference to contextual semantic features. However, most models cannot obtain semantic features directly through entity characters. Welding terminology related to welding design, process, production, inspection, and maintenance is collected, and the sentence-level data is segmented into words via the CRF model. Word embedding methods are considered to convert characters to numeric vector features. We trained the words with the skip-gram model of Word2vec and obtained the corresponding word vectors to characterize the semantic features. In addition, we have customized the dictionary of domain-specific terms to ensure that essential terms are not separated. The training parameters are shown in TABLE 1.

Standardized results documents and production data are unique data resources organized into tabular form via fixed tabs. In addition, label encoding is employed to convert

character features to numeric features for standardized result files. The normalized method avoids erroneous feature gradients caused by category and numerical differences.

C. RESULTS FILE ATTRIBUTE SIMPLIFICATION

Welding process specifications (WPS) are essential documents for welding manufacturing and are used to support the task of extracting relationships from standardized result files. The critical attributes that make up the WPS include weld method, weld joint, weld groove, assembly parameters, base material parameters, preheat, and other information in the welding process design. We collected standard WPS files for bogie welding fabrication of high-speed trains. Selected welding position (Position), blunt edge range (Blunt), assembly gap (Assembly), preheating temperature (Preheat), and gas flow rate (Flow) as decision attributes to complete the property reduction. The detailed information is listed in TABLE 2. The conditional attributes have 9 categories: weld method (Method), weld type (Type), weld groove (Groove), the base material 1 (Base-1), the base material 2 (Base-2), the thickness of base material 1 (Min-1, Max-1), and thickness of base material 2 (Min-2, Max-2). Significance (Sig) and weights (Weight) characterize the influence of conditional attributes on decision attributes. As shown in TABLE 2, the invalid attribute’s significance and weight are denoted as “-,” defined by a significance threshold of 0.01.

D. MODEL TRAINING AND EVALUATION

We train models based on the trained word embedding in the supervised condition to accomplish unstructured document relationship extraction. The 1832 sentence-level data related to welding manufacturing are divided into training, validation, and test sets and contain five relationship categories (belong_to, reference, requirement, applicable_to, unknown). The detailed information is shown in TABLE 3. Furthermore, we run programs written via the python programming language (version:3.7) in the TensorFlow framework (version: 1.14.0).

Accuracy is the commonly used evaluation metric in most conditions. However, in classification problems, the accuracy calculation relies on large sample categories and has low

TABLE 2. Property simplification results.

Attributes	Position		Flow		Preheat		Assembly		Blunt	
	Sig	Weight	Sig	Weight	Sig	Weight	Sig	Weight	Sig	Weight
Method	-	-	0.0545	0.3158	0.0182	0.1818	-	-	0.0182	0.0476
Type	-	-	-	-	-	-	-	-	-	-
Groove	0.1455	0.6957	-	-	0.0182	0.1818	0.1000	0.5000	0.1273	0.3333
Base-1	-	-	0.0545	0.3158	-	-	0.0182	0.0909	0.0727	0.1905
Base-2	-	-	0.0364	0.2105	-	-	-	-	0.1091	0.2857
Min-1	-	-	-	-	-	-	-	-	0.0182	0.0476
Max-1	0.0182	0.0870	-	-	0.0182	0.1818	-	-	-	-
Min-2	-	-	0.0182	0.1053	-	-	0.0182	0.0909	-	-
Max-2	0.0455	0.2173	-	-	0.0273	0.2727	0.0636	0.3182	-	-

TABLE 3. Relationship classification information.

Categories	Train	Validation	Test	Total	Example
belong_to	298	37	37	372	The welding position contains PA. (PA, belong_to, welding position)
reference	357	45	45	447	Bogie welding reference EN 15085. (Bogie welding, reference, EN 15085)
requirement	266	33	33	332	Welding grade CP CA requires defect grade B. (CP CA, requirement, defect grade B)
applicable_to	243	30	30	303	ISO 17638:2016 applies to the magnetic particle inspection in the non-destructive testing of welds. (ISO 17638:2016, applicable_to, magnetic particle inspection)
unknown	302	38	38	378	Relationships cannot be defined via belong_to, reference, requirement, and applicable_to.
Total	1466	183	183	1832	

confidence for small sample categories due to different sample sizes. Hence, considering the effect of the sample category employing Precision (P), Recall (R), and F1-score (F) as model evaluation metrics. We can calculate these indicators according to True Positive (Both true and predicted categories are positive examples, TP), False Positive (The true category is negative and the predicted category is positive, FP), and True Negative (Both true and predicted categories are negative examples, TN), and False Negative (The true category is positive and the predicted category is negative, FN). The calculation process is listed in (16) and (17). Macro-averaging effectively avoids metric errors caused by category imbalances and is obtained by calculating the arithmetic mean of each metric for each category to improve the credibility of the metrics. The calculation process is shown in (18) and (19).

$$P = \frac{TP}{TP + FP}, R = \frac{TP}{TP + FN} \quad (16)$$

$$F = \frac{2 \times P \times R}{P + R} \quad (17)$$

$$\text{Macro_P} = \frac{1}{n} \sum_{i=1}^n P_i, \text{Macro_R} = \frac{1}{n} \sum_{i=1}^n R_i \quad (18)$$

$$\text{Macro_F} = \frac{1}{n} \sum_{i=1}^n F_i \quad (19)$$

IV. RESULTS AND DISCUSSION

A. RELATIONSHIP EXTRACTION RESULTS

Unstructured data relationships in weld manufacturing are extracted, and the extracted models are verified by the test set with Precision, Recall, and F1-score metrics. Considering the effect of different relationship categories on the accuracy, we calculate the macro averages of the corresponding indicators. The actual result data are listed in TABLE 4.

As shown in TABLE 4, BiLSTM+Attention achieves better macro-average metrics results than CR-CNN. For BiLSTM+Attention, the F1-score of the “belong_to” category is the highest, i.e., 0.912, while the category reference F1-score is the lowest, i.e., 0.688. The high score of category “belong_to” may be due to the independent entities that make the sentence relationship feature clear, such as “CP C1 belong_to the weld quality level”. The low scores in category reference may be because most target entities are composed of multiple independent entities, making the sentence relationship characteristics ambiguous, such as “Arc bolt welding of metallic materials reference ENISO14555”. Hence, enhanced entity features may positively affect the extraction of welding manufacturing relationships.

B. RELATIONSHIP COMPLEMENT RESULTS

The 110 actual production welding procedures were collected as a sample to support the validation of the model. The sample

TABLE 4. Models results and baseline.

Models	Metrics	belong_to	reference	requirement	applicable_to	unknown	Macro-average
BiLSTM+Attention	Precision	0.838	0.568	0.806	0.857	0.872	0.788
	Recall	1.000	0.875	0.625	0.878	0.850	0.846
	F1-score	0.912	0.688	0.704	0.867	0.861	0.816
CR-CNN	Precision	0.912	0.704	0.670	0.818	0.850	0.791
	Recall	1.000	0.792	0.763	0.878	0.638	0.814
	F1-score	0.954	0.745	0.713	0.847	0.729	0.802

TABLE 5. Data on conditional attributes.

ID	Method	Type	Groove	Base-1	Base-2	Min-1 (mm)	Max-1 (mm)	Min-2 (mm)	Max-2 (mm)
1	t135	BW	HY	S355J2G3-EN10025	S355J2W-EN10025	35	35	14	14
2	t135	BW	HY	S355J2G3-EN10025	S355J2W-EN10025	30	30	18	18
3	t135	BW	HY+a	S355J2W-EN10025	S355J2W-EN10025	14	14	12	18
4	t135	BW	HY+a	S355J2W-EN10025	S355J2G3-EN10025	14	14	35	35
5	t135	BW	HY+a	S355J2W-EN10025	S355J2W-EN10025	14	18	12	12
6	t135	BW	HY+a	S355J2G3-EN10025	S355J2W-EN10025	20	20	14	14
7	t135	FW	a	S355J2W+N-EN10025-5	S355J2W+N-EN10025-5	9	9	14	14
8	t135	FW	a	S355J2W+N-EN10025-5	S355J2W+N-EN10025-5	10	14	12	18
...				
110	v135	BW	U/V	S355J2H-EN10210	S355J2H-EN 10210	15	15	15	15

attributes were divided into conditional attributes (Method, Type, Groove, Base-1, Base-2, Min-1, Max-1, Min-2, Max-2) and decision attributes (Position, Assembly, Blunt, Flow, Preheat). Conditional attributes can be considered known information, and decision attributes can be considered target information. The detailed data information is presented in TABLE 5 and TABLE 6.

As shown in TABLE 5 and TABLE 6, 9 conditional attributes were selected to support the relationship acquisition of 5 decision attributes. Conditional attributes have positive or negative information gain for each decision attribute. Therefore, the neighborhood rough set is employed to approximate the useless attributes to improve the accuracy of the attribute relationships. Moreover, this reduced attribute forms an inference relationship with the decision attribute. The attribute simplification results are presented in TABLE 3, and the analysis of the results can be summarized as follows:

(1) For a certain structure and assembly conditions of welding, category Groove contains important structural information, such as V-bevels are generally butt welds. Under certain conditions, it should be preferred to the welding position PA. Therefore, category Groove has extensive information gained for selecting the weld position.

(2) Suitable welding assembly gap is one of the critical factors in ensuring welding quality, which significantly correlates with the geometric form of the weld. Categories Groove,

Max-2, and Max-1 contain important geometric information that positively impacts the selection of attribute Assembly.

(3) The blunt edge refers to the part without a groove in the thickness direction, which is used to prevent welding penetration. Groove and partial plate thickness information are considered factors affecting the selection of blunt edges. And Category Method, Base-1, and Base-2 also influence the choice of a blunt edge due to different base materials and welding methods with different melting depths.

(4) The choice of welding gas flow rate directly affects the quality of welding production. In practice, the gas flow rate is related to the welding method, the material of the welded part, and some plate thickness information. Therefore, the simplification results have credibility.

(5) Preheating before welding effectively controls welding quality, incredibly thick plate welding. Maximum plate thickness information positively influences the selection of preheating temperature. In addition, different bevel geometries and heat flow densities will result in different preheating, which makes property Method and property Groove influence property Preheat.

The simplification results are highly similar to the actual production experience based on the above information. Therefore, we extracted the relationship between different decision attribute values and conditional attribute values

TABLE 6. Data on decision attributes.

ID	Position	Assembly (mm)	Blunt (mm)	Flow (l/min)	Preheat
1	PA	0-1	2	10-15	150°C
2	PA	0-1	2	10-15	150°C
3	PA+PB	2	0-1	10-15	/
4	PA+PB	2	0-1	10-15	150°C
5	PA+PB	0-1	0	10-15	/
6	PA	0-1	0.5-1	10-15	/
7	PB	0-1	0	10-15	/
8	PB	0-1	0	10-15	/
...		
110	PA	0-0.5	2	15-20	/

TABLE 7. The results of rule and relationship extraction.

Categories	Rules	Relationships	Confidence
Position	{'HY+a'}=>{'PA+PB'}	(HY+a, requirement, PA+PB)	0.9032
	{'HY'}=>{'PA'}	(HY, requirement, PA)	1.0000
	{'a'}=>{'PB'}	(a, requirement, PB)	1.0000
	{'Max-2: 18', 'HY+a'}=>{'PA+PB'}	(Max-2: 18 and HY+a, requirement, PA+PB)	0.9333
Assembly	{'Base-1: S355J2W+N-EN10025-5'}=>{'0-1'}	(S355J2W+N-EN10025-5, requirement, 0-1)	0.9200
	{'Min-2: 14'}=>{'0-1'}	(Min-2: 14, requirement, 0-1)	0.9286
Blunt	{'a', 't135'}=>{'0'}	(a and t135, requirement, 0)	1.0000
	{'a'}=>{'0'}	(a, requirement, 0)	1.0000
Flow	{'Base-2: Q345E-GB/T 1591'}=>{'15-20'}	(Q345E-GB/T 1591, requirement, 15-20)	0.9260
	{'Base-2: Q345E-GB/T 1591', 't135'}=>{'15-20'}	(Q345E-GB/T 1591 and t135, requirement, 15-20)	0.9260
	{'Base-2: S355J2W+N-EN10025-5', 't135'}=>{'10-15'}	(S355J2W+N-EN10025-5 and t135, requirement, 10-15)	0.9583
	{'Min-2: 14', 't135'}=>{'10-15'}	(Min-2: 14 and t135, requirement, 10-15)	1.0000
Preheat	{'Max-2: 12'}=>{'/'}	(Max-2: 12, requirement, /)	0.9333
	{'Max-1: 10'}=>{'/'}	(Max-1: 10, requirement, /)	0.9375
	{'Max-2: 12', 't135'}=>{'/'}	(Max-2: 12 and t135, requirement, /)	0.9286
	{'t135', 'Max-1: 10'}=>{'/'}	(t135 and Max-1: 10, requirement, /)	0.9333

based on the simplification results via association rules. The extracted relations are described in the form of a triplet whose easy-to-implement knowledge graph embedding. Detailed rules and related information are listed in TABLE 7, where nodes with the same value are considered the same.

TABLE 7 summarizes the relationship data extracted from the engineering documents and lists the extraction results' corresponding rules and confidence levels. It can be seen from the results that the extracted relationship has high quality and reliability, and the accompanying confidence is higher than 0.9. A possible explanation is that deleting redundant attributes will reduce the negative gain of data on the extraction results. Another possible explanation we found is a strong correlation between the effective attribute and the

result attribute. Therefore, we think the proposed method can effectively handle the relationship extraction of many redundant data and provide a reference for engineering data processing similar structures.

C. KNOWLEDGE GRAPH EMBEDDING

A knowledge graph is a networked form of data storage that positively impacts transforming knowledge and relationships into practical engineering applications. Relationships as an essential factor in knowledge graph construction are focused on in this paper. The relationship extraction in welding manufacturing differs from traditional extraction methods due to the complexity and specialization of engineering data. This study divided the relationship extraction task into three

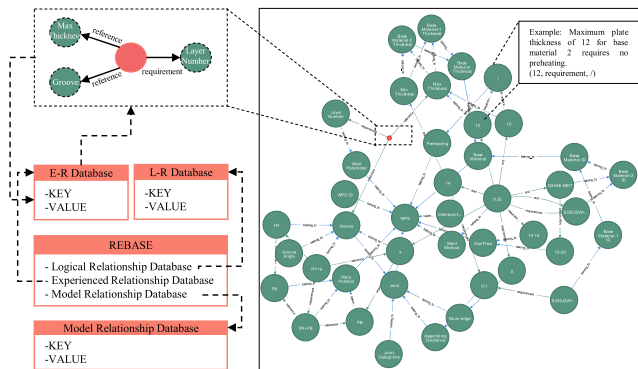


FIGURE 5. Knowledge graph of welding process specification.

subtasks: unstructured data extraction, relationship complementation, and experience relationships processing. We have employed and trained the BiLSTM+Attention model for such relationship extraction. Production result documents such as welding plans and welding process specifications contain much relational knowledge, which is also unique to welding products in data. Neighborhood rough set-based association rules are proposed to transform the result file into a structural relationship of the form (entity, relation, entity). In addition, relational databases are constructed to support knowledge graph embedding and knowledge reasoning of empirical relationships.

The data collected in the welding production of high-speed train bogies are used for the research of relationship extraction. The knowledge graph based on WPS data is described in FIGURE 5. Relationships in unstructured data and result files are characterized via the entity node-relationship-entity node form. Empirical relationships are characterized through the entity node-relationship node-entity node form, and the relationship nodes correspond to the content of the associated relationship base.

As shown in FIGURE 5, the left side is the empirical relationship process; the right side is the knowledge graph about welding process specifications. The results could be applied to solve complex problems in welding manufacturing, such as knowledge retrieval, intelligent question and answer, and decision-making.

(1) Knowledge retrieval. For example, we only enter the weld position, and this correlation and node can be obtained. Based on the search results, we can also obtain valid information such as the weld position containing PA, PB, and PA+PB, an element of the welding process specifications.

(2) Intelligent question and answer. For example, for a question: What is the preheating temperature for a welding method of t135 and a maximum plate thickness of 12 mm for base material 2? The answer (Preheat temperature of "/>) can be obtained according to the knowledge graph.

(3) Decision-making. For example, for a joint with a bevel "V" and a plate thickness of 12, a 3-layer weld layer can be acquired based on the corresponding empirical relationship. Furthermore, welding positions, assembly gaps,

blunt edge preheating, etc., can also be obtained through the knowledge graph, which can be used to make decisions on welding process design.

V. CONCLUSION

This work describes methods for extracting and processing relationships under different data types and structures in welding manufacturing.

(i) Unstructured data relationships are extracted by employing the BiLSTM+Attention model. This model is trained based on pre-trained word vectors under supervised conditions and has good Macro-average metrics (0.788 for Precision, 0.846 for Recall, and 0.816 for F1-score).

(ii) The actual engineering file relationships are extracted through a neighborhood rough set-based association rule model. Several relationship rules are obtained from 110 engineering data and are consistent with engineering experience.

(iii) Relational databases and relational nodes are introduced to implement knowledge graph embeddings of empirical relationships with positive engineering application effects.

The proposed method can complete the extraction of welding relations, especially suitable for processing a large number of redundant data. The domain knowledge graph based on the extracted relationship can support the solution of complex engineering problems such as domain knowledge retrieval, intelligent question answering, and expert decision-making. Furthermore, our research may extend data applicability, improve model accuracy, and efficient engineering applications based on obtained results.

REFERENCES

- [1] A. Theorin, K. Bengtsson, J. Provost, M. Lieder, C. Johnsson, T. Lundholm, and B. Lennartson, "An event-driven manufacturing information system architecture for industry 4.0," *Int. J. Prod. Res.*, vol. 55, no. 5, pp. 1297–1311, Mar. 2017, doi: [10.1080/00207543.2016.1201604](https://doi.org/10.1080/00207543.2016.1201604).
- [2] A. K.-F. Lui, Y.-H. Chan, and M.-F. Leung, "Modelling of destinations for data-driven pedestrian trajectory prediction in public buildings," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2021, pp. 1709–1717, doi: [10.1109/BigData52589.2021.9671813](https://doi.org/10.1109/BigData52589.2021.9671813).
- [3] J. Chen, Z. Zhang, X. Jia, and J. Geng, "Research on ship welding process planning with case-based reasoning," in *Proc. 2nd IEEE Int. Conf. Inf. Manage. Eng.*, Apr. 2010, pp. 605–608, doi: [10.1109/ICIME.2010.5478182](https://doi.org/10.1109/ICIME.2010.5478182).
- [4] C. Favi, R. Garziera, and F. Campi, "A rule-based system to promote design for manufacturing and assembly in the development of welded structure: Method and tool proposition," *Appl. Sci.*, vol. 11, no. 5, p. 2326, Mar. 2021, doi: [10.3390/app11052326](https://doi.org/10.3390/app11052326).
- [5] T. W. Liao, "Classification of welding flaw types with fuzzy expert systems," *Expert Syst. Appl.*, vol. 25, no. 1, pp. 101–111, 2003, doi: [10.1016/S0957-4174\(03\)00010-1](https://doi.org/10.1016/S0957-4174(03)00010-1).
- [6] J. Feng, Y. Yao, S. Lu, and Y. Liu, "Domain knowledge-based deep-broad learning framework for fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 68, no. 4, pp. 3454–3464, Apr. 2021, doi: [10.1109/TIE.2020.2982085](https://doi.org/10.1109/TIE.2020.2982085).
- [7] J.-J. Cui and D.-Y. Wang, "Application of knowledge-based engineering in ship structural design and optimization," *Ocean Eng.*, vol. 72, pp. 124–139, Nov. 2013, doi: [10.1016/j.oceaneng.2013.06.013](https://doi.org/10.1016/j.oceaneng.2013.06.013).
- [8] D. Sagheer and F. Sukkar, "A template-based information extraction system for text understanding," *Int. J. Comput. Appl.*, vol. 182, no. 28, pp. 28–33, Nov. 2018, doi: [10.5120/ijca2018918167](https://doi.org/10.5120/ijca2018918167).
- [9] J. Pereira, "Supervised learning for relationship extraction from textual documents," M.S. thesis, Dept. Perustieteiden Korkeakoulu, Aalto Univ., Helsinki, Finland, 2013.

- [10] Z. Zhang, "Mining relational data from text: From strictly supervised to weakly supervised learning," *Inf. Syst.*, vol. 33, no. 3, pp. 300–314, May 2008, doi: [10.1016/j.is.2007.10.002](https://doi.org/10.1016/j.is.2007.10.002).
- [11] K. Xu *et al.*, "Semantic relation classification via convolutional neural networks with simple negative sampling," *Comput. Sci.*, vol. 71, no. 7, pp. 941–949, Jun. 2015, doi: [10.48550/arXiv.1506.07650](https://doi.org/10.48550/arXiv.1506.07650).
- [12] L. Yang *et al.*, "A dependency-based neural network for relation classification," *Comput. Sci.*, vol. 2, pp. 285–290, Jul. 2015, doi: [10.3115/v1/P15-2047](https://doi.org/10.3115/v1/P15-2047).
- [13] S. Zheng, Y. Hao, D. Lu, H. Bao, J. Xu, H. Hao, and B. Xu, "Joint entity and relation extraction based on a hybrid neural network," *Neurocomputing*, vol. 257, pp. 59–66, Sep. 2017, doi: [10.1016/j.neucom.2016.12.075](https://doi.org/10.1016/j.neucom.2016.12.075).
- [14] C. N. D. Santos, X. Bing, and B. Zhou, "Classifying relations by ranking with convolutional neural networks," *Comput. Sci.*, vol. 86, pp. 132–137, May 2015, doi: [10.48550/arXiv.1504.06580](https://doi.org/10.48550/arXiv.1504.06580).
- [15] L. Wang, Z. Cao, G. De Melo, and Z. Liu, "Relation classification via multi-level attention CNNs," in *Proc. 54th Annu. Meeting Assoc. Comput. Linguistics*, vol. 1, 2016, pp. 1–10, doi: [10.18653/v1/P16-1123](https://doi.org/10.18653/v1/P16-1123).
- [16] M. Miwa and M. Bansal, "End-to-end relation extraction using LSTMs on sequences and tree structures," in *Proc. 54th Annu. Meeting Assoc. Comput. Linguistics*, vol. 1, 2016, pp. 1–13, doi: [10.18653/v1/P16-1105](https://doi.org/10.18653/v1/P16-1105).
- [17] G. Ji, K. Liu, S. He, and J. Zhao, "Distant supervision for relation extraction with sentence-level attention and entity descriptions," in *Proc. Nat. Conf. Artif. Intell.*, 2017, vol. 31, no. 1, pp. 1–7, doi: <https://doi.org/10.1609/aaai.v31i1.10953>.
- [18] A. Carlson *et al.*, "Toward an architecture for never-ending language learning," in *Proc. 24th AAAI Conf. Artif. Intell.*, Atlanta, GA, USA, 2010.
- [19] N. Milosevic and W. Thielemann, "Relationship extraction for knowledge graph creation from biomedical literature," 2022, *arXiv:2201.01647*.
- [20] P. Zhou, W. Shi, J. Tian, Z. Qi, B. Li, H. Hao, and B. Xu, "Attention-based bidirectional long short-term memory networks for relation classification," in *Proc. 54th Annu. Meeting Assoc. Comput. Linguistics*, vol. 2, 2016, pp. 1–6, doi: [10.18653/v1/P16-2034](https://doi.org/10.18653/v1/P16-2034).
- [21] T. Mikolov *et al.*, "Efficient estimation of word representations in vector space," *Comput. Sci.*, vol. 2013, pp. 1–12, Jan. 2013, doi: [10.48550/arXiv.1301.3781](https://doi.org/10.48550/arXiv.1301.3781).
- [22] Q. Hu, D. Yu, J. Liu, and C. Wu, "Neighborhood rough set based heterogeneous feature subset selection," *Inf. Sci.*, vol. 178, no. 18, pp. 3577–3594, 2008, doi: [10.1016/j.ins.2008.05.024](https://doi.org/10.1016/j.ins.2008.05.024).
- [23] R. Srikant and R. Agrawal, "Mining quantitative association rules in large relational tables," *ACM SIGMOD Rec.*, vol. 25, no. 2, pp. 1–12, Jun. 1996, doi: [10.1145/235968.233311](https://doi.org/10.1145/235968.233311).
- [24] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735).



KAINAN GUAN was born in Luoyang, China, in 1994. He received the M.S. degree in materials science and engineering from Dalian Jiaotong University, where he is currently pursuing the Ph.D. degree in materials science and engineering.

His research interests include welding production informatization and intelligence, application of knowledge engineering in traditional manufacturing, expert decision algorithms, and data support. Some of his research results have been successfully applied to practical welding production.



LIANG DU received the M.S. degree in materials science and engineering from Dalian Jiaotong University, in 2022.

His research interests include welding informatization and digitization. Some of his research results have been successfully applied to practical welding production.



XINHUA YANG received the Ph.D. degree in mechanical engineering from the Dalian University of Technology, in 2005. He is currently a Professor and the Doctoral Supervisor with the College of Materials Science and Engineering, Dalian Jiaotong University. He has authored over 30 papers in peer-reviewed international journals and conferences, such as *Entropy*, *Material Science and Engineering A*, *Soft Computing*, *Applied Soft Computing Journal*, *Kybernetes*, *Symmetry-Basel*, and *Transactions of the China Welding Institution*. His research interests include welding mechanics, structural reliability analysis, and intelligent technology and application.

...