

Article

Product Quality Detection through Manufacturing Process Based on Sequential Patterns Considering Deep Semantic Learning and Process Rules

Liguo Yao ^{1,2}, Haisong Huang ^{1,*} and Shih-Huan Chen ²

¹ Key Laboratory of Advanced Manufacturing Technology, Ministry of Education, Guizhou University, Guiyang 550025, China; yaoliguo1990@163.com

² Department of Industrial Engineering and Management, Yuan Ze University, Taoyuan 32003, Taiwan; zxc980071@gmail.com

* Correspondence: hshuang@gzu.edu.cn; Tel.: +86-183-0265-1543

Received: 28 May 2020; Accepted: 26 June 2020; Published: 28 June 2020



Abstract: Companies accumulate a large amount of production process data during product manufacturing. Sequence data from the mining production process can enable a company to evaluate the manufacturing process, to find the key factors affecting product quality, and to improve product quality. However, the production process mainly exists in the form of text. To solve this problem, we propose a novel frequent pattern mining algorithm (EABMC) based on the text context semantics and rules of the manufacturing process to remove redundant sequences and to obtain good mining results. In this algorithm, first, we use embeddings from language models (ELMo) to improve the process of text similarity matching and to classify similar semantic processes into one class. Then, the manufacturing process unit (MPU) is proposed by extracting the characteristics of manufacturing process data according to the constraints of the manufacturing process and other conditions. The above two steps cause the complex manufacturing process sequence to merge and simplify. Once again, a frequent pattern mining algorithm (CloFAST) is used to explore the important manufacturing process relationships behind a large amount of manufacturing data. In addition, taking the data from a production enterprise in Guizhou Province as an example, the validity of the method is verified. Compared with other methods, this method is shown to have greater mining efficiency and better results and can find out the key factors that affect product quality, especially for text data.

Keywords: product quality detection; manufacturing process diagnostics; deep semantic learning; manufacturing process rule; frequent pattern mining

1. Introduction

With the advancement of sustainable manufacturing technologies and the development of the world's manufacturing industry, manufacturing companies are paying more and more attention to the connection between product and manufacturing data [1]. Through the information upgrade of manufacturing equipment and software, data in the domain fields can be acquired, stored, and analyzed simultaneously in the manufacturing process and then is fed back to production to improve the production efficiency and yield, to shorten the manufacturing cycle, and to improve product quality. This has become a trend in the manufacturing industry [2,3]. Historical data, such as design information and manufacturing information in an enterprise, contain a rich product design and manufacturing knowledge. Mining and analysis have become an important means for enterprises to enhance their competitiveness [4]. In the manufacturing industry, process planning is a kind of experience-based complex knowledge application activity. The mining of process planning data can ensure product

quality. Therefore, determining how to turn the data into useful knowledge that helps manufacturing process diagnostics and improves product quality has become the focus of research [5].

Data mining technology is one of ten emerging technologies that is predicted to “change the world in the 21st century”. Data mining refers to the discovery of hidden information or knowledge from large amounts of data by combining different technologies such as artificial intelligence, statistical analysis, computer science, machine learning, pattern recognition, expert systems, databases, and graph visualization. Different data mining methods, such as association rule mining [6], sequence pattern mining [7], classification [8], clustering [9], text mining [10], and knowledge transfer [11], will also have different effects on data processing and mining results. Therefore, it is especially important to choose the appropriate mining method [12].

CAPP (Computer Aided Process Planning) refers to the use of computer software and hardware technology to support the environment, where the computer is used to perform numerical calculations, to make logical judgments, and to carry out reasoning functions to formulate parts of the machining process [13]. In current manufacturing cases, the technology combines CAPP with data mining to get good results. Implicit process experience and knowledge can be obtained from these historical manufacturing information data. The records of daily manufacturing activities are stored in the manufacturing database of the company, which has important attributes such as the manufacturing ID, the date of manufacturing, the equipment ID, the manufacturing type, etc. [14]. Due to the numerous constraints of process planning problems and the influence of process planning personnel, the design and optimization of process planning methods are difficult points of process planning problems. To obtain more suitable process planning mining results for the manufacturing industry, we should build relevant models according to the characteristics of process planning data. Designing special domain models for process planning problems can improve the efficiency and precision of mining to achieve good process planning knowledge.

Nowadays, many studies are devoted to discovering factors that cause the quality of products in the manufacturing process to decline. Some experts have used the association rule [6] for mining process planning knowledge in the past. However, the association rule pays more attention to the relationships among transactions, ignoring the chronological order of events. Sequential pattern mining [7] is focused on timing-based event-related mining, so it is more suitable than the association rule for the discovery of process planning knowledge. Due to the special nature of process knowledge, often in the form of text, some process steps that are combined and cannot be presented separately could have good mining results. Therefore, the article researches natural language processing technology and encapsulates these processes in the form of a manufacturing process unit (MPU). As an event, the excavation time can be reduced and the excavation accuracy can be improved by using MPU.

This article takes the manufacturing of a wheel hub produced by an enterprise in Guizhou Province, China as the research object and builds an unstructured quality analysis data set by integrating the text data of each production link of wheel manufacturing and proposes a novel frequently closed sequential pattern mining algorithm based on the manufacturing process text contextual semantics and manufacturing rules (EABMC) to obtain key factors affecting product quality. By performing sequence pattern mining on the quality analysis data set, it can help the company to discover the abnormal quality and its influencing factors in the manufacturing process of wheel hub products and to find out the correct sequence relationship that affects quality. It not only can accurately locate quality problems but also help companies improve process parameters.

The rest of the sections in this paper are organized as follows: Section 2 reviews previous work related to developments and applications of similarity measurements of text semantics, different algorithms for the manufacturing process, and frequent sequential pattern mining. Section 3 introduces the proposed algorithm, EABMC. Section 4 illustrates a case study to verify the feasibility of the proposed method and the performance of different thresholds on results and compares them with other methods. The results and analysis of the case study are also presented in this section. Finally, Section 5 provides major conclusions and points out the future directions of this paper.

2. Literature Review

2.1. Text Mining and Semantic Similarity Measurement

Text similarity measurement is an important task in natural language processing. Salton and Buckley [15] proposed the term frequency-inverse document frequency (TF-IDF), which converts the text into a high-dimensional and sparse word matrix and then uses cosine similarity to calculate the text similarity. Mikolov et al. [16] proposed Word2vec, which improved the Skip-gram model to obtain higher quality vectors faster. Blei et al. [17] proposed latent Dirichlet allocation (LDA), which is a generative probabilistic model that is used for the collection of text corpora data. Pennington et al. [18] proposed a new clear and interpretable language model, GloVe, to form word vectors. The essence of the model is to integrate the latest matrix factorization and word2vec at that time and to use nonzero data in the global word cooccurrence matrix for training instead of using only the local window information of a word. Latent Semantic Analysis (LSA) [19] is an algorithm that obtains a semantic representation of words or paragraphs through statistical calculation. By mapping the high-dimensional document–word vectors to the latent semantic space, the concepts related to documents and terms are extracted and the relationship between documents and terms is analyzed. The core of LSA is to extract the subject of the document based on the Singular Value Decomposition (SVD) matrix decomposition method, which effectively solves the problem of polysemy and polysemy that traditional vector models cannot handle. ELMo [20] is a new word vector representation model based on a deep learning framework. This model not only can represent the grammatical and semantic features of vocabulary but also changes with the change of context. The model is essentially a combination of internal hidden state features of a two-way language model based on large-scale corpus training.

2.2. Application of Different Algorithms in the Manufacturing Process

Manufacturing process planning is one of the key aspects of a product's lifecycle [21]. Decision support methodology will help companies to improve their production efficiency [22]. Many scholars have applied advanced algorithms to develop a good manufacturing process to help companies with their decision-making. Su et al. [23] proposed a genetic algorithm based on edge selection to solve the optimal sequence of processing operations with minimum processing cost and satisfying all priority constraints. The strategy based on edge selection can generate feasible solutions during initialization and can ensure that each feasible solution is generated with an acceptable level of probability, thereby improving the convergence efficiency of the Genetic Algorithm (GA). Phung et al. [24] presented an improved clustering algorithm to optimize the operation sequence. The key concept of this method is to first check the priority constraints to select all possible following operations for the last operation in the sequence and then to compare their driving costs to select the best feasible operation with the minimum driving cost in the sequence.

Wang et al. [25] proposed a general sorting method for machining features, which is used in the process planning of complex parts to solve one-to-many mapping between machining features and machining operations that leads to an increase in non-cutting tool paths. Milošević et al. [26] proposed a system for a distributed and collaborative environment that can help manufacturing companies and experts discuss, recommend, evaluate, and select the best process plan for manufacturing a series of parts. Cheng and Wang. [27] used a data-driven matching method for processing parameters as a process knowledge service. The ability to quickly set process parameters for a new, complex product responds to the needs of customers in a process manufacturing environment.

2.3. Applications of Sequential Pattern Mining Algorithm

Sequential pattern mining (SPM) was first applied to shopping data to determine customers' purchase rules. Subsequently, it has been widely used for travel recommendations, movie recommendations, supply chain diagnosis, and many other practical problems [28]. Huang et al. [29] used frequent closed sequence mining technology (ClaSP) to analyze cargo

transportation data to help transportation companies to discover potential factors that reduce the quality of cargo during transportation. The results showed that this method can be used to determine the causes of low-quality transportation services. Amiri et al. [30] proposed a new prediction model based on sequential pattern mining. This model considers the correlations among different resources and extracts the application's behavioral pattern independently of the fixed pattern length, thereby clearly indicating that the pattern-based mining model can provide novel and useful perspectives to solve some of the problems involved in predicting application workloads. Tsai et al. [31] proposed an effective sequence classification method based on two-stage SPM. In the first stage, during the sequential pattern mining process, if a pattern is subsequent to other sequential patterns, the redundant sequential pattern will be identified. A list of compact sequential modes (not including redundant modes) is generated and used as a representative function of the second stage. Huynh et al. [32] proposed a parallel method, Multiple threads CM-SPADE (MCM-SPADE), for multi-core processor systems. This method uses the multithreading technology of the SPM database, which can improve the performance of SPADE and Co-occurrence MAP-SPADE (CM-SPADE).

Many kinds of literature have proposed ways to combine other algorithms with sequential pattern mining to achieve better results. Tarus et al. [33] proposed a knowledge-based hybrid recommendation system based on ontology and sequential pattern mining for recommending e-learning resources to learners. In the proposed recommendation method, ontology is used to model and express domain knowledge about learners and learning resources while the SPM algorithm discovers the learner's sequential learning model. Ding et al. [34] proposed a spatial sequence model of coastal land use based on association rules to mine interesting sequential patterns of land use along the sea and land along the coastal zone. Yuan et al. [28] proposed an SPM algorithm to mine the failure sequence pattern in text data. The algorithm aims to solve the problem of poor structure of text data and the existence of multiple forms of text expression in the same concept. The traditional SPM algorithm cannot be directly applied to text data. Experiments show that the algorithm can effectively mine sequential patterns in text data.

Sequential pattern mining is a data mining method for obtaining frequent sequential patterns in a sequential database. The above researchers studied different algorithms for the manufacturing process to optimize the operation sequence and used a sequential pattern mining algorithm to mine different types of data. However, they did not combine their methods of seeking the rules behind manufacturing features from the perspective of discovering quality reasons. A novel, frequently closed sequential pattern mining algorithm based on the manufacturing process of text contextual semantics and the manufacturing process unit was proposed to obtain key factors of product quality.

3. Wheel Hub Quality Data Analysis

For manufacturing companies, their product quality control is very critical. In the manufacturing process of the wheel hub, the company has accumulated a large amount of production data and inspection data. In the era of big data, how to use the mining technology of industrial big data to find the law of quality transfer from massive time series production and manufacturing data to achieve effective control and improvement of product quality is a new problem faced by manufacturing enterprises. Therefore, quality data analysis has become an important requirement of industrial big data. For the collected wheel production process data, the traditional probability statistical method and data mining algorithm are used to build a complete and targeted analysis model, and the key factors that affect the quality of the wheel hub can be found through correlation analysis for subsequent quality improvement and production. Process improvement provides reasonable data support. Figure 1 shows the framework of the product quality analysis process based on manufacturing data.

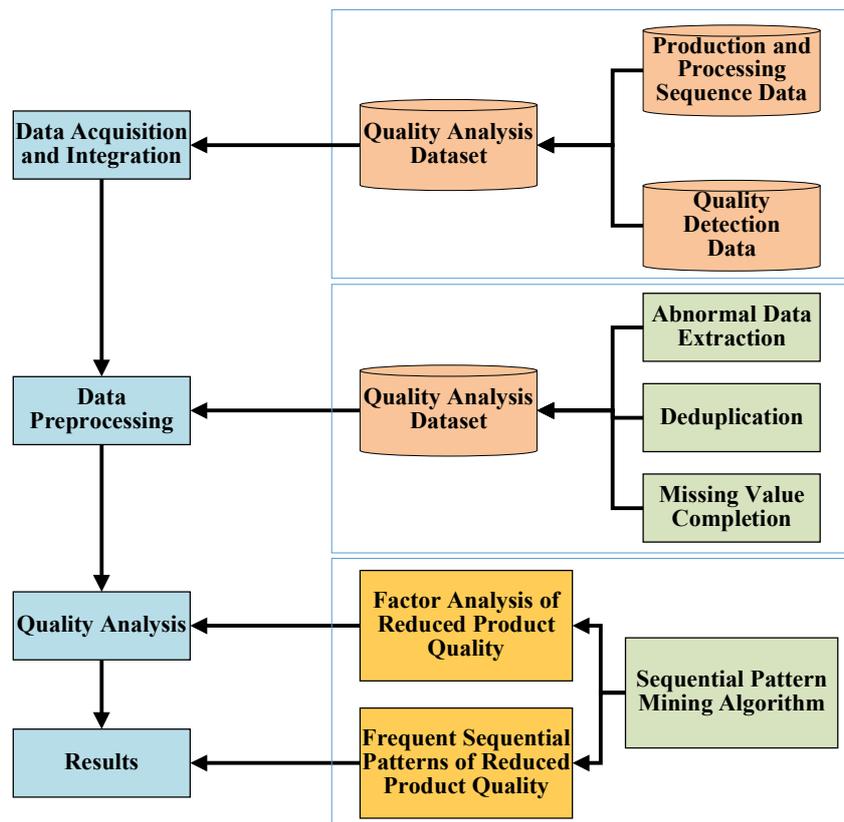


Figure 1. Product quality analysis process based on manufacturing data.

The wheel hub production process flow is incoming materials→forging→X-ray inspection→appearance inspection→deburring→dissolution treatment→aging treatment→lathe surface→lathe shape→fine slot→dovetail slot→deburring→drilling→milling inner cavity→tapping→drilling→cleaning→detection→anodizing→painting→drying→sampling detection→painting→shipping.

The product quality data is defined in Equation (1):

$$Quality_data = (Product_id, Quality_result) \quad (1)$$

where *Product_id* is the product ID and *Quality_result* is the product test results.

Manufacturing data are the relevant data generated during the production of the product. The manufacturing data for a single process are defined as shown in Equation (2):

$$Product_data = (Product_id, Equip_id, Product_time, Shift, Operator, Product_parameters) \quad (2)$$

where *Product_id* is the product ID; *Equip_id* is the device ID; *Product_time* is the machining time; *Shift* is the team ID; *Operator* is the operator ID; and *Product_parameters* is the processing parameter set.

By combining the above two data with the *Product_id* as the association, the product quality data based on the process were obtained as shown in Equation (3):

$$Product_Quality_data = \langle (Product_id, Equip_id, Product_time, Shift, Operator, Product_parameters), Quality_result \rangle \quad (3)$$

The wheel hub manufacturing data is the relevant data generated during the production procedure of the product, including the intermediate status generated during the manufacturing, such as

temperature, pressure, and other information; the equipment number; the coding information of the tooling or fixture used; equipment alarm information; qualified and unqualified product information; manufacturing time; quality detection data; and product material traceability information.

Product quality factor analysis can use data mining methods to find the influencing factors that cause bad products. In this paper, sequential pattern mining is used as the algorithm for wheel hub quality analysis. The names and contents of the processes in the production of the wheel hub, the characteristic values of the process parameters (mold temperature, forming temperature, billet temperature, and dissolution time), forging operator number (FO_ID), heat treatment operator number (HTO_ID), production equipment number (PE_ID), production shift number (PS_ID), production workshop number (PW_ID), and batch number (B_ID) are used as input to the sequential pattern mining algorithm, and the product quality detection results are divided into qualified and unqualified. The sequential pattern mining algorithm will output a series of frequent sequences that satisfy the support and confidence as the mining results. Because the traditional sequential pattern mining algorithm is not suitable for unstructured data such as text, to effectively solve the mining and analysis of the enterprise hub quality data, this paper improves the sequential pattern mining algorithm.

4. EABMC Sequential Patterns Mining Method

4.1. Measurement of Text Semantic Similarity Based on Contextual Word Embedding

Manufacturing process text is unstructured data with domain knowledge that contains a large number of manufacturing-related terms. However, because different operators have different understandings and applications of related terms, often, the same professional term is presented by different words. In text mining, different words are often represented by different labels, which not only increase the amount of data but also make the semantics process less readable. Traditional machine learning methods cannot effectively solve the problem of semantics term and context connection. Therefore, this article uses the method of contextual word embedding to represent all text vocabulary in the form of word vectors. In the process of word representation, this method learns the semantic correlations between words and uses the appearance of the words to get a contextual connection. Table 1 shows the characteristics of different language embedding models.

Table 1. Comparison of characteristics of the language embedding models.

Functions	Models	Bag-of-Words Model [15]	Topic Model [17]	Word Embedding Distance Model [16]	Latent Semantic Analysis Model [19]	Contextual Word Embedding Model [20]
Word embedding		No	Yes	Yes	Yes	Yes
Suitable for short text		No	Yes	Yes	Yes	Yes
Implies contextual semantic		No	No	No	Yes	Yes
Solve the problem of polysemy		No	No	No	Yes	Yes

The bag-of-words model uses one-hot encoding. In this model, the word has a value of 1 when it appears and the rest is 0. The vector dimension of words is the number of all occurrences of words. When the number of texts increases, it will increase the vector dimension and increase the amount of calculation. Any two words are independent of each other, so they cannot reflect the semantic relationship of the text. The topic model can give the theme of each document in the document set in the form of a probability distribution so that, after analyzing some documents to extract their theme distribution, they can perform text similarity work according to the theme distribution. The word embedding distance model is based on word2vec technology. After converting all the words into a vector, the cosine value between each word is calculated to obtain the similarity between the texts. Since words and vectors are in a one-to-one relationship, the problem of polysemy cannot be solved. The latent semantic analysis model is to reduce the dimension using singular value decomposition (SVD) in the concurrence matrix of word and document so that the similarity of text can be measured

by the cosine similarity of two low-dimensional vectors. The contextual word embedding model is no longer just a vector correspondence but a trained model. When in use, input a sentence or a paragraph into the model and the model will infer the word vector corresponding to each word according to the online text. One of the obvious benefits after doing this is that, for polysemous words, the polysemy can be understood in combination with the context before and after. Therefore, this paper uses a contextual word embedding model in the follow-up study.

ELMo is a pretrained contextual word embedding model, shown in Figure 2, which uses a bidirectional long-term short-term memory network (LSTM) language model consisting of a forward and a backward language model. The model solves two problems: one is the semantic and grammatical complexity of word usage, and the other contains the local changes. These usages should change as well. The representation of each word is a function of the entire input sentence. The specific method involves training a bidirectional LSTM model with a language model as the target on a large corpus and then uses the LSTM to generate the word representation. Given a sequence with N tokens (t_1, t_2, \dots, t_N) , the objective function is the maximum log-likelihood of a bidirectional (forward and backward) language model, as shown in Equation (4).

$$\sum_{i=1}^N (\log p(t_k|t_1, t_2, \dots, t_{k-1}; \Theta_x, \vec{\Theta}_{LM}, \Theta_s) + \log p(t_k|t_{k+1}, t_{k+2}, \dots, t_N; \Theta_x, \overleftarrow{\Theta}_{LM}, \Theta_s)) \quad (4)$$

where Θ_x represents the parameters represented by the token, $\vec{\Theta}_{LM}$ and $\overleftarrow{\Theta}_{LM}$ represent the LSTM parameters in the bidirections, and Θ_s represents the parameters of the softmax layer. For token i in each layer, we can calculate $(2L + 1)th$ representations, as shown in Equation (5).

$$R_k = \left\{ x_k^{LSTM}, h_{k,j}^{\rightarrow LSTM}, h_{k,j}^{\leftarrow LSTM} \mid j = 1, \dots, L \right\} = \left\{ h_{k,j}^{LSTM} \mid j = 0, \dots, L \right\} \quad (5)$$

where R_k is the representation of token k and $h_{k,j}^{LSTM}$ is the hidden layer, which is equal to $\begin{bmatrix} \rightarrow LSTM & \leftarrow LSTM \\ h_{k,j} & h_{k,j} \end{bmatrix}$. The ELMo representation of token i is calculated by Equation (6).

$$ELMo_k^{task} = E(R_k; \Theta^{task}) = \gamma^{task} \sum_{j=0}^L s_j^{task} h_{k,j}^{LSTM} \quad (6)$$

where γ^{task} is a scalar factor that adjusts the vector scale according to the characteristics of a specific task and s_j^{task} is the normalized weight.

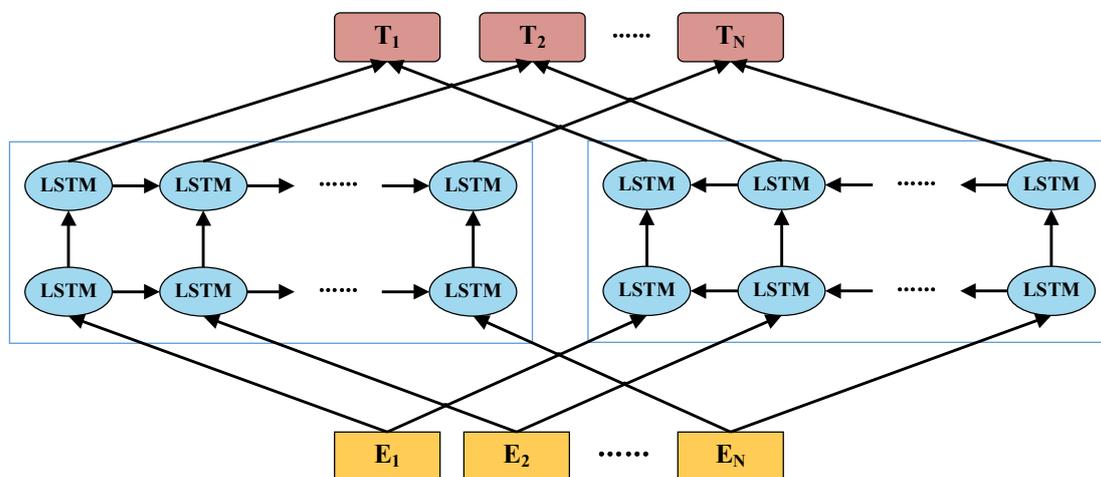


Figure 2. Structure of ELMo.

In natural language processing, the data involved are often contextual, and traditional feedforward neural networks are unable to process such data well. A recurrent neural network (RNN) is a typical neural network structure applied to sequence data. This network processes sequence data by introducing directed loops. The structure of the RNN is divided into three layers, namely the input, the middle, and the output. The middle layers can be connected back and forth, so that the information of the current state can be passed to the next state as part of the input of the next state. In this way, the nodes in the sequence can obtain previous information. However, when the sequence data becomes longer, the RNN is not able to handle this problem well. As a special RNN, the long-term short-term memory network (LSTM) selectively retains context information through a specially designed gate structure, which can effectively solve the problem of gradient explosion and gradient disappearance when RNN processes long sequence data. Equations (7)–(11) show the operation mechanism of the LSTM.

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \quad (7)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \quad (8)$$

$$C_t = f_t C_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (9)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \quad (10)$$

$$h_t = o_t \tanh(C_t) \quad (11)$$

where σ is the activation function sigmoid; \tanh is the hyperbolic tangent activation function; x_t is the unit input; i_t , f_t , and o_t are the input, forget, and output gates at time t ; w, b is the weight matrix and offset vector of the input gate and forget gate; c_t is the state at time t ; and h_t is the output at time t .

Words before and after each word will affect it, so we must fully consider the context of the text. Therefore, this paper uses a bidirectional long-term short-term memory network (BiLSTM) for feature extraction. In the BiLSTM, \vec{h}_t and \overleftarrow{h}_t are the intermediate states of the forward and backward LSTM outputs, respectively. Then, the total intermediate state of the BiLSTM output at time t is $h_t = [\vec{h}_t, \overleftarrow{h}_t]$. Besides, we connect an attention layer with the BiLSTM layer. In this layer, we calculate the attention score by the *query vector* (Q), *key vector* (K), and *value vector* (V) through Equation (12).

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (12)$$

where d_k is the dimension of the dot product of Q and K scaled by k . Q , K , and V are calculated from the same input, which is represented by BiLSTM in this model. Cosine similarity is used to measure the similarity of text. Equation (13) represents the text similarity.

$$sim(A, B) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{k=1}^n A_k \times B_k}{\sqrt{\sum_{k=1}^n A_k^2} \times \sqrt{\sum_{k=1}^n B_k^2}} \quad (13)$$

where A_k and B_k are the vectors where the similarity calculation will be performed.

Then, we proposed the structure of the sentence similarity model based on ELMo and the Attention-BiLSTM (EAB), the structure flow is shown in Figure 3. Table 2 shows a comparison of three algorithms used on the CCKS2018 dataset. The results show that our proposed EAB algorithm's performance is better than that of others. Table 3 shows five sentences, and Figure 4 shows the clustering results of 2D space.

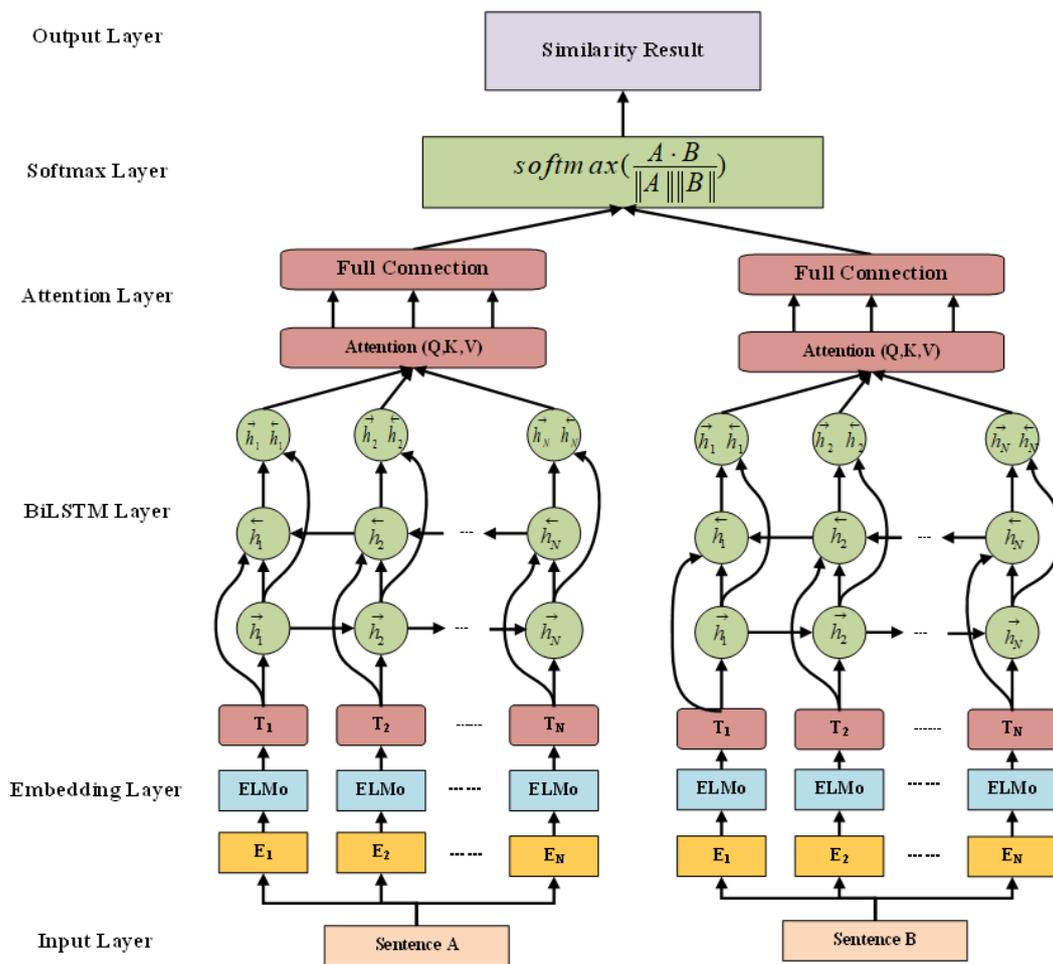


Figure 3. Structure of the sentence similarity model based on EAB.

Table 2. Comparison of three algorithms on the CCKS2018 dataset.

Model	Precision	Recall	F1
Word2vec	0.721	0.612	0.557
ELMo	0.820	0.804	0.781
EAB	0.861	0.854	0.843

Table 3. Five sample sentences.

Sentence ID	Sentence Content
A	Check that the tooling model is in good condition and complete and that the preparation of chilled iron, core sand, and alloy meet the process requirements.
B	Check that the tooling model has no offset and that the surface of the mold is tight.
C	Check that the tooling model is in good condition.
D	Fluorescence inspection of the cabin according to HVJ40-23001.
E	Vibration aging of cabin according to Ez2082-34002.

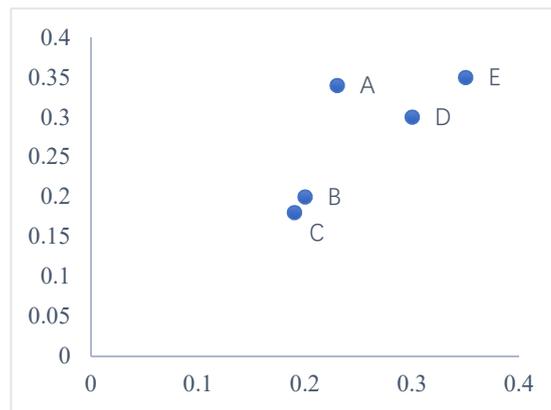


Figure 4. Clustering result of 2D space based on the sentences in Table 2.

4.2. Manufacturing Process Unit

By defining the MPU, the part process planning becomes a problem of sorting and optimizing the MPUs, and the ordering and optimization must satisfy the constraints and the ordering rules.

Process constraints exist in every section of the part's manufacturing activities, including knowledge constraints, resource constraints, technical constraints, and demand constraints.

Knowledge constraints: Knowledge constraints mean that all manufacturing methods and manufacturing sequences must be selected to conform to process knowledge, process rules, and process standards. When designing a process plan, the process rules to be followed include roughing, precision, benchmarking, clustering, and some specific criteria, such as threading, which are usually arranged after the outer round before the rough grinding of the outer circle.

Resource constraints: Resource constraints refer to the manufacturing conditions, manufacturing equipment, manufacturing materials, and other material conditions that are available inside the enterprise. The MPU may be available with a variety of manufacturing resources, which allows for more alternatives when selecting manufacturing resources, but it also introduces complexity into manufacturing decisions.

Technical constraints: Technical constraints refer to the specific shape of the part's geometry and technical conditions (shape tolerance, surface roughness, accuracy grade, etc.) and are the basis for selecting the manufacturing method. For example, the rotary parts are mainly used for turning and the contour type parts are mainly used for milling.

Order constraints: Order constraints refer to the specific requirements of the supply contract with the customer for the product, such as the order time, order method, and so on. The content of the contract has an important impact on the organization of production, process flow, technical specifications of the implementation, and so on.

From the above-described manufacturing feature information model, it is known that process data with a unit as a carrier is packaged in each MPU. Process sequencing is the sequential arrangement of all MPUs to form a parts processing sequence. To this end, there are the following process sequencing criteria.

Process ordering rule 1—general guidelines: this refers to the knowledge constraints that must be met when sorting MPUs.

Process ordering rule 2—customization criterion: in addition to this general criterion, when the topology of parts is too complex or the processing conditions of enterprises are limited, technicians need to make some artificial regulations on the sequence of MPUs according to the current process conditions.

The part feature refers to a combination of a series of information including a certain structural shape, manufacturing accuracy, and assembly requirements of the part. Part features are generally divided into two categories: (1) basic features, which are features that build part of geometry topologies and are not capable of secondary splitting, such as planes, holes, etc., and (2) auxiliary features of

the main features, which can be split twice, such as threads, keyways, etc. For example, Part A has a total of n features and the parts can be expressed as $A = (a_1, a_2, \dots, a_n)$, where a_i represents the i th processing feature, $1 \leq i \leq n$.

The MPU is the basic unit that constitutes the feature of the part, for example, completion of dimensions and tolerance for hole features $\varnothing 60K6_{+0.002}^{+0.021}$ represents the hole features, and the MPU is expressed as *Rough Turning* \rightarrow *Half-Fine turning* \rightarrow *Rough grinding* \rightarrow *Fine grinding*. For part A, feature φ_i is $\varphi_i = (\omega_1, \omega_2, \dots, \omega_n)$, where ω_i represents the i th process unit, $1 \leq i \leq n$.

In part processing, the processing resources corresponding to each process unit are different. Processing resources mainly refer to machine tools, cutters, fixtures, etc. Therefore, the processing unit can be seen as a collection of processing resources. We assume that machine toolsets are $X = (x_1, x_2, \dots, x_m)$, cutter sets are $Y = (y_1, y_2, \dots, y_n)$, and fixture sets are $Z = (z_1, z_2, \dots, z_o)$. Among them, m , n , and o are the sum of machine tools, cutters, and fixtures in the manufacturing sector. For part A, process unit ω_i in feature φ_i is $\omega_i = (x_\alpha, y_\beta, z_\gamma)$, $1 \leq \alpha \leq m$, $1 \leq \beta \leq n$, $1 \leq \gamma \leq o$, where $x_\alpha, y_\beta, z_\gamma$ represents machine tools, cutters, and fixtures needed for manufacturing process units ω_i , respectively.

A part consists of different features, and each of them has several serial or parallel relationships. Many manufacturing features have evolved. Figure 5 shows the manufacturing process model with six MPUs. For MPU 1, the processing unit 101 is in the front and processing unit 102 is in the back; MPU 2, MPU 3, and MPU 4 belong to a parallel relationship. A directed acyclic graph consists of several nodes and arcs, and there is no closed-loop; this is called the sequence of process units.

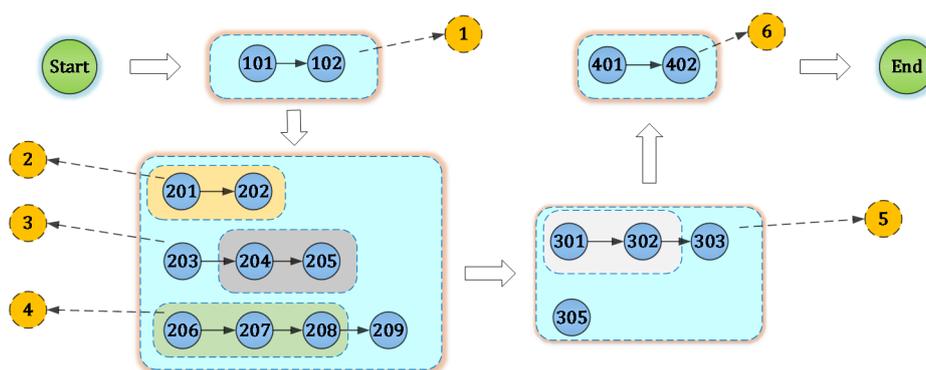


Figure 5. Manufacturing process model with six manufacturing process units (MPUs).

4.3. EABMC Sequential Pattern Mining Algorithm

Process knowledge is an important part of manufacturing. To improve manufacturing precision and to shorten the manufacturing cycle, it is necessary to obtain the process knowledge that is urgently needed by the current processing enterprises from their historical process data. The structure of mechanical parts is composed of a limited number of typical manufacturing features, which are reassembled according to the functions of the parts. Therefore, the current process planning of mechanical parts mainly focuses on the selection of manufacturing methods and manufacturing equipment and the arrangement of the manufacturing sequence. After the long-term accumulation of these works, the historical process data become the experience and rules that are often used in the process planning of mechanical parts in a period, which is embodied in the typical process sequence and process decision rules. There are a large number of process-related data in the CAPP system, such as material types, part features, processing methods, processing equipment, tools, etc. The data related to the part processing process represent the process of knowledge. Similar parts tend to have similar manufacturing processes. This paper proposes the use of the frequently closed sequential pattern mining algorithm based on the text contextual semantics of the manufacturing process (EAB) and manufacturing process units (MPU). The algorithm is shown in Figure 6, which includes the construction of the process knowledge requirement model, construction of the process knowledge

data model, and recommendation of frequent key MPU patterns. Through the analysis and mining between the wheel hub manufacturing data and the wheel inspection unqualified data, the influencing factors of the production process with a high unqualified rate are found, and then, the main factor sequence collection of wheel quality is sorted out.

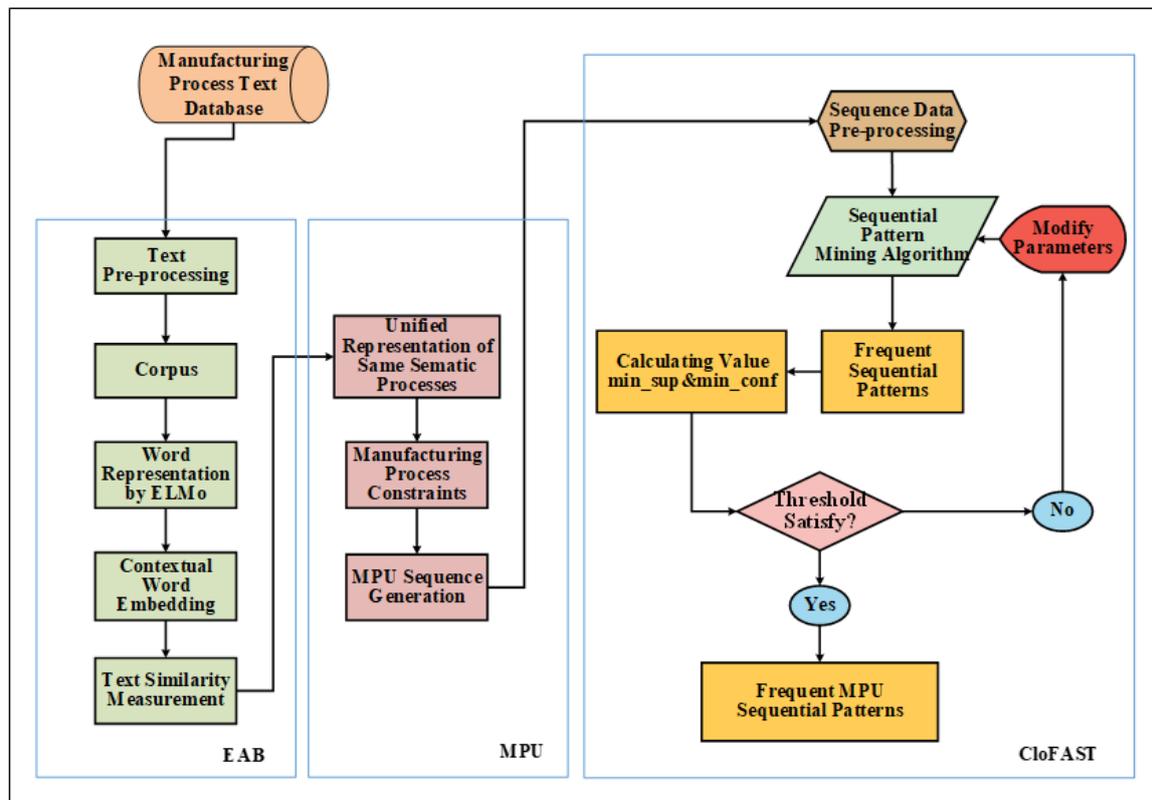


Figure 6. EABMC sequential patterns mining model.

The previous association rules are used to mine the process knowledge, and after removing the manufacturing time, manufacturing equipment, and other data, the processing sequence is obtained. The transaction set T is a collection of processes of a component, $T = \{t_1, t_2, \dots, t_n\}$, where t_n is the process data for each item. Sequence pattern mining and association rule mining are similar in many respects, but sequence pattern mining pays more attention to the sequential relevance of data. The objects and results of the sequential pattern mining are ordered, that is, the entries of each sequence in the data set are ordered in time or space and the output results are also ordered. Therefore, it is more suitable for mining of a typical process sequence and reasoning of process decisions than association rules.

The pattern of sequential pattern mining is as follows: $X \rightarrow Y$, where $X \subseteq I$, $Y \subseteq I$, and $X \cap Y = \emptyset$. X is the former term, and Y is the latter term. The probability that items contained in itemset X and itemset Y appear simultaneously in the transaction set T is recorded as the support of the sequence pattern $Sup(X \Rightarrow Y)$, $Conf(X \Rightarrow Y)$ as shown in Equations (14) and (15) and is an important indicator of the sequence pattern.

$$Support_D(X \Rightarrow Y) = \frac{|\{(sid, s) | (sid, s) \in D \wedge (X \rightarrow Y \subseteq s)\}|}{|D|} \quad (14)$$

$$Confidence_D(X \Rightarrow Y) = \frac{Support_D(X \Rightarrow Y)}{Support_D(X)} \quad (15)$$

In the transaction containing itemset X , the conditional probability of the occurrence of item set Y is the confidence of the sequence pattern, which is a measure of the accuracy of the sequence pattern and is used to measure the strength of the sequence pattern

The minimum support degree min_sup and the minimum confidence level min_conf are set. If the support degree of the item set is greater than or equal to the minimum support degree min_sup , it is called a frequent itemset; if the confidence level of the frequent itemset is greater than or equal to the minimum confidence level min_conf , then the frequent itemset is called strong rules. Therefore, the problem of mining the sequence pattern in the transaction database can be divided into the following two processes: (1) find all itemsets that satisfy the minimum support, that is, obtain frequent itemsets, and (2) generate strong association rules through frequent itemset.

Sequential pattern mining is the subject of data mining. It involves finding statistically relevant patterns between data examples. In these examples, the values are passed in order. It is generally assumed that these values are discrete, so time series mining is closely related, but it is generally considered to be a different activity. Sequential pattern mining is a special case of structured data mining [35].

In this section, we first introduce some preliminary concepts and then formalize the closed sequential pattern mining problem.

Definition 1. (Itemset) An itemset is a set containing m different items, referred to as $I = \{i_1, i_2, \dots, i_n\}$.

Definition 2. (Sequence) Abbreviated as SID, a sequence is a complete stream of information. Sequence Y is referred to as $Y_1 \rightarrow Y_2 \rightarrow \dots \rightarrow Y_L$, where $Y_i (1 \leq i \leq n)$ is the i item.

Definition 3. (Sequence attribute) Each sequence has a unique identifier (Sid), and each item set for each sequence has a temporary item set identifier (Eid), which is a Timestamp. The Eid of a sequence is unique.

Definition 4. (Subsequence) For two sequences $a = (a_1, a_2, \dots, a_i)$ and $b = (b_1, b_2, \dots, b_j)$, if a positive integer $m_1, m_2, \dots, m_i, 1 \leq m_1 < m_2 < \dots < m_i \leq j$ exists, let $a_1 \subseteq b_{m_1}, a_2 \subseteq b_{m_2}, \dots, a_i \subseteq b_{m_i}$, where a is a subsequence to b , referred to as Sub-S.

Definition 5. (Frequent sequence) Given the minimum support threshold, if sequence Y in the sequence database support is not less than the threshold value, a frequent sequence called sequence Y is referred to as FS.

Definition 6. (Frequently closed sequence) If the sequence and super sequence support are not the same and it is a frequent sequence, the sequence is a frequently closed sequence (FCS).

There are several algorithms for mining sequential patterns, such as FAST [36], GSP [37], SPADE [38], and PrefixSpan [39]. These algorithms show good performance in databases that contain short frequent sequences or support thresholds that are not very low. Closed sequence mining [40] aims to reduce the number of sequences that exceed the threshold and to pick long sequences to reduce the amount of calculation and time. The closed sequence pattern mining algorithm has better performance than other sequence pattern mining algorithms and is favored by more and more users.

The closed FAST sequence mining algorithm based on sparse ID lists (CloFAST) is a novel algorithm that is used for mining closed frequent sequences of an itemset, as cited by Fumarola et al. [41]. The EABMC (details are shown in the Algorithm 1) combines natural language processing, process rules, and sparse ID list technologies based on sparse ID lists and vertical ID lists. Its theoretical properties are studied to quickly count the support of sequential patterns with a novel one-step technique to both check sequence closure and to prune the search space. EABMC is better than other closed sequential pattern mining algorithms.

In the first database scan, EABMC was used to find a frequent itemset and to establish its sparse ID list (line 2). Then, it also found frequent closed itemsets and built its sparse ID list (line 4). This was

achieved by constructing a closed itemset enumeration tree (CIET) based on a modified version of the FAST algorithm [36], which integrates the marking and pruning techniques proposed in Moment [42]. Lines 5 to 12 initialize the first level of the closed sequence enumeration tree (CSET). Each node in the first level represents a (candidate) closed sequence of size 1, for which the only element is a closed frequent item set. The vertical id-list (VIL) of the first-level node can be directly calculated according to the sparse id-list (SIL) of the closed frequent itemset. Starting from the first layer, according to the depth-first search strategy, the nodes in CSET are regarded as sequence expansion. During the mining process, the current closed sequential pattern set is stored in CSET. Finally, the EAMBC returns the complete set of closed sequential patterns in CSET.

The EAMBC algorithm proceeds in two steps:

(1) It generates a subset of frequent sequences (FS) and a superset of closed frequent sequences (CFS), called closed frequent candidates (CFC), and this subset is stored in the main memory.

(2) It performs the post-pruning stage to eliminate all non-closed sequences from the FCC to finally obtain accurate CFS (closed frequent sequences).

Algorithm 1. EAMBC (E-MSDB, min_sup)

Input: EAB-MPU Sequence database E-MSDB, int min_sup
Output: Complete set of EAMBC frequent sequences (CEFS);
Data: CSET T = new Tree (), Frequent Items (FI),
 Closed Frequent Itemset (CFI), Node n
 1: // Identify frequent 1-itemsets and establish their SILs;
 2: FI = loadFrequentSILs (E-MSDB, min_sup);
 3: // Identify closed frequent frequent itemsets and their SILs
 4: CFI = mineClosedFItemset (FI, min_sup);
 5: **for each** cfi \in CFI **do**
 6: // Create VIL_{cfi} from SIL_{cfi}
 7: vil = createVil(cfi);
 8: // Create CSET node associated to cfi
 9: n = createNode(cfi, vil);
 10: labelNodeAs(n, "closed");
 11: addChildNode(T, root(T), n);
 12: **end for**
 13: **for each** child \in children (T, root(T)) **do**
 14: // start the depth first search
 15: sequenceExtension (T, child, min_sup);
 16: **end for**
 17: **return** closedSequentialPatterns(T).

5. Implementation and Experiment Results

5.1. Data Preprocessing

All the results reported in this paper were obtained with a PC with an AMD Ryzen 7 1800X Eight-Core processor with 32G RAM, and the analysis language used was Java. The manufacturing process dataset was used in the example stored in a large manufacturing database of a foundry company from Guizhou Province, China. In the study, a foundry product was used as an example. There are 29,687 history records from the years 2017 to 2018. By performing data preprocessing, relevant demand data were obtained. The part of the manufacturing process carried out by EAB and MPU is shown in Table 4.

Table 4. The part of the manufacturing process by EAB and MPU.

ID	Process Name	Process Content	MPU ID
3	Modeling	First, make the bottom plate shape. Then, wet the core shape on the bottom plate. At the same time, the outer mold and the cover shape are formed.	B
4	Checking	Check that the chilled iron has no offset and that the surface of the mold is tight.	
7	Repair modeling	Dress the core and exterior mold, and print the furnace number on the wet core furnace number.	C
...
37	Vsr	Vibration aging of cabin according to Ez2082:34002.	R
38	Checking	Inspection cabin vibration aging operation meets process requirements.	
40	Product delivery	Make products into the library.	T

In sequential pattern mining and analysis, some key sequences are often selected and frequent sequences are judged by the number of occurrences of key sequences. Sparse data is generated when there are many key sequences, but each sequence contains a small number of key sequences. The sparseness of the data can cause deviations or even errors in the mining results. The processing unit proposed in this paper can solve the sparseness problem of the data. By combining the processes, the quasi-class of the process is reduced, and the key sequence in each sequence data is increased. The input data description is shown in Table 5.

Table 5. Input data description.

Time	Manufacturing Process Sequences	Product Quality
20171021	Incoming materials→Forging→X-ray inspection→Appearance inspection→Deburring→Dissolution treatment→Aging treatment→Lathe surface→Fine slot→Dovetail slot→Deburring→Drilling→Milling inner cavity→Tapping	Qualified
20171201	Deburring→Drilling→Milling inner cavity→Tapping→Drilling→Cleaning→Detection→Anodizing→Painting→Drying→Sampling detection→Painting→Shipping.	Unqualified
...
20170824	Dissolution treatment→Aging treatment→Lathe surface→Lathe shape→Fine slot→Dovetail slot→Deburring→Drilling→Milling inner cavity→Tapping→Drilling→Cleaning	Unqualified
20181228	Dovetail slot→Deburring→Drilling→Milling inner cavity→Tapping→Drilling→Cleaning→Detection→Anodizing→Painting→Drying→Sampling detection→Painting→Shipping	Qualified

5.2. Discussion of Minimum Support Count

The minimum support in the data mining algorithm is the threshold that will affect the accuracy of the mining result. A high threshold will lose much of the significant information, and a low threshold will increase the workload. Choosing the right threshold is critical for sequential pattern mining.

The data set obtained by the preprocessing steps was used as a specific data set for determining the support threshold, the CloFAST algorithm was applied, and then the judgment result of the support threshold was obtained. The results are shown in Figure 7.

The experiment set $min_sup = 0.01$ as the initial value, and as min_sup increased, the number of closed sequential patterns decreased. When $min_sup = 0.04$, there were 241 closed frequent sequential patterns, which was equal to the count of closed frequent sequential patterns for $min_sup = 0.05$. Thus, an optimal support threshold of $min_sup = 0.04$ was finally obtained.

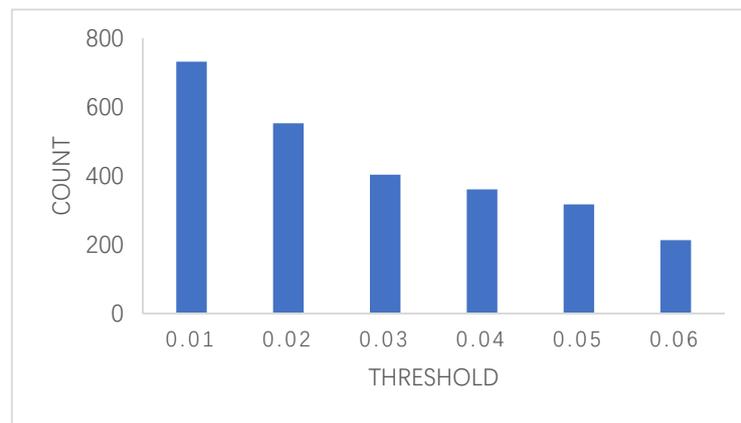


Figure 7. The number of EABMC frequent sequential patterns of datasets under minimum supports.

The Chi-square test [43] is a method that is commonly used in statistics for data analysis. It is mainly used to compare two or more sample rates (composition ratios) and to carry out a correlation analysis of two categorical variables. This method classifies data into different parts to ensure independence among the categorical data points.

The dataset was divided into two data sets, D_1 and D_2 , randomly. The CloFAST algorithm was applied to obtain the results shown in Figure 6. When the dataset was reduced to half of the original size; D_1 showed the first extreme point of value m at $min_sup = 0.04$, while D_2 showed the first extreme point of value m at $min_sup = 0.04$. Then, the dataset was randomly divided into four data sets: D_3 , D_4 , D_5 , and D_6 . The CloFAST algorithm was applied, and it showed that the first extremum point appeared when $min_sup = 0.04$ for all four data sets, as shown in Figure 8. Therefore, based on the observations from Figure 7, the minimum support value was set as 0.4 in the following experiment.

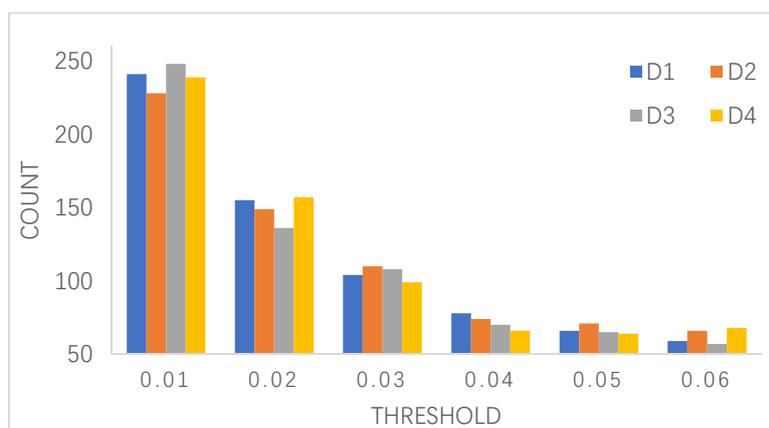


Figure 8. The number of EABMC frequent sequential patterns of 4 datasets under different minimum supports.

5.3. Experiment Results

Table 6 shows the number of sequences produced under different methods. Through our method, we were able to remove the repetitive, synonymous, and redundant sequences, which provided a good premise for the next step of sequential pattern mining.

Table 6. The number of sequences under different methods.

Model	Original	EAB	EAB+MPU
No.	29687	25373	20281

To verify that the performance of the improved method is superior to the traditional algorithm, we compared the proposed EABMC with CloFAST, MPU-CloFAST, Word2Vec-CloFAST, ELMo-CloFAST, and Word2Vec-MPU-CloFAST in terms of the accuracy rate (shown in Figure 9), running time (shown in Figure 10), and memory consumption (shown in Figure 11), according to the dataset configuration and after varying the support threshold. In terms of the three main indexes of data mining, EABMC generally outperformed all of the other systems for almost every support value when the number of frequent sequences was higher.

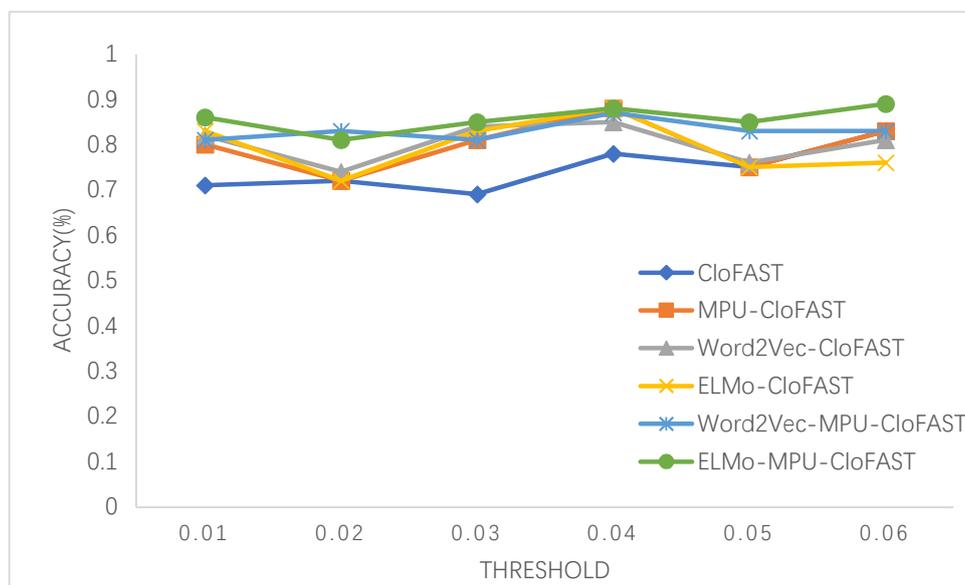


Figure 9. Accuracy comparison of different algorithms.

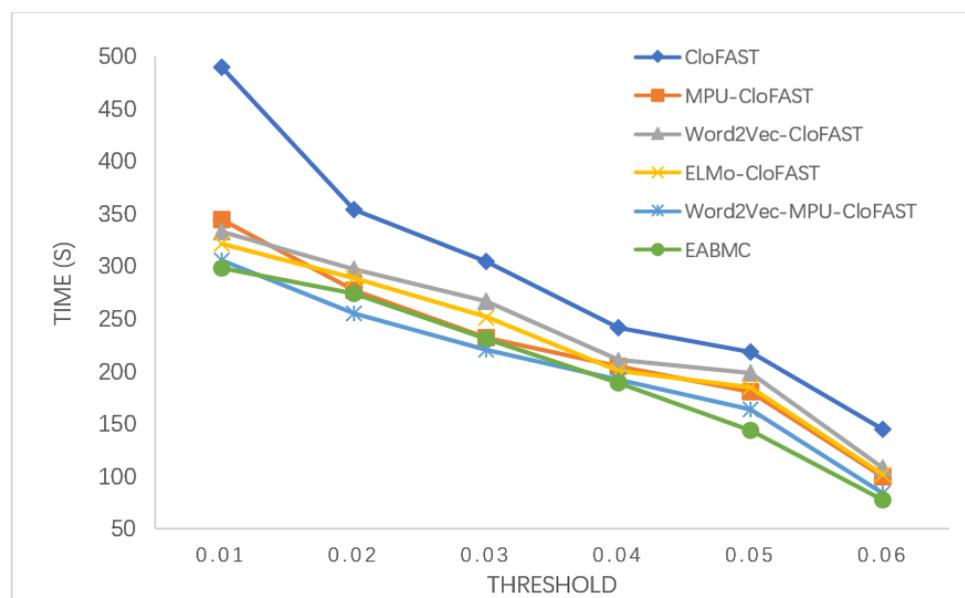


Figure 10. Running time comparison of different algorithms.

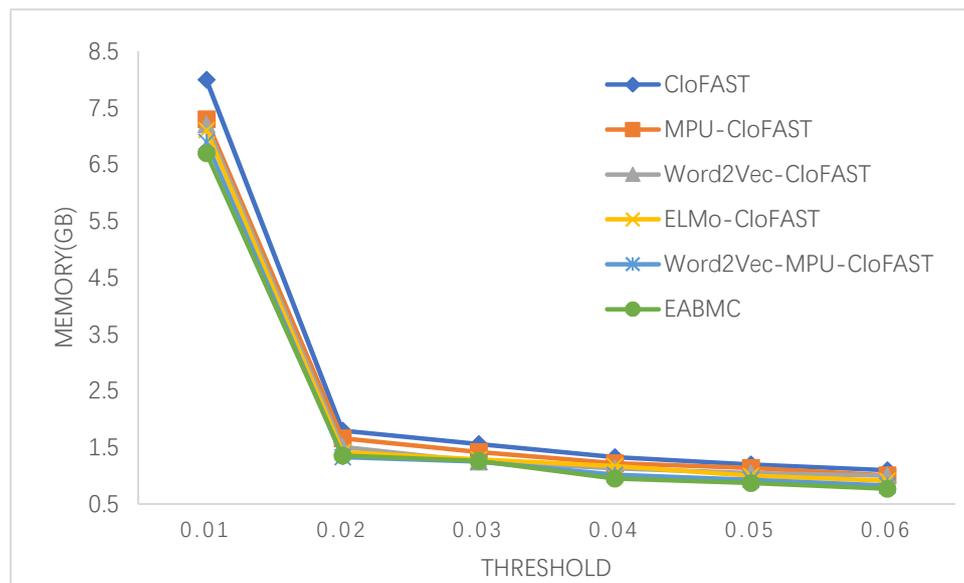


Figure 11. Memory consumption comparison of different algorithms.

We analyze the sequence pattern set, combined the wheel hub quality fault diagnosis knowledge and process knowledge, summarized and sorted out the main factors affecting the wheel hub quality and the process parameter optimization rules, and provided data support for product quality improvement and process optimization.

The sequence rules derived from the sequence pattern mining algorithm can help enterprises to analyze the abnormal product quality data and to determine the potential sequence rules that lead to the degradation of product quality. Taking the data of a factory in Guizhou as an example, the data included 20,281 product processing sequences. Through quality inspection, 20,152 items were qualified, 129 items were unqualified, and the product yield was 99.36%. The EABMC algorithm was used to mine the sequence pattern of 129 unqualified sequences, and 124 frequent sequences were obtained. Table 7 shows some of the mining results. Through the analysis of these 124 sequences, the enterprise could obtain the potential causes of product disqualification.

Table 7. Part of EABMC frequent sequential patterns for the manufacturing process.

EABMC Sequential Patterns	Support
2-item closed frequent sequential patterns	
Incoming materials→Forging	759
Incoming materials→Dissolution treatment	715
3-item sequential patterns	
Forging→X-ray inspection→Drilling	536
Deburring→Dissolution treatment→Fine slot	502
4-item sequential patterns	
Forging→Appearance inspection→Dissolution treatment→Fine slot	346

To more accurately determine the causes of quality degradation, we further analyzed the following: incoming materials→forecasting and incoming materials→solution treatment is highly supported in the 2-item assessment, indicating that, in these two processes, the product quality will be unqualified. From this frequent sequence, it is known that the forging and dissolution treatment are the key procedures that affect the quality of the wheel hub, so companies should focus on these two procedures. Forging→X-ray inspection→drilling means that, after the forging, the X-ray inspection and drilling procedures are the important reasons that will cause the quality of the wheel hub to decrease. When the wheel hub manufacturing sequence includes the deburring→dissolution treatment→fine slot sequence,

it will cause the quality of the wheel to decrease. With these frequent sequence patterns, the company can focus on these sequence patterns when formulating the wheel manufacturing process, and avoiding these sequence patterns can improve the quality of the hub.

5.4. Sequence Relation Visualization

The sequence relations obtained by the EABMC algorithm were visualized, and the sequence relation visualization is shown in Figure 12. The visualization revealed a clear temporal direction between the manufacturing processes. The knowledge graph includes all procedures that lead to the reduction of wheel hub quality and the direct connection of each procedure, which can make the enterprise more intuitively understand the factors that lead to the reduction of product quality.

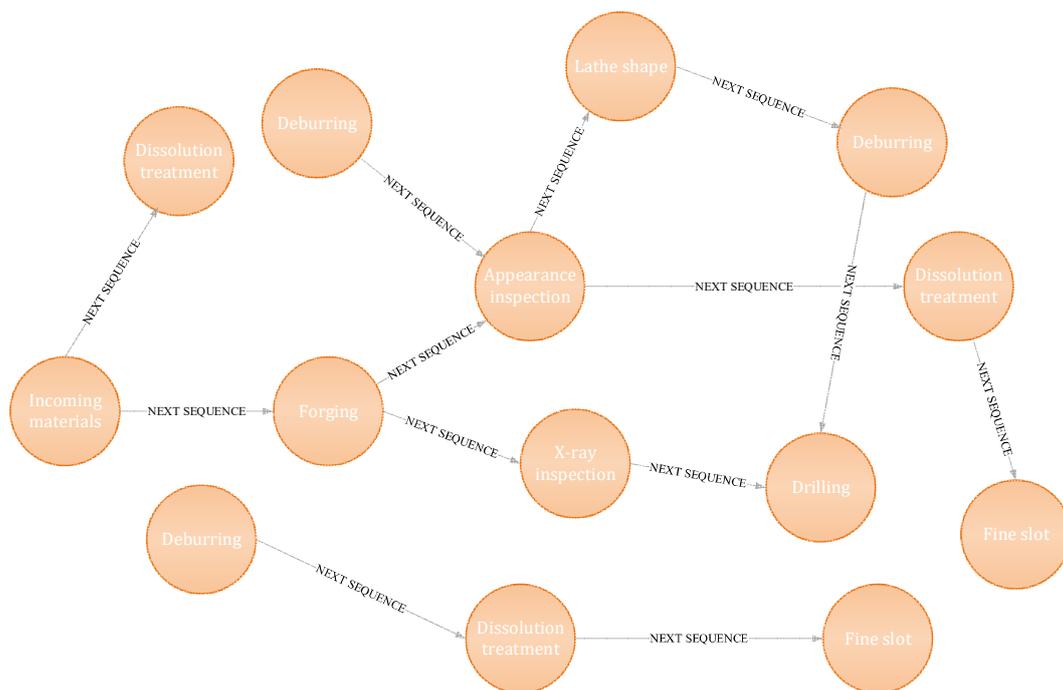


Figure 12. Sample product quality degradation sequence relation visualization based on EABMC.

After correcting the factors that lead to the reduction of wheel hub quality obtained by the sequential pattern mining algorithm, the wheel hub is manufactured again. The analysis software is used to compare the effect before (shown in Figure 13a) and after (shown in Figure 13b) improvement. From the average grain size distribution in Figure 13b, it can be seen that, compared with the original scheme, the uniformity of hub grain refinement in the optimization scheme is significantly improved.

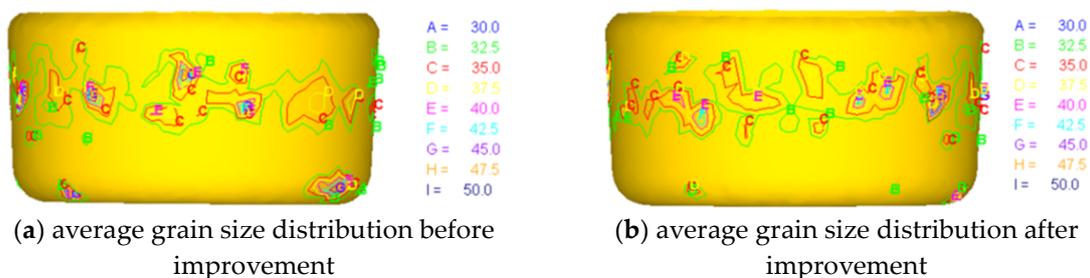


Figure 13. Comparison of grain refinement effect of the wheel hub.

6. Conclusions and Future Work

This paper has proposed a frequent closed sequential pattern mining algorithm based on the text contextual semantics of the manufacturing process and the manufacturing process rules (EABMC). This algorithm aims to obtain frequent sequence relations of product manufacturing processes to help with the identification of factors affecting product quality and to improve the product quality. We used EAB to merge semantically similar sequential texts, and we utilized the MPU to deal with sequences containing simultaneous occurrences and to reduce impurities in the manufacturing sequence. To get a good mining result, particularly when the volume of the processed data is large, we chose a closed sequential pattern mining approach to decrease the number of sequences beyond the threshold and to pick out long sequences and we proposed the use of a processing unit that consists of common match processes to reduce the amount of computing and calculating time required. A longer sequence contains a higher amount of manufacturing information. Process design relies heavily on the designer's process knowledge and related experience. The suggested method tries to avoid these human factors. The method sets threshold support values according to the mapping relationship between the account of closed frequent sequences and the support threshold *min_sup*. In the proposed manufacturing process method, the threshold support value is dynamically adjusted for manufacturing situations, such as for different categories, according to the data connected to the manufacturing database. The generated closed sequential patterns may change depending on the different manufacturing situations, e.g., the category, precision, or personalization. Compared with other methods, this method has a higher level of efficiency and better performance. This paper will help the manager make decisions to improve product quality and find important factors related to production and manufacturing that affect product quality.

This model is suitable for text-based sequence data; if structured data is the main, it is not very suitable. If it is in a specific field such as medical treatment and product recommendation, it is necessary to fine-tune the model with relevant field text data to improve the accuracy of the text similarity in the model, to better compress the semantic similar data, and to improve the operation efficiency. In sequence pattern mining, the degree of support determines the number of frequent sequences generated after mining. The degree of support is low, the frequent patterns generated are many, the degree of support is high, and the frequent patterns generated are few. Too few frequent patterns are not conducive to discovering the relationship between elements, and too frequent patterns will analyze many useless relationships. When applying this model in different fields, it is necessary to select appropriate support parameters according to the actual situation.

Future work will optimize the use of sequential pattern mining to detect the temporal relationships among sequential patterns in manufacturing records. Through integration with intelligent manufacturing technology, a complete knowledge graph of the process sequence relationship will be constructed and research on product quality analysis will be based on the knowledge graph.

Author Contributions: Project administration, L.Y.; conceptualization, L.Y. and H.H.; methodology, L.Y. and H.H.; software, L.Y.; validation, L.Y.; investigation, H.H. and L.Y.; data curation, L.Y. and S.-H.C.; writing—original draft preparation, L.Y. and H.H.; writing—review and editing, L.Y., H.H., and S.-H.C.; visualization, H.H.; supervision, H.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by The National Natural Science Foundation of China under grant No. 51865004, by Science and Technology Project of Guizhou Province under grant No. Talents [2018]5781, and by Major Project of Science and Technology in Guizhou Province under grant No. [2017]3004. Talent Project in Guizhou Province under grant No. KY [2018]037.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Muthalagu, I. *Plm Manufacturing Change Order and Data Enrichment Collaboration for Engineering Industries Manufacturing*; Social Science Electronic Publishing: New York, NY, USA, 2017.
- Fei, T.; Cheng, J.; Qi, Q.; Meng, Z.; He, Z.; Sui, F. Digital twin-driven product design, manufacturing and service with big data. *Int. J. Adv. Manuf. Tech.* **2018**, *94*, 3563–3576.
- Wu, Y.; He, F.; Zhang, D.; Li, X. Service-oriented feature-based data exchange for cloud-based design and manufacturing. *IEEE T. Serv. Comput.* **2018**, *11*, 341–353. [[CrossRef](#)]
- Sadati, N.; Chinnam, R.B.; Nezhad, M.Z. Observational data-driven modeling and optimization of manufacturing processes. *Expert Syst. Appl.* **2017**, *93*, 456–464. [[CrossRef](#)]
- Youngs, H.; Somerville, C. Best practices for biofuels: Data-based standards should guide biofuel production. *Science* **2014**, *344*, 1095–1096. [[CrossRef](#)]
- Agrawal, R.; Srikant, R. Fast Algorithms for Mining Association Rules. In Proceedings of the 20th Int. Conf. very Large Data Bases, VLDB, Santiago de Chile, Chile, 12–15 September 1994; pp. 487–499.
- Agrawal, R.; Srikant, R. Mining sequential patterns. In Proceedings of the Eleventh International Conference on Data Engineering, IEEE, Washington, DC, USA, 6–10 March 1995; pp. 3–14.
- Quinlan, J.R. Induction of decision trees. *Mach. learn.* **1986**, *1*, 81–106. [[CrossRef](#)]
- Jain, A.K.; Dubes, R.C. Algorithms for clustering data. *Technometrics* **1988**, *32*, 227–229.
- Tan, A.-H. Text mining: The state of the art and the challenges. In Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases, Beijing, China, 16–18 April 1999; pp. 65–70.
- Argote, L.; Ingram, P. Knowledge transfer: A basis for competitive advantage in firms. *Organ. Beh. Hum. Dec. Pro.* **2000**, *82*, 150–169. [[CrossRef](#)]
- Köksal, G.; Batmaz, İ.; Testik, M.C. A review of data mining applications for quality improvement in manufacturing industry. *Expert Syst. Appl.* **2011**, *38*, 13448–13467. [[CrossRef](#)]
- Hitomi, K. *Manufacturing Systems Engineering: A Unified Approach to Manufacturing Technology, Production Management and Industrial Economics*; Routledge: London, UK, 2017.
- Hallac, D.; Vare, S.; Boyd, S.; Leskovec, J. Toeplitz inverse covariance-based clustering of multivariate time series data. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, Halifax, NS, Canada, 13–17 August 2017; pp. 215–223.
- Salton, G.; Buckley, C. Term-weighting approaches in automatic text retrieval. *Inform. Process. Manag.* **1988**, *24*, 513–523. [[CrossRef](#)]
- Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.S.; Dean, J. Distributed representations of words and phrases and their compositionality. *Adv. Neural Inf. Process. Syst.* **2013**, *2*, 3111–3119.
- Blei, D.M.; Ng, A.Y.; Jordan, M.I.J. Latent dirichlet allocation. *J. Mach. Learn.* **2003**, *3*, 993–1022.
- Pennington, J.; Socher, R.; Manning, C.D. Glove: Global vectors for word representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1532–1543.
- Deerwester, S.; Dumais, S.T.; Furnas, G.W.; Landauer, T.K.; Harshman, R. Indexing by latent semantic analysis. *J. Am. Soc. Inf. Sci.* **1990**, *41*, 391–407. [[CrossRef](#)]
- Peters, M.E.; Neumann, M.; Iyyer, M.; Gardner, M.; Clark, C.; Lee, K.; Zettlemoyer, L. Deep contextualized word representations. *arXiv* **2018**, arXiv:1802.05365.
- Yang, B.; Qiao, L.; Cai, N.; Zhu, Z.; Wulan, M. Manufacturing process information modeling using a metamodeling approach. *Int. J. Adv. Manuf. Tech.* **2018**, *94*, 1579–1596. [[CrossRef](#)]
- Trojanowska, J.; Kolinski, A.; Galusik, D.; Varela, M.L.R.; Machado, J. A methodology of improvement of manufacturing productivity through increasing operational efficiency of the production process. In *Advances in Manufacturing*; Springer: New York, NY, USA, 2018.
- Su, Y.; Chu, X.; Chen, D.; Sun, X. A genetic algorithm for operation sequencing in capp using edge selection based encoding strategy. *J. Intell. Manuf.* **2018**, *29*, 313–332. [[CrossRef](#)]
- Lan, X.P.; Tran, D.V.; Hoang, S.V.; Truong, S.H. Effective method of operation sequence optimization in capp based on modified clustering algorithm. *J. Adv. Mech. Des. Syst. Manuf.* **2017**, *11*, JAMDSM0001.
- Wang, W.; Li, Y.; Huang, L. Rule and branch-and-bound algorithm based sequencing of machining features for process planning of complex parts. *J. Intell. Manuf.* **2018**, *29*, 1329–1336. [[CrossRef](#)]

26. Milošević, M.; Lukić, D.; Antić, A.; Lalić, B.; Ficko, M.; Šimunović, G. E-capp: A distributed collaborative system for internet-based process planning. *J. Manuf. Syst.* **2017**, *42*, 210–223. [[CrossRef](#)]
27. Cheng, J.; Wang, J. Data-driven matching method for processing parameters in process manufacturing. *CIMS* **2017**, *23*, 2361–2370.
28. Yuan, X.; Chang, W.; Zhou, S.; Cheng, Y.J.S. Sequential pattern mining algorithm based on text data: Taking the fault text records as an example. *Sustainability* **2018**, *10*, 4330. [[CrossRef](#)]
29. Huang, H.; Yao, L.; Tsai, C.-Y. Transportation service quality improvement through closed sequential pattern mining approach. *Cyb. Infor. Tech.* **2016**, *16*, 185–194. [[CrossRef](#)]
30. Amiri, M.; Mohammad-Khanli, L.; Mirandola, R. A sequential pattern mining model for application workload prediction in cloud environment. *J. Netw. Comput. Appl.* **2018**, *105*, 21–62. [[CrossRef](#)]
31. Tsai, C.-Y.; Chen, C.-J. A pso-ab classifier for solving sequence classification problems. *Appl. Soft Comput.* **2015**, *27*, 11–27. [[CrossRef](#)]
32. Huynh, B.; Trinh, C.; Huynh, H.; Van, T.-T.; Vo, B.; Snasel, V. An efficient approach for mining sequential patterns using multiple threads on very large databases. *Eng. Appl. Artif. Intell.* **2018**, *74*, 242–251. [[CrossRef](#)]
33. Tarus, J.K.; Niu, Z.; Yousif, A. A hybrid knowledge-based recommender system for e-learning based on ontology and sequential pattern mining. *Future Gener. Com. Sy.* **2017**, *72*, 37–48. [[CrossRef](#)]
34. Zhi, D.; Liao, X.; Su, F.; Fu, D. Mining coastal land use sequential pattern and its land use associations based on association rule mining. *Remote Sens.* **2017**, *9*, 116.
35. Papaioannouab, G. The evolution of cell formation problem methodologies based on recent studies (1997–2008): Review and directions for future research. *Eur. J. Oper. Res.* **2010**, *206*, 509–521. [[CrossRef](#)]
36. Mabroukeh, N.R.; Ezeife, C.I. A taxonomy of sequential pattern mining algorithms. *ACM Comput. Surv. (CSUR)* **2010**, *43*, 3. [[CrossRef](#)]
37. Salvemini, E.; Fumarola, F.; Malerba, D.; Han, J. *Fast Sequence Mining Based on Sparse Id-Lists*; Springer: New York, NY, USA, 2011.
38. Srikant, R.; Agrawal, R. Mining sequential patterns: Generalizations and performance improvements. In Proceedings of the International Conference on Extending Database Technology, Avignon, France, 25–29 March 1996.
39. Zaki, M.J. Spade: An efficient algorithm for mining frequent sequences. *Mach. Learn.* **2001**, *42*, 31–60. [[CrossRef](#)]
40. Pei, J.; Han, J.; Mortazavi-Asl, B.; Pinto, H.; Chen, Q.; Dayal, U.; Hsu, M.-C. *Prefixspan: Mining Sequential Patterns Efficiently by Prefix-Projected Pattern Growth*, ICCCN; IEEE: Piscataway, NJ, USA, 2001; p. 0215.
41. Yan, X.; Han, J.; Afshar, R. Clospan: Mining: Closed sequential patterns in large datasets. In Proceedings of the 2003 SIAM International Conference on Data Mining, SIAM, San Francisco, CA, USA, 1–3 May 2003; pp. 166–177.
42. Fumarola, F.; Lanotte, P.F.; Ceci, M.; Malerba, D. Clofast: Closed sequential pattern mining using sparse and vertical id-lists. *Knowl. Inf. Syst.* **2016**, *48*, 429–463. [[CrossRef](#)]
43. Yun, C.; Wang, H.; Yu, P.S.; Muntz, R.R. Catch the moment: Maintaining closed frequent itemsets over a data stream sliding window. *Knowl. Inf. Syst.* **2006**, *10*, 265–294.

