Association Analysis-Based Aggregation Method for Resource Virtualization in Cloud Manufacturing

Xuhao Li and Youling Chen(✉)

College of Mechanical and Vehicle Engineering, Chongqing University, Chongqing, China
18742543184@163.com

Abstract. Cloud manufacturing resource virtualization is the basis for cloud manufacturing services to be carried out, where resource virtualization aggregation is an important way to improve the efficiency of subsequent resource allocation and scheduling. However, current resource aggregation methods mostly construct qualitative rules in a task-oriented manner with insufficient reusability and low quantification. To this end, this paper proposes a bottom-up resource virtualization aggregation method based on association analysis using data mining algorithms. Firstly, we construct a target transaction model for manufacturing activities. Secondly, the multiple minimum support MS-Apriori algorithm is employed to extract frequent resource item sets and association relationships within virtualization resources from historical manufacturing data transaction sets, and the virtualized resources are aggregated according to related resource information fusion. Specifically, the concept of multiple minimum support and the support tuning model are introduced to optimize the insufficient objectivity of support threshold setting in the traditional Apriori algorithm. Finally, we apply the resource aggregation process with examples to verify the effectiveness and feasibility of proposed method.

Keywords: Cloud Manufacturing · Virtualization · Resource Aggregation · Data Mining

1 Introduction

Recently, the manufacturing industry is transforming and upgrading to smart manufacturing and intelligent manufacturing. Cloud manufacturing as a new concept of service-oriented manufacturing proposed by Li has received widespread attention [1]. The idea of cloud manufacturing is developed on the previous cloud computing and aims to realize the sharing of manufacturing resources in the cloud. Based on the concept of manufacturing resources as services, the scattered resources are centralized management and the centralized resources are decentralized for utilization. The cloud manufacturing paradigm starts with the virtualization description of a large number of scattered manufacturing resources break the strong coupling between resource entities and services

X. Yuan et al. (Eds.): ICEKIM 2023, AHCS 13, pp. 1841–1854, 2023.
https://doi.org/10.2991/978-94-6463-172-2_204

and realize the cloud-based invocation of resources, then the virtual resources are subsequently classified to form a cloud pool with same category of resources and services, and finally matched with the corresponding decomposition granularity tasks. Therefore, the virtualization of resources is the basis of cloud manufacturing [2] but the simple description of individual resources is not sufficient to achieve the subsequent goals. On the one hand, a manufacturing service often requires the collaboration of multiple categories of resources, on the other hand, the efficiency of services tasks matching will greatly reduce if they are sunk to a single resource granularity, and each level of decline in resource service granularity will increases task matching complexity exponentially. Therefore, it is of great practical importance to improve the granularity of services through a rule or algorithm to pre-aggregate the initial single resources, so as to improve the efficiency and accuracy of resource allocation.

Many scholars have conducted relevant studies on resource aggregation. Peng [3] used the process flow and the collaboration relationship between resources as constraints to related combinations of fine-grained manufacturing resources through homogeneous and heterogeneous aggregation. Zheng [4] proposed a resource aggregation method based on the manufacturing capability of manufacturing activities by analyzing the manufacturing capability of manufacturing activity units after task decomposition and aggregating the related matching resources accordingly. Pang [5] extracts the association collaboration laws among manufacturing resources from manufacturing data to abstracts functional unit and then encapsulates them into manufacturing services. Dong [6] draws on total quality management theory to classify resources in each domain according to people, machines, materials, methods, and environment, and to aggregate and encapsulate the corresponding resources based on actual manufacturing task activities. Based on the workflow logs and the configuration scheme of activities and resources in the workflow model, Li [7] mines the usage frequency of the physical manufacturing resource portfolio and provides a basis for mapping the physical resource portfolio to virtual resources. Zhu [8] combined with the workflow model to reflect the dependency relationship between resources by constructing the correlation degree matrix, and the resources with strong dependency relationship are combined together to complete the aggregation of resources.

All the above studies have explored the aggregation methods of multiple manufacturing resources in the cloud manufacturing environment, which largely promote the development of aggregation methods and improve the efficiency of service matching of resources. However, the current resource aggregation methods still have the following deficiencies. 1) Aggregation rules in resource aggregation methods rely heavily on expert experience and lack the use of objective historical data. 2) Resource aggregation is mostly task-oriented and top-down to bring together manufacturing resources with corresponding functions for packaging and combination, which leads to high difficulty and high computational effort in combining resource services. The actual cloud manufacturing tasks are diverse and uncertain, so the aggregation of resources for a specific task is often difficult to apply to other tasks that come later, resulting in inefficient service matching. In fact, the manufacturing service process will generate a large amount of manufacturing data, which contains the process of cooperation and association between resources. Thus, the information of association between resources can be extracted from these data and used to aggregate virtual resources to solve the above problem. Therefore, inspired by the idea of data mining, this paper proposes to mine the association relationships among resources from the historical manufacturing data of resources, so as to complete the aggregation among resources in a bottom-up way.

2 Problem Analysis

As the basis and key of cloud manufacturing, virtualization of resources refers to the description, encapsulation and classification of various physical resources, thus completing the mapping between physical and virtual resources to form a standardized and unified virtual resource cloud pool to meet the subsequent cloud service invocation of resources. The detailed process is shown in Fig. 1. Resource aggregation as part of virtualization aims to address the lack of functional expressiveness and the high complexity of service matching caused by single-grained resources.

In the actual generative manufacturing process, manufacturing resources are often combined with each other to complete certain manufacturing tasks, and some of these resources are always used together in fixed combinations due to their intrinsic relevance. For example, a certain processing equipment is always used by a specific operator or with a fixed auxiliary processing resource for manufacturing processing. If these resources with high dependency relationships are aggregated to form resource aggregator (multi-category resource unit) to provide services externally as a combination, tasks can be quickly matched to corresponding resources. How to get this correlation between resources, and then aggregate the corresponding resources to form a resource aggregator with larger granularity to provide services to the outside, so as to improve the efficiency of resource services, is the problem to be studied in this paper.

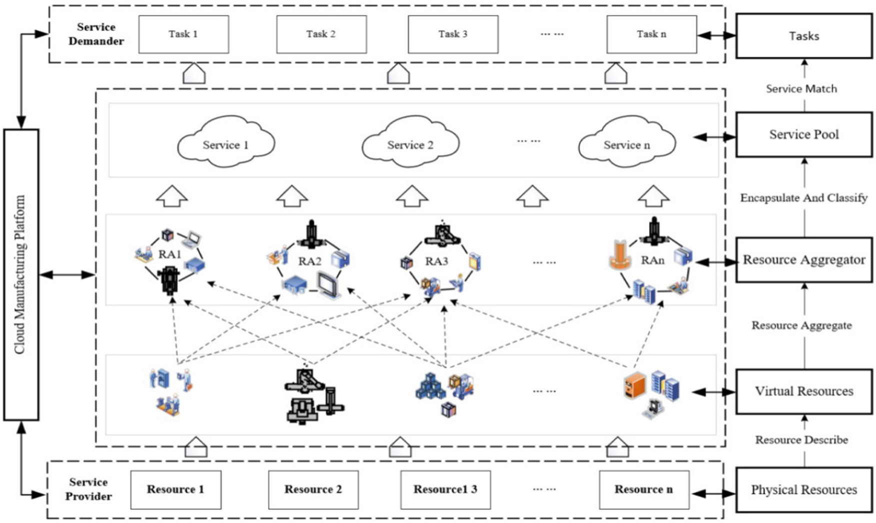

Fig. 1 Virtualization framework diagram

3 Resource Virtualization Aggregation Method Based on Correlation Analysis

In the actual manufacturing process of an enterprise, a large amount of manufacturing data is generated, which contains various manufacturing transactions and corresponding manufacturing resources. Association analysis can extract the interdependencies between resources from a large amount of transaction data, and these relationships are reflected in the form of frequent item sets and association rules [9]. Resource aggregation starts with processing the original data to obtain a normalized set of transaction data; then iterative loops are performed by using association rule mining algorithms to obtain association relationships among resources; finally, the existing resource information is combined to build and improve the resource aggregator to converge to the resource cloud pool to form a collection of various resource services. The resource virtualization aggregation method based on association analysis has the following steps: 1) constructing a transaction model for manufacturing activities; 2) mining resource association relationships; 3) aggregating associated resources. The flow of resource aggregator acquisition is shown in Fig. 2.

3.1 Manufacturing Activity Transaction Model Construction

Manufacturing activity data is the specific record of related manufacturing resources to complete the processing task, including the target processing object, processing time, processing process, and the specific process involving the use of various types of equipment and fixtures, etc. However, the workshop manufacturing activity data is cumbersome and redundant, requiring pre-processing such as attribute filtering and generalization. Combined with the actual association mining requirements, the filtered and simplified manufacturing activity transaction model is shown in Fig. 3. The manufacturing resource part is used as the main mining information to discover the association relationship between resources, and the manufacturing task part is used as auxiliary information to facilitate the extraction of relevant functions after the completion of subsequent resource aggregation.

The attributes involved in the above manufacturing activity transaction model, such as equipment type and tooling model, are often presented as multi-value type in the

Fig. 2 Resource virtualization aggregation process

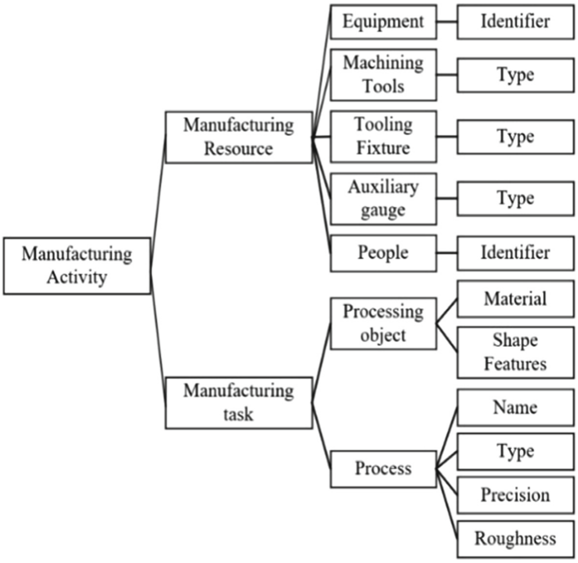

Fig. 3 Manufacturing activity transaction model

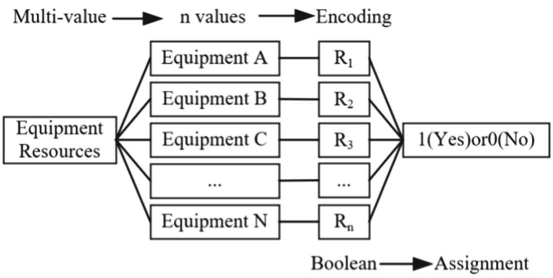

Fig. 4 Multi-value mapping of device resources

actual manufacturing activity transaction instance, and the current mainstream classical association rule mining algorithms are often Boolean-oriented, so it is necessary to encode the activity transaction data from a multi-value type to multiple Boolean data to facilitate the subsequent mining of frequent item sets. A schematic of the mapping transformation regarding device resources is shown in Fig. 4. Multiple devices are uniquely coded and assigned values based on their presence or absence in the active transaction. The rest of the resource mapping conversions are the same and are not listed here.

3.2 Resource Association Mining

Resource Association Mining Analysis

The object of association rule mining is a transaction set $T = \{t_1, t_2, t_3, t_4, \ldots\}$ that contains a large number of transaction instances, where each transaction t consists of several items I, that is, $t = \{I_1, I_2, \ldots\}$. The goal of rule mining is to analyze and extract the dependencies between items I from a large collection of transactions. In this paper, the manufacturing activities are treated as transactions, and the resources involved in manufacturing activities are treated as items for mining and analysis to obtain the association relationship between resources. Relevant indicators and rules in correlation analysis are defined as follows:

Support: For the total set of transactions T, the proportion of transactions containing a particular item set C is the support of that item set. The support degree is used to indicate the frequency of item set occurrence and is denoted by $Sup(c)$, for which we define the minimum support as *Minsup*, which is used to filter the item set. The support degree is calculated as follows:

$$Sup(c) = Count(T(c))/Count(T) \tag{1}$$

Confidence: For the transactions that contain item set C_1 in the full set of transactions T, if a certain proportion also contains item C_2, the proportion is the confidence of the association rule $C_1 \rightarrow C_2$ and is denoted by $Conf(c_1 \rightarrow c_2)$. We define the minimum confidence as *Minconf* and use this to measure the strength of the association rule. The confidence is calculated as follows:

$$Conf(c1 \rightarrow c2) = Sup(c1 \cup c2)/Sup(c1) \tag{2}$$

However, the confidence level has a limitation that it only considers the support of some terms such as C_1 and does not consider the effect of the frequency of the correlation term C_2. In fact, if the support of the association term C_2 is high, it will raise the confidence index. As a result, the importance of an association rule is often misestimated, thus the strong association rules filtered by support and confidence alone often do not have a true logical meaning. For this reason, we introduce the *lift* as a criterion for judging the logical significance of association rules.

***Lift*:** The ratio of the probability of containing item set C_1 in a transaction while containing item C_2 to the probability of containing item set C_2 is the *lift*. The *lift* reflects the correlation between C_1 and C_2 in the association rule. A higher $lift > 1$ indicates a higher positive correlation, a lower $lift < 1$ indicates a higher negative correlation, and a $lift = 1$ indicates no correlation. The *lift* is calculated as follows:

$$Lift(c1 \rightarrow c2) = Conf(c1 \rightarrow c2)/Sup(c2) = Sup(c1 \cup c2)/(Sup(c1) * Sup(c2)) \tag{3}$$

The process of association mining is to obtain the set of frequent items that meet the minimum *support* requirement through an iterative loop and filter them by *confidence* and *lift*, and finally arrive at the set of resources with high association strength to build resource aggregator.

Improve Apriori Algorithm Based on Multiple Support Degrees

As a classical association rule mining algorithm, Apriori algorithm [10] is mainly applied to Boolean data. The basic process is to get all frequent K_itemsets by iterating in the transaction database from low to high, which are those whose support is not lower than the set minimum *support* threshold. then build frequent $(K + 1)$_itemsets by "join" and "prune" and iterate through the cycle until no new frequent item-sets can be generated.

From the above introduction, it can be seen that the mining of frequent itemsets relies heavily on the setting of the minimum *support* threshold, which is set too high will filter out some important but less frequent itemsets, and set too low will produce a large number of worthless rules to affect the mining efficiency. Moreover, in the actual data mining process, the data volume of the target database is often different, and the mining results generated by a single artificial subjective setting of *support* threshold is obviously less convincing. For this reason, we introduce multiple-minimum support to optimize the traditional association rule mining Apriori algorithm and realize the dynamic tuning of minimum support by combining the statistical analysis of the support of candidate target item set, so as to weaken the influence of human subjectivity to some extent and better discover the association relationship between resources.

1) Multiple minimum support tuning model

For the differences of itemsets in different stages of the mining process, different minimum support degrees $Minsup_k$ are set. Moreover, we introduce the *preselection* degree to measure the appropriateness of the minimum support, and propose a *Minsup* tuning strategy based on the dichotomous method. The tuning model involves relevant definitions as follows:

Average Support *(Avesup)*: The average of the support of each item of the candidate item set C_k in each layer. The average support is calculated as follows:

$$Ave\sup k = (\sum Sup(c))/Count(Ck)\; c \in Ck \tag{4}$$

Preselection *(R)*: The ratio of the number of preselected frequent itemsets to the number of candidate itemsets obtained at a given minimum support threshold. The preselection degree is calculated as follows:

$$Rk = Count(Lk)/Count(Ck) \tag{5}$$

The minimum support tuning strategy is as follows:

$$\underset{new}{Min\sup} = \begin{cases} Min\sup - 0.5 \times (Min\sup - v)\; R_k < a \\ Min\sup + 0.5 \times (w - Min\sup)\; R_k > b \end{cases} \tag{6}$$

Minsup, $Minsup_{new}$ are the minimum support before and after adjustment respectively; *(v, w)* are the upper and lower limits of the current minimum support that can be adjusted, which are updated with each iteration of adjustment, and initially are the minimum and maximum support of each candidate set in the current hierarchy respectively; *(a, b)* are the moderate range of the established preselection degree R_k.

2) Multiple minimum support tuning process

In practical applications, for mining frequent itemsets at different levels, the *Avesup* of each candidate itemset at that level is first calculated and used as the *Minsup* to calculate the current preselection degree R_k, which reflect the moderate degree of frequent itemset. Then determine whether R_k is within the moderate range *(a,b)* and adjust minimum support according to the tuning strategy. The adjusted support $Minsup_{new}$ is used as the minimum support threshold to iterate cyclically until R_k falls to the range *(a,b)* or reaches the maximum number of iterations n. Finally, the current support or the optimal support after successive tuning is output as the minimum support threshold of the current hierarchy for subsequent frequent itemset mining. The algorithm model diagram of tuning is shown in Fig. 5.

Resource Association Mining Process Based on Improved MS-Apriori Algorithm. To address the problems inherent in the traditional association rule mining Apriori algorithm, this paper proposes an improved MS-Apriori algorithm, which is optimized by introducing the concept of multiple minimal support and formulating a dynamic tuning strategy for the support degree. The support threshold that can be dynamically adjusted weakens the subjective selectivity of human to a certain extent, and the association rules mined on this basis can better reflect the true relationship of the data. The specific mining process is shown in Fig. 6.

Step1: Initialize relevant data and parameters. The transaction set database D and the values of the parameters involved in the iteration of the algorithm should be confirmed, including the minimum confidence Minconf, the preselection degree moderate range (a, b) and the number of support tuning n.

Step2: Scan the database and construct the initial candidate resource set C_1.

Step3: Confirm the minimum support threshold and generate the frequent $K\_$ item set L_k.

By invoking the multilayer support tuning model described in the previous section, the minimum support threshold is confirmed at each stage; the support of each item in

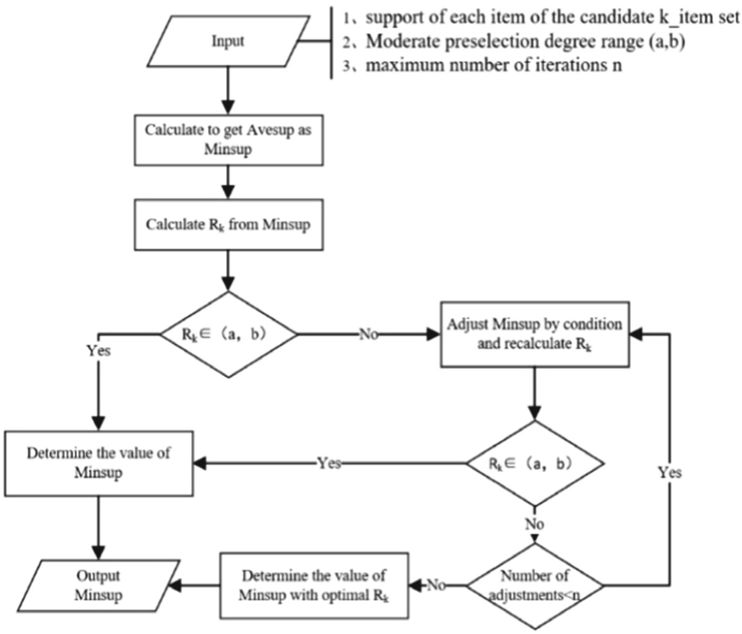

Fig. 5 Support tuning process

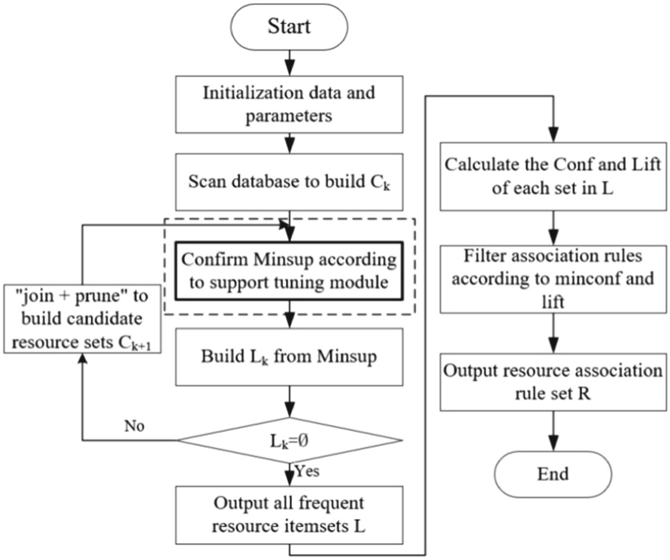

Fig. 6 Resource association mining process based on improved MS-Apriori algorithm

the candidate resource set is obtained, and the set of items meeting the minimum support requirement is screened to form the corresponding frequent item set.

Step4: "join + prune" to build a list of candidate resource sets C_{k+1}.

The list of candidate resource sets C_{k+1} indicates that the number of resources contained in each resource set in the list is k + 1. It is composed by cross-linking the resource sets in the list of frequent resource sets L_k generated earlier. The cross-linking rule is as in Eq. 7: where *m*, *n* are the resource sets containing k resources in L_k respectively, and $m \cup n$ denotes the resource sets including both m and *n* resources.

$$Ck = \{m \cup n|m, n \in Lk\} \tag{7}$$

When the size of the resource set is large, the list of candidate resource sets C_{k+1} needs to be "pruned" to avoid generating too many candidate data items and slowing down the algorithm. Without calculating the support, delete the set of resources in C_{k+1} for whose subset are not frequent sets in L_k according to the a priori principle of the Apriori algorithm.

Step5: Iteratively loop step 3 and step 4 until no new frequent itemsets can be generated. Output all frequent resource itemsets *L*.

Step6: Association rules filtering analysis.

Frequent resource set association rules are extract and the confidence and lift are calculated to filter the association rules that meet the predefined threshold requirements. The corresponding frequent resource set is the final aggregation object.

Associated resource aggregation.

Resource aggregator is functional manufacturing resources that shield the underlying physical resource information after the resources have been processed by virtualization

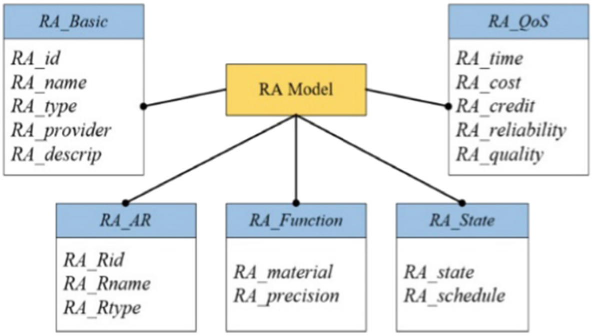

Fig. 7 Resource Aggregator Information Model

aggregation. After obtaining the set of resources with strong association through association mining analysis, it is necessary to further integrate the relevant raw resource physical property information and capability information to form directly invokable services. To this end, the structured information model of the resource aggregator is shown in Fig. 7:

The model normalizes the description of the resource aggregator in five aspects: basic information, associated resource information, functional information, status information, and quality information. The above model can be formally represented as follows:

$$RA =< RA\_Basic, RA\_AR, RA\_Function, RA\_State, RA\_QoS >$$

where *RA* denotes a resource aggregator; *RA_Basic* is the basic information of this resource aggregator, including the resource aggregator unique identifier (*RA_id*), name (*RA_name*), type (*RA_type*), provider (*RA_provider*) and related description (*RA_descrip*); *RA_AR* is the resource information associated with this resource aggregator, such as manufacturing equipment, operators, auxiliary equipment, etc., including the resource unique identifier (*RA_Rid*), resource name (*RA_Rname*) and type (*RA_Rtype*); *RA_Function* is the functional information of this resource aggregator, including the processing material requirements (*RA_ material*) and processing precision (*RA_ precision*); *RA_State* is the status information of this resource aggregator, including the current task state (*RA_state*) and execution progress (*RA_schedule*); *RA_QoS* is the quality information of this resource aggregator, including manufacturing time (*RA_time*) and cost (*RA_cost*), Credibility (*RA_ credit*), reliability (*RA_ reliability*) and quality compliance rate (*RA_ quality*).

4 Case Study

In this paper, the workshop production data of a manufacturing company is used as an example to verify the proposed resource aggregation method. First, the data is preprocessed based on the manufacturing activity transaction model to construct the transaction dataset. And part of the manufacturing activity transaction set is shown in Table 1.

Table 1 Set of manufacturing activity transactions

	{Equipment, machining tools, auxiliary tooling, gauges, person}	{task, material, precision…}
A1	CK61100 × 5000 CNC horizontal lathe, adjustable three-jaw self-centering chuck, YW2 turning tool, Vernier calipers, turner #23	rough turning outer circle, 5CrMnMo, low precision…
A2	TK6920A/130 × 40 CNC floor milling and boring machine, four-jaw compound chuck, bore alloy boring tool, micrometer, boring worker #16	fine-boring, 4Cr5MoSiV1, high precision…
A3	SKD40 CNC double column vertical lathe, adjustable three-jaw self-centering chuck, YT30 turning tool, Vernier caliper with table, turner #2	fine turning outer circle, 4Cr5MoSiV1, high precision…
A4	HTIIIG250 × 120/200-NC CNC lathe grinder, four-jaw compound chuck, YBG205 indexable machine chuck insert, Vernier caliper with table, turner #18	semi-finish turning outer circle, 2Cr13, medium precision…
A5	HTIIIG250 × 120/200-NC CNC lathe grinder, four-jaw compound chuck, WA46K white corundum grinding wheel, Vernier caliper with table, grinding worker #8	grinding, 2Cr13, high precision…

Subsequently, all manufacturing activities and the resources are coded, the transaction set database D is constructed, and the parameters are initialized and assigned values. The *Minconf* was set to 0.65, the maximum number of support tuning iterations $n = 10$, and the moderate preselection range (0.33, 0.67). Some of the association rules obtained from association analysis mining are shown in Table 2. The first of these rules indicates that CK61100 × 5000 CNC horizontal lathe, adjustable three-jaw self-centering chuck, YW2 turning tool and auxiliary gauge Vernier-calipers are often cooperated with turner #23 to complete the machining tasks. By tracing the information of the manufacturing activity model, it was found that mainly about rough turning of outer circle related work.

Finally, the resources involved in each rule are aggregated according to the aggregator information model by combining the corresponding manufacturing task information in the transaction database and the resource information in the resource database. The resource aggregator is shown in Fig. 8.

Table 2 Rules for association between resources

Number	X	Y	Sup	Conf	Conf
1	CK61100 × 5000 CNC horizontal lathe, adjustable three-jaw self-centering chuck	YW2 turning tool, Vernier caliper, turner #23	0.52	0.68	1.04
2	FAF-260 CNC floor milling machine	cylindrical milling cutter, milling worker #2	0.66	0.657	1.13
3	HTIIIG250 × 120/200-NC CNC lathe grinder	Four-jaw compound chuck, YBG205 indexable machine chuck insert	0.57	0.68	1.02

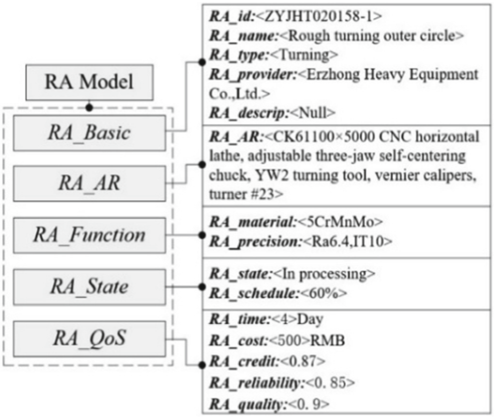

Fig. 8 Example of resource aggregator

5 Conclusion

With the current trend of manufacturing industry changing information-based intelligent manufacturing accelerating, the cloud manufacturing model continues to receive attention, and resource virtualization as the key process of cloud manufacturing realization and its related research has significant value. By analyzing the resource virtualization process, this paper proposes a bottom-up resource aggregation method based on the improved MS-Apriori correlation analysis algorithm for the resource aggregation problem in the virtualization process, combined with the idea of data mining. Firstly, the resource association mining model is constructed, then the data mining algorithm is used to mine and analyze the association relationship between resources from historical manufacturing data, and finally the information model of resource aggregator is constructed to complete the relevant resource aggregation, which realizes the bottom-up aggregation of resources. Among them, multiple minimum support and support tuning model are introduced for optimization for the lack of objectivity of traditional Apriori association rule mining algorithm. Finally, the feasibility of this method is verified with examples.

This study lays the foundation for the subsequent improvement of resource service matching efficiency and the realization of efficient resource allocation and scheduling, and provides a reference direction as well as a theoretical basis for the study of resource virtualization aggregation methods.

Acknowledgements. This work was partly funded by Project supported by National key research and development plan project (No. 2018YFB1703002).

References

1. LI Bohu, ZHANG Lin, WANG Shilong, et al. Cloud manufacturing: a new service-oriented networked manufacturing model [J]. Computer Integrated Manufacturing Systems,2010,16(1):1–7,16.
2. REN Lei, ZHANG Lin, ZHANG Yabin, et al. Resource virtualization in cloud manufactur-ing [J]. Computer Integrated Manufacturing Systems,2011,17(03):511–518.
3. PENG Gongzhuang, WU Youqi, He Anrui. Research on modeling and scheduling method of multi-granularity manufacturing service[J]. Journal of Huazhong University of Science and Technology (Natural Science Edition),2020,48(5):80–85.
4. ZHENG Jun, WANG Lihang. Manufacturing capability modeling and measurement model based on manufacturing activity[J]. Computer Integrated Manufacturing Systems,2018,24(12):3038–3049.
5. PANG Shibao, GUO Shunsheng, WANG Lei, et al. Constrution Method of Workshop Manufacturing Service Collaboration Chains for Mass Personalization Manufacturing [J]. China Mechanical Engineering,2020,31(17):2104-2111.
6. DONG Yuanfa, WU Zhengjia, DU Xuan, et al. Meta-activity Model Driven Grain Layered Service Encapsulation and Retrieval for Multi-domain Manufacturing Resources [J]. China Mechanical Engineering,2018,29(12):1475-1484.
7. LI Haibo. Approach to multi-granularity resource composition based on workflow in cloud manufacturing [J]. Computer Integrated Manufacturing Systems,2013,19(1):210-216.
8. ZHU Yinjuan, LI Haibo. Resource composition based on correlation degree in cloud manufacturing [J]. Computer Engineering and Applications,2016,52(5):255-261.
9. Bi Jianxin, Zhang Qishan. Survey of the Algorithms on Association Rule Mining [J]. Strate-gic Study of CAE,2005(04):88-94.
10. Rao Zhengchan, Fan Nianbai. A review of associative rule mining Apriori algorithm [J]. Computer Era, 2012(09):11-13.