Chemometrics and Intelligent Laboratory Systems 205 (2020) 104119

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemometrics

Maximal structure generation of superstructure for semantic triple generated by DEVS ontology in the process industry

Jian Cao [a,b], Yan-Lin He [a,b,**], Qun-Xiong Zhu [a,b,*]

[a] College of Information Science & Technology, Beijing University of Chemical Technology, Beijing, 100029, China
[b] Engineering Research Center of Intelligent PSE, Ministry of Education of China, Beijing, 100029, China

ARTICLE INFO

Keywords:
Superstructure
P-Graph
DEVS Ontology
Process industry

ABSTRACT

In the process industry, a large number of multi-source heterogeneous data are generated. Ontology combined with resource description framework (RDF) is a great way to describe multi-source heterogeneous data. Due to the characteristics of free extension, no model limitations and few constraints in the calculation process, semantic data have been increasingly used in industrial systems. However, semantic data have not been fully utilized to realize automatic modeling. Lots of time and experience from knowledge workers is still required in the feedstock scheduling. Therefore, this article presents a method of constructing the RDF triple graph data based on the discrete event system specification ontology. Based on the advantage of graph data, the procedure of superstructure generation is greatly simplified by the proposed RDFMSG (maximal structure generation based on resource description framework data) algorithm. In the RDFMSG algorithm, the original structure is no longer pruned and then reorganized; the superstructure can be generated by the traversing RDF data set only once. The effectiveness and efficiency of the proposed methodology are verified through three cases. In addition, the representation of the discrete event system specification ontology can achieve good performance. Furthermore, the combination explosion caused by set operation can be greatly reduced, indicating that the strategy of the RDFMSG plays an important role in the solution structure generation algorithm and the accelerated branch and bound algorithm.

1. Introduction

In 2013, McKinsey released a report titled Outlook 2025, Determining 12 Major Subversion Technologies for the Future Economy, knowledge worker automation is taken as the second disruptive technology and estimated the economic impact would increase about 5.2–6.7 billion US dollars by 2025. The German government proposed the concept of Industry 4.0 as the German 2020 High Technology Strategy [1]. As one of the ten future projects, the technical foundation of Industry 4.0 is the Cyber-Physical System (CPS) and the Internet of Things [2]. Industry 4.0 refers to the use of the CPS to digitize supply, manufacturing, and sales information in production, and finally a fast, effective, and personalized product supply is achieved. The vision of Industry 4.0 is to establish a knowledge-based industrial intelligent Internet of Things(IOT) information system. In the system, the complex industrial processes are optimized to achieve the purpose of reducing energy consumption and improving efficiency.

Knowledge-based CPS systems rely heavily on knowledge data storage and management. Industrial systems are confronted with important issues such as unstructured, big data, slow retrieval speed and poor scalability. The most fundamental and urgent problem is how to store multi-source heterogeneous data in a unified manner or build a good storage framework [3,4]. Resource description framework (RDF) proposes a standard data model for the fusion of multi-source heterogeneous data, and adds data semantic information at the same time [5]. RDF is a good choice for data representation, data fusion and data extension but lacks organization. So the discrete event system specification (DEVS) and ontology are applied to constrain and regulate the RDF data [6]. Meanwhile, simple reasoning and interpretation can be implemented through strategies consistent with human thinking.

With the enrichment of semantic data in industry, this paper designs an algorithm for solving the superstructure modeling based on RDF data for process synthesis problem [7]. In the proposed algorithm, a graph storage structure based on RDF data is realized within the scale

* Corresponding author. College of Information Science & Technology, Beijing University of Chemical Technology, Beijing, 100029, China.
** Corresponding author. College of Information Science & Technology, Beijing University of Chemical Technology, Beijing, 100029, China.
E-mail addresses: heyl@mail.buct.edu.cn (Y.-L. He), zhuqx@mail.buct.edu.cn (Q.-X. Zhu).

https://doi.org/10.1016/j.chemolab.2020.104119
Received 6 June 2020; Received in revised form 13 July 2020; Accepted 20 July 2020
Available online 5 August 2020

Table 1
RDFs class, property.

Name	Type
Rdfs:Resource	Class
Rdf:Property	Class
rdfs:Class	Class
rdf:type	Property
rdfs:subClassof	Property
rdfs:subProperty	Property
rdfs:range	Constraint
rdfs:domain	Constraint

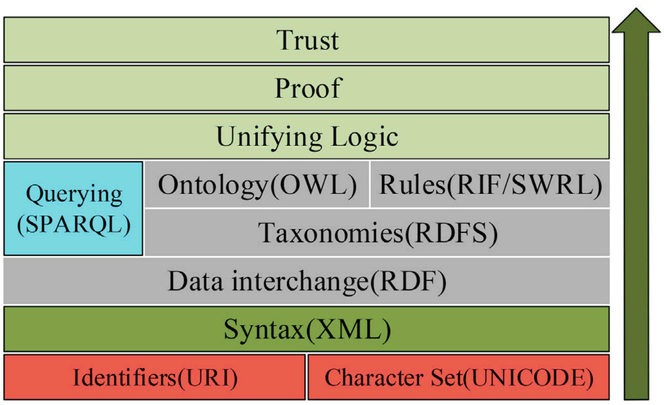

Fig. 1. Semantic language hierarchy.

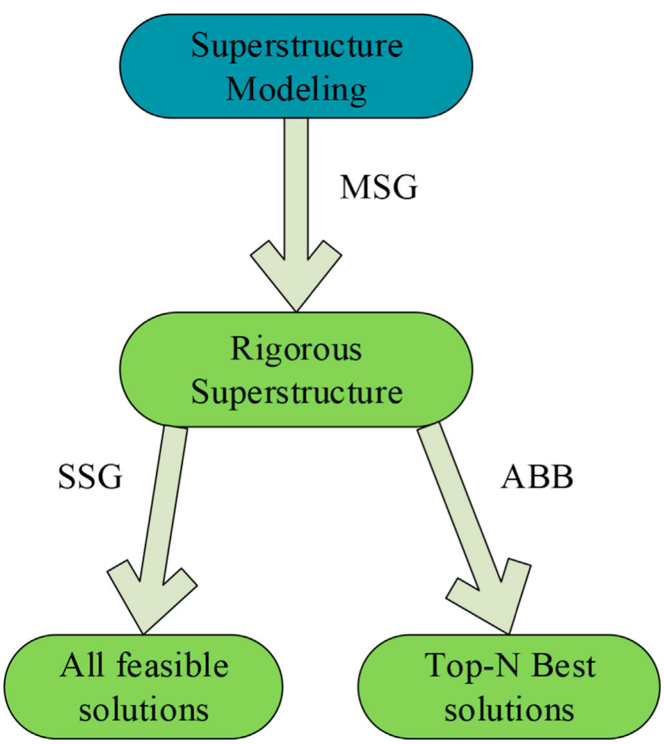

Fig. 2. Application process of the P-graph theory.

controllable range. The maximal structure generation (MSG) algorithm in the P-graph and the corresponding ideas of solution structure generation (SSG), accelerated branch and bound (ABB) algorithms are given. This work provides important reference for the subsequent application based on semantic data.

This paper focuses on designing the MSG algorithm in the P-graph

$$P \subseteq M,\ R \subseteq M$$

$$O \subseteq (\zeta(M) \times \zeta(M))$$

$$O \cap M = \varnothing,\ P \cap R = \varnothing$$

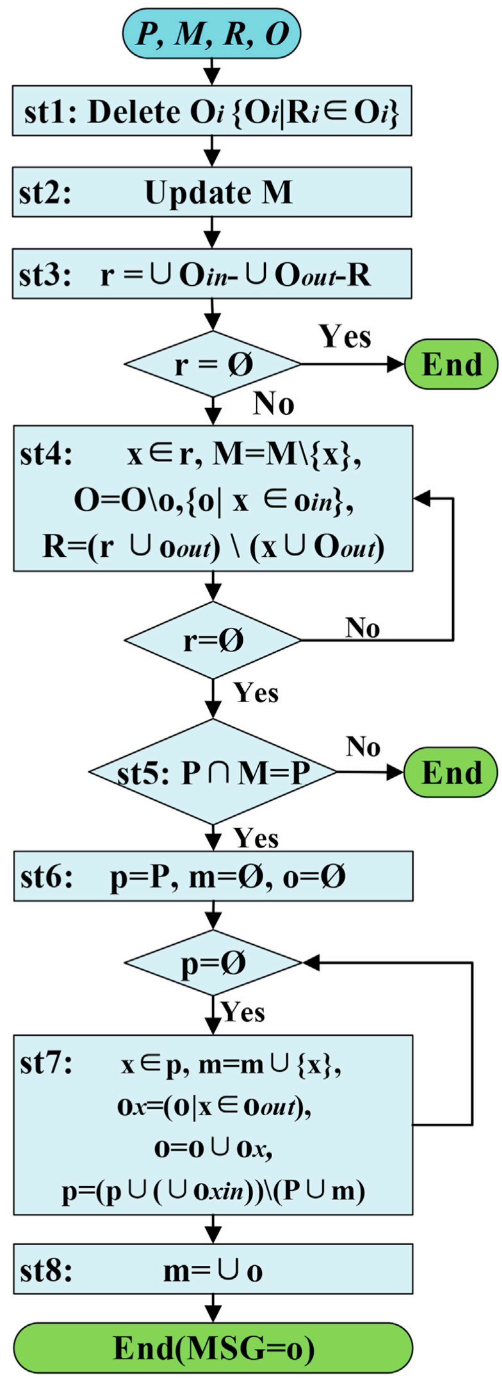

Fig. 3. MSG flow chart.

Table 2
Operation unit is convert into RDF <s, p, o>.

Operation Unit	RDF Format<s, p, o>
O1: {C, D, F}{A}	C, in, O1 D, in, O1 F, in, O1 O1, out, A

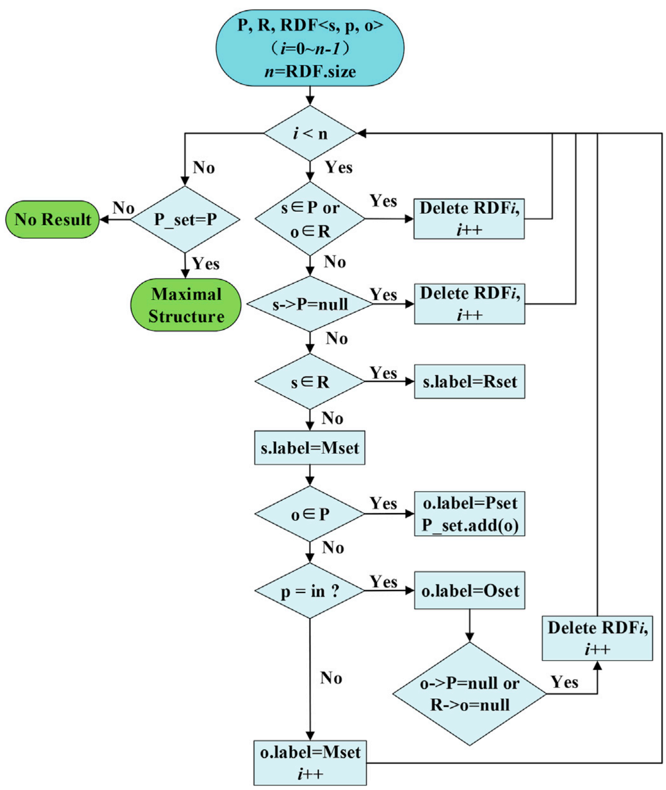

Fig. 4. RDFMSG flow chart.

based on the RDF data. In section 2, the advantages and application scenarios of the RDF semantic data based on DEVS ontology are introduced. In section 3, the specific industrial field application of the P-graph and the framework of the RDFMSG algorithm are proposed. In Section 4, the correctness and feasibility of the proposed algorithm are verified by two cases, and advantages and disadvantages of the algorithm are analyzed. Another Case is a DEVS model expansion in time perspective. Finally, the main work of this paper and ideas for future research are summarized.

2. Related work

2.1. Discrete event system specification

DEVS is a hierarchical organization of the modeling proposed by Zeigler in 1976 [8], and has the ability to express all dynamic systems, formal description of discrete and continuous systems. DEVS uses component hierarchies to express relationships between components of the modelling system and supports hierarchical, modular model development. DEVS consists of two components: the atomic DEVS model and the couple DEVS model. The atomic model is the most basic constituent element of the discrete system, mainly used to define the behavior of the system, express the state change. Besides, specific functions of the model under the influence of external events and corresponding output information can both be produced in the atomic DEVS model. The couple DEVS model consists of several atomic DEVS models.

As DEVS has developed over decades, it has been well applied in modeling and simulation. Alshareef designed a way to apply UML activities to specify DEVS model behaviors [9], the semantics are combined to meet with time-accurate simulation requirements, and to some extent the cost of model reconstruction is reduced. The E-CD Boost library was applied to build a real-time DEVS model and implement in hardware [10]. With the room occupancy as the input, this model can realize automatic control of lighting and other emergency systems. As a result, energy consumption and greenhouse gas emissions were greatly reduced. Celik used the DEVS-suite simulator tool to model problems that may occur when developing infrastructure solutions for mobile ad hoc networks (MANETs) [11]. Then the ant colony based load-balancing scheme was used to verify that the model can maintain the memory consumption stability despite the rapid growth of traffic and node. Jarrah proposed a DEVS-based approach to simulate photovoltaic arrays, wind turbines, storage devices, and load requirements in a smart grid [12]. Renewable resources with information technology are combined to make accurate design decisions for optimal and clean power generation. For modeling and reuse of the simulation model, Peng combined the ontology and principles of the DEVS simulation model, and recorded the change of the simulation model, the ontology atomic operations and model modification operations are defined simultaneously [6]. Based on the JMASE DEVS scheduling engine, a frame-based real-time event scheduling strategy and a multi-threaded real-time event scheduling algorithm were proposed to meet the needs of real-time simulation [13], which provides effective suggestions for our work. Compared with Petri nets, the DEVS has advantages of display parallelism and separation of modelling and simulation. Based on the existing two-stage mining method, Wang designed a mapping method to directly transform the transition system into the development system [14]. DEVS formalism is also applied to model and simulate the car parking problem [15], and intelligent parking with flow management, space management and traffic system development are combined. Thus a smart city-based intelligent parking solution is created, problems such as unreliable on-demand distribution and insufficient public transportation with the increase of vehicles have been well solved.

2.2. Resource description framework and ontology

RDF, a data model expressed by XML (Extensive Markup Language) syntax, was proposed by the World Wide Web Consortium, i.e. w3c in 1999 [16,17]. RDF is used to describe the characteristics of web resources and relationships between resources. The RDF-based data model mainly includes three object types: (1) resources, which are unique types of entities described by the model and named by Uniform Resource Identifier (URI); (2) attributes, which are used to describe the characteristics of a resource or relationship with other resources; (3) statements, a specific resource is stated by an attribute and a corresponding attribute value. Thus, a RDF can be defined, whereas the resource is defined as subject, the attribute as predicate, and the attribute value as object. In other words, the RDF can be expressed in a triple form of <subject, predicate, object>, which can be simplified as <s, p, o>. Data pattern of the RDF can be described as a graph model, wherein the subject and object are vertices. Actual storage is defined for the subject and the object labels so as to be distinguished, and the predicate is defined as a directed edge in the middle. The triple form is a basic semantic expression, and RDFs (RDF Schema) modeling language is also generated with broader attribute values and constraints extended on the basis of RDF, as shown in Table 1.

With the continuous development of semantic representation, a complete semantic representation language stack has been formed as shown in Fig. 1. The encoding form of the resource is specified by Unicode, and the URI sets a unique identifier for each resource. XML sets the content of the resource and data structure, but XML has no semantic description capability. RDF and RDFS are mainly used to describe the

```
Comment: MSG based on RDF
Input: R_set, P_set, RDF set<s, p, o>
Output: RDF set result <s, s.Label, p, o, o.Label> // s.lable, o.label∈{Pset, Mset, Rset, Oset}
---------------------------------------------------------------
P_temp_set //store temporary P

For i each RDF：//traverse the RDF to exclude data that do not meet the proposed axioms
  if  s ⊂P_set  or  o⊂R_set  //location semantic information does not follow axioms
     Delete RDF_i
  if  s⊂R_set
     s.label=Rset  // add P-graph element labels to RDF data
  else
     s.lable=Mset
     if s->P_set==NULL //delete this RDF data when there is no full path
          Delete RDF_i
  if  o⊂P_set
     o.label=Pset
     P_temp_set.add(o) //save when the object position element is a product
  elif p==in
     o.label=Oset
     if o->Pset==NULL or Rset->o ==NULL //delete this RDF data when there is no full path
        Delete RDF_i
  if(P_temp_set != P_set)
     then stop //according to the original rules of msg, there is no maximal structure
  else
      maximal structure = RDF
```

M= {A, C, B, D, E, F, G, H, I, J, K, L, M, N, Q, T, U, V},

P= {B},

R= {F, H, M, T},

O={({C,D,F},{A}),({D},{B,G}),({E},{B,U}),({F,G},{C,D}),({G,H},{D}), ({H,I},{E}), ({J,K},{E}),({M},{G}),({N,Q},{H}),({T,U},{I}), ({V},{J})}

relationship between resources by the way of <subject, predicate, object>. SPARQL (SPARQL Protocol and RDF Query Language) is a query language for querying RDF data. OWL (Web Ontology Language) adds constraints on attributes and classes on the basis of RDF to facilitates further classification. Besides, OWL is also conducive to RIF (Rule Interchange Format) to achieve inference. Based on the above content, the layers of Unifying Logic, Proof and Trust can ensure reasonable, accurate and trusting interaction with users.

Graph data based on the RDF organization is stored in a NoSQL data base, and relational data are verified to reduce the storage space to a certain extent. Considering the data complexity, efficiency and diversity of queries, graph pattern is used to manage semantic data as efficient systems. For example, the TDB in the Jena framework uses B+ trees to maintain three basic forms of Triple index. Node table and other information is stored in the format of SPO, POS and OSP respectively, and the TDB is similar to RDF-3X, a stand-alone RDF data management engine. Neo4j is also a NoSQL high-performance graph database, triple data are stored on the network map, with embedded, lightweight and other advantages. Neo4J has been widely used in industrial systems. The SPARQL is an officially published and recommended RDF query language in 2013 [18]. The SPARQL is a query language based on subgraph mode, in the form of 'select ?s in {?s p o}', that is, the query attribute and the attribute value are p and o, whereas ?s is the variable that you want to query. Even there are millions of RDFs in the NoSQL database, triple collections can also have high query efficiency. One SPARQL query syntax, a representation model framework and the transformation algorithm with SPARQL is effectively applied to solve problem of the temporal data query and management [19]. Besides, a probabilistic SPARQL

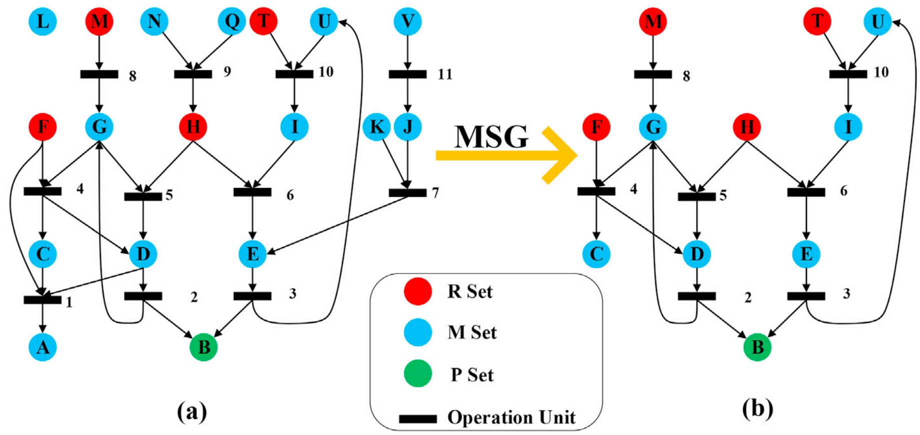

Fig. 5. MSG procedure in literature.

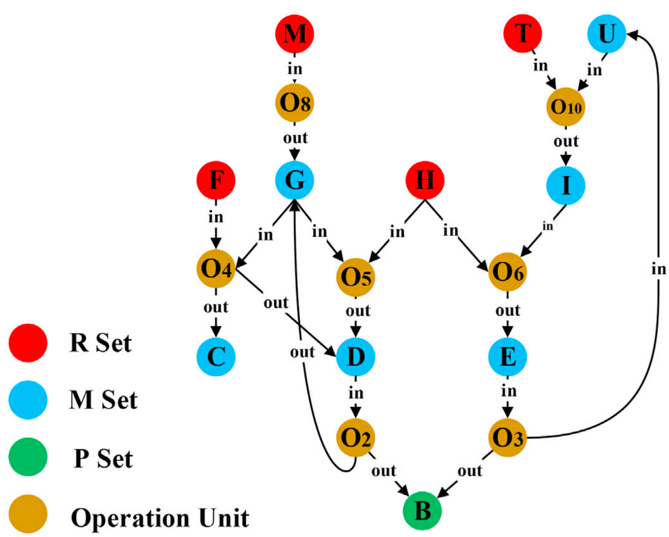

Fig. 6. RDFMSG Algorithm stored in Neo4j.

Table 3
Add RDF triples for testing axioms.

RDF triples	Axiom
<X,in,O_{12}>	S2
<B,in,O_{12}>	S4

query language framework is proposed to solve the uncertain query problem in RDF data [20]. Li et al. proposed a methodology to realize subgraph pattern matching through fuzzy RDF graph modeling [21], and the problem of low query efficiency under new conditions and constraints in the matching process is solved simultaneously.

Generation of RDF data depends on the logical description of the real word by knowledge workers. In industrial systems, the data model can describe the units, devices, systems, materials, products and equipment. As these subsets contain rich relationships, domain experts need to abstract these entity models and relationships to construct an abstract framework for RDF semantic data generation. Ontology is proposed by Gruber as a formal description of the conceptualization of things [22]. Based on the RDF-based construction method, computer can store and calculate data according to human understanding. In specific application practice, the ontology is instantiated [23]. The above process will generate a large number of RDF data organized according to the ontology, which can independently express the precise information content. Knowledge maps of specific vertical domain can be formed, and the application is very extensive, but further research and development is still needed. Based on the basic framework principle of the DEVS, the existing simulation-agnostic SysML of domain enterprise information system was transformed into executable simulation code, and simulation results were integrated through model driven architecture combined with expert knowledge [24]. Karhela proposed a solution based on semantic data modeling and simulation, and integrated the database system with ontology mapping technology to utilize its advantages [25]. Based on domain ontology driver, the simulation model of aviation scene was constructed and the automatic transformation into XML schema had been realized simultaneously [26].

2.3. Maximal structure generation algorithm (MSG)

The construction process of the superstructure combines multiple parallel alternatives to form a redundant network structure, enterprises and governments can be guided to realize the construction, reconstruction or expansion of industrial process systems. Modern industrial systems have many types, and obviously are different from each other. Therefore, it is difficult to construct a unified large-scale system superstructure. So far, superstructures are generally aimed at specific subsystems, mature superstructure models are the stage-wise superstructure proposed by Grossmann et al. and the state-space superstructure proposed by Bagajewicz [27–29]. A special form of superstructure modeling methodology named the P-graph (Process graph) theory by Friedler et al. [30]. The data set of a P-graph model is mainly composed of the raw set (R), material set (M), product set (P) and the operation unit (O). An operation unit refers to a specific device or a concrete operation, mainly contains an input set and an output set. Input set and output set of one operation unit are respectively a subset of M. The procedure of constructing a P-graph is mainly divided into three steps: (1) all feasible structures need to be defined in the problem domain, after that, a rigorous superstructure can be obtained by the MSG algorithm; (2) all feasible structural solution set can be generated based on the SSG algorithm; (3) Top-N best solutions can be selected from the set of solutions generated in step (2) through the ABB algorithm. The whole calculation process and the application process of above three algorithms are shown in Fig. 2.

P-graph has achieved excellent results in process synthesis and optimization scheduling. By using P-graph and Monte Carlo simulation method, Tan et al. obtained the approximate optimal network with minimum CO_2 emission [31,32]. Cabezas used the P-graph framework to solve the supply chain problems in sustainable energy production [33]. How introduced ANN into the application process of P-graph [34], and the problem with linear default parameters in the P-graph software was solved. Süle took the P-graph as the interface between the process flow

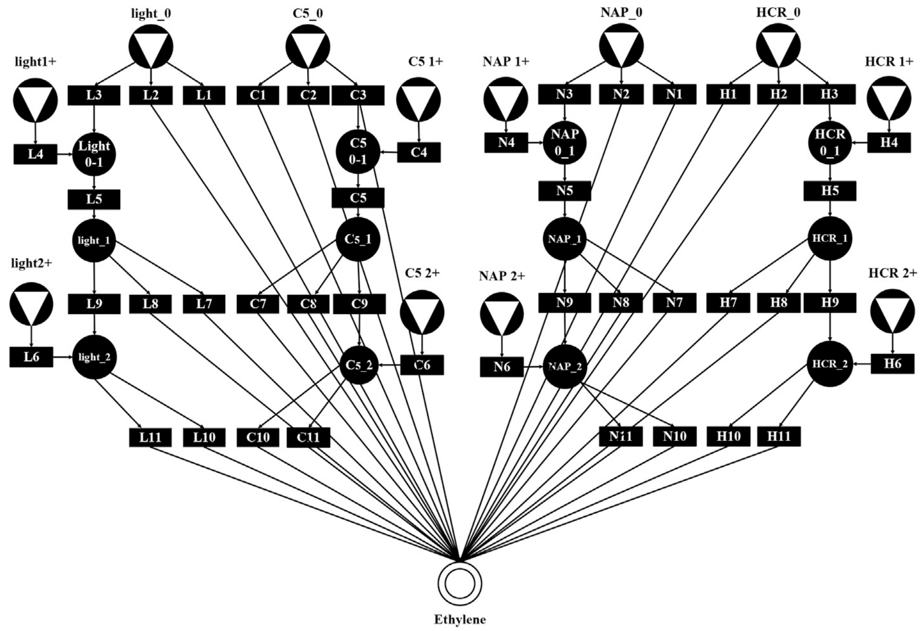

Fig. 7. The superstructure of four raw materials distribution for two plants in three months.

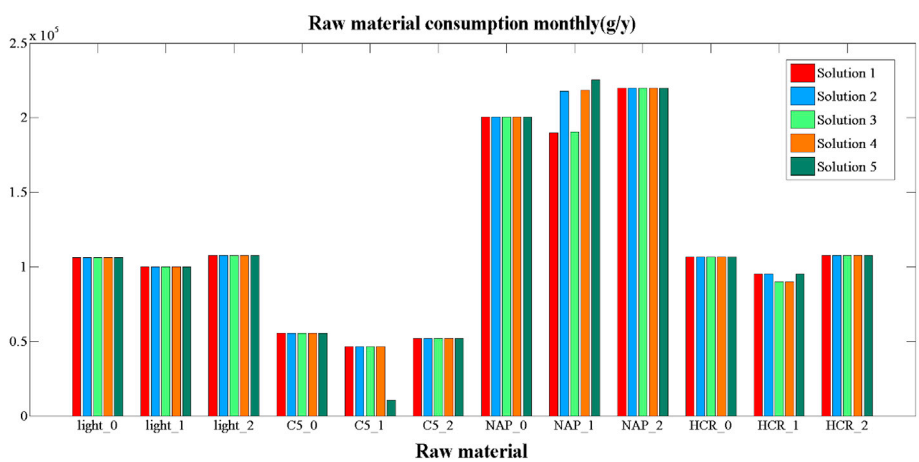

Fig. 8. Raw material consumption(g/y).

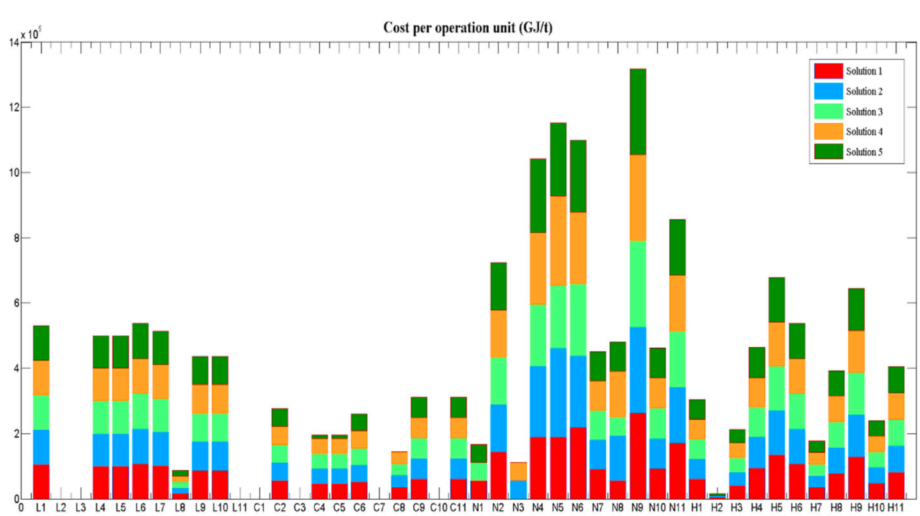

Fig. 9. Atomic model instance of Operation unit consumption.

chart and the risk model [35], and proposed a new multi-objective optimization method, which effectively evaluated the criticality between units and subsystems. Chin solved the problem of saving energy and water in heat-integrated water network with the P-graph method [36].

In addition to MSG, SSG and ABB algorithms, Friedler also proposes five axioms as follows. **S1**: Each product is represented in the superstructure. **S2**: M node has no input if it is only a raw material. **S3**: Each O-node must belong to the Cartesian set of the material set. **S4**: Each O-node must have a path connected to the product, and each path must include an end product. **S5**: If a node belongs to the model network, it must be an input or output of an O-type node. This paper researches on the MSG algorithm based on RDF data, mainly including the reduction and combination procedures. The flow chart of the MSG algorithm is shown in Fig. 3. The inputs are defined as P, M, R, O sets; the O set is the Cartesian product of the M set.

$$P \subseteq M, \cdot R \subseteq M \tag{1}$$

$$O \subseteq (\zeta(M) \times \zeta(M)) \tag{2}$$

$$O \cap M = \varnothing, \cdot P \cap R = \varnothing \tag{3}$$

Although practical application of the algorithm is very efficient, the semantic information of the graph structure is not fully utilized. SSG and ABB algorithms contain a large number of mathematical combination calculations, combinatorial explosion problem often occurs in SSG. Even the tool P-Graph Studio commonly used in the optimization of superstructure often fails to give results due to the combinatorial explosion problem. Therefore, in order to solve the above problems, the semantic information of RDF triple data is fully applied in this paper. Simple storage with rich semantic information in triple data makes the MSG consume a small amount of calculation work on path traversal, but overall procedure of the algorithm is simplified.

3. RDFMSG algorithm based on RDF data generated by DEVS ontology

Under the framework of P-graph theory, ontology is constructed for the operation unit in the process synthesis problem based on DEVS, and an atomic model ontology containing sets of input and output [37]. The atomic model guarantees that RDF data conform to axioms in the P-graph, then the RDF set is optimized.

The atomic model of the operation unit

M $<X, Y, S, \delta_{int}, \delta_{ext}, \lambda, ta>$
X input event set
Y output event set
S state set
δ_{int} internal state transfer function
δ_{ext} external state transfer function
ta time advance function

Coupled model of P-graph

W $< X, Y, M, E_{IC}, E_{OC}, IC, SN >$
X external event input interface
Y external event output interface
M atomic set
E_{IC} external event input connected with M
E_{OC} external event output connected with M
IC coupled relationship between atomic models
SN selection function that determines the order in which states occur when more than one state changes

In process industry, the current state of the plant will change with the adjustment of time. For example, the feedstock scheduling of the plant will change with the operation efficiency. RDF data are generated by DEVS ontology based on above atomic model, as shown in Table 2.

RDF data <s, p, o> are based on the graph pattern, and RDF format is rich in semantic information. In order to make full use of semantic information and speed up the optimization process of superstructure, it is necessary to define the unique axioms and constraints as follows:

S1: Every product should be displayed in the RDF set.
S2: The elements in the product set P can only be in the o position of the RDF, and elements in the set R can only be in the s position of the RDF data.
S3: RDF triples contains all elements in the product set P.
S4: RDF set and any s in RDF format <?s in O>, must have triple <O out ?o>, the correct operation unit with input and output to be guaranteed.
S5: In addition to elements in the raw material set R and the product set P, other elements in the M set must have paths to one element in the product set P.
S6: At least one raw material in the R set has a path to the operation unit in the current RDF.
S7: The operation unit in the current RDF has at least one path to the product set P.

The proposed RDFMSG iteratively traverses each triple data of RDF, according to the positional semantic information of <s, p, o> and basic rules of S1~S7, each RDF data are judged one by one whether to be saved, and labels are added to elements p and o simultaneously. The semantic information of RDF is enriched, and the storage of data, retrieval and application calculation of algorithms all will have great convenience.

The flow chart of RDFMSG is shown as Fig. 4:

Based on the above axioms and rules, the pseudo code of the RDFMSG algorithm is summarized as follows:

RDFMSG algorithm has obvious advantages as the following: (1) the ontology based on DEVS atomic model ensures rationality of RDF data, the generation of unreasonable data in RDF set is reduced. When RDF data are large in scale, the graph data storage system has many advantages, for example, semantic information can be expressed flexibly and easily extended; (2) The RDFMSG algorithm is no longer a method of set theory, instead, it is implemented from the perspective of the network path. The RDFMSG results can be obtained by traversing the RDF set, and at the same time, the label of each type of node is given without pruning the structure; (3) In the representation of RDF, semantics are used as edges in the graph database, the efficiency of graph data retrieval can greatly be improved in this way. At the same time, SPARQL, as a unique RDF retrieval language, has completed a large amount of research work on retrieval optimization, so the retrieval time consumption can also greatly be reduced [38–41]. However, SSG algorithms are all traversed from the perspective of mathematical combination. After the storage system of RDF graph data is selected, the SSG algorithm only needs to define the raw material node and the product node. After traversing the path and combining the repetition, all feasible solutions of the SSG can be formed, the risk of combinatorial explosion in SSG is reduced.

4. Cases study

In this section, we will compare the RDFMSG algorithm proposed in this paper with the MSG algorithm in the literature through two cases of process synthesis problem. The applicability and correctness of the proposed algorithm is verified. After that, the algorithm calculation process, advantages and costs are analyzed. Case 3 is used to verify the feasibility of DEVS model construction method in P-graph.

Case 1:

We apply the Case in the reference [30], the data are as follows:

The procedure of MSG algorithm is described in the reference [30]. Two main steps are the reduction and composition. Fig. 5(a) is the original superstructure before MSG runs, material subset {A, C, J, K, N, Q, L} and the operation unit subset {({C, D, F}, {A}), ({N, Q}, {H}), ({J, K}, {E})} is subtracted, the rigorous superstructure is shown in Fig. 5(b). Finally, superstructure model of the process synthesis problem is formed.

Before the RDFMSG algorithm is applied, we first need to construct an RDF data set based on the DEVS ontology, the format is shown in Table 2. At the same time, due to the special definition logic form of the operation unit ontology, only R set and P set data need to be provided, so the single node form in M set can be automatically excluded, such as the node L. Neo4j-based graph data storage system adds labels and indexes to the node and edge, which can speed up the path retrieval process. The superstructure model formed by the RDFMSG algorithm as shown in Fig. 6.

Case 2:

For the S1–S7 axioms and constraints proposed in section 3, opposite cases are respectively constructed. RDF<s, p, o>, the triple data set are added, such as the RDF data shown in Table 3. Because other axioms are tested in Case1, after the RDF data in Table 3 are tested, the correct rigorous superstructure can also be generated. Therefore, the proposed RDFMSG is proved to be correct and effective.

Case 3:

Considering feedstock of ethylene production, light material(light), C5, naphtha (Nap), and hydrogenation tail oil (HCR) between two plants scheduling as an example [42,43]. Objective function is to allocate to minimum production cost consumption of same yield ethylene, three months as a cycle. Referring to the construction process of above DEVS atomic model, each plant monthly distribution is atomic model input, ethylene production is output, four raw materials in each month form a couple model. Input port is four raw materials, output port is the ethylene, and three couple model of three months extends from above couple model in time dimension, finally a couple model of the upper layer is formed. Overall the input port is the first month supply of the four raw materials, and the output port is ethylene. The superstructure is built as shown in Fig. 7, each layer represents one month, supposing a surplus after monthly allocation, and materials are supplemented before the next month. In Fig. 7, the subscript 0–1 indicates the intermediate state between month 0 and month 1.

Using statistics data of two factories in actual ethylene production, we can find the optimal distribution of top five solutions by ABB algorithm in PNS software. Each raw material consumption per month as shown in Fig. 8, i (0–2) indicates that i month raw material consumption, when its value is zero, means that this month did not consume this kind of raw material to produce ethylene.

The cost consumption of each atomic model in the superstructure is shown in Fig. 9. When the value of the atomic model is 0, it means that the atomic model instance does not exist, that is, the solution does not contain this atomic model instance. It can be seen from the figure that which operating unit has higher energy consumption, and specific energy saving measures can be implemented according to this result.

5. Conclusions

In order to improve the automation of process industry, this paper proposes RDFMSG algorithm based on RDF data generated by the DEVS ontology. Because of the RDF graph storage, the capacity of the database is greatly saved through the reduction of redundant data, and accurate semantic information is provided to the data. Through the verification of the above three cases, conclusions can be drawn as follows. First, the semantic information in RDF graph data is fully applied in the RDFMSG algorithm, especially the format like <s, p, o>. The problem of combinatorial explosion is transformed into graph data management. Second, the complexity of MSG is O ($l \cdot dm \cdot do$), the complexity of RDFMSG depends on the size of RDF triples. The consumption in path retrieval procedure is greatly reduced through the index built by the graph storage for nodes and edges. Third, the implementation process of RDFMSG has been simplified; SSG and ABB are also suitable for the same strategy. In summary, the semantic data in industry have been applied efficiently; at the same time, expert knowledge work is reduced based on the proposed methodology. The work in this paper provides a good reference for knowledge work automation in industrial system.

CRediT authorship contribution statement

Jian Cao: Methodology, Writing - original draft, Software. **Yan-Lin He:** Writing - review & editing, Validation, Investigation. **Qun-Xiong Zhu:** Writing - review & editing, Investigation, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research is supported by the National Natural Science Foundation of China under Grant No. 61533003 and 61703027.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chemolab.2020.104119.

References

[1] S. Park, J. Huh, Effect of cooperation on manufacturing IT project development and test bed for successful industry 4.0 project, Processes 6 (88) (2018) 1–31. Safety Management for Security.
[2] I. Zambon, M. Cecchini, G. Egidi, M.G. Saporito, A. Colantoni, Revolution 4.0: industry vs. Agriculture in a future development for SMEs, Processes 7 (36) (2019) 1–16.
[3] E.M. Eisman, M. Navarro, J.L. Castro, A multi-agent conversational system with heterogeneous data sources access, Expert Syst. Appl. 53 (2016) 172–191.
[4] Y.L. He, Y. Tian, Y. Xu, et al., Novel Soft Sensor Development Using Echo State Network Integrated with Singular Value Decomposition: Application to Complex Chemical Processes, Chemometrics and Intelligent Laboratory Systems, 2020, p. 103981.
[5] N.M. Elzeina, M.A. Majida, I.A.T. Hashemb, I. Yaqoobc, F.A. Alabad, M. Imran, Managing big RDF data in clouds: challenges, opportunities, and solutions, Sustain. Cities Soc. 39 (2018) 375–386.
[6] Y. Peng, J. Huang, K. Huang, Ontology-based DEVS model maintenance and reuse, in: International Conference on Computer Application and System Modeling, ICCASM, 2010.
[7] J. Li, S.E. Demirel, M.M.F. Hasan, Building block-based synthesis and intensificaiton of work-heat exchanger networks (WHENS), Processes 7 (23) (2019) 1–28.
[8] B.P. Zeigler, H. Praehofer, T.G. Kim, Theory of Modeling and Simulation, Academic Press, USA, 2000.
[9] A. Alshareef, H.S. Sarjoughian, B. Zarrin, Activity-based DEVS modeling, Simulat. Model. Pract. Theor. 82 (2018) 116–131.
[10] B. Earle, K. Bjornson, G. Wainer, et al., Design and Implementation of a Building Control System in Real-Time DEVS, Society for modeling & Simulation International, Tucson, AZ, USA, 2019. April 29-May 2.
[11] F. Celik, DEVS-M, A discrete event simulation framework for MANETs, J. Comput. Sci. 13 (2016) 26–36.
[12] M. Jarrah, Modeling and simulation of renewable energy sources in smart grid using DEVS formalism, Procedia Comput. Sci. 83 (2016) 642–647.
[13] X. Zhang, G. Li, P. Wang, et al., Design and implementation of real-time DEVS event scheduling algorithm, in: 3rd International Conference on Mechanical,Control and Computer Engineering, 2018, pp. 163–168.
[14] Y. Wang, G. Zacharewicz, D. Chen, et al., A Proposal of Using DEVS Model for Process Mining, 27th European Modeling &Simulation Symposium (Simulation in Industry), Bergeggi, Italy, Step, 2015, pp. 403–409.
[15] A. Dominici, E.D. Gentili, L. Capocchi, Smart-Parking: integration and data management by modeling and simulation using connected objects according to the DEVS formalism, in: IEEE 4th International Conference on Universal Village, 2018. Session 8a-3.
[16] N.M. Elzein, M.A. Majid, I.A.T. Hashem, Managing big RDF data in clouds: challenges, opportunities, and solutions, Sustain. Cities Soc. 39 (2018) 375–386.

[17] M. Achichi, Z. Bellahsene, M.B. Ellefi, et al., Linking and disambiguating entities across heterogeneous RDF graphs, Web Semantics: Science, in: Services and Agents on the World Wide Web, vol. 55, 2019, pp. 108–121.
[18] SPARQL. http://www.w3.org/TR/rdf-sparql-query/.
[19] F. Zhang, K. Wang, Z. Li, et al., Temporal data representation and querying based on RDF, IEEE Access 7 (2019) 85000–85023.
[20] H. Fang, pSPARQL, A querying language for Probabilistic RDF data, Complexity 2019 (2019) 8258197. Mar.
[21] G. Li, L. Yan, Z. Ma, An approach for approximate subgraph matching in fuzzy RDF graph, Fuzzy Set Syst. 376 (2019) 106–126.
[22] T.R. Gruber, A translation approach to portable ontology specificaitons, Knowl. Acquis. 5 (2) (1993) 199–220.
[23] F.T. Imam, Application of Ontologies in Cloud Computing: the State-Of-The Art. Distributed, Parallel, and Cluster Computing, 2016, pp. 1–13.
[24] A. Tsadimas, G.D. Kapos, V. Dalakas, et al., Simulating simulation-agnostic SysML models for enterprise information systems via DEVS, Simulat. Model. Pract. Theor. 66 (2016) 243–259.
[25] T. Karhela, A. Villberg, H. Niemisto, Open ontology-based integration platform for modeling and simulation in engineering, Int. J. Model. Simulat. Sci. Comput. 3 (2) (2012) 1–36.
[26] S. Jafer, B. Chhaya, J. Updegrove, U. Durak, Schema-based ontological representation of a domain-specific scenario modeling language, J. Simulat. Eng. 1 (2018) 1–15.
[27] I.E. Grossmann, R.W.H. Sargent, Optimum design of heat exchanger networks, Comput. Chem. Eng. 2 (1978) 1–7.
[28] M.J. Bagajewicz, V. Manousiouthakis, Mass/Heat-exchange network representation of distillation networks, AIChE J. 38 (11) (1992) 1769–1800.
[29] M.J. Bagajewicz, R. Pham, V. Manousiouthakis, On the state space approach to mass/heat exchanger network design, Chem. Eng. Sci. 53 (14) (1998) 2595–2621.
[30] F. Friedler, L. TarJan, Y.W. Huang, L.T. Fan, Graph-theoretic approach to process synthesis: polynomial algorithm for maximal structure generation, Comput. Chem. Eng. 17 (9) (1993) 929–942.
[31] R.R. Tan, K.B. Aviso, D.C.Y. Foo, P-graph and Monte Carlo simulation approach to planning carbon management networks, Comput. Chem. Eng. 106 (2017) 872–882.
[32] K.B. Aviso, J.I.A. Arogo, A.L.O. Janairo, et al., P-graph approach to planning biochar-based carbon management networks, Chem. Eng. Trans. 70 (2018) 37–42.
[33] H. Cabezas, A. Argoti, F. Friedler, et al., Design and engineering of sustainable process systems and supply chains by the P-graph framework, AIChE 37 (2) (2018) 624–636.
[34] B.S. How, S.Y. Teng, W D L, et al., Non-linear programming via P-graph framework, Chem. Eng. Trans. 76 (2019) 499–504.
[35] Z. Süle, J. Baumgartner, J. Abonyi, et al., P-graph-based multi-objective risk analysis and redundancy allocation in safety-critical energy systems, Energy 179 (2019) 989–1003.
[36] H.H. Chin, D.C.Y. Foo, H.L. Lam, Simultaneous water and energy integration with isothermal and non-isothermal mixing – a P-graph approach, Resour. Conserv. Recycl. 149 (2019) 687–713.
[37] G. Sirin, C.J.J. Paredis, B. Yannou, E. Coatanéa, E. Landel, A model identity card to support simulation model development process in a collaborative multidisciplinary design environment, IEEE Syst. J. 9 (4) (2015) 1151–1162.
[38] A. Katib, V. Slavov, P. Rao, Riq, Fast processing of SPARQL queries on RDF quadruples, Web Semant. Sci. Serv. Agents World Wide Web 37–38 (2016) 90–111.
[39] E.G. Kalayci, T.E. Kalayci, D. Birant, An ant colony optimization approach for optimizing SPARQL queries by reordering triple patterns, Inf. Syst. 50 (2015) 51–68.
[40] M. Bakhshi, M. Nematbakhsh, M. Mohsenzadeh, A.M. Rahmani, Data-driven construction of SPARQL queries by approximate question graph alignment in question answering over knowledge graphs, Expert Syst. Appl. 146 (2020) 113205.
[41] F. Alkhateeb, J.F. Baget, J. Euzenat, Extending SPARQL with regular expression patterns (for querying RDF), Web Semant. Sci. Serv. Agents World Wide Web 7 (2009) 57–73.
[42] P. Mu, X.B. Gu, Q.X. Zhu, Modeling and optimization of ethylene cracking feedstock scheduling based on P-graph, CIE J. 70 (2) (2019) 556–563.
[43] J. Wang, Y.L. He, Q.X. Zhu, Energy and production efficiency optimization of an ethylene plant considering process operation and structure, Ind. Eng. Chem. Res. 59 (3) (2020) 1202–1217.