An evolutionary model for dynamic and adaptive service composition in distributed environment

Jiawei Lu, Huan Zhou, Jun Xu, Gang Xiao
School of Computer Science and Technology
Zhejiang University of Technology, Hangzhou, China
Email: {viivan, zhouchuan, xujun, xgj}@zjut.edu.cn

Haibo Pan
Whzhen avenue Technology Co., Ltd
Hangzhou, China
Email: 345383630@qq.com

Abstract—Service composition is an important mean for integrating the individual Web services to create new value-added systems that satisfy complex requirements. Therefore, how to effectively analyze different types of services and find out the matching similarity between services to efficiently substitute failed services in a distributed and dynamic environment becomes crucial to service composition. In this paper, we propose a novel approach based on a data cell evolution model (DCEM) to support the dynamic adaptation of service compositions. The model combines data service information and biological cell behavior analysis to encapsulate data services into data cells. In order to reach optimum adaptations, we analyzed the static and dynamic structure of data cells based on bigraph theory to guarantee the consistency of service evolution. To evaluate the proposed approach, a series of simulation experiments and comparisons are conducted to demonstrate the effectiveness of service composition.

Keywords—Service Composition; Data Service; Bigraph; Data Cell; Service Evolution.

I. INTRODUCTION

Data as a Service (DaaS) is a new cloud computing service model that provides consumers on demand with data through different protocols on the Internet in a timely and low cost manner. However, as the function of a single Web service is simple and limited, it is difficult to meet the various requirements in a complex network environment. How to effectively model evolution of service composition and analyze its service behavior has become an issue that existing research must deal with.

Due to Web service with multi-source, heterogeneous, autonomous and dynamic characteristics, the evolution of service composition is different from traditional software evolution and faces more serious challenges. Many scholars have researched in this field through formal methods [1], semantic [2] or combinatorial models [3]. However, the above studies mainly focus on abstract behaviors and semantic analysis of services in specific field, without considering the dynamic contexts. Consequently, they easily lead to performance degradation and composition failure when the environment changes.

Biological cell is a precise structural, functional, and evolutionary unit, which structure can change dynamically with the environment during growth, differentiation and physiological process. Comparing the dynamic behavior of service composition with biological cell, they have similarities in some aspects. It is possible to combine data service with biological cell to analyze the evolution of service composition [4,5].

In this paper, we encapsulate data services into data cells and analyze the dynamic behavior at the cell level. The data cell, like biological cell has a strong hierarchical structure. Thus, a formal method is needed to effectively explain the static and dynamic information of data cells and reflect the important characteristics such as functions and location interconnectivity of services. Bigraph [6] is a graphical formalization theory tool. It has more extensive applications in formal modeling and consistent evolutionary analysis [7].

In order to reason about the evolution of service composition better, we propose a data cell evolution model (DCEM) to increase the flexibility of service in Web system. The main contributions of this paper can be summarized as follows:

- We use bigraph theory to encapsulate data services into data cells, and construct evolution model for data cell to describe different information of service such as service name, service quality, service context, and so on.
- Because the composition processes may be canceled, or services may be moved or withdrawn, it is necessary to recombine it to provide a more powerful service, so that the service dynamic behavior based on bigraphical reactive system is analyzed. Meanwhile, we propose a self-healing algorithm which is used to dynamically adjust available service to substitute the failed ones.

II. DATA CELL MODELING BASED ON BIGRAPH THEORY

A. The Bigraph Theory

Bigraph is proposed by Milner and other scholars in 2001 to emphasize the location and connection of computing (physical or virtual) [6]. Bigraph is a 2-tuple $B = \langle B^P, B^L \rangle$. $B^P$ is the place graph and $B^L$ is the link graph. The place graph is used to represent the location of nodes, which are nested with each other in bigraph. The link graph ignores the nested relationship and only indicates the connection between nodes. Fig. 1 shows a bigraphical structure.
Definition

\[ C = \langle CN, CT, P, CL, U \rangle \]

Definition 1 (Data Cell). A bipartite definition of data cell is a 5-tuple \( DC = \langle S, E, Ctrl, C^0, C^1 \rangle : \langle m, X \rangle \mapsto \langle n, Y \rangle \), where:

- \( S \) is a limited set of services in a data cell, \( \forall s \in S \) is called a data service;
- \( E \) is a set of finite edges, \( \forall e \in E \) is called a connecting edge;
- \( Ctrl: S \to C \) is a mapping relation between services and service controls;
- \( C^0 \) is the place graph to represent the location of services and \( C^1 \) is the link graph to represents service dependencies;
- The inner interface \( \langle m, X \rangle \) indicates that the bipartite has \( m \) sites and a set of inner names \( X \). The outer interface \( \langle n, Y \rangle \) indicates that the bipartite has \( n \) regions and a set of outer names \( Y \).

Definition 2 (Service Control). A service control is a 5-tuple \( C = \langle CN, CT, P, CL, U \rangle \), where:

- \( CN \) is a control name of service. \( CT \) is used to specify the type of this service, whether atomic or composite;
- \( P \) is a limited set of ports, which describes the inputs and outputs of service, \( \forall p \in P \) is called a service port;
- \( CL = \langle DL, CN \rangle \) is the dependency status of current service, including \( DL \) which is the dependent level with \( CN \) from other service;
- \( U \) is a probability value which represents the service reliability.

Definition 3 (Bipartite Reactive System). A bipartite reactive system for data cell is a 5-tuple \( BS = \langle BC, R, BC' \rangle : BC \to BC' \), where:

- \( BC \) is the reactants and \( BC' \) is the products, which are corresponding to data cells with the bipartite structures;
- \( R \) is a set of reaction rules and specifies the reaction process from \( BC \) to \( BC' \).

B. Data Cell Modeling

Based on the bipartite theory, we constructed DCEM that maps the structures and message interaction of services to bipartite, so as to formalize the services and their compositions by process calculus. DCEM mainly consists of two layers: data cell and data cell cluster. The form is defined as follows:

As examples depicted in Fig. 2, the bipartite definition of data cell aims to represent the structural relationship and data characteristics in services. The core element correspondence between data cell and bipartigraphic structure is also shown in Table 1.

### Table 1. Data Cell Structure Definition

<table>
<thead>
<tr>
<th>Element in data cell</th>
<th>Element in bipartigraph</th>
<th>The example</th>
</tr>
</thead>
<tbody>
<tr>
<td>( DC )</td>
<td>Root</td>
<td>0, 1, ...</td>
</tr>
<tr>
<td>( m )</td>
<td>Site</td>
<td>( X ) ...</td>
</tr>
<tr>
<td>( S )</td>
<td>Node</td>
<td>( x_0, x_2, x_3 ) ...</td>
</tr>
<tr>
<td>( CL )</td>
<td>Edge</td>
<td>( e_0, e_2, e_3 ) ...</td>
</tr>
<tr>
<td>( C )</td>
<td>Node control</td>
<td>( C_i, C_2, C_3 ) ...</td>
</tr>
<tr>
<td>( pC )</td>
<td>Node Ports</td>
<td>...</td>
</tr>
</tbody>
</table>

In practice, in order to meet the increasingly complicated requirements of users, it is necessary to select appropriate services from the network and combine them according to certain business rules to construct a scalable, loosely coupled combination. The data cell cluster is based on four kinds of structures (sequence, conditional, parallel, and loop) in service composition and combines a plurality of data cells with bipartigraph operations. The relevant forms of data cell cluster proposed in this paper are as follows:

Definition 4 (Data Cell Cluster). A bipartite definition of data cell cluster is a 3-tuple \( DCC = \langle DCS, CL, LinkS \rangle \), where:

- \( DCS \) is a limited set of data cells;
- \( CS \) is a limited set of composite structures in data cells;
- \( LinkS \) is a limited set of link ports in the cluster, \( \forall Link \in LinkS \) is called a connection between two ports.

The term language [13] is the basis for the formal specification and verification of dynamic evolution in bipartigraph. Fig. 3 shows the structure of \( DCC \) based on different workflows. Taking the parallel structure (Case 3) as an example, the cluster has three data cells \( DC_0, DC_1, \) and \( DC_2 \). The place graph indicates the positional relationship of services and other information (e.g., the number and distributivity of the cells). The link graph shows the dependency relationship of the services.

To adapt the dynamic environments to complex requirements, the data interaction between cells is constantly changing, forming new cell clusters or modifying the original cell cluster structure. Ensuring the structural integrity of data cells during this interactive process, while increasing the effectiveness of service composition, requires serious consideration. A checking technique for verifying the data flow of the process model and re-adjusting the model...
The algorithm contains two phases and takes into account time, QoSs, and service context information constraints. In the first phase, we take an initial bigraph \( BC \) and a set of reaction rules \( R \). For each reaction rule \( r \) in \( R \), the method \( \text{isMatch}(BC,r) \) is called to determine whether the elements in the bigraph can be matched. In the second phase, if \( r \) is matched and the constraint is satisfied, the matching part in the bigraph will be replaced by products in \( r \). \( \text{isMatch}(BC,r) \) is a recursive method that is iteratively executed until the last node in the bigraph has been checked.

\section*{Algorithm 1. BigraphMatch}

\textbf{Input:} bigraph \( BC \) (an initial bigraph), a set of reaction rules \( R \)

\textbf{Output:} a new bigraph \( BC' \)

1: \( \text{if} \ R = \text{Null} \) \( \text{then} \)
2: \( \text{return} \ BC' \)
3: \( \text{else} \)
4: \( \text{for} \) each reaction rule \( r \) in \( R \) \( \text{do} \)
5: \( \text{if} \ \text{flag} = \text{isMatch}(BC, r) \) \( \text{then} \)
6: \( \text{if}(\text{flag} = \text{TRUE} \& \& \text{timeConstraints} = \text{true}) \) \( \text{then} \)
7: \( \text{if the match is successful and satisfies the time constraint, the reaction proceeded} \)
8: \( \text{end if} \)
9: \( \text{end for} \)
10: \( \text{return} \ BC' \)
11: \( \text{end if} \)

\textbf{Procedure isMatch(BC, r)}

\textbf{Input:} bigraph \( BC \), term \( r \) in \( R \)

\textbf{Output:} a flag to indicate whether the match is found

1: \( \text{for} \) each service \( s \) in \( BC \) \( \text{do} \)
2: \( \text{if}(r \text{ contains } s) \) \( \text{then} \)
3: \( \text{if}(s.C.CN \leftarrow r.C.CN \| s.C.CT \leftarrow r.C.CT \| s.C.P \leftarrow r.C.P) \) \( \text{then} \)
4: \( \text{// Service control matching} \)
5: \( \text{continue} \)
6: \( \text{else} \)
7: \( \text{for} \) each port \( p \) in \( C \) \( \text{do} \) // Service port matching
8: \( \text{if}(s.p.pl \leftarrow r.p.pl \| s.p.pN \leftarrow r.p.pN \| s.p.pT \leftarrow r.p.pT) \| s.p.pC \leftarrow r.p.pC) \) \( \text{then} \)
9: \( \text{continue} \)
10: \( \text{return} \ True \)
11: \( \text{end if} \)
12: \( \text{end for} \)
13: \( \text{end if} \)
14: \( \text{end if} \)
15: \( \text{continue} \)
16: \( \text{end if} \)
17: \( \text{end for} \)
18: \( \text{return} \ False \)

\section*{III. EVOLUTIONARY ANALYSIS OF SERVICE DYNAMIC BEHAVIOR}

In actual use, service composition may face situations such as service failure and service composition disorder. These cause data cell variation, thus losing the original functional properties and structural stability.

In our work, DCEM periodically tests the service availability through a data cell self-healing algorithm (Algorithm 2), allowing the structural variation of the cells to be fixed. This enables the service composition to restore the expected functions, and ultimately achieves the effect of self-repair to improve the service adaptability. The basic process of Algorithm 2 is as follows:

\section*{Algorithm 2. Data Cell Self-Healing}

\textbf{Input:} bigraph \( BC \) (an initial bigraph)

\textbf{Output:} a new bigraph \( BC' \)

1: \( \text{for} \) each service \( s \) in \( BC \) \( \text{do} \)
2: \( \text{flag} = \text{isInvalid}(s) \) //indicate whether the services is invalid
3: \( \text{if}(\text{flag} = \text{true}) \) \( \text{then} \)
4: \( \text{a broken bigraph } BC' = \text{BigraphReplaced}(BC, s) \)
5: \( \text{//Choose the most reliable service MDCC} \)
6: \( \text{MDCC = Max(CR(DCC))} \)
7: \( \text{generate new reaction rules } R = \text{create}(s, MDCC) \)
8: \( \text{// use algorithm 1 to ensure the structural integrity of DC} \)
9: \( \text{end if} \)
10: \( \text{end for} \)

- Call method isInvalid(s) at a specified time interval to check whether the service in data cell is fail.
- If the service fails then use Algorithm 3 to adjust the bigraphical structure, such as delete nodes and control belong to the failed service. Otherwise adjust the time interval to continue testing.
- Select the most reliable service from the similar service clustering to replace the fail one. There are many clustering algorithms; we mainly use the tag clustering algorithm in reference [15] and execute the aggregation process to construct the similar service clustering. Then the chosen service is considered as a reaction in rules \( R \).
- Based on the previous steps and the bigraphical system, generate the appropriate reaction rules \( R \), and call Algorithm 1 to verify the rationality of reaction. Finally, a new bigraph is generated with a new reliable composition.

When service is detected as failed, Algorithm 3 needs to find out the failed service in the data cell. For given bigraphical information and a failed service \( s \), we traverse each node in bigraph. If there is a surjection relationship
between the node and s, the corresponding structure is deleted and a broken bigraph is given. The specific algorithm is described as follows:

**Algorithm 3** BigraphReplace

**Input:** Bigraph BC (an initial bigraph), s (a failed service)

**Output:** a broken bigraph BC*

1. for each service s' in BC do
   // Indicate whether the service is match
2. flag = node_conMatch(s', s)
3. if (flag == true) then
   // Delete structures belong to s'
4. BC* = deleteBigraph(BC, s')
5. if (s' C.C.L.D.L = single) then
   // Delete dependency belong to s'
6. BC* = deleteDependent(BC*, s')
7. end if
8. end if
9. end for

### IV. CASE STUDIES AND VERIFICATION

To illustrate the effectiveness of DCEM in service composition, we introduce a composite service that supports online book shopping at Orange Country Bookstore (OCB) [16] (depicted in Fig. 4).

![Fig. 4. Composite service for online book shopping](image)

The business process in OCB includes (1) Look for books, including Search Book, Show Book Info, Show Related Titles, which are all part of the Barnes & Noble Books composite service. (2) Add books to shopping cart in a loop, e.g., Barnes & Noble Shopping Cart. (3) Authentication and payment at checkout, such as Google Authentication or Payment Calculator. (4) Email and invoice service, such as invalid Card, UPS Shipping Web, and E-mail Invoice.

The service composition created by this instance may change the contextual events during the actual operation, resulting in service failure. In addition, the system requirements state that the service composition must maintain a main workflow after the evolutionary adjustment (for example, always ensuring that books are first searched and then added to the cart).

Here, we fully consider the evolution possibility of each service by analyzing its context. First, we construct the data cells based on the different business processes, as listed in Table II, and then further evolve the data cells into cell clusters according to functional attributes and requirements. Table III lists the data cell clusters related to the main workflow of the shopping cart, searching for books, and payment in the OCB website.

When Barnes & Noble Books is unavailable, causing the functionality of s1 to go missing (see Fig. 5), the system detects the failed service and traverses the data cell cluster according to Algorithm 3 to alter the structure. It then finds the substitute service Amazon Books (DCCa) from the similar service clustering via Algorithm 2 to repair the structure.

**Table II. DATA CELL MODELING**

<table>
<thead>
<tr>
<th>Data Service</th>
<th>Data Cells</th>
<th>Term Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search Book</td>
<td>![s1]</td>
<td>![S1]</td>
</tr>
<tr>
<td>Show Book Info</td>
<td>![s1]</td>
<td>![S1]</td>
</tr>
<tr>
<td>Show Related Titles</td>
<td>![s1]</td>
<td>![S1]</td>
</tr>
</tbody>
</table>

**Table III. DATA CELL CLUSTER MODELING**

<table>
<thead>
<tr>
<th>Data Services</th>
<th>Data Cell Clusters</th>
<th>Term Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barnes &amp; Noble Books Shopping Cart</td>
<td>![s2]</td>
<td>![S2]</td>
</tr>
<tr>
<td>Amazon Books</td>
<td>![s3]</td>
<td>![S3]</td>
</tr>
<tr>
<td>Credit Card Payment shopping</td>
<td>![s4]</td>
<td>![S4]</td>
</tr>
</tbody>
</table>

![Fig. 5. Data cell cluster on OCB when Barnes & Noble Books has failed](image)

![Fig. 6. Self-healing process in data cell cluster with Amazon Books](image)
Finally, as shown in Fig. 6, DCCSRM is absorbed by the data cell cluster, retaining an unchanged workflow order such as a loop structure with Book Searching, Book Description, Related Titles, and Shopping Cart. $s_{10}$, $s_{11}$, and $s_{12}$ are added in the site of $s_5$. After that, the output port of $s_{12}$ is connected to the input port of $s_7$ by $e_{15}$. Tables IV summarize the term language operational sets as $EO_1$ to formalize this process of evolution.

<table>
<thead>
<tr>
<th>DC</th>
<th>$EO_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCC</td>
<td>$EO_1$</td>
</tr>
<tr>
<td>$\phi$</td>
<td>$\phi$</td>
</tr>
<tr>
<td>$s_{12}$</td>
<td>$s_{21}(s_{20},s_{22})\rightarrow s_{21}(s_{20},s_{22})$</td>
</tr>
<tr>
<td>$s_{10}$</td>
<td>$s_{10}(s_{11},s_{12})\rightarrow s_{10}(s_{11},s_{12})$</td>
</tr>
<tr>
<td>$s_{11}$</td>
<td>$s_{11}(s_{12},s_{12})\rightarrow s_{10}(s_{11},s_{12})$</td>
</tr>
<tr>
<td>$s_{12}$</td>
<td>$s_{12}(s_{11},s_{12})\rightarrow s_{10}(s_{11},s_{12})$</td>
</tr>
</tbody>
</table>

These actions express how to reorganize elements in the composition model to re-select suitable data cells from the same cluster when the Barnes & Noble Books composite service fails. The failed units are replaced by Amazon Books and Related Titles (according to the business process from Fig. 4).

The self-healing process in Fig. 6 is formalized by the following term language:

$e_1$: $e_2/\sigma_{e_2}/e_2/e_2/\sigma_{e_2}/\sigma_{e_2}/e_2/\sigma_{e_2} / (s_{10}(s_{11},s_{12})\rightarrow s_{10}(s_{11},s_{12}))$

$e_2$: $s_{10}(s_{11},s_{12})\rightarrow s_{10}(s_{11},s_{12})$

$e_3$: $s_{11}(s_{12},s_{12})\rightarrow s_{10}(s_{11},s_{12})$

$e_4$: $s_{12}(s_{11},s_{12})\rightarrow s_{10}(s_{11},s_{12})$

V. EXPERIMENT ANALYSIS

To illustrate the effectiveness of the algorithm described in this paper, we conducted a series of experiments in different settings. The PC configuration was as follows: Intel Core i5-8250U CPU (1.6 GHz), Windows 8 and 6 GB RAM. We used all 12 of the atomic services discussed in Section 4. We then extracted their parameters and used them as a seed to randomly generate an extended dataset with 500-2000 services. Each service contained basic information such as the service name, input, output, and success probability. The experiments compared three kinds of algorithm: (1) The algorithm (DPSRM) based on the Dynamic Software Product Line approach [17]; (2) The algorithm (IASRM) based on the process ontology and multiple recovery [18]; (3) The data cell self-healing algorithm (DCSRM) proposed in this paper. In addition, we randomly set the effective service that failed as a variant $\nu$ during the composition, which automatically triggers the service substitution. To ensure unbiased statistical results, all algorithms were executed independently 20 times.

Fig. 7a presents the results obtained using only DCSRM. The service number varies from 100 to 500 and the number of variants increases gradually from 1 to 8. We found that when the number of variants is small, an increase in the service number produces a steady increase in the reliability. However, as the number of variants increases, the reliability becomes relatively low, especially when the number of services is small. This is because there are fewer similar services in the small service set, resulting in no suitable service being found when the number of variants grows. Fig. 7b shows results for $\nu = 9$ and the service number varying from 100 to 2000. Initially, IASRM gives the highest reliability, but this obviously decreases as the number of services increases. Furthermore, DCSRM always outperforms DPSRM.

VI. RELATED WORK

A. Bigraphs and their application

Bigraphs were proposed by Milner and other scholars in 2001 to emphasize the location and connection of computing units (physical or virtual) [6], and has now become the tool of choice for many service-adaptive and software-reconfigurable systems because of its complete formal theory and dynamic mobility. For example, Lian et al. [7] simulated modeling using a bigraphical reactive system to analyze the mobile cloud. Calder et al. [8] proposed a model based on checking predicates from user-initiated and network events by extension to a bigraphical reactive system

B. Evolution of service composition

Ensuring the rationality of business process structures after evolution is an important problem in the composition of services. If the evolution operation is not implemented properly, it may cause problems such as a logical deadlock or component service unreachability [20]. This section analyzes the evolution of service composition in terms of the rationality of service evolution operations and the current solutions for failure recovery.

1) Rationality of service evolution operations
The rationality of the service composition process and the correctness of the data flow are usually guaranteed by defining the evolution criterion in the operation process, thereby avoiding complex verification processes after evolution. To ensure the rationality of the service composition process, Zeng et al. [9] proposed a set of basic evolutionary operations for adjusting the service workflow structure based on workflow network modeling, which guarantees the rationality of the internal process logic during the business process. Urbiet et al. [3] propose an adaptive service composition framework that supports the dynamic reasoning of user requests and service behaviors in the smart city. Khanfar et al. [2] propose a framework for automatic generation and publishing of service descriptions by using OWL-S semantic annotations, the purpose of it is to analyze the process modeling and the choreography of service composite. In service behavior analysis, the research is mainly used by formal methods such as Petri net, process algebra, and π calculus, etc.

2) Current solutions to failure recovery

Self-adaptation is the ability of a system to adapt to changes in its environment to maintain the original functionality, and is used in different problem domains. Many recent studies have focused on enabling adaptation for BPEL processes. For instance, the monitoring mechanism embedded in the BPEL engine can be used to capture fault messages [10], allowing existing processes to be directly deployed without any modifications. Some scholars perform fault monitoring and recovery through transaction attributes of object states. Ettazi et al. [11] fulfilled user requirements in mobile environments by focusing on transactional aspects of context-aware services. In addition, there has been some research based on security monitoring and self-healing systems [12,19]. Asim [12] presented a framework that automates the monitoring of business processes and reports the compliance violations at runtime.

VII. CONCLUSION

Service composition is an important technology for integrating information to create new value in systems that satisfy complex requirements. In this paper we propose a novel approach for service composition with data cell modeling, which is inspired by biological cell and guided by the bigraph theory. This enables users to efficiently analyze the services in a dynamic distributed environment.

However, the preliminary data cells and clusters from model still need to be manually configured by analyzing and extracting the important characteristics from services. So, for future work, some tools may be applied to extract service features automatically. Another problem for future research is that more constraints from specific customer requirements, including service rating, service price and so on, need be considered to optimize the model in different scenes.

ACKNOWLEDGMENT

This work is supported by the Science and Technology Key Research Planning Project of Zhejiang Province, China (NO.2018C01064), and Zhejiang Natural Science Foundation, China (No. LY19F020034).

REFERENCES


