International Journal of Control and Automation
Vol. 10, No. 2 (2017), pp.129-136
http//dx.doi.org/10.14257/ijca.2017.10.2.11

An Approach to Web Service Faults Diagnosis Based on Conditional Random Fields

Yong Liu[1, 2] and Ling Qiu[3]

[1] *School of Automation and Electronic Information, Sichuan University of Science and Engineering, Zigong 643000, China*
[2]*Artificial Intelligence Key Laboratory of Sichuan Province, Sichuan University of Science and Engineering, Zigong 643000, China*
[3] *School of Computer Science, Sichuan University of Science and Engineering, Zigong 643000, China*
yongliusichuan@163.com

Abstract

Compared to single web service, web service composition is able to meet users' complex requirement. However, faults often happen when compound web services serve users. To diagnose the faults, this paper proposes an approach based on conditional random fields. Firstly, message transition matrix and behavior transition matrix are built via analyzing communication between single web services. Secondly, a diagnosis model is built based on conditional random fields. Thirdly, doubtful message sequence fragment is found and replaced with normal message by comparing the similarity between observed message sequence and normal message sequence in diagnosis model. At last, doubtful behavior transition sequence is picked out when message transition sequence is inputted into Viterbi algorithm. By comparing doubtful behavior transition sequence to observed behavior transition sequence, the web service providing fault behavior is picked out. The experiments show that our approach has higher precision than those based on historical data.

Keywords*: Web Service; Faults Diagnosis; Conditional Random Fields*

1. Introduction

Nowdays, e-commerce has been a virtual commerce schema. More people, especially young people, are enjoying e-commerce than traditional commerce. Hence, internet service providers are trying to invent various web services to attract their customers. As demand of customer becomes more and more complex, single web service is hard to meet demand of customer. Meanwhile, the cost of developing complex web services is high. An feasible approach is that existed web services can be composed to meet users' demand. Hence, the concept of web services composition is proposed. Web services composition aims to integrate multiple web services into a coherent whole [2].

However, the current situation is that web services on internet increase rapidly. It makes the web service composition complex that the process scale of web service composition becomes larger. Web service composition is being faced with some problems such as poor availability, low reliability and poor ability of fault-tolerant [3]. In addition, uncertainty of internet context is also an important factor which disturbs the application of web service composition. These problems lead to exception or interrupt of web service composition process. Obviously, one or more faults happen when web service composition works. Faults will be accumulated and spread via the inter operation between web services. Hence, it is necessary to diagnose the faults between web service composition processes.

ISSN: 2005-4297 IJCA

So far, a number of approaches have been proposed to diagnose faults. Some are based on models [4-6] and others are based on historical data [7-12]. For approaches based on models, the quality of diagnosis heavily depends on the completeness of web service composition system. Unfortunately, web service composition system is usually incomplete in real world. So, approaches based on models is restricted by the completeness of web service composition system. Approaches based on historical data release from the completeness of web service composition system. People try to find faults by analyzing historical behaviors. However, noisy historical data leads to imprecise or even false conclusion. By our survey, the fault recognition precision from approaches based on historical data is not high. So in our paper, a conditional random fields model is applied to web service faults diagnosis in order to improve the fault recognition precision.

This paper is organized as follow. We introduce related works in section 2. Conditional random fields and Viterbi algorithm are described in section 3. Model and algorithms of fault diagnosis of web service is highlighted in section 4. Experiments analysis is given in section 5 and conclusion and the next work are arranged in the last section.

2. Related Work

For the first time, Ardissono, *et al*, applied the approach based on model to web service diagnosis [4]. Ardissono proposed a hierarchical diagnosis model which was composed of a global diagnosis component and a set of local diagnosis components. When a fault was thrown, global diagnosis component sent a message to the local diagnosis component which corresponded to the fault. The local diagnosis component tried to explain the fault and then reported the interpretation to global diagnosis component. Later, Ardissono extended their work [14]. Ardissono appended a fault repair component which was used to repair fault according to the interpretation. Bocconi improved the algorithm of fault location [15]. The improved algorithm located the fault more accurately. Again, Ardissono integrated fault diagnosis component into fault repair component. Because the interaction between global diagnosis component and local diagnosis component was reduced, the efficiency of diagnosis was improved [16]. Li proposed an approach based on colored Pertri network [13]. Based on local diagnosis component, web service model and fault model were converted into colored Pertri network. To represent the status of data, three typical dependence relations [4] to color reproduction functions were defined. Based on color reproduction functions, diagnosis service component computed fault iteratively until the computed result was same to the observed result. Because each task was assigned to a local diagnosis component, the approach can not only ensure the privacy of web service but also relieve the pressure of global diagnosis component. However, the actual fault was hard to be predicted. So, the approach was more suited for the situation that the type of fault was known in advance. Yan converted business processes of web services composition into a synchroniaed automata [5]. Based on automata, diagnosis component synchronized observed system path and found the track of process. By analyzing the features of the track of process, diagnosis component deduced the fault location and type. Against the unpredictability of actual fault, Mayer proposed an approach of web fault diagnosis from incomplete knowledge base [6]. Mayer built a behavior set and a condition set. Input and output were defined strictly in behavior set. Crucial and independent conditions were defined in condition set. And then, the behavior set and the condition set were used to reduce fault set. The approach can reduce constraint conditions and diagnose multiple faults at the same time. However, it was usually difficult to build a complete a behavior set and a condition set.

Approaches based on historical data tries to learn rules from historical data to diagnose fault. Compared to the approaches based on model, the approaches based on historical data save the cost of logical relation computing between behaviors so that the

complexity of algorithm was decreased dramatically. Zhu attached historical data to Bayes network [9]. And then posterior probability of fault diagnosis was calculated based on probability of correct web service and conditional probability of of incorrect web service. Dai collected the previous behavior set of all abnormal behavior path [7]. And then the probability of each behavior which was viewed as fault root was calculated. By calculating the similarity between the actual path of web service process and the path of incorrect web service process, the fault set with maximum similarity was decided as diagnosis result. To increase the ratio of correctness of fault diagnosis, Mosrfaoui improved Dai's approach [10]. Mosrfaoui proposed a hierarchical model to web service diagnosis. The hierarchical model was composed by instance layer, component layer and composition layer. For every layer, the core code was isolated from fault tolerance component. Han [11] clustered faults based on structures of log files. These clusters were distinguished from private data set to similar data set. And then Han exploited machine learning algorithm to build Bayes network. The fault type with maximum similarity between actual web service process and learned fault process from Bayes network was decided as final fault type.

3. Conditional Random Fields

3.1. Related Definitions

Definition 1 Let $G=(V, E)$ be a graph such that $Y=(Y_v)_{v\in V}$, so that Y is indexed by the vertices of G. Then (X,Y) is a **conditional random field** in case, when conditioned on X, the random variables Y_v obey the Markov property with respect to the graph: $P(X_v|X, Y_w, w\neq v)= P(X_v|X, Y_w, w\sim v)$, where $w\sim v$ means that w and v are neighbors in G[1].

According to probability theory, CRF is a MRF (Markov Random Field) with a set of observations. MRF is a random field which has restricted Markov property. Markov property means that for any random variable v, it is same that probability distribution of v is computed from other random variables or all neighbour random variables. Hence, the joint probability of graph $G=(V, E)$ is equal to the product among potential functions of all maximum cliques of G. Hence, the joint probability of graph $G=(V, E)$ is represented as formula (1).

$$p(\boldsymbol{x})=\frac{1}{Z}\prod_{c\in C}\Psi_c(x_c) \tag{1}$$

C is the set of maximum cliques.

$P(y|x)$ is calculated according to formula (2).

$$P(\boldsymbol{y}\mid\boldsymbol{x})=\frac{P(\boldsymbol{x},\boldsymbol{y})}{P(x)}=\frac{P(\boldsymbol{x},\boldsymbol{y})}{\sum_{y'}P(\boldsymbol{y}',\boldsymbol{x})}=\frac{\frac{1}{Z}\prod_{c\in C}\Psi_c(x_c,y_c)}{\frac{1}{Z}\sum_{y'}\prod_{c\in C}\Psi_c(x_c,y_c')}=\frac{\prod_{c\in C}\Psi_c(x_c,y_c)}{\sum_{y'}\prod_{c\in C}\Psi_c(x_c,y_c')} \tag{2}$$

where Z is a normalization constant. $\Psi_c(x, y_c)$ is potential function of clique c . y_c is the random variable of clique c. Hence, the general representation of CRF is shown in formula (3).

$$p(y|\boldsymbol{x})=\frac{1}{Z(x)}\prod_{c\in C}\Psi_c(x,y_c) \tag{3}$$

where

$$\Psi_c(x,y_c)=\exp\left(\sum_k\theta_k f_k(c,y_c,x)\right) \tag{4}$$

Where $f_k(c, y_c, x)$ is feature function, θ_k is weight of f_k. Finally. the final representation of CRF is shown in formula (5).

$$p(y|\boldsymbol{x}) = \frac{1}{Z(x)} \exp\left(\sum_{c \in C} \sum_{k} \theta_k f_k (c, y_c, x) \right) \tag{5}$$

where

$$Z(x) = \sum_{y'} \exp\left(\sum_{c \in C} \sum_{k} \theta_k f_k (c, y_c, x) \right) \tag{6}$$

3.2. Viterbi Algorithm

Viterbi algorithm is used to find an optimal sequence of states from CRF. Formally, for a sequence of observations $V=(V_1,V_2,\ldots,V_T)$ and a CRF C, Q^* is the optimal sequence of states iff $\forall Q \in perm(Q)$, $p(Q|V, C) \leq p(Q^*|V, C)$ holds.

Let $\delta_t(i,j)$ be max probability of observation V emitted by states path $q_1, q_2, \ldots, q_{t-1}=S_i$, $q_t=S_j$ at time point t. $\delta_t(i,j)$ can be calculated by Formula (7). In a similar way, $\delta_t(i,j)$ can be calculated by Formula (8).

$$\delta_t (i, j) = \max_{q_1, \ldots q_{t-1}} P\left(q_1, \ldots, q_{t-1} = S_i, q_t = S_j, V \mid \lambda\right), 1 \leq i, j \leq N, 2 \leq t \leq T \tag{7}$$

$$\delta_{t+1}(j,k) = \max_{1 \leq i, j \leq N} [\delta_t (i, j) a_{ijk}] b_{ijt+1}, 1 \leq j, k \leq N, 2 \leq t \leq T-1 \tag{8}$$

Viterbi algorithm is described as follow.

Step 1: initialization

$$\delta_2 (i, j) = \pi_i a_{ij} b_{ii} b_{ij2}, 1 \leq i, j \leq N \tag{9}$$

$$\Psi_2 (i, j) = 0, 1 \leq i, j \leq N \tag{10}$$

Step 2: recursion

$$\delta_{t+1}(i, j) = \max_{1 \leq i, j \leq N} [\delta_t (i, j) a_{ijk}] b_{ijt+1}, 1 \leq j, k \leq N, 2 \leq t \leq T-1 \tag{11}$$

$$\Psi_{t+1}(j,k) = \arg\max_{1 \leq i \leq N} \left[\delta_t (i, j) a_{ijk} \right], 1 \leq i, j \leq N, 2 \leq t \leq T-1 \tag{12}$$

Step 3: terminal

$$P^* = \max_{1 \leq i, j \leq N} [\delta_T (i, j)] \tag{13}$$

$$q_T^* = \arg\max_{1 \leq i, j \leq N} [\delta_T (i, j)] \tag{14}$$

Step 4: finding an optimal sequence of states

$$q_{t-1}^* = \Psi_{t+1}\left(q_t^*, q_{t+1}^*\right), t = T-1, T-2, \ldots, 2 \tag{15}$$

4. Model and Algorithms of Fault Diagnosis of Web Service

4.1. Definitions of Fault Diagnosis of Web Service

Definition 2: An **object system** S is a 3-tuples (*SD*, *COMPS*, *OBS*), where

- *SD* is a set of descriptions of behaviors.
- *COMPS* is a set of characters, which represent components.
- *OBS* is a set of observation data.

Convention 1: For any $c \in COMPS$, $\neg ab(c)$ means that c works well and $ab(c)$ means that c is out of order.

Definition 3: $C_S=\{c|\neg ab(c), c \in COMPS\}$is called as a **diagnosis** of an object system (*SD*, *COMPS*, *OBS*).

Definition 4: Conflict set MC_S is said to be a **minimum diagnosis** iff for any conflict set C_S, $|MC_S| \leq |C_S|$ holds.

Definition 5: According to formulas (3) and (4), **diagnosis model** *DM* is defined as 4-tuples $DM=(y, x, f, \theta)$, where

- $y=\{y_1, y_2, \ldots\}$ is a set of faults
- $x=\{x_1, x_2, \ldots\}$ is a set of messages sent by web service.
- $f=\{f_1, f_2, \ldots\}$ is a set of features which is distinguished from independent features f_{in} and dependent features f_{de}.
- θ is weights of features.

4.2. Model Training

For a given training set $D=\{(x_1, y_1), (x_2, y_2), \ldots, (x_M, y_M)\}$, the essential task of model training is to evaluate the tuple θ according to Formula (16)[1].

$$L(\theta)=\sum_{i=1}^{M}\log P\left(y^i \mid x^i\right)=\sum_{i=1}^{M}\log\left(\frac{\exp\left(\sum_{t=1}^{T}\sum_{k=1}^{F}\theta_k f_k\left(y_{t-1}^i, y_t^i, x^i, t\right)\right)}{\sum_{y\in Y}\exp\left(\sum_{t=1}^{T}\sum_{k=1}^{F}\theta_k f_k\left(y_{t-1}^i, y_t^i, x^i, t\right)\right)}\right) \tag{16}$$

- $t\in T$ is a time point.
- F is set of feature functions. For any $f_k\in F$, f_k is a independent features f_{in} or dependent features f_{de}. For a given time point t, f_k is assigned to 1 if f_k is a independent features f_{in} and x^i is found from the observation set of behavior y^i. Otherwise, f_k is assigned to 0. f_k is assigned to 1 if f_k is a dependent features f_{de} and y^i_{t-1} is adjacent to y^i_t. Otherwise, f_k is assigned to 0.

4.3. Fault Diagnosis

For a given message sequence $x=\{x_1, x_2, \ldots, x_T\}$, a behavior sequence with maximum possibility $y^*=\{y_1, y_2, \ldots, y_T\}$ is decided according to training model. Formally, y^* is described as $y^*=\text{argmax}_y P(y|x, \theta)$.

$$y^* = argmax_y\, p\left(y|\boldsymbol{x},\theta\right) = argmax_y \frac{1}{Z(x)}\exp\sum_{k=1}^{F}\theta_k f_k\left(y,x\right) = \underset{y}{argmax}\exp\sum_{k=1}^{F}\theta_k f_k\left(y,x\right) \tag{17}$$

And then, y^* is calculated according to Veterbi algorithm.

5. Experiments

5.1. Experiments Setup

We generate 100 behavior sequences as raw data randomly from booking component of airline. 80 behavior sequences are extracted as training data and others are used as test data. We repeat experiment for 10 times.

Two criterion are used to evaluate the performance of our approach, *ar*(accuracy rate) and *nr*(nois rate). The operations of noise generation are parted into three types. (1) deletion a node from a behavior sequence. (2) exchange between two nodes in a behavior sequnce. (3) replace a node with a error node in a behavior sequence.

The formulas are given in formula (18) and (19). N_a is the number that web service faluts are diagnoised correctly. N is the number that web service faluts are diagnoised. D_n is the number of historical behavior sequences with noise. D is the number of historical behavior sequences.

$$ar = N_a / N \tag{18}$$

$$nr = D_n / D \tag{19}$$

5.2. Experiment Evaluation

We compare the accuracy rate based on Yan [5], Dai [7] with our approach on *nr*=10%, 20% and 30%. The experiment result is shown in Figure 1, Figure 2 and Figure 3.These figures show that our approach is superior to Yan and Dai. The avaerage accuracy rate of our approach is about 7.1% than Yan and 7.9% than Dai when *nr* is 10%. Also, the avaerage accuracy rate of our approach is about 3.6% than Yan and 5.9% than Dai when *nr* is 20%. The avaerage accuracy rate of our approach is about 2.4% than Yan and 2% than Dai when *nr* is 30%.

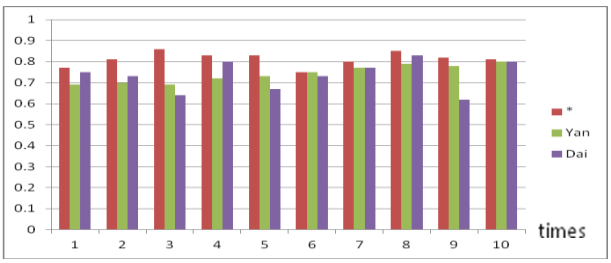

Figure 1. The Accuracy of *, Yan and Dai on nr=10%.

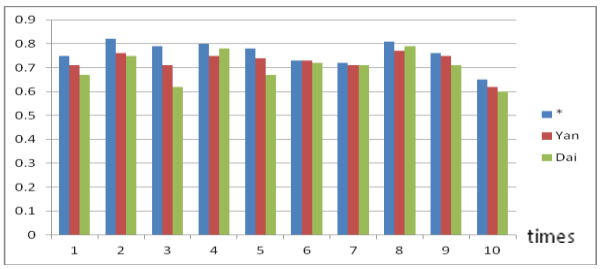

Figure 2. The Accuracy of *, Yan and Dai on nr=20%.

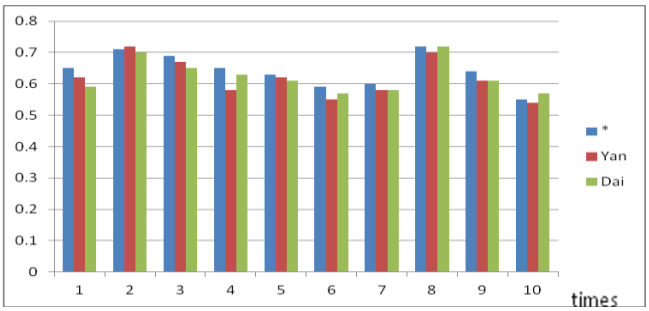

Figure 3. The Accuracy of *, Yan and Dai on nr=30%.

6. Conclusion and Future Work

In this paper, we applied condition random fields to fault diagnosis of web service. The relationship between the probability and the model's historical states is considered reasonably. Compared to previous approach, our approach shows higher precision by experiments.

The experiment result shows the advantage of our approach. However, some details attract us. The cost of time of our approach is higher than Yan and Dai. In addition, the

advantage of our approach is weakened with raise of nosie rate. In future work, we will optimize our approach from both performance and accuracy rate.

Acknowledgements

We thank all reviews for their useful comments to improve the paper. This work was supported by The Open Project Program of Artificial Intelligence Key Laboratory of Sichuan Province (2013RZJ02).

References

[1] J. Lafferty, A. McCallum and F.C. Pereira, "Conditional random fields: probabilistic models for segmenting and labeling sequence data", Proceeding of International Conference on Machine Learning, **(2001)**, pp. 282-289.
[2] K. S. M. Chan, J. Bishop and J. Steyn, "A fault taxonomy for web service composition", Service-Oriented Computing - ICS0C 2007 Workshops, Vienna, Austria, **(2007)**, pp. 363-375.
[3] S Bruning, S Weissleder and M. Malek, "A fault taxonomy for service—oriented architecture", 10th IEEE High Assurance Systems Engineering Symposium (HASE* 07), Piano, TX, **(2007)**.
[4] L. Ardissono, L. Console and A. Goy, "Enhancing web services with diagnostic capabilities", Proceedings of the Third European Conference on Web Services (ECOffS' 05), Vaxjo, Sweden, **(2005)**, pp. 182-191.
[5] Y. Yan, P. Dague and Y. Pencole, "A model-based approach for diagnosing faults in web service processes", The International Journal of Web Services Research (JWSR), vol. 6, no. 1, **(2009)**, pp. 87-110.
[6] W. Mayer, G. Friedrich and M. Stumptner, "Diagnosis of service failures by trace analysis with partial knowledge", Service-Oriented Computing, vol. 6470, **(2010)**, pp. 334-349.
[7] Y. Dai, L. Yang and B. Zhang, "Exception diagnosis for composite service based on error propagation degree", 2011 IEEE International Conference on Services Computing (SCC 2011). Washington, DC, USA, **(2011)**, pp. 160-167.
[8] L. Ardissono, S. Bocconi and L. Console, "Enhancing web service composition by means of diagnosis", Business Process Management Workshops, Milano, Italy, **(2008)**, pp. 468-479.
[9] Z. Zhu and W. Dou, "QoS-based probabilistic fault-diagnosis method for exception handling", New Horizons in Web-Based Learning: ICWL 2010 Workshops, Berlin, **(2011)**, pp. 227-236.
[10] G. K. Mostefaoui, Z. Maamar and N. C. Narendra, "On modeling and developing self-healing web services using aspects", Proceedings of the 2007 2nd International Conference on Communication System Software and Middleware and Workshops (C0MSWARE 2007), Bangalore, **(2007)**, pp. 1-8.
[11] X. Han, Z. Shi and W. Niu, "Similarity-based Bayesian learning from semi-structured log files for fault diagnosis of web services", 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), Toronto, Canada, **(2010)**, pp. 589-596.
[12] Z. Jia and R. Chen, "LBD4WS: log-based diagnosis for web service", Journal of Theoretical and Applied Information Technology, vol.48, no.1, **(2013)**, pp. 247-253.
[13] Y. Li, L. Ye and P. Dague, "A decentralized model-based diagnosis for BPEL services", Proceedings of the 2009 21st IEEE International Conference on Tools with Artificial Intelligence (ICTAI' 09), Newark, NJ, **(2009)**, pp. 609-616.
[14] L. Ardissono, R. Furnari and A. Goy, "Fault tolerant web service orchestration by means of diagnosis", Proceedings of the Third European conference on Software Architecture (EWSA* 06), Nantes, France, **(2006)**, pp. 2-16.
[15] S. Bocconi, C. Picardi and X. Pucel, "Model-based diagnose ability analysis for web services", 10th Congress of the Italian Association for Artificial Intelligence, Rome, Italy, **(2007)**, pp. 24-35.
[16] L. Ardissono, S. Bocconi and L. Console, "Enhancing web service composition by means of diagnosis", Business Process Management Workshops, Milano, Italy, **(2008)**, pp. 468-479.

Authors

Yong Liu, he is a lecturer, received the bachelor degree from in Sichuan University of Science & Engineering in 2003. The main research directions: Semantic Web, Ontology, Artificial Intelligence.

Ling Qiu, she is an associate professor, received the master degree from University of Electronic Science and Technology of China in 2010. The main research directions: Computer Application, Artificial Intelligence.