

The Trustworthiness Analyzing of Interacting Business Process Based on the Induction Information

Xianwen Fang^{1,2}, Changjun Jiang¹, Zhixiang Yin², and Xiaoqin Fan¹

¹ Key Lab of Embedded System and Service Computing,
Ministry of Education, Tongji University,
Shanghai 201804, China
{fangxianwen, cjjiang,xqfan}@hotmail.com

² Department of Information and Computing Science,
Anhui University of Science and Technology,
Huainan 232001, China
{fangxianwen, zxyin62}@hotmail.com

Abstract. Under the open environments, it is very difficult to guarantee the trustworthiness of interacting business process using traditional software engineering methods, at the same time, for dealing with the influence of external factors, some proposed business process mining methods are only effective 1-bounded business process, and some behavior dependent relationships are ignore. A behavior trustworthiness analysis method of business process based on induction information is presented in the paper. Firstly, aimed to the internal factors, we analyze the consistent behavior relativity to guarantee the predictable function. Then, for the external factors, in order to analyze the behavior change of business process, we propose a process mining methods based on induction information. Finally, experiment simulation is given out, and compares our method with genetic process mining methods. Theoretical analysis and experimental results indicate that our method is better than the genetic process mining method.

Keywords: trustworthiness, consistent behavior relativity, business process, process mining.

1. Introduction

With the tremendous growth of information and communication technologies, it is advantageous to design and implement the complex inter-enterprise business processes. One of the major innovations is the concept of service-oriented architectures (SOA) which considers software systems as being made up with autonomous, dynamic, loosely coupled and service-based components. For interacting business processes, especially those that run

over the Internet, behavior trustworthiness plays a crucial role. The operating and development environment of business process turn from the traditional closed static environment and extend to an open, dynamic, ever-changing network environment [1]. In the network computing environment, the behavior of the entity is uncontrollable and uncertain, which posed a severe challenge for behavior trustworthiness of the running processes [2]. Therefore, the guarantee of the behavior trustworthiness of the business under the network environment has become the industry's research focus.

Analysts often see lack of trustworthy methods as a major impediment to the adoption of web services in building agile business process. Software behavior is a running operation series, which can describe the software some feature when interacting with others. By analyzing the definition of trustworthiness, that the behavior can be predictable is a key property. So, achieving behavior trustworthy goals often relies on an appropriate use of software behavior. However, the notion of behavior trustworthiness is often neglected in business process, which usually concentrates on modeling the process in a way that functional correctness and security can be shown, either manually or using formal proof tools like model checking and some static validation methods. In contrast to features that are crucial for functional correctness and security features, and so on, behavior trustworthiness is typically integrated in an application, which mainly assures the execution behavior and computing result of the business process to be consistent with the expected effects. The integration of trustworthy features into a business process is not well understood. Especially, the trustworthy features are difficult to describe, and to analyze quantitatively [3]. So, the problem need further investigation, and seek new methods to deal with the behavior trustworthiness.

The interacting business process may run into two kinds of problems, one problem is the interactive influence between the business processes in the building process; one component business process may impact the other, which leads to the function of composite business process to be improper. The other problem is that the composite business process may be influenced by the outer factors. Aimed to the internal interactive influence, the interaction behavior relativity of business process need be studied. For the outer factors, we use the business process mining methods to construct the actual behavior model based on the running logs. At present, some process mining methods [4-6] have been proposed, but the existing methods are deficient to solve the problem. In the paper, our contributions are given out the analyzing methods of the behavior trustworthiness about business process based on the induction information; the methods can deal with k-bounded business process, and take the indirect dependent relationships into account.

The rest of this paper is organized as follows. Section 2 introduces related work. In Section 3 discusses motivating examples and research framework. Section 4 details propose the analyzing of the composite model of business process based on Petri net. Section 5 presents Building the K-bounded process behavior model based on new process mining methods. Simulation

experiments are given out in the Section 6. Finally, Conclusions and acknowledgements are in Section 7 and Section 8 respectively.

2. Related Works

The study of business process mainly takes advantage of some methods that described the interaction between different components of the system. Research on business process modeling has recently started to encode business-process diagrams into a formal model that can be given a suitable semantics, usually based on interacting state machines [7], Petri nets [8], or graph grammars [9]. Models that explicitly address the incorporation of security issues in the design process are typically extensions of a fragment of UML that can be given the desired semantics. They address more general notions of security than in the standards like multi-level security [10] or role-based access control [11]. A concrete counterexample is given in [12]. Being more precise, there is no guarantee that properties of these abstractions are also valid for the concrete implementation. In the literature [13], the paper presents a general methodology for integrating arbitrary trusted requirements in the development of business processes in a both elegant and rigorous way, and show how trusted relationships between different parties and their respective security goals can be reflected in a specification, which results in a realistic modeling of business processes in the presence of malicious adversaries. Special attention is given to the incorporation of cryptography in the development process with the main goal of achieving specifications that are sufficiently simple to be suited for formal verification, yet allow for a provably secure cryptographic implementation. [14] described a business process driven framework (called the BPD-ACS) for developing both the model and formulating the access decision rules. The model used is the Role Based Access Control (RBAC) model and the access decision rules are based on temporal business associations. In the literature [15], analysis of workflow dynamic changes based on Petri net is proposed, but did not take the behavior relativity into account. [16] proposed the logic-based approach about specification and verification of declarative open interaction of business process, in order to guarantee the trustworthiness and reliability of the developed model, as well as to ensure that the actual executions of the system effectively comply with it. [17-18] proposed a business process to be trustworthy if the behavior of all services is trustworthy, and gave out a verification mechanism through which the trustworthiness of a business process can be verified. Collaborative business processes often consist of services provided by multiple business entities which agree to join business collaboration. To enable trustworthy and secure consumption and provisioning of services across organizational boundaries, security requirements must be carefully defined so as to be coherent, consistent, and in compliance with designed business processes. However, managing security requirements in collaborative environments is error-prone, effort

inefficient, and hard to be verified. [19] discussed the architecture specific to this issue, as an add-on to a trustworthy service-oriented architecture, and proposed suitable formal notations and formal analysis in the construction of this automated facility.

All works mentioned above did not study the behavior trustworthiness of business process, only the function correctness or security, so it is necessary to study new methods to analyze the behavior trustworthiness of interactive business processes.

3. Motivating Examples and Research Framework

3.1. Motivating Examples

In order to satisfy the user' demand, it is necessary though compositing several business processes to realize the complex function demand, and satisfies certain constraints. Under open environment, some composite business process cannot be anticipated after the interacting, but the behavior predictability of the composite business process is a key to the trustworthiness study of business process. At present, some behavior conformance methods is proposed, the methods need to build the dynamic model based on the running log, and compare the dynamic model with the original theoretical model, obtain the behavior conformance by some metrics methods. In building the dynamic model, some process mining methods is proposed, such as @-algorithm [4], genetic mining methods[20,21]. These methods can solve effectively safe business process, but they are deficient to deal with K-bounded business process.

The ultimate objective of this paper is to study the behavior trustworthiness based on the process mining and induction information. The notion establishes whether the behavior of interacting business process is predictable and it can be employed to suitably address the following issues:

1) The process behavior models are built by some event dependent relationships and the k-bounded value. For example, the process behavior model may be not 1-bounded, so it is can not implement by @-algorithm. The region-based mining methods can implement the k-bounded model, but the methods obtain the k-bounded value by analyzing the transition system, which may lead to the error value and cost a mount of time-consuming.

2) Some indirect dependent relationships should take into account in the process of business process mining, but the region-based methods ignore some indirect dependent relationships.

3) For analyzing the behavior predictability under outer factors, the behavior conformance of two models is a key method, which is determined by structure fitness and behavior appropriateness, the effect should be distinguishable.

We can hence outline the main features that a suitable notion of behavior predictability of business process in open and changing environments should have, that is, the function correctness and behavior conformance of process behavior model. Indeed, it is reasonable that the correct process behavior model can be built by new process mining methods. The behavior conformance between the theoretical composite models with the building process behavior model is also key problem for analyzing the behavior predictability.

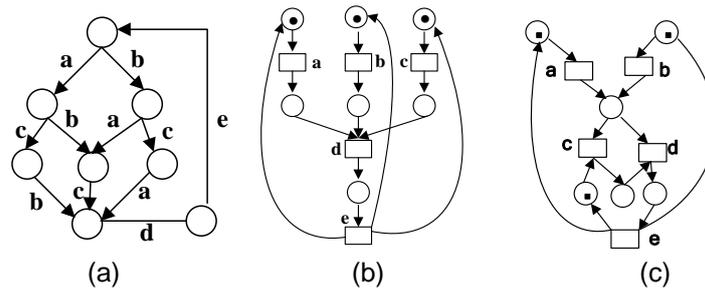


Fig.1. The TS and corresponding to process mining models.

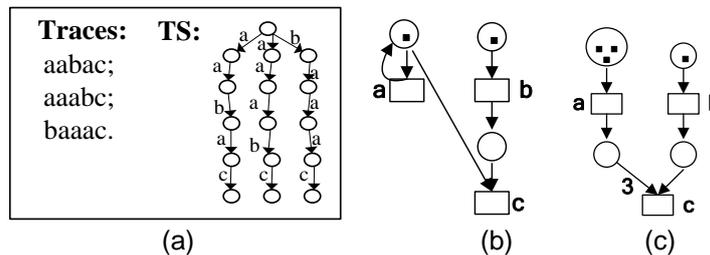


Fig.2. The log traces and corresponding to process mining modes

In Fig.1, Fig.1(a) is the transition system(TS) from the log traces of business process[22], which indicates the running state of business process, but it is transition running sequences without be handled, not the complete behavior relationships. So, it is necessary to mine the behavior relationships and to build the behavior model, in order to analyze the behavior conformance. Using the @-algorithm, we can obtain a soundness business process, the model is showed in the Fig.1 (b), because @-algorithm only can mine the 1-bounded model, the Fig.1 (b) is a 1-bounded model, and it can include all traces of the Fig.1 (a), but it generate more traces, for example the trace “cabde”, the Fig.1(a) can not generate it. In the Fig.1(c) is a 2-bounded business process model, its traces is the same as the Fig.1(a), so the Fig.1(c) is the better than the Fig1(b).

In Fig.2, Fig.2 (a) is the transition system from the log traces of business process, Fig.2(b) is 1-bounded process mining model by @-algorithm or others, Fig.2(c) is a 3-bounded process miming model. Obviously, the

language of Fig.2(b) is $(a^*||b)c$, it is much greater than the traces in the Fig.2(a). The language of Fig.2(c) is $(a^3||b)c$, it is greater than the traces in the Fig.2(a), but it is less than the language of Fig.2(b), and it is more accurate one, so the model in Fig.2(c) is better than the model in Fig.2(b).

Viewed from the Fig.1 and Fig.2, when we mine the process model, the bounded value is also key factor. It is necessary to analyze the bounded value before business process mining.

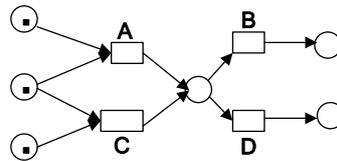


Fig.3. The process mining model of composite business process

In the Fig.3, for the log traces AB, CD of business process and the corresponding to transition systems $TS\{AB\}$ and $TS\{CD\}$. When composing these business processes, it is easy to see that the minimal regions of each are compatible with the empty region of the other; hence the minimal regions of the composite transition system are simply the union of the two sets of minimal regions. The composite business process can be mined by region-based method proposed in the literature[22], the process mining model is showed in the Fig.3, we find that we can no longer just consider minimal regions, but also have to consider composite regions (in this case the region where A and C enter and B and D exit). So, the new region-based process mining methods proposed in the literature [22] can mined the k-bounded process model, but it has a drawback: incrementally considering only minimal regions does not always give a correct answer. This means that it can still construct the proper set of regions by considering the compatible ones, but we cannot simply always use the minimal regions of the components of the transition system.

3.2. Research Framework

Under the open and dynamic environments, owing to software evolution continually, it is very difficult to guarantee the software quality using traditional software engineering method; meanwhile, for dealing with external factors, there is a big limitation using the existing program verification and analyzing methods [12]. Because the existing formal verification methods mainly aimed to the program correctness in close environment, the composite software system may not be behavior trustworthy by some functional correctness software interacting with each other. With the increase of software scale and complex, and the process dynamic evolution, the traditional software testing technology is difficult to discover and locate

The Trustworthiness Analyzing of Interacting Business Process Based on the Induction Information

software vulnerability point. So it is difficult to guarantee that software is controllable, manageable, and preventable in open environment, also very difficult to realize the predictable behavior and effect. The software running behavior can reflect comprehensively the software behavior changing situation in the dynamic evolution and under the complex environment.

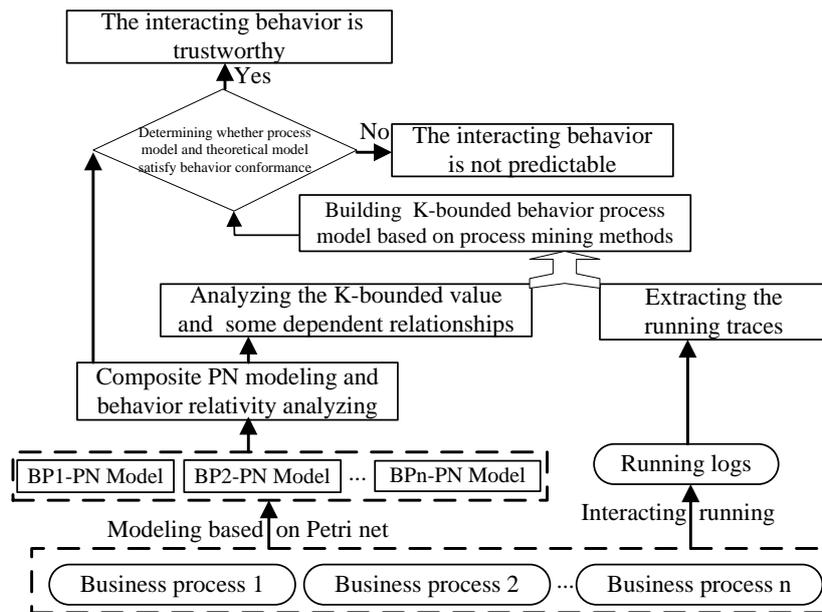


Fig.4. Research framework

By taking into account the requirements comprehensively outlined in this motivating section and the requirements of behavior trustworthiness, viewed from the software behavior, we propose a behavior-aware software trustworthiness research frame, showed in Fig.4. The main research idea is as follows: firstly, we model the software components based on Petri net, which can describe the system's dynamic characteristic concisely from the behavior angle, and has great advantage on describing some important activities and phenomena such as concurrence, conflict and synchronization, etc. Secondly, for analyzing the interacting influence of the software components, we take advantage of the analyzing methods of interaction behavior relativity to judge whether the interacting software are behavior consistent relativity. Finally, for dealing with the influence of external factors, we extract the dynamic behavioral model based on process mining, compare the dynamic behavior model with theoretical composite model, and judge whether the two models satisfy behavior congruence, if they are behavior congruence, then the component interacting process is not to be influenced by outside factors, such as virus, vicious program, and network environment and so on. The implementation process can be controllable, manageable,

and preventable, its behavior and result may be also predictable. If they are not complete behavior congruence, we analyze the behavior fitness and behavior appropriateness, and select an optimization composite behavior model.

4. The Analyzing of the Composite Model of Business Process Based on Petri Net

4.1. Basic Concept

Here, we only introduce several conceptions correlating with the paper close, other Petri Nets terms in the literature [23,24].

Definition 1^[24]. Let X be a finite alphabet, $Y \subseteq X$, let $\Gamma_{X \rightarrow Y} : X^* \rightarrow Y^*$, is called a projection from X to Y , if and only if $\forall \sigma \in X^*$, and $\Gamma_{X \rightarrow Y}(\sigma)$ is the residue string as all characters excluded from $X-Y$. The function $\Gamma_{X \rightarrow Y}$ is called the Projection from X to Y .

Definition 2^[24]. Let X be a finite alphabet, $Y \subseteq X$, let L_X, L_Y are the language of X and Y respectively.

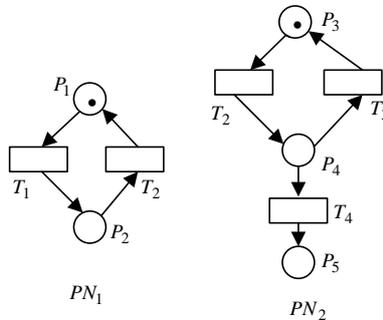


Fig.5. An example of consistent behavior relativity

If $\Gamma_{X \rightarrow Y} : (L_X) = \{\Gamma_{X \rightarrow Y}(\sigma) \in Y^* \mid \forall \sigma \in L_X\}$, then $\Gamma_{X \rightarrow Y} : (L_X)$ is called as the projection language of L_X from X to Y .

Definition 3^[24]. Let $PN_i = (P_i, T_i; F_i, M_{0i}) (i=1,2)$ are two basic Petri nets, $PN = PN_1 O_T PN_2$. if $\Gamma_{T \rightarrow T_i} (L(PN)) = L(PN_i) (i=1,2)$, then called the behavior of PN_1 is consistent with PN_2 , represented as $Be(PN_1) = Be(PN_2)$.

In Fig.5, $\Gamma_{T \rightarrow T_1}(L(PN)) = \text{pref}((t_1 t_2)^*) = L(PN_1)$,
 $\Gamma_{T \rightarrow T_2}(L(PN)) = \text{pref}((t_2 t_3)^* + t_2 t_4) = L(PN_2)$, SO $Be(PN_1) = Be(PN_2)$.

4.2. The Function Predictability Analyzing of Composite Business Process Based on the Behavior Relativity of Petri Net

Component behavior relativity [24] mainly refers to one model behavior may be influence by others when component interacting, leading to some model behavior function occur to change, even some interactions are insignificance. The interaction behavior relativity has four kinds: consistent behavior relativity, interactive behavior relativity, controlled behavior relativity, exclusive behavior relativity. Consistent behavior relativity is one kind of good interactive behavior relationship, which indicates two interaction models accomplishing the function requirements and the behavior of themselves are not influenced. We can analyze whether the function and behavior of composite business process is influenced after interacting with each other, if they satisfy the consistent behavior relativity, the interaction process is good, which indicates the function and behavior are predictable. The decision algorithm of consistent behavior relativity is showed in algorithm 1.

Algorithm 1: The decision algorithm of consistent behavior relativity
 Input: two Petri net models
 Output: the result of consistent behavior relativity

Let $PN_i = (P_i, T_i; F_i, M_{0i}) (i=1,2)$ are two Petri nets, $PN = PN_1 O_T PN_2$, and $\Delta = T_1 \cap T_2$, $X_{ij_i} (j_i = 1, 2, \dots, q_i; i=1,2)$ are all the minimum T-invariant of PN_i , $X_{ij_i}^\Delta = \Gamma_{T_i \rightarrow \Delta}(X_{ij_i})$, $j_i = 1, 2, \dots, q_i, q_i \leq q_i'; i=1,2$ are the non-zero projection vector of the minimum T-invariant of PN_i , q_1 and q_2 are the number of PN_1 and PN_2 respectively.

(1) According to the definition of Petri net incidence matrix, computing the incidence matrix of PN_1 and PN_2 .

(2) Computing the minimum T-invariant $X_{ij_i} (j_i = 1, 2, \dots, q_i'; i=1,2)$ of PN_1 and PN_2 .

(3) Computing the projection vector $X_{ij_i}^\Delta = \Gamma_{T_i \rightarrow \Delta}(X_{ij_i})$,

$j_i = 1, 2, \dots, q_i, q_i \leq q_i'; i=1,2$ of the minimum T-invariant of PN_1 and PN_2 on the share transition Δ .

(4) Computing the equation $X_{ij_i}^\Delta = \sum_{j_{3-i}=1}^{q_{3-i}} k_{3-ij_{3-i}} X_{3-ij_{3-i}}^\Delta$, $j_i \in \{1, 2, \dots, q_i\}$,

$j_{3-i} = 1, 2, \dots, q_{3-i}, 0 \leq k_{3-ij_{3-i}} \leq 1, i = 1 \vee 2$. If the equation has non-zero solutions, the $k_{3-ij_{3-i}}$ is not all zero, which means that $X_{ij_i}^\Delta$ can be non-negative linear expressed by other some vectors, then $b(X_{ij_i}^\Delta) = 1$, otherwise, there only has zero solution, that is $k_{3-ij_{3-i}} = 0$, which means that $X_{ij_i}^\Delta$ can not be non-negative linear expressed by other some vectors, then $b(X_{ij_i}^\Delta) = 0$.

(5) Determine whether the solution is all 1, if it is "yes", then the projection of minimum T-invariant of $PN_1 (PN_2)$ can be expressed by linear combination of the projection of minimum T-invariant of $PN_2 (PN_1)$, so they satisfy the consistent behavior relativity, else, the result is not consistent behavior relativity.

For convenience, we develop a behavior relativity analyzer based on Petri net. The analyzer mainly including drawing Petri net, the matrix of Petri net, the language of Petri net, the composite Petri net and the analyzing of the interaction behavior relativity. The composite result of Fig.5 in analyzer is showed in the Fig.6.

4.3. Analyzing the K-bounded Value and Some Dependent Relationships

4.3.1. Behavior dependent relationship

For convenience, in this paper, the traces of event log are corresponding to the legal sequences of Petri net model, and taking the properties of Petri nets into account.

Definition 4^[25](Basic order relationship) Let W is a event log, $PN = (P, T; F)$ is the building Petri net of business process, here, $W = L(PN), a, b \in T$

- (1) $a >_w b$ iff $\exists \sigma = t_1 t_2 t_3 \dots t_n \in W, i \in \{1, \dots, n-1\} : t_i = a \wedge t_{i+1} = b$;
- (2) $a \Delta_w b$ iff $\exists \sigma = t_1 t_2 t_3 \dots t_n \in W, i \in \{1, \dots, n-2\} : t_i = t_{i+2} = a \wedge t_{i+1} = b$;
- (3) $a \diamond_w b$ iff $a \Delta_w b \wedge b \Delta_w a$;
- (4) $a \rightarrow_w b$ iff $a >_w b \wedge \neg(b >_w a) \vee a \diamond_w b$;
- (5) $a \#_{ww} b$ iff $\neg(a >_w b) \wedge \neg(b >_w a)$;
- (6) $a \parallel_w b$ iff $a >_w b \wedge b >_w a \wedge \neg(a \diamond_w b)$;

The Trustworthiness Analyzing of Interacting Business Process Based on the Induction Information

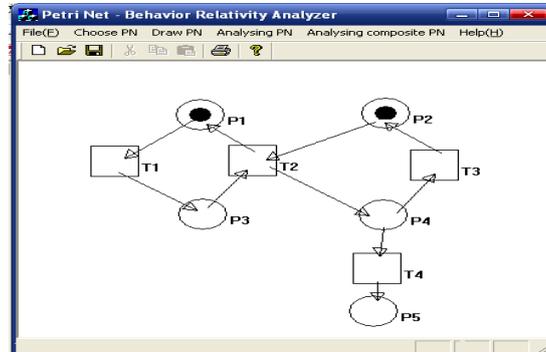


Fig.6. The composite result of Fig.5 in analyzer

As can be seen from the definition 4, $>_w$ and Δ_w are the basic order relationship, the former represents the two tasks occur one after another, the latter represents two tasks can generate a specific piece of the cycle track. These can be used to difference the relationships of two tasks. (3) (4) (5) (6) corresponding to cycle, sequence, select, and concurrency relations respectively.

Definition 5^[25] (Direct dependent relationship) $PN = (P, T; F)$ is the building Petri net of business process, for arbitrary $a, b \in T$, if there exists direct dependent relationship between a and b iff

- (1) $a \bullet \cap \bullet b \neq \Phi$;
- (2) There exists reachability mark $s \in [PN, [i] >$ to make $(PN, s)[a >$ and $(PN, s - \bullet a + a \bullet)[b >$

Definition 6^[25] (Indirect dependent relationship) $PN = (P, T; F)$ is the building Petri net of business process, for arbitrary $a, b \in T$, if there exists indirect dependent relationship between a and b iff

- (1) $a \bullet \cap \bullet b \neq \Phi$ and arbitrary $p \in a \bullet \cap \bullet b$, p is not the implicit place;
- (2) There do not exist reachability mark $s \in [PN, [i] >$, to make $(PN, s)[a >$ and $(PN, s - \bullet a + a \bullet)[b >$;
- (3) There exists reachability mark $s \in [PN, [i] >$, to make $(PN, s)[a >$ and exists reachability mark $s' \in [PN, s - \bullet a + a \bullet >$, to make $(PN, s')[b >$.

4.3.2. The basic Petri net structure with the same sequence

The business process mining, is different from the existing workflow net mentioned in some literatures, because the business process may be not soundness, that is, the token number of the initial mark is also less than or equal to 1, but business process may be not 1 - bounded in the running

process, may remain some unused tokens after running. Business process mining can borrow some ideas from the workflow mining, but the structure of business process has its own specialty, the same sequence may correspond to different models. Now basic Petri net structures with the same sequence are given, it is convenient to learn new models in HMM-based process mining.

(1) If there exist AB and AC in two different traces, then it may be have two Petri net structures shown in Fig.7. Fig.7 (a) the corresponding structure has concurrency relationships between B and C, Fig.7(b) the corresponding structure has choice relationships between B and C.

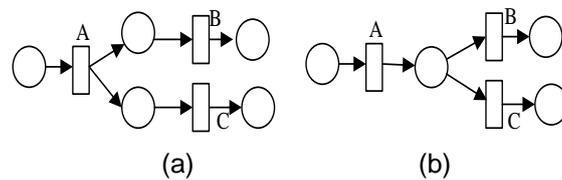


Fig.7. The corresponding Petri net structures of the sequences of AB and AC.

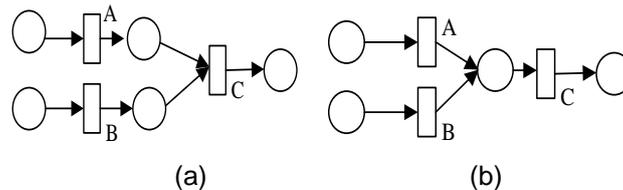


Fig.8. The corresponding Petri net structures of the sequences of AC and BC.

(2) If there exist AC and BC in two different traces, then it may be have two Petri net structures shown in Fig.8. In Fig.8 (a) the corresponding structure has concurrency relationships between A and B, Fig.8(b) the corresponding structure has not concurrency relationships between A and B.

(3) If there exist ABD and ACD in two different traces, then it may be have two Petri net structures shown in Fig.9. In Fig.9(a) the corresponding structure has not concurrency relationships between B and C, Fig.9(b) the corresponding structure has concurrency relationships between B and C.

(4) If there exist ABD and ACD, ADB and ADC in the different traces, and satisfy the prerequisites, then it may be have Petri net structures shown in Fig.10, the corresponding structures have not concurrency relationships between B and C, have concurrency relationships between D and B or D and C.

(5) If there exist ACD and BCD in two different traces, then it may be have two Petri net structures shown in Fig.11. In Fig.11(a) the corresponding structure has concurrency relationships between B and C or A and C, Fig. 11(b) the corresponding structure has not concurrency relationships between A and B.

The above basic structures are some expansion structures to the direct dependent relationship of the basic order relationship. It can improve the learning process by using the corresponding basic Petri net structure with the same behavior sequence, but there exist some exceptional situation, a kind of indirect dependent relationship has not effect by using the basic structure. For example[25], the case sequences are {ACD, BCE}, if using a ordinary workflow mining method, corresponding model is shown in Fig. 12(a), but Fig. 12(a) model has the actual transition sequences are{ACD, BCE, ACE, BCD}, they are more than the existing sequences {ACD, BCE},so it is not a good mining result. But, if we take the indirect dependent relationship into account, and carry out process mining, the result is shown in Fig.12(b), which the actual transition sequences is the same with the case sequences are {ACD, BCE}. For the model with indirect dependent relationship, its corresponding basic Petri net structure is shown in Fig.13.

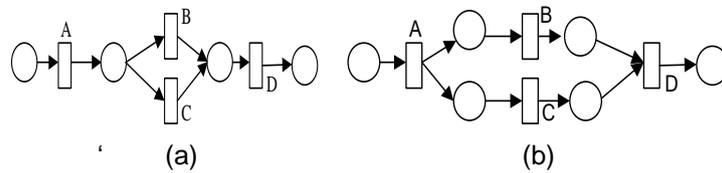


Fig.9. The corresponding two Petri net structures of the sequences of ABD and ACD.

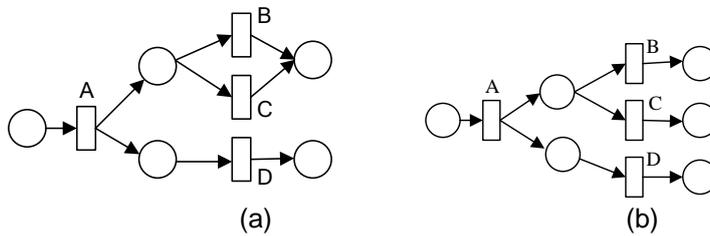


Fig.10. The corresponding Petri net structures of the sequences of ABD and ACD,ADB and ADC

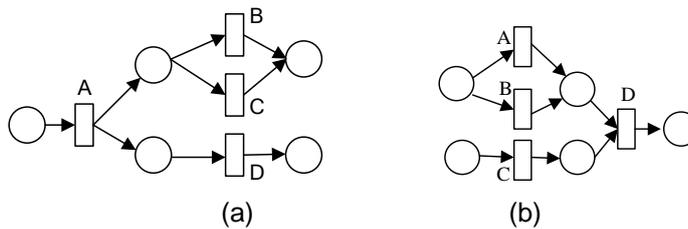


Fig.11. The corresponding Petri net structures of the sequences of ABD and BCD.

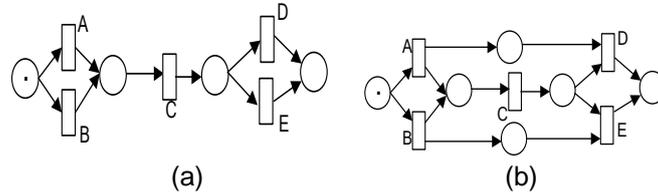


Fig.12. The corresponding Petri net model of the sequences of ACD and BCE.

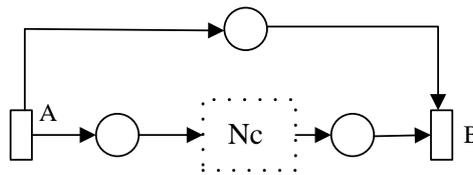


Fig.13. The corresponding Petri net structure which A and B are indirect dependent relationships

4.3.3. Analyzing the K-bounded value of the composite business process

In this paper, we need analyze the bounded value of the composite Petri net model, that is, determine the largest token number among all place of the composite Petri model, so we can adopt the reachability graph methods [23], firstly, build the corresponding reachability marked graph of Petri net, then computing the maximum value of all quantity of marking, the maximum value is the bounded value. The solving algorithm is as follows:

Algorithm 2: the bounded value of composite business process

Input: the Petri net model of business process

Output: the bounded value

- (1) Aimed to the initial marking M_0 , firing each enabled transition;
- (2) Assume the current marking is M , firing the transition t of the current marking, and obtaining the marking M' , adjusting every quantity of the marking, then continue to do;
- (3) If M has appeared in the reachability graph, then draws a directed arc from M to the existing marking, if M has not appeared in the reachability graph t , but only one quantity of M marking is bigger than (smaller than) the corresponding quantity of the existing marking, then judges whether M increase (or decrease) by the times of running number. If “yes”, then abstract the M and the existing marking into a kind marking, the corresponding quantity is noted as “ n ”. Else, it is a new marking M' , draws a directed arc from M to M' , and adjusts the corresponding quantities;

(4) If the Petri net model has not the enabled transitions or can not generate new marking, then terminates the generating process;

(5) In all marking, solves the largest quantity among the reachability marking, and returns to this value, it is the k-bounded value. If the value is n, indicates that the Petri net model is unbounded.

5. Building the K-bounded Process Behavior Model Based on New Process Mining Methods

5.1. The Metrics of Behavior Conformance about the Petri Net Model

Generally, people use exemplary behavior to compare two models, the exemplary behavior can be obtained on the basis of real process executions, and the paper assumes this exemplary behavior to be recorded in an event log. To quantify differences between two models, we introduce structure fitness, behavior precision behavior recall and behavior redundancy metrics [20,26]. Structure fitness is used to measure the structure similarity of two models. Behavior precision is used to measure the cover degree which the second model's behavior includes the first model's behavior. Behavior recall is used to measure the cover degree which the first model's behavior includes the second model's behavior. Behavior redundancy is used to evaluate the occupancy of the high frequency traces. Using the mentioned-above metrics, we take some process mining tactics to obtain the better behavior model, which has high behavior precision and behavior recall based on the event log.

Definition 7 (The structure incident matrix of Petri net model) Let $PN = (P, T; F, M)$ is a Petri net, $P = (p_1, p_2, \dots, p_m)$, $T = (t_1, t_2, \dots, t_n)$, the structure incident matrix of Petri net model can be represented by a $n \times n$ matrix, the $A = (a_{ij})_{n \times n}$ (n is the transition number)

$$a_{ij} = \begin{cases} 1 & |t^\bullet| = 1, t_i^\bullet \cap t_j^\bullet \neq \Phi \\ 1/k & |t^\bullet| = k, t_i^\bullet \cap t_j^\bullet \neq \Phi \\ 0 & \text{others} \end{cases}$$

A is called as the structure incident matrix of Petri net model.

Definition 8 (Structure fitness) let structure incident matrix of PN_1 and PN_2 are $A_{n \times n}$, $B_{m \times m}$ respectively, suppose $n \geq m$, then the structure

fitness of PN_1 and PN_2 is $1 - \sqrt{\sum (a_{ij} - b_{ij}^*)^2} / n$, here, $(b_{ij}^*)_{n \times n} = \begin{cases} b_{ij} & 1 \leq i, j \leq m \\ 0 & \text{others} \end{cases}$. If

the structure fitness value is bigger, the structure fitness is better.

Definition 9^[20] (Behavioral precision and recall). Let (N_1, M_1) and (N_2, M_2) be marked Petri nets and let $L \in B(T^*)$ be a multi-set over T.

$$\begin{aligned}
 &precision((N_1, M_1), (N_2, M_2), L) = \\
 &\left(\sum_{\sigma \in L} \frac{L(\sigma)}{|\sigma|} \left(\sum_{i=0}^{|\sigma|-1} \frac{|enabled(N_1, M_1, hd(\sigma, i)) \cap enabled(N_2, M_2, hd(\sigma, i))|}{|enabled(N_2, M_2, hd(\sigma, i))|} \right) \right) / |L| \\
 &recall((N_1, M_1), (N_2, M_2), L) = \\
 &\left(\sum_{\sigma \in L} \frac{L(\sigma)}{|\sigma|} \left(\sum_{i=0}^{|\sigma|-1} \frac{|enabled(N_1, M_1, hd(\sigma, i)) \cap enabled(N_2, M_2, hd(\sigma, i))|}{|enabled(N_1, M_1, hd(\sigma, i))|} \right) \right) / |L|
 \end{aligned}$$

Here, $hd(r, k) = \langle a_1, a_2, \dots, a_k \rangle$, i.e., the sequence of just the first k elements.

Definition 10 Assume the firing probability of all transition sequences is the same, the largest frequency of each sequence is the MAX, $MAX = \max \{L(\sigma) | \sigma \in (N, M) \ \& \ \sigma \in L\}$, then the redundancy degree of Petri net based on the event log is

$$RED(N, M, L) = \left(\sum_{\sigma \in L} \frac{MAX - L(\sigma)}{|\sigma|} \left| \{i \in \{0, |\sigma| - 1\} | \sigma(i+1) \notin enabled(N, M, hd(\sigma, i))\} \right| \right) / |L|$$

Here, $hd(r, k) = \langle a_1, a_2, \dots, a_k \rangle$, i.e., the sequence of just the first k elements.

In Fig.14, The structure incident matrix of Petri net model is

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad B = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The structure fitness is $1 - \sqrt{\frac{(\frac{1}{2})^2 + (\frac{1}{2})^2}{4}} = 0.823$, its behavior appropriateness are: behavior precision is 1, behavior recall is 1. In the literature [20], the three indexes are all 1, so the two models are complete the same, but their behavior are not the same, obviously, the measure methods have some difficult to differentiate the two models, our methods can differentiate them.

In Fig.15, The structure incident matrix of Petri net model is

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad B = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The structure fitness is $1 - 1/7 = 0.857$, its behavior appropriateness are: behavior precision is 1, behavior recall is 1. Using the methods In the literature [20], the three indexes are all 1, so the two models are complete the same, but their behavior are not the same.

The Trustworthiness Analyzing of Interacting Business Process Based on the Induction Information

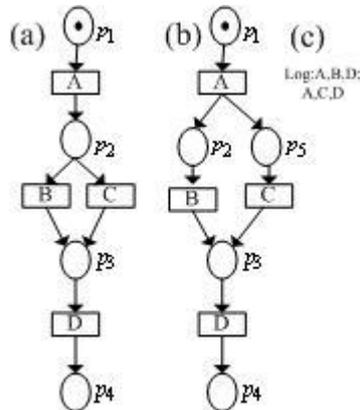


Fig.14. two models (a) and (b) which can accept the log (c).

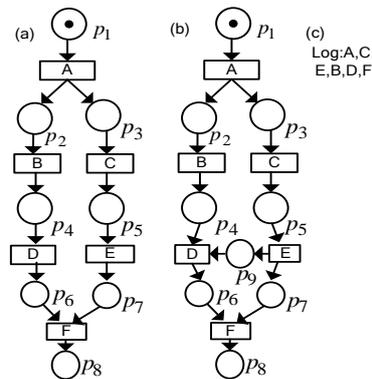


Fig.15. two models (a) and (b) which can accept the log (c).

5.2. Building K-bounded process behavior model based on the induction information

According to the motivating examples and the research framework, the business process mining is important. If we adopt the existing mining methods, it is difficult to mine the exact model which is behavior conformance with the theoretical composite model. Since each mining algorithm has its limitations, it is necessary to find new methods to solve this problem. For example, α -series algorithms can only mine 1- bounded nets, but the composite business process may not be 1- bounded. The region-based mining methods can mine K-bounded ($k > 1$) model, but it is difficult to deal with the behavior dependent relationships.

Due to the particularity of the problem, we analyze the behavior conformance between the mined models with the theoretical composite

model. In the analysis process, we can obtain some behavior dependent information and K - bounded value by analyzing the theoretical composite model in advance, which can induced the business process mining.

In process mining, we can possibly apply the K - bounded value and behavior dependent relationships as induction information to induce the process behavior model's generation.

When $K = 1$, we propose a business process mining algorithm based behavior dependent relationships, it is showed in algorithm 3. We mainly consider behavior dependent relationships in process mining, and build the optimized model according to the behavior conformance metrics of the Petri nets model.

When $K > 1$, it is difficult to control the bounded value by adopting the general mining methods, the region-based mining method [22] is a better method, but the method can tackle the behavior dependent relationship well. An improved K-bounded process mining methods is proposed in algorithm 4.

Algorithm3:1-bounded process mining algorithm based on behavior dependent relationships

Input: log traces, induction behavior dependent relationships.

Output: the Petri net model of business process.

(1) Log sequence must be pretreated first; mostly the same sequences will be merged. Then analyzing the direct dependent relationships between all tasks (or transition);

(2) Building the initial model λ_0 , according to the pretreated sequences, using the direct dependent relationships to build the initial Petri net model. In building the model, we do not consider the metrics of behavior conformance and other factors, only consider the behavior dependent relationships of tasks (transition), the built model is called as λ_0 ;

(3) Checking whether the initial model λ_0 accepts the sequence, if some sequences can be accepted, then learn the typical behavior sequences based on the frequency of sequences (i.e. the behavior sequences of large frequency should be accepted), obtaining the new model by learning process;

(4) According to the definition of structure fitness, behavior precision and behavior recall, computing the metrics value of model λ_0 and model λ based on the log sequences, if $(fitness(\lambda, L) \geq fitness(\lambda_0, L) \text{ and } precision(\lambda, \lambda_0, L) \geq \Delta_1)$ or $(fitness(\lambda, L) \geq fitness(\lambda_0, L) \text{ and } recall(\lambda, \lambda_0, L) \geq \Delta_2)$, then return to step (5); else, the initial model λ_0 keeps unchanging, repeat the step(3), continue to learn and select the next typical behavior sequences, the selecting consider the different structures with the same sequences, the basic Petri net structure with the same firing sequences can be referenced.

(5) If $recall(\lambda, \lambda_0, L) \geq precision(\lambda, \lambda_0, L)$ then $\lambda_0 = \lambda$, return to the step (6); else, computing the redundancy degree $RED(\lambda_0)$ and $RED(\lambda)$ of the

The Trustworthiness Analyzing of Interacting Business Process Based on the Induction Information

model λ_0 and the model λ respectively, if $\frac{fitness(\lambda, L)}{RED(\lambda)} \geq \frac{fitness(\lambda_0, L)}{RED(\lambda_0)}$

then $\lambda_0 = \lambda$, return to the step (6); else, determine whether the model λ has the indirect dependent relationship of tasks (transitions), if there exist the indirect dependent relationship, the learning process by the basic structure of indirect dependent relationship, and return to the step (4);

(6) If $fitness(\lambda_0, L) \geq \Delta_3$ && $\frac{fitness(\lambda_0, L)}{RED(\lambda)} \geq \Delta$, then the learning

process end, and the maximum possibility is λ ; else, return to the step (4) and continue to learn, until the learning process end.

(7) Output the model λ .

Algorithm4: K-bounded ($K > 1$) process mining algorithm based on induction information

Input: log traces, induction behavior dependent relationships.

Output: the Petri net model of business process.

(1) Transforming the log traces into the corresponding transition system, then analyzing the direct dependent relationships between all tasks (or transition);

(2) Using the Generate Minimal Regions algorithm [22] to generate the minimal region R of the TS;

(3) Let $TS = (S, E, A, s_{in})$, $PN = (P, T; F, M_0)$, for each $r \in R$, each r is corresponding to a place p, $p_i = r_i, |p| = |r|$;

(4) For each $r \in R$, $M_0[r] = r(s_{in})$, each e is corresponding to a place t, $t_i = e_i, |t| = |e|$;

(5) Let $ER(e) = \{s \mid \exists s' : (s, e, s') \in E\}$, $SR(e) = \{s \mid \exists s' : (s', e, s) \in E\}$, if $ER(e) \subseteq r$, then linking a directed arc from the p which corresponding to r to t which corresponding to e. if $SR(e) \subseteq r$, then linking a directed arc from the t which corresponding to e to the p which corresponding to r. So, a Petri net model can be obtained, and the model is called as λ_0 ;

(6) Using the induction behavior dependent relationships, selecting one of dependent relationship to adjust the model λ_0 , and obtaining the new model λ ;

(7) Computing the metrics value of model λ_0 and model λ based on the log sequences, if $(fitness(\lambda, L) \geq fitness(\lambda_0, L) \text{ and } precision(\lambda, \lambda_0, L) \geq \Delta_1)$ or $(fitness(\lambda, L) \geq fitness(\lambda_0, L) \text{ and } recall(\lambda, \lambda_0, L) \geq \Delta_2)$, then $\lambda_0 = \lambda$; else, the initial model λ_0 keeps unchanging, then considering the next induction information;

(8) If the induction information has been learned end or the model λ is not change, then the modeling process end, and output the model λ .

6. Simulation Experiments

Based on the methods proposed in the paper, we use the ProM framework to analyze the behavior trustworthiness of component interaction. The ProM framework [5] is an open-source tool and it can be downloaded at www.processmining.org, specially tailored to support the development of process mining plug-ins. In ProM, plug-ins can be categorized. The plug-ins based on data in the event log only is called discovery plug-ins because they do not use any existing information about deployed models. The plug-ins that checks how much the data in the event log matches the prescribed behavior in the deployed models is called conformance plug-ins. Finally, the plug-ins that needs both a model and its logs to discover information that will enhance this model are called extension plug-ins. In our methods, we develop (i) behavior relativity plug-ins to analyze the behavior relativity of interaction components (in Fig.16), (ii) behavior conformance plug-ins to analyze conformance between the building model and the theoretical model (in Fig.17), and (iii) extension plug-ins to use the genetic process mining methods to optimize the process, and to develop performance analyzing tool to compare different process mining methods.

Experiment environment: CPU is Intel dual 3.2 0GHz, Memory is 2.00GB, and operation system is Windows XP. Our plus-ins are developed based on the ProM with version 5.2.

Test 1. For the same event logs from the large Benchmarks, giving some known induction information, using our methods and genetic methods presented in the literature [20] respectively, we compare the three indexes including cost, fitness and appropriateness, in order to analyze the relationships between the mined models with the logs by two process mining methods.

We compare our methods with genetic process mining methods which are studied in literature [20]. In Table 1, with the case increase, the costs of two methods are both increasing quickly, but our methods is better than the methods of genetic process mining. The reason is that our methods have smaller solution space because of using the basic behavior structure. The fitness and appropriateness of two methods are irrelevant with the case number, but our methods are better than the genetic process mining methods.

The Trustworthiness Analyzing of Interacting Business Process Based on the Induction Information

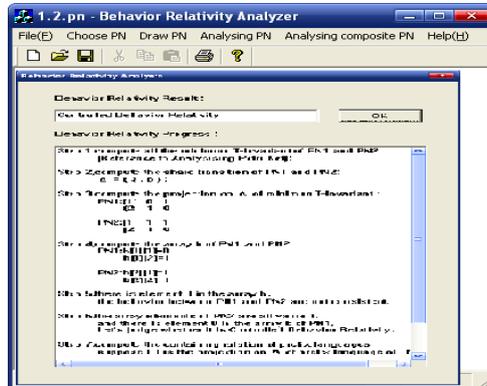


Fig.16. A screenshot of behavior relativity analysis

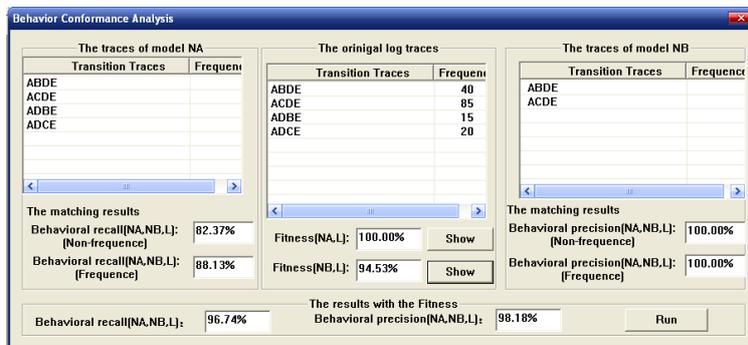


Fig.17. A screenshot of behavior conformance analysis

Table 1. The results of two methods based on large Benchmarks

Log		Log5-1		Log5-2		Log5-3		Log5-4		Log5-5	
		Case	K	Case	K	Case	K	Case	K	Case	K
		100	1	300	1	500	1	700	2	900	2
Our Method	Cost	5.32s		7.68s		10.57s		14.34s		17.84s	
	Fitness	1		1		0.992		0.981		0.958	
	App.	0.996		0.993		0.985		0.972		0.963	
Genetic Method	Cost	5.74s		8.07		11.46s		16.17 s		21.90s	
	Fitness	1		0.997		0.982		0.893		0.887	
	App.	0.994		0.989		0.970		0.882		0.876	

Test 2. For the component cases from the large Benchmarks, giving some known induction information, such as the K-bounded value and some dependent relationships, using our methods and genetic methods presented in the literature [20] respectively, we compare the some indexes using two methods under different interacting business process.

Table 2. The result of two methods based on benchmarks

Index methods	Interacting BP	(Bp1, Bp2)	(Bp1, Bp3)	(Bp1, Bp4)	(Bp1, Bp5)	(Bp2, Bp3)	(Bp2, Bp4)
		K-bounded	1	1	2	2	2
Our Methods	K-bounded	1	1	2	2	2	2
	Structure Fitness	1	1	1	1	0.996	0.994
	Behavior Precision	1	1	0.985	0.978	0.971	0.963
	Behavior Recall	0.992	0.986	0.978	0.966	0.959	0.952
Genetic Methods	K-bounded	1	1	1	1	1	1
	Structure Fitness	1	1	0.993	0.989	0.984	0.977
	Behavior Precision	0.996	0.994	0.975	0.958	0.920	0.901
	Behavior Recall	0.969	0.971	0.937	0.915	0.895	0.886

In Table 2, a tuple (B_{pi}, B_{pj}) represents the interaction between the business B_{pi} and the business process B_{pj}. By analyzing, the K-bounded value is given. The experiment results show that our methods are better than the genetic mining methods, especially, when the K-bounded value of composite business process is larger than 1, the difference is larger.

7. Conclusions

At present, the trustworthiness of business process is a study focus, the essence characteristics of trustworthy business process are the execution result and behavior can be predictable. Under the open and dynamic changing environment, some complex demands need several business processes interacting to implement, these cause the composite business process to behave uncontrollably, uncertainly, and unpredictably. The paper presents a behavior trustworthiness analysis method of business process based on induction information. Firstly, aimed to the internal factors, we analyze the consistent behavior relativity in order to guarantee the predictable function of business process. Then, for the external factors, in order to analyze the behavior change of business process, we propose a process mining methods based on induction information. Finally, using methods above mentioned, we test our methods in large Benchmark, and compare our methods with genetic process mining. Theoretical analysis and

experimental results indicate that our method is better than the methods of genetic process mining.

Based on the theoretical analysis and experimental results, the innovation and advantage of the paper are: 1) In order to analyze the predictable interaction function, consistent behavior relativity based on Petri net is presented. 2) An effective measure methods of behavior conformance is proposed, the methods can solve these problems such as Fig.14 and Fig.15, but these problems can not be solved in literature in [20]. 3) Business process mining methods based on induction information is proposed, which can take into account induction information from the theoretical composite model, and avoid the blindness of building process model.

In the future, we would like to study the trustworthy evaluation of networked software, and study the adaptation methods of non-consistent behavior relativity. Moreover, it is also one of our future works to study the intelligence and dynamic behavior analyzing methods of complex business process.

Acknowledgement. We would like to thank the support of the National Natural Science Foundation of China under Grant No.90818023, No.60873144 and No.60973050, the Program for Changjiang Scholars and Innovative Research Team in University, the Major State Basic Research Development Program of China under Grant No. 2010CB328101, the Natural Science Foundation of Educational Committee of Anhui Province of China (KJ2011A086, KJ2010B310 and KJ2009A50), the Natural Science Foundation of Shanxi (Nos. 2011021013-2).

References

1. Zheng Z.M., Ma S.L.: Dynamical characteristics of software trustworthiness and their evolutionary complexity. *Science in China Ser F-Inf Sci*, 52(8), 1328–1334. (2009)
2. Wang H.M., Tang Y.B., Ying G.: The trustworthy mechanism of internet software. *Chinese Science (E series)*, 36(10), 1156–1169. (2006)
3. Liu K, Shan Z.G., Wang J.: The general study on the major project of trustworthy software. *Chinese Science Foundation*, 22(3), 145-151. (2008)
4. W.M.P. van der Aalst, T. Weijters and L. Maruster, *Workflow Mining: Discovering process models from event logs*, *IEEE Transactions on Knowledge and Data Engineering*. 16(9), 1128-1142. (2004)
5. W.M.P. van der Aalst, B.F. van Dongen, J. Herbst, L. Maruster and G. Schimm, *Workflow Mining: A Survey of Issues and Approaches*, *Data and Knowledge Engineering*. 47(2), 237-267. (2003)
6. W.M.P. van der Aalst, H.A. Reijers and A.J.M.M. Werjters, *Business process mining: An industrial application*, *Information System*. vol 32, 713–732. (2007)
7. J. Koehler, G. Tirenni, and S. Kumaran. From business process model to consistent implementation: A case for formal verification methods. In *Proc. 6th IEEE International Enterprise Distributed Object Computing Conference (EDOC)*, pages 96–106, 2002.

8. E. Kindler and T. Vesper. A temporal logic for events and states. In Proc. 19th International Conference on Application and Theory of Petri Nets, volume 1420 of Lecture Notes in Computer Science, pages 365–384. Springer, 1998.
9. C. Klauck and H.-J. Mueller. Formal business process engineering based on grammar graphs. *International Journal on Production Economics*, 50:129–140, 1997.
10. J. Jürjens. Towards development of secure systems using UML. In Proc. Fundamental Approaches for Software Engineering (FASE), pages 187–200, 2001.
11. T. Lodderstedt, D. Basin, and J. Doser. Secure UML: A UML-based modeling language for model-driven security. In Proc. 5th International Conference on the Unified Modeling Language, volume 2460 of Lecture Notes in Computer Science, pages 425–441. Springer, 2002.
12. Du YY, Jiang CJ, Zhou MC, et al. Modeling and monitoring of E-commerce workflows. *INFORMATION SCIENCES*, 179(7):995-1006. (2009)
13. Michael Backes, Birgit Pfizmann, Michael Waidner. Security in Business Process Engineering, LCNS 2678, pages 168 - 183, June 2003.
14. Hassan T , Morteza A , Rasool J . Trust-based user-role assignment in role-based access control//Proceedings of the AICCSA. Amman , Jordan , 807-814.(2007)
15. Ping Sun, Changjun Jiang. Analysis of workflow dynamic changes based on Petri net. *Information and Software Technology*, v 51, n 2, p 284-292, February 2009
16. Montali, M.: Specification and verification of declarative open interaction models: A logic-based approach. *Lecture Notes in Business Information Processing*, v 56 LNBI, p 1-402, (2010)
17. Alam, Masoom ,Nauman, Mohammad; Zhang, Xinwen: Behavioral attestation for web services based business processes. *International Journal of Web Services Research*, v 7, n 3, p 52-72, (2010)
18. Alam, Masoom Nauman, Mohammad; Zhang, Xinwen; Ali, Tamleek; Hung, Patrick C.K. Behavioral attestation for business processes. 2009 IEEE International Conference on Web Services, ICWS 2009, p 343-350, (2009)
19. Wan, Kaiyu (East China Normal University, Shanghai, China); Muhammad, Mubarak; Alagar, Vasu, A formal model of business application integration from Web services. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, v 5404 LNCS, p 656-667, (2009)
20. A.K. Alves de Medeiros, W.M.P. van ,A.J.M.M. Weijters, Quantifying process equivalence based on observed behavior, *Data & knowledge engineering*, 64,55–74 (2008)
21. K. A. de Medeiros, A. J. M. M. Weijters and W. M. P. van der Aalst, Genetic process mining: an experimental evaluation, *Data Mining and Knowledge Discovery*. 14,245–304. (2007)
22. J. Carmona, J. Cortadella and M. Kishinevsky, New Region-Based Algorithms for Deriving Bounded Petri Nets, *IEEE Transactions on Computers*. 59(3),371-384. (2010)
23. T. Murata, Petri nets: Properties, analysis and applications. *Proc. IEEE*, 77(4) (1989) 541–580.
24. C.J. JIANG, Behavior theory and its application of Petri net, (Higher education press, 2003).

The Trustworthiness Analyzing of Interacting Business Process Based on the Induction Information

25. Wen Lijie, Wil M. P. van der Aalst and Wang Jianmin etc, Mining Process Models with Non-Free-Choice constructs, Data Mining and Knowledge Discovery. 15(2), 145-180.(2007)
26. Rozinat and W.M.P. van der Aalst, Conformance Testing: Measuring the Fit and Appropriateness of Event Logs and Process Models, Workshop Business Process Intelligence (BPI '05). vol1,1-12. (2005)

Xianwen Fang received his M.A. degree from Shandong University of Science and Technology, China, in 2004, and PhD. degree in the key Lab of Service Computing at Tongji University in 2011. He has been an associate processor with Anhui University of Science and Technology, China, since 2007. His research interests include Petri net, trustworthy software and Web services.

Changjun Jiang received the PhD. degree from Chinese Academy of Sciences, Beijing, China, in 1995. He is currently a Professor with the Department of Computer Science and Engineering, Tongji University, China. His current areas of research are concurrent theory, Petri net, and formal verification of software. He has published more than 100 papers in domestic and international academic journals and conference proceedings. These papers are embodied more than 150 times by SCI and EI and are cited more than 200 times by others.

Zhixiang Yin received her M.A. degree from Nanjing Normal University of China in 1996, and PhD. degree in the Huazhong University of Science and Technology in 2003. His research interests include Petri net and trusted software.

Xiaoqin Fan received her M.A. degree from TaiYuan University of Science and Technology, China, in 2002, and PhD. degree in the key Lab of Service Computing at Tongji University in 2010. Her research interests include Petri net and Web services.

Received: April 11, 2010; Accepted: January 19, 2011.

