http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 13 (1): 86-93, 2014
ISSN 1812-5638 / DOI: 10.3923/itj.2014.86.93

An Optimized Method of Business Process Mining Based on the Behavior Profile of Petri Nets

[1]Xianwen Fang, [1]Junzhi Wu and [2]Xiangwei Liu
[1]Department of Information and Computer Science, Anhui University of Science and Technology, Huainan, Anhui, 232001, China
[2]Department of Management, Anhui University of Science and Technology, Huainan, Anhui, 232007, China

Abstract: A good business process is the key role to enhance the operational efficiency and the service quality in Business Process Management (BPM). So the purpose of the study is to establish an algorithm to approve the business process. Especially, the analyzing method is presented based on behavior profiles of Petri net, while most of the existing methods are based on the analysis of the structural properties of Petri nets and ignoring the behavioral properties. First, the behavioral order relations are constructed and analyze their behavior profile to establish initial process model; then present the measurement methods of consistency analysis and other evaluation criteria to optimize the initial process model and last get the optimized model according to the algorithm. Finally, simulation analysis is conducted on the PROM5.2 platform and through comparing this method with another, the result show the advantages of the method.

Key words: Petri net, mining of business process, behavior profiles, consistency

INTRODUCTION

Business Process Management (BPM) not only can help to maintain sustainable competitive between enterprises but also can shorten the cycle time of service operation. Each service operation is conducted by the operation of a series of business processes, i.e., Business Process Management must be flexible to adopt business processes to respond to the changing needs of the business problems. The efficiency of business process management depends directly on the merits of business processes and business process mining is a relatively new application in the field of business process management about data mining, analyzing logs generated by the operation of business processes, restoring the true process, then analyzing and optimizing it. For the problem of process mining, the existing mining methods are mainly about behavioral semantics and behavioral processes. About behavioral semantics, (Zhang and Yu, 2012), a mining method was proposed to mine trust relationships based on semantics. Literature (Tiwari and Kaushik, 2013) established a new method to find significant places from a geo-spatial region according to semantics of the places. A lot of researches based on behavioral semantics have been conducted but the aspect of behavioral processes is relatively small. An β algorithm to dig reasonably structured workflow nets from event logs was proposed in (Wen *et al*., 2007), which contain both start and completed events and also proposes an β algorithm to mine reasonable workflow nets containing non-free-choice constructs. Unstructured process mining was presented in (Fang *et al*., 2011), meanwhile, conducting the mining of non-free-choice constructs by using the "default exists" and "default does not exist" ideological. Literature (Xiaohui and Fengjuan, 2010) presented a mining method of workflow based on rough Petri nets, which gathered rough set theory and the characteristics of Petri nets. In Sun *et al*. (2011), a method to mine processes from fragmentations after distributed execution of workflow was proposed, according to optimizing storage fragmentation and distribution to shorten time and improve efficiency of service. In Carmona *et al*. (2008), using the region theory to mine processes was proposed, mainly making use of a bounded Petri nets to over-approximate the behavior of an event log. According to logistic regression model to discover the direct relationship between events of the reality and incomplete logs with noise was proposed by Maruster *et al*. (2002). According to Gao *et al*. (2009), there were three methods: default mining, mining based on activity similarity and mining based on case similarity. In Gaaloul and Godart (2005), the purpose of the process mining was

Corresponding Author: Xiangwei Liu, Department of Management, Anhui University of Science and Technology, Huainan Anhui, 232007, China

placed in the transactional-dependent behavior rather than control flow aspects. Most of the above-mentioned are built on the workflow nets and their methods are based on the behavioral dependencies between activities, while the aspect of business processes is less studied and most do not conduct consistency measurement with the process behavior for the merits of mining degree.

Based on the above background, a mining method is proposed to dig out the appropriate business processes from execution logs generated by running programs. On the basis of the behavioral dependencies, behavior profile is considered in the modeling process. According to the structures of Petri nets corresponding to behavior profiles in the section 4.1, the mining models are continually modified and established. In this process, the consistency of behavior profiles is taken into account to select the better process. Naturally, fitness and appropriateness between logs and processes are also used to optimize the mining models. Eventually the model is achieved, which has reasonable behavior and well-structured business process. Finally, the analysis of an example proves this approach is practical and valid.

BASIC CONCEPTIONS

In this part, two major concepts are from literature (Weidlich, 2011).

Behavioral profile is a description of microscopic information of network system, specifically, it capturing the potential order of occurrence between transitions. Behavioral profile is defined on the basis of the concept of weak ordering; two transitions t_1 and t_2 are in a weak ordering, if there is a occurrence sequence staring in the initial marking such that t_1 occurred before t_2.

Definition 1 (Weak order): Let $S = (N, M_0)$ be a net system with $N = (P, T; F)$ and $T' \subseteq T$ a set of transitions. A pair of transitions $(x, y) \in (T' \times T')$ is in the weak order relation $\succ$ over T', iff there exists a firing sequence $\sigma = t_1t_2...t_n$ such that (N, M_0) [> and indices $j, k \in N$, $1 \leq j < k \leq n$ for which holds $t_j = x$ and $t_k = y$.

According to the weak order relationship between transition pairs, three relationships of behavior profiles are defined on the set of the transitions.

Definition 2 (Behavior profile): Let $S = (N, M_0)$ be a net system with $N = (P, T; F)$ and $T' \subseteq T$ a set of transitions. A pair of transitions $(x, y) \in (T' \times T')$ can be in the following profile relations:

- The strict order relation $\rightarrow$, iff $x \succ y$ and $y \not\succ x$
- The exclusiveness relation $+$, iff $x \not\succ y$ and $y \not\succ x$

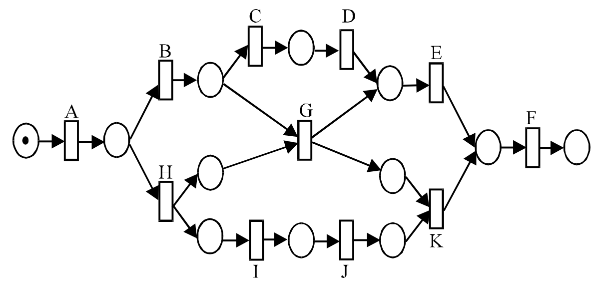

Fig. 1: A simple illustrational process model M

- The interleaving order relation $||$, iff $x \succ y$ and $y \succ x$

$B = \{\rightarrow, +, ||\}$ is the behavior profile over T'.

If an activity pair (x, y) meets strict order relation, i.e., $x \rightarrow y$, the activity pair (y, x) meets inverse strict order relation, i.e.,$y \rightarrow^{-1} x$.

The process model M in Fig. 1 illustrates the relations of the behavior profile. For example, the activity pair (A, K) is the strict order relation but note that K does not necessarily occur, as behavioral profile describes the potential order of occurrence; the activity pair (B, H) is the exclusiveness relation, thus B and H can not occur in the same sequence; the activity pair (G, J) is the interleaving order relation, as their occurrence order can not be determined, i.e., G can occur before J, also can occur after J.

CONSISTENCY ANALYSIS AND EVALUATION METHODS

Consistency analysis and evaluation between two process models: In the business process management system, consistency of two aligned process models must be maintained between different stakeholders; therefore, consistency analysis and evaluation of two aligned business process models become a problem that can not be ignored. In this study, behavior profile is used to analyze consistency, i.e., consistency of behavior profile (Weidlich and Mendling, 2012; Weidlich *et al*., 2011). It is studied based on the relationship of behavior of the corresponding transitions, unlike the constraint of analyzing consistency of trace equivalence being injective, it can being 1: n, even n: m. In addition, the consistency of behavior profile is studied on the basis of behavior relationship of the corresponding actives, this correspondence do not containing overlapping correspondence (Rozinat and van der Aalst, 2006).

Definition 3 (Behavior profile consistent transition pairs): Let $S_1 = (P_1, T_1; F_1, M_1)$ and $S_2 = (P_2, T_2; F_2, M_2)$ be two Petri net systems with $B_1 = \{\rightarrow_1, +_1, \|_1\}$ and $B_2 = \{\rightarrow_2, +_2, \|_2\}$ their behavior profiles and $\sim \subseteq T_1 \times T_2$ a correspondence relation. Let $R_1 \in B_1 \cup \{\rightarrow_1^{-1}\}$ and $R_2 \in B_2 \cup \{\rightarrow_2^{-1}\}$. The set of behavior profile consistent transition pairs $C\tilde{T}_1 \subseteq \tilde{T}_1 \times \tilde{T}_1$ for S_1 contains all pairs (t_x, t_y), such that:

- if $t_x = t_y$, then $\forall t_s \in \tilde{T}_2$ with $t_x \sim t_s$ it holds $(t_x R_1 t_x \wedge t_s R_2 t_s) \Rightarrow R_1 = R_2$
- if $t_x \neq t_y$, then $\forall t_s, t_t \in \tilde{T}_2$, $t_x \neq t_s$ with $t_x \sim t_s$ and $t_y \sim t_t$ it holds either (1) $(t_x R_1 t_y \wedge t_s R_2 t_t) \Rightarrow R_1 = R_2$ (2) $t_x \sim t_t$ and $t_y \sim t_s$

The set $C\tilde{T}_2$ for S_2 is defined analogously.

Definition 4 (Degree of behavior profile consistency): Let $S_1 = (P_1, T_1; F_1, M_1)$ and $S_2 = (P_2, T_2; F_2, M_2)$ be two Petri net systems with $B_1 = \{\rightarrow_1, +_1, \|_1\}$ and $B_2 = \{\rightarrow_2, +_2, \|_2\}$ their behavior profiles and they are aligned about $\sim$, $C\tilde{T}_1 \subseteq \tilde{T}_1 \times \tilde{T}_1$ and $C\tilde{T}_2 \subseteq \tilde{T}_2 \times \tilde{T}_2$ a consistent set of transition pairs. The degree of behavior profile consistency of $\sim$based on the set of transitions is defined as:

$$MBP(S_1, S_2) = \frac{\sum_{(t_x,t_y)\in C\tilde{T}_1 \cup C\tilde{T}_2} \left|(t_x, t_y)\right|}{\sum_{(t_x,t_y)\in(\tilde{T}_1,\tilde{T}_1)\cup(\tilde{T}_2,\tilde{T}_2)} \left|(t_x, t_y)\right|}$$

Consistency analysis between the logs and the model: The consistency analysis between two process models biases towards "theory", while the consistency analysis between the logs and the model compared with the former more focused on "practice". It is that through "practice" to test whether a model is an optimized model by using the consistency analysis between the logs and the model.

Basic knowledge: Here, related knowledge of behavior profiles of logs is introduced. The behavior profile of logs is defined on the basis of the extended order relation based on behavior profiles. This definition can reflect the behavior relationship of two adjacent activities in a log to a greater extent, laying the foundation to mine processes from logs.

Definition 5 (Extended order relation based on behavior profile): Let L be a log and NP = (P, T; F, M_0) be a Petri net corresponding business processes, i.e., L = L (N). Two adjacent activities a, b $\in$ T can be in the following relations:

- The recurrence relation: $a \mapsto_L b$ iff $a \succ b$ and $b \nsucc a$
- The concurrent relation: $a \oplus_L b$ iff $a \succ b$ and $b \succ a$
- The select relation: $a \oplus_L b$ iff $a \nsucc b$ and $b \nsucc a$
- The co-occurrence relation: $a >_L b$ iff $a \nsucc b$ and $b \succ a$

A process log containing four traces is given as L = {ABDE, ACDE, ADBE, ADCE}. By definition 5, in all adjacent activity pairs, there are $A \mapsto_L B$, $A \mapsto_L C$, $A \mapsto_L D$, $B \mapsto_L E$, $C \mapsto_L E$ and $D \mapsto_L E$ being the recurrence relation; there are $B \oplus_L D$ and $C \oplus_L D$ being the concurrent relation, i.e., activities B and D can appearing as BD or DB, activities C and D can appearing as CD or DC; there is $B \oplus_L C$ being the select relation, i.e., activities B and C can not occurring in succession or that the two not connected; for the co-occurrence relation, for example, $B >_L D$ is co-occurrence relation, i.e., the occurrence of B implying the occurrence of A. On the basis of the order relation extended based on behavior profiles, the concept of behavior profile of logs is proposed.

Definition 6 (Behavior profile of logs): Let L be a log and NP = (P, T; F, M_0) be a Petri net corresponding business processes, i.e., L = L (N). Two adjacent activities a, b $\in$ T can be in the following relations:

- The recurrence relation: $a \mapsto_L b$ iff $a \succ b$ and $b \nsucc a$
- The concurrent relation: $a \oplus_L b$ iff $a \succ b$ and $b \succ a$
- The select relation: $a \oplus_L b$ iff $a \nsucc b$ and $b \nsucc a$
- The co-occurrence relation: $a >_L b$ iff $a \nsucc b$ and $b \succ a$

$B_L = \{\mapsto_L, \oplus_L, \otimes_L, >_L\}$ is the behavior profile of logs over L.

Note that, the recurrence relation and the concurrent relation can be observed in a trace but the select relation can be found should be combined with other traces. In L = {ABDE, ACDE, ADBE, ADCE}, activities B and C can not occur in succession because of $B \otimes_L C$; there is $A \mapsto_L B$ in trace ABDE, while in ACDE, it holds that $A \not\mapsto_L B$ and $A \mapsto_L C$, so B and C are the select relation.

Evaluation criteria of consistency between the logs and the model: Firstly, replay of logs in the model can evaluate fitness of a model (fitness can verify whether the process behavior meets control flows defined in the process model, i.e., whether the execution logs can run in the model). The judging criterion of fitness (Rozinat and van der Aalst, 2006) is given below. Let k be the number of different traces from the aggregated logs. For each log trace I ($1 \le i \le k$), n_i is the number of process instances in trace I; m_i is the number of missing tokens in trace i; r_i is the number of remaining token in trace I; c_i is the number of consumed tokens in trace i and p_i is the number of produced tokens during log replay of the trace i:

$$Fitness = \frac{1}{2}\left(1 - \frac{\sum_{i=1}^{k} n_i m_i}{\sum_{i=1}^{k} n_i c_i}\right) + \frac{1}{2}\left(1 - \frac{\sum_{i=1}^{k} n_i r_i}{\sum_{i=1}^{k} n_i p_i}\right)$$

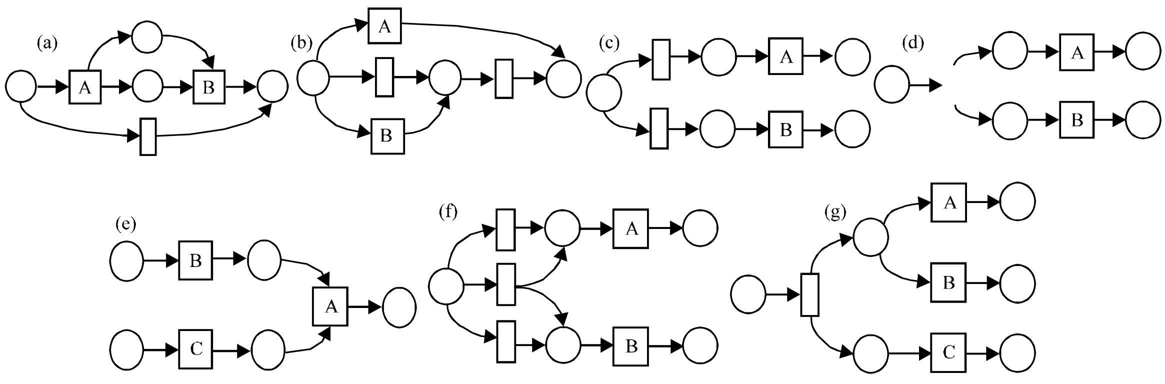

Fig. 2(a-g): Basic structures based on the behavioral profile of Petri nets: (a) Structure of order relation, (b) (c) Structure of select relation, (d) (e) Structure of concurrent relation, (f) Mixed structure of concurrent relation and select relation

Secondly, in the case that fitness close to 1, consider. Behavioral appropriateness is the accuracy of the behavior observed in the model, while structural appropriateness makes the model simpler in the case of meeting requires. Comparison between the two, apparently, behavioral appropriateness is more important than important than structural appropriateness. Two evaluation criteria are given below:

$$a_S = \frac{|T| + 2}{|T| + |P|}$$

And, T is the set of transitions; P is the set of places; 2 means the star place and the end place:

$$a_B = 1 - \frac{\sum_{i=1}^{k} n_i (x_i - 1)}{(m-1)\sum_{i=1}^{k} n_i}$$

And, k is the number of different traces from the aggregated logs; for each log trace I ($1 \leq i \leq k$), n_i is the number of process instances in trace I; x_i is the mean number of enabled transitions during log replay of the trace I; m is the number of labeled tasks (does not include invisible tasks and assuming m>1).

According to the three evaluation criteria above, the order of the comparative analysis should be: fitness$\rightarrow a_B \rightarrow a_S$.

Mining method of business process: Process mining in this article is based on logs generated during the actual execution of a process to build process models with the logs as input and then the output is process models. In this process, firstly, make sure that the uniqueness of the log, which does not contain repetitive operations; then create a relationship table (Fang *et al*., 2011) containing all activities in a log. The basic order relationships between activities can be drawn from the relationship matrix and thus get their behavior profiles according to definition 7. Finally, build process models in the light of the basic structure in Fig. 2.

A log containing five activities is given as L = {ABDE, ACDE, ADBE, ADCE}. The relationship table is built according to Definition 7 as follows:

Note that, for any two activities X and Y, integer m, it holds that N (X, Y) = m, which means the form XY appearing m times in all execution logs. For example, N (B, D) in the Table 1 means the form BD appearing one time in the execution log; N (D, B) means that the form DB does not appear in the execution log.

A definition method of weak behavior profile of logs is given according to the relationship table of activities above.

Definition 7 (Weak behavior profile of logs): Let L be a log, for any two activities X, Y $\in$ L. The relationship between X and Y can be in the following profile relations:

- The strict order relation $\rightarrow_L$, if N (X, Y)$\neq$0$\wedge$ N (X, Y) = 0, denoted X $\rightarrow_L$ Y
- The interleaving order relation $\|_L$, if N (X, Y)$\neq$0$\wedge$ N (Y, X) = 0, denoted X $\|_L$ Y
- The exclusiveness relation $+_L$, if N (X, Y) = 0$\wedge$ N (Y, X) = 0, denoted X $+_L$ Y

$B_L = \{\rightarrow_L, \|_L, +_L\}$ is the weak behavior profile of logs over L. (Note that, this process begins with appearing different activities in different sequences, as shown in Table 1, A and E are not the exclusiveness order relation, although N (A, E) = 0$\wedge$ N (E, A) = 0).

Table 1: Relationship table of five activities in L

Variable	A	B	C	D	E
A	0	1	1	1	0
B	0	0	0	1	1
C	0	0	0	1	1
D	0	1	1	0	1
E	0	0	0	0	0

In Table 1, A and B are the strict order relation because of N (A, B) = 1 ≠ 0 but N (B, A) = o; C and D are the interleaving order relation because of N (C, D) = 1 ≠ 0 and N (D, C) = 1 ≠ 0; B and C are the exclusiveness relation because of N (B, C) = N (C, B) = 0.

MINING METHOD OF BUSINESS PROCESS BASED ON THE BEHAVIORAL PROFILE OF PETRI NETS

Basic structures based on the behavioral profile of Petri nets: Business process mining is different from workflow mining, which consists of one or more workflows. To validate merits of a mined model, its fitness is important to verify; then consider the behavioral appropriateness and structural appropriateness after meeting the fitness. The basic structures based on the behavioral profile of Petri nets as follows:

- In Fig. 2a, that activities A and B are the strict order relation of the behavior profile corresponds to the structure of order relation of Petri nets
- In Fig. 2b, c, that activities A and B are the exclusiveness relation of the behavior profile, i.e., not accruing at the same time, corresponds to the structure of select relation of Petri nets, selecting one of the branches to run
- In Fig. 2d, that activities A and B are the interleaving order relation of the behavior profile, i.e., A occurring before B or after B, corresponds to the structure of concurrent relation of Petri nets and both branches can run at the same time
- In Fig. 2e, that activities B and C are the interleaving order relation of the behavior profile corresponds to the structure of concurrent relation of Petri nets; it is noteworthy that if A occurs, activities B and C should occur at the same time
- In Fig. 2f, activities A and B can occur in isolation and also can occur simultaneously; they are the exclusiveness relation of the behavior profile when they occur in isolation corresponding to the structure of select relation of Petri nets; they are the interleaving order relation of the behavior profile when they occur in simultaneously corresponding to the structure of concurrent relation of Petri nets
- In Fig. 2g, activities A and B are the exclusiveness relation of the behavior profile, i.e., one of them will occurring; the occurred activity and C are the interleaving order relation; correspond to the structure of Petri nets, first, the structure of concurrent relation (the occurrence of A (or B) and C), then, the structure of select relation (the occurrence of A or B)

Mining algorithms of business process: In this method, take the structures based on the behavior profile of Petri nets as basic; analyze the traces of logs; take the traces with the most frequency in logs as the standard, supplemented by the traces with a smaller frequency, ensuring the traces with the most frequency conforming to the model. A mining algorithm of business process based on the behavior profile of Petri nets is given as below.

Algorithm

Process mining based on the behavior profile of Petri nets:

Input: Worked execution logs
Output: Petri net model
Step 1: First of all, do pretreatment for all the event logs and merge the same logs to avoid repeating
Step 2: Create the initial model λ_0; analyze the behavior profile of the traces (at least 3 traces) with the most frequency in the worked event logs; build the relationship table of activities in logs; build the preliminary Petri net model by using directly the relationship table. In this process of modeling, evaluation indexes are not considered
Step 3: Compute the fitness of the model according to event logs; judge the result of the replay of the logs according to the fitness; adjust the control flows of the model λ_0 according to the traces with the most frequency in the remaining logs to get a new model λ_1
Step 4: Compute the behavioral appropriateness a_B and the structural appropriateness a_S in the case that fitness close to 1. If $|a_B(\lambda_0, L)-a_B(\lambda_1, L)|<0.1$ && $a_S(\lambda_0, L) \leq a_S(\leq\lambda_1, L)$ or $a_B(\lambda_0, L)-a_B(\lambda_1, L)> 0.1$, turn into Step 5; otherwise, turn into Step 3
Step 5: Adjust the control flows of the model λ_1 according to the traces with the most frequency in the remaining logs to get a new model λ_2
Step 6: Compute, respectively the degree of behavior profile consistency MBP (λ_0, λ_1) and MBP (λ_0, λ_2) between λ_0 and λ_1, λ_0 and λ_2. If MBP (λ_0, λ_1)< MBP (λ_0, λ_2), let $\lambda_0 = \lambda_2$; otherwise, let $\lambda_0 = \lambda_1$ and $\lambda_1 = \lambda_2$. Turn into Step 3

Step 7: The model λ_0 is the mined model when the replay of all the logs is completed

In the algorithm, some values such as 0.1 are derived from the trial experiences. First, the implementation of this algorithm can guarantee the success of the replay of the log sequences and restore it to the path; and second, consistency can make the model achieve the desired objectives.

CASE STUDY

In this section, a simple example is used to illustrate the feasibility of the mining algorithms of business process. Given 4 event logs, Table 2 contains the traces and the number of instances (or frequency). Merger the log and order the traces in accordance with the size of the number of instances.

Consider the traces with bigger frequency in the logs and select the first three to build a relationship table like (a) of Table 3; then compute the behavior profile of the activities like (b) of Table 3 according to definition 7.

The initial model M_0 like Fig. 3 can be drawn with the behavior profile derived from (b) of Table 3 and the basic structure of Petri nets in Section 4.1.

According to the Section 3.2.2, get the value fitness $(M_0)\approx 0.9137$ and this value is not very good as not closely to 1.According to the remained worked logs, from which find the trace with the most frequency and adjust M_0 to get M_1 like Fig. 4.

From Fig. 4, the value fitness $(M_1)\approx 0.9826$ got is close to 1; following that, consider the behavioral appropriateness and structural appropriateness and behavioral appropriateness: a_B $(M_0, L)\approx 0.9327$, a_B $(M_1, L)\approx 0.9611$, structural appropriateness: a_S $(M_0, L)\approx 0.5909$, a_S $(M_1, L)\approx 0.5652$. If a model's behavior is more complex, it is unlikely to have a simple structure; so in the case of little difference between behavioral appropriateness, there is a tending to choose the model with larger structural appropriateness; if there is a big difference between behavioral appropriateness, it is reasonable to select the model with a larger value.

Additionally, an efficient process mining method based on Discrete Particle Swarm Optimization (DPSO) is proposed in (Fang *et al*., 2011), the method in this article denoted as BCO and then compare them. Experimental environment: CPU is 1.60 GHz Intel dual-core; memory is 2.00 GB; the operating system is Windows XP. The mining plus-ins is developed based on the ProM with version 4.2.

Getting the same logs from the large standard testing procedures, mine process models from the logs by using the method in this article and DPSO method, respectively.

Table 2: Traces and the instances No. of event logs

Event Log L_1		Event Log L_2	
No. of instances	Log traces	No. of instances	Log traces
3428	ABGEF	2012	ABGEF
1246	ABCDEF	968	ABCDEF
245	AHIJGKF	286	AHGIJKF
236	AHIGJKF	26	ABGF
Event Log L_3		**Event Log L_4**	
No. of instances	Log traces	No. of instances	Log traces
1205	ABIJGKF	224	ABCDEF
21	AHIJKF	156	AHIJGKF
15	BGEF	34	BCDE
3	AHKF	12	HGKF

Work all the event logs to get the traces and the No. of instances as follows: ABGEF (5440), ABCDF (2438), AHIJKF (1606), AHIGJKF (236), AHGIJKF (286), BCDE (34), ABGF (26), AHIJKF (21), BGEF (15), HGKF (12), AHKF (3)

Table 3(a-b): Process of building behavior profile: (a) Relationship table for activities in the event logs, (b) Behavior profile for activities based on relationship table

(a)	A	B	C	D	E	F	G	H	I	J	K
A	0	2	0	0	0	0	0	1	0	0	0
B	0	0	1	0	0	0	1	0	0	0	0
C	0	0	0	1	0	0	0	0	0	0	0
D	0	0	0	0	1	0	0	0	0	0	0
E	0	0	0	0	0	2	0	0	0	0	0
F	0	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	1	0	0	0	0	0	1
H	0	0	0	0	0	0	0	0	1	0	0
I	0	0	0	0	0	0	0	0	0	1	0
J	0	0	0	0	0	0	1	0	0	0	0
K	0	0	0	0	0	1	0	0	0	0	0

(b)	A B C D E F G H I J K
A	$\rightarrow_L$ $\rightarrow_L$
B	$\rightarrow_L$ $\rightarrow_L$ $+_L$
C	$\rightarrow_L$ $+_L$
D	$\rightarrow_L$ $+_L$
E	$\rightarrow_L$
F	
G	$+_L$ $+_L$ $\rightarrow_L$ $\rightarrow_L$
H	$\rightarrow_L$ $\rightarrow_L$
I	$\parallel_L$ $\rightarrow_L$
J	$\parallel_L$ $\rightarrow_L$
K	$\rightarrow_L$

For a better comparison of the two methods by the relationship between the mining models, compare the time-cost, behavioral fitness, behavioral appropriateness and structural appropriateness of the two.

From Table 4 in the aspects of behavioral appropriateness and structural appropriateness, there is little difference; for the time-cost, both are increasing with the increase of the number of instances but the rate of increase by the method in this article is slightly less than the DPSO method. The reason is that the BCO mining methods can analyze the behavioral profile between the transitions comprehensively but the DPSO mining methods only analyze the behavior relationships from the model structure.

Table 4: Computational results based on two mining methods BCO and DPSO

		a22f0n00-1		a22f0n00-2		a22f0n00-3		a22f0n00-4		a22f0n00-5		a22f0n00-6	
	Log	Case	[S]	Case	[S]	Case	[S]	Case	[S]	Case	[S]	Case	[S]
Methods	Index	100	227	300	657	500	917	700	1123	900	786	1100	1009
BCO	Cost	5.140 sec		7.130 sec		11.020 sec		14.010 sec		19.230 sec		26.780 sec	
	Beh-Fit	1.000		0.987		0.947		0.915		0.892		0.811	
	Beh-App	0.972		0.954		0.927		0.920		0.899		0.817	
DPSO	Cost	5.190 sec		7.340 sec		12.050 sec		14.970 sec		20.090 sec		28.070 sec	
	Beh-Fit	1.000		0.990		0.944		0.910		0.897		0.820	
	Beh-App	0.980		0.956		0.933		0.918		0.900		0.814	

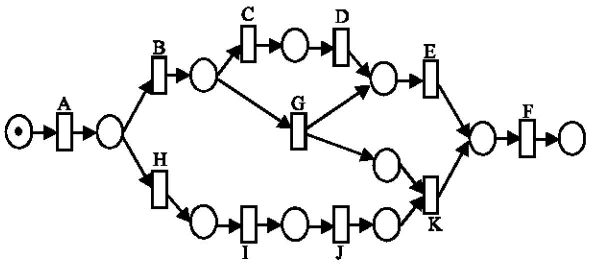

Fig. 3: Initial process model M_0

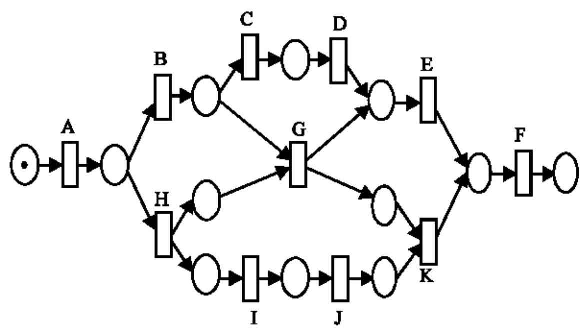

Fig. 4: Adjusted process model M_1

CONCLUSION

In order to adapt to the rapid development of business managements, optimizing business process model has a certain practical significance. In this study, the basic structures based on the behavioral profile of Petri nets are presented firstly; and in the mining algorithm, construct a model to meet the replay of logs by using a plurality of indicators, such as fitness, behavioral appropriateness and structural appropriateness; then optimize models with consistency. Lastly, it shows that the algorithm can find the optimal model with a simulation experiment.

In the future, the method will be applied to mine models which contain cyclic structures or invisible tasks to meet more needs of business management.

ACKNOWLEDGMENT

We would like to thank the support of the National Natural Science Foundation of China under Grant No. 61272153, 61170059 and 61170172, the Natural Science Foundation of Educational Government of Anhui Province of China (KJ2012A073, KJ2011A086), Anhui Provincial Natural Science Foundation (1208085MF105), Anhui Provincial Soft Science Foundation (12020503031).

REFERENCES

Carmona, J., J. Cortadella and M. Kishinevsky, 2008. A region-based algorithm for discovering Petri nets from event logs. Proceedings of the 6th International Conference on Business Process Management, September 2-4, 2008, Milan, Italy, pp: 358-373.

Fang, X., X. Gao, Z. Yin and Q. Zhao, 2011. An efficient process mining method based on discrete particle swarm optimization. Inform. Technol. J., 10: 1240-1245.

Gaaloul, W. and C. Godart, 2005. Mining workflow recovery from event based logs. Proceedings of the 3rd International Conference on Business Process Management, September 5-8, 2005, Nancy, France, pp: 169-185.

Gao, A., Y. Yang, M. Zeng, J.L. Zhang and Y.W. Wang, 2009. Organizational structure mining based on workflow logs. Proceedings of the International Conference on Business Intelligence and Financial Engineering, July 24-26, 2009, Beijing, China, pp: 455-459.

Maruster, L., A.J.M.M. Weijters, W.M.P. van der Aalst and A. van den Bosch, 2002. Process mining: Discovering direct successors in process logs. Proceedings of the 5th International Conference on Discovery Science, November 24-26, 2002, Lubeck, Germany, pp: 364-373.

Rozinat, A. and W.M.P. van der Aalst, 2006. Conformance testing: Measuring the fit and appropriateness of event logs and process models. Proceedings of the International Workshops on Business Process Management Workshops, September 5, 2005, Nancy, France, pp: 163-176.

Sun, S.X., Q. Zeng and H. Wang, 2011. Process-mining-based workflow model fragmentation for distributed execution. IEEE Trans. Syst. Man Cybernetics, Part A: Syst. Humans, 41: 294 -310.

Tiwari, S. and S. Kaushik, 2013. Mining popular places in a geo-spatial region based on GPS Data using semantic information. Databases Networked Infor. Syst., 7813: 262-276.

Weidlich, M. and J. Mendling, 2012. Perceived consistency between process models. Inform. Syst., 37: 80-98.

Weidlich, M., 2011. Behavioural profiles: A relational approach to behaviour consistency. Ph.D. Thesis, Institutional Repository of the University of Potsdam, Germany.

Weidlich, M., A. Polyvyanyy, N. Desai, J. Mendling and M. Weske, 2011. Process compliance analysis based on behavioural profiles. Inform. Syst., 36: 1009-1025.

Wen, L., W.M.P. van der Aalst, J. Wang and J. Sun, 2007. Mining process models with non-free-choice constructs. Data Min. Knowledge Discovery, 15: 145-180.

Xiaohui, W. and G. Fengjuan, 2010. Workflow process mining based on rough Petri net. Proceedings of the International Conference on Advances in Energy Engineering, June 19-20, 2010, Beijing, China, pp: 277-280.

Zhang, Y. and T. Yu, 2012. Mining trust relationships from online social networks. J. Comput. Sci. Technol., 27: 492-505.