

# A Method of Process Mining Based on Model Merged\*

Pei Hua, Xianwen Fang  
School of Science  
Anhui University of Science and Technology  
Huainan Anhui, China  
{497457097,280060673}@qq.com

Xiangwei Liu  
School of Management  
Anhui University of Science and Technology  
Huainan Anhui, China  
992450828@qq.com

**Abstract**—In the area of Business process mining, there are many methods are available to discover process models from event logs. A new process mining method is proposed based on model merged. Firstly, sub-models are established respectively according to the behavioral profiles of each event log. Then finding out the maximum matched regions of the sub-models, merging the maximum matched regions and adding the rest of the transitions into the suitable place. Thus, the merged model is obtained. The method considers the maximum matched regions as a whole, greatly simplifying the process of modeling. Finally, a simple example is listed to show the feasibility of the method.

**Keywords**— process mining; behavioral profiles of Petri net; model merged

## I. INTRODUCTION

In the process of enterprise merger or reconstruction, there will be several model variants for the same business process, which need to be merged to facilitate the management of enterprises, thus model merged techniques arises at the historic moment. In recent years, many scholars do a lot of research with regard to model merged, such as in the literature [1], Shuang sun, Akhil Kumar et al proposed a workflow merged method, which introduces sequential merge, parallel merge, conditional merge, iterative merge and complex merge. Marcello La Rosa proposed a method for merging multiple process variants and extracting the digest in the literature [2]. In addition, the model merged techniques is typically used to discover configurable process models because the merged model must contain all the behaviors of the variants. Professor Aalst has made brilliant achievements in the field of process mining. The method of mining configurable process models from event logs is presented in the literature [3].

Typically, there are three main types of process mining, that is process discovery, conformance checking, and model enhancement [4]. Discovery techniques learn a model from event logs. Conformance checking techniques can be used to identify the differences between observed behavior and modeled behavior. After conformance checking, model and log are aligned and information from the logs may be used to enhance the model. In addition, model repair as a new type of process mining was introduced in [5]. If model N conforms to L (i.e., the observed behavior can be fully explained by the model), then there is no need to change N. However, if parts

of N do not conform to L, these parts can be repaired using the model repair technique.

In this paper, a new process mining method is proposed based on model merged, and the concept of region behavioral profiles is proposed based on behavioral profiles. The event logs are used as the starting point of the analysis, and sub-models are established respectively according to the behavioral profiles of each event log. Then finding out the maximum matched regions based on matched transitions and merging maximum matched regions according to the region behavioral profiles. Finally, putting residual transitions into merged model according to the behavioral profiles between residual transitions and the transitions in maximum matched regions.

In the remainder of the paper, we first introduce the basic concepts in section II. Section III introduces the basic structures of process models based on region behavioral profiles, the method of process mining based on model merged and gives an example to verify the feasibility of the method. Finally, section IV concludes the paper.

## II. BASIC CONCEPTS

In this section, we will give the basic concepts needed in the paper.

**Definition 1 (Matched transitions)**<sup>[6]</sup> Let  $PM_1$  and  $PM_2$  be two Petri net models,  $T_1, T_2$  be the set of transitions respectively and  $\sim \subseteq T_1 \times T_2$  be the correspondence relation between them. Thus the matched transitions set of  $T_1$  is  $T_1^- = \{t_1 \in T_1 \mid \exists t_2 \in T_2, t_1 \sim t_2\}$ , and the matched transitions set of  $T_2$  is  $T_2^- = \{t_2 \in T_2 \mid \exists t_1 \in T_1, t_1 \sim t_2\}$ .

**Definition 2 (Maximum matched regions)** Let  $mmr(PM_1)$  and  $mmr(PM_2)$  be the part of  $PM_1$  and  $PM_2$  respectively,  $T_1' \in T_1^-$  and  $T_2' \in T_2^-$  be the transitions set of  $mmr(PM_1)$  and  $mmr(PM_2)$  respectively and  $F_1, F_2$  be flow relation. Thus,  $mmr(PM_1) \sim mmr(PM_2)$  if and only if:  $\forall t_i \in T_1', \exists t_i' \in T_2'$ , such that  $t_i \sim t_i' \wedge F_1 = F_2$ . The maximum matched regions set of  $PM_1$  is  $MMRS(PM_1) = \{mmr(PM_1) \mid \exists mmr(PM_2), mmr(PM_1) \sim mmr(PM_2)\}$ , and the maximum matched regions set of

$PM_2$  is  $MMRS(PM_2) = \{mnr(PM_2) | \exists mnr(PM_1), mnr(PM_1) \sim mnr(PM_2)\}$ .

The method of this paper is based on the behavioral profiles, and the knowledge of the weak order relation please refer to [7]. The formal definition of the behavioral profiles and the region behavioral profiles is given as follow:

Definition 3<sup>[7]</sup> (Behavioral Profiles (Process Model)) Let  $PM$  be a Petri net model,  $T$  be the set of transitions. A pair  $(x, y) \in (T \times T)$  is in one of the following three relations:

- strict order relation  $x \rightarrow y$ , if  $x \succ y$  and  $y \not\succeq x$ ;
- interleaving order relation  $x \parallel y$ , if  $x \succ y$  and  $y \succ x$ ;
- Exclusiveness relation  $x + y$ , if  $x \not\succeq y$  and  $y \not\succeq x$ .

The set of the three relations is the behavioral profiles of  $PM$ .

Definition 4<sup>[8]</sup> (Behavioral Profiles (Log)) Let  $PM$  be a Petri net model, and  $L$  be a log of  $PM$ . A pair  $(x, y) \in (T_L \times T_L)$  is in one of the following two relations:

- strict order relation  $x \rightarrow y$ , if  $x \succ y$  and  $y \not\succeq x$ ;
- interleaving order relation  $x \parallel y$ , if  $x \succ y$  and  $y \succ x$ ;

The set of the two relations is the region behavioral profiles of  $L$ .

Definition 5 (Region behavioral profiles) Let  $mnr_x, mnr_y \in MMRS(PM)$ . A pair of the maximum matched regions  $(mnr_x, mnr_y) \in MMRS(PM) \times MMRS(PM)$  is in one of the following three relations:

- strict order relation  $mnr_x \rightarrow mnr_y$ , if  $mnr_x \succ mnr_y$  and  $mnr_y \not\succeq mnr_x$ ;
- interleaving order relation  $mnr_x \parallel mnr_y$ , if  $mnr_x \succ mnr_y$  and  $mnr_y \succ mnr_x$ ;
- Exclusiveness relation  $mnr_x + mnr_y$ , if  $mnr_x \not\succeq mnr_y$  and  $mnr_y \not\succeq mnr_x$ .

The set of the three relations is the region behavioral profiles of  $PM$ .

### III. A METHOD OF PROCESS MINING BASED ON MODEL MERGED

#### A. the Basic Structure of Process Model Based on Region Behavioral Profiles

The model merged method proposed in this paper is based on behavioral profiles, including the behavioral profiles of event logs, the behavioral profiles of process models and the region behavioral profiles. The basic structures of process models based on behavioral profiles see [9]. The basic structures of process model based on region behavioral profiles as shown in figure 1, where  $mnr1$  and  $mnr2$  are maximum matched regions.

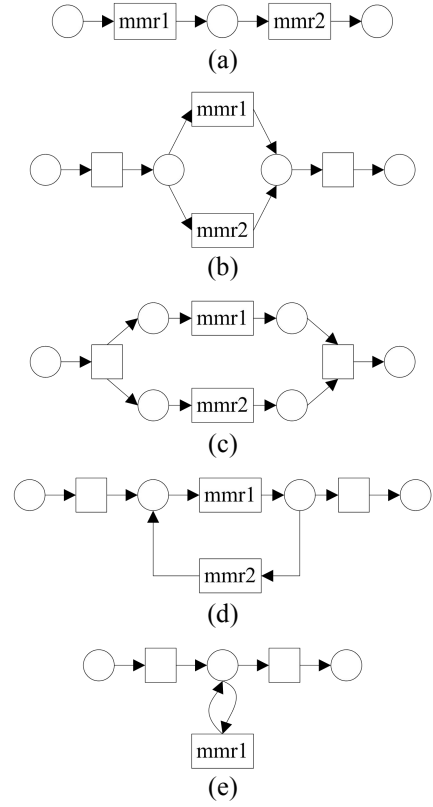


Fig. 1. the basic structures of process model based on region behavioral profiles

In figure 1(a), the pair of maximum matched regions ( $mnr1, mnr2$ ) is strict order relation, that is  $mnr1 \rightarrow mnr2$ . In figure 1(b), the pair of maximum matched regions ( $mnr1, mnr2$ ) is exclusiveness relation, that is  $mnr1 + mnr2$ . In figure 1(c), (d), the pair of maximum matched regions ( $mnr1, mnr2$ ) is interleaving order relation, that is  $mnr1 \parallel mnr2$ . However, they form a cycle as shown in figure 1(d) if they can happen many times repeatedly. The individual maximum matched region forms a self-cycle as shown in figure 1(e) if it can happen many times repeatedly.

#### B. the Method of Model Merged Based on Behavioral Profiles

Before merging the models, we should match the models first and identify matched transitions and maximum matched regions. In this paper we think transitions with the same label are matched. In order to determine the maximum matched regions, we first select a pair of matched transitions as maximum matched region of their respective models. Then seeing the transitions that adjacent with it whether in matched transitions set or not. If so, seeing whether the flow relations in the maximum matched region have matched flow relations. Proceeding until all the matched transitions are in a maximum matched region.

It's time to merge models after determining the maximum matched regions. We see the maximum matched regions as a whole. At first, initial merged model is obtained by merging the maximum matched region according to the region behavioral profiles; then calculating the behavioral profiles

relations between transitions that do not belong to any maximum matched regions and the behavioral profiles relations between these transitions and maximum matched regions based on the behavioral profiles of process models and event logs; Finally, the final merged model is obtained by adding these transitions into the initial merged model based on the behavioral profiles of process models and event logs.

### C. Process Mining Algorithm Based on Model Merged

In the last part, the basic idea of process mining method based on model merged is introduced, and the algorithm is given as follow:

Step 1: Each event logs must be pretreated and the incomplete log sequences should be removed to improve the quality of the merged model

Step 2: According to the behavioral profiles of each event logs, the sub-models  $PM_1$ ,  $PM_2$ ,  $PM_3$  and so on are established respectively. These sub-models constitute the model set  $S(PM)$

Step 3: For arbitrary  $PM_x, PM_y \in S(PM)$ , finding out their maximum matched regions sets  $MMRS(PM_x)$  and  $MMRS(PM_y)$ . Then merging the maximum matched regions according to the region behavioral profiles. The merged model is called as  $mpm(PM_x, PM_y)$

Step 4: Calculating the behavior relation of transition pair  $(a, b) \in T_r \times T_r$ , where  $T_r = (T_1 - T'_1) \cup (T_2 - T'_2) \cup \bullet MMRS(PM) \cup MMRS(PM) \bullet$ ,  $\bullet MMRS(PM)$  is the source transitions set and  $MMRS(PM) \bullet$  is end transitions set of the maximum matched regions. According to the behavioral profiles relations, inserting the rest transitions that do not included in any

maximum matched region into  $mpm(PM_x, PM_y)$ , The merged model is called as  $MPM(PM_x, PM_y)$

Step 5: Then  $S(PM) = (S(PM) - \{PM_x, PM_y\}) \cup \{MPM(PM_x, PM_y)\}$ , and return to step 3

Step 6: Output the model  $MPM$  until all the sub-models in  $S(PM)$  are merged

### D. Case Study

In order to verify the feasibility of the algorithm, in this part we will give a simple example, and the recorded event logs shown in Table 1. Capital letters denote the following activities: (A) enter the website, (V) enter the shop, (B) search goods, (C) add goods, (D) ready to checkout, (E) old user login, (F) new user registration, (G) VIP, (H) non-VIP, (I) ten percent discount, (J) VIP price, (K) regular price, (M) calculate the total price, (O) place order, (Q) payment (S) cancel the order, (U) exit website.

TABLE I. EVENT LOGS

EVENT LOG $L_1$	EVENT LOG $L_2$
ABVCFIMU	AVBCDFIMOQU
ABVCDEGJMU	AVBCDFIMOSU
BVCDE	AVBCDEHKMOQU

Firstly, the log sequence  $BVCDE$  is removed because it is incomplete. Then the behavioral profiles of event log  $L_1$  is calculated, and the key behavior relations are listed as follows:  $A \rightarrow B$ ,  $B \rightarrow V$ ,  $V \rightarrow C$ ,  $C \rightarrow D$ ,  $D \rightarrow E$ ,  $D \rightarrow F$ ,  $E + F$ ,  $F \rightarrow G$ ,  $G \rightarrow J$ ,  $F \rightarrow I$ ,  $J \rightarrow M$ ,  $I \rightarrow M$ ,  $M \rightarrow U$ . According to the behavioral profiles of  $L_1$ , constructing sub-model  $PM_1$  as shown in Figure 2. Similarly, constructing sub-model  $PM_2$  as shown in Figure 2 according to the behavioral profiles of  $L_2$ .

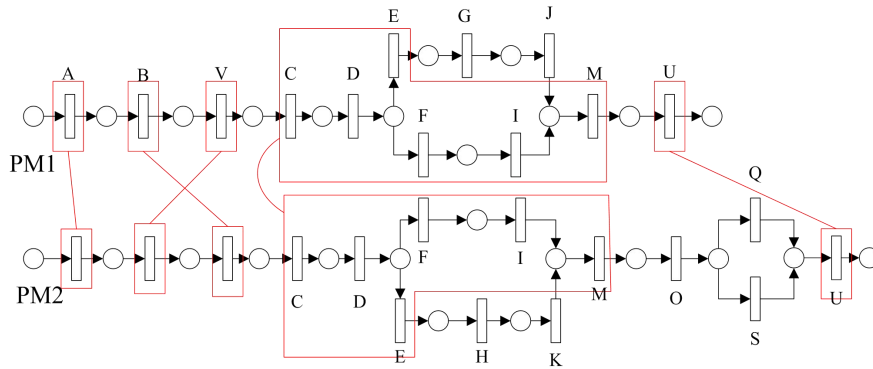


Fig. 2. the process model of  $L_1$  and  $L_2$

In Figure 1, the regions with a red frame marked are the maximum matched regions, and the red connecting lines between maximum matched domains represent their matched relation. The maximum matched regions of  $PM_1$  from left to right denoted as  $mmr_1$ ,  $mmr_2$ ,  $mmr_3$ ,  $mmr_4$ ,  $mmr_5$ , and in

$PM_2$ , denoted as  $mmr_1$ ,  $mmr_3$ ,  $mmr_2$ ,  $mmr_4$ ,  $mmr_5$ . Thus we can get the region behavioral profiles as follows:  $mmr_1 \rightarrow mmr_2$ ,  $mmr_1 \rightarrow mmr_3$ ,  $mmr_2 \parallel mmr_3$ ,  $mmr_3 \rightarrow mmr_4$ ,  $mmr_2 \rightarrow mmr_4$ ,  $mmr_4 \rightarrow mmr_5$ . The merged model  $mpm(PM_1, PM_2)$  shown in figure 3 is obtained by merging

maximum matched regions.

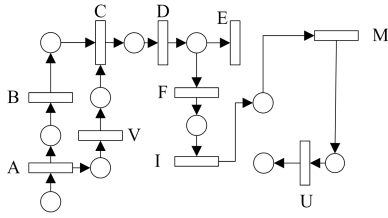


Fig. 3. Merged model  $mpm(PM_1, PM_2)$

Next, calculating the behavior relations of transition pair  $(a, b) \in T_r \times T_r$ , where  $T_r = \{A, B, V, C, E, M, O, Q, S, U\}$ . The result is,  $A \rightarrow B$ ,  $B \parallel V$ ,  $B \rightarrow C$ ,  $V \rightarrow C$ ,  $C \rightarrow E$ ,  $E \rightarrow G$ ,  $E \rightarrow H$ ,  $G + H$ ,  $G \rightarrow J$ ,  $H \rightarrow K$ ,  $J \rightarrow M$ ,  $K \rightarrow M$ ,  $O \rightarrow Q$ ,  $O \rightarrow S$ ,  $Q + S$ ,  $S \rightarrow U$ ,  $Q \rightarrow U$ . According to the behavioral profiles, inserting the rest transitions that do not included in any maximum matched region into  $mpm(PM_x, PM_y)$ . The final merged model is called as  $MPM(PM_1, PM_2)$ , as shown in figure 4.

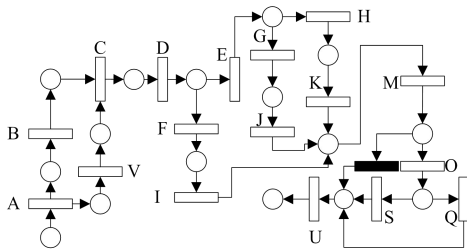


Fig. 4. Merged model  $MPM(PM_1, PM_2)$

Since all complete log sequences of event log  $L_1$  and  $L_2$  can be replayed by the model and the model does not produce too many other sequences, the process mining method based on model merged proposed in this paper is feasible.

#### IV. CONCLUSIONS

In this paper, the model merged method is based on the behavioral profiles, including the region behavioral profiles, the behavioral profiles of event logs and the behavioral profiles of the process models. This method can effectively solve the problem of behavior constraints. However, we only

mine the model based on the model merged without optimization. The next step is to evaluate and optimize the model from different angles. In addition, discovering configurable models based on behavioral profiles is also a problem worthy of further study.

#### ACKNOWLEDGMENT

We would like to thank the support of the National Natural Science Foundation of China under Grant No.61272153, No.61402011, No.61340003 and No.61170059, Anhui Provincial Natural Science Foundation (1508085MF111), Anhui Provincial Soft Science Foundation (12020503031), the Natural Science Foundation of Educational Government of Anhui Province of China (KJ2012A073, KJ2014A067), the youth academic backbone foundation of AUST, the Academic and Technology Leader Foundation of Anhui Province.

#### REFERENCES

- [1] Shuang Sun, Akhil Kumar, John Yen. Merging workflows: A new perspective on connecting business processes. *Decision Support Systems*, 2006, vol. 42, pp. 844–858.
- [2] Marcello La Rosa, Marlon Dumas, Reina Uba, Remco Dijkman. *Business Process Model Merging: An Approach to Business Process Consolidation*. *ACM Transactions on Software Engineering and Methodology*, 2013, vol. 22.
- [3] J.C.A.M. Buijs, B.F. van Dongen, W.M.P. van der Aalst. Mining Configurable Process Models from Collections of Event Logs. *Business Process Management Lecture Notes in Computer Science*, 2013, pp. 33–48.
- [4] W. van der Aalst, *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Springer-Verlag, Berlin, 2011.
- [5] D. Fahland, W.M.P. van der Aalst, *Repairing process models to reflect reality*, *BPM 2012, Lecture Notes in Computer Science*, vol. 7481, Springer, 2012, pp. 229 – 245.
- [6] Matthias W et al. Efficient consistency measurement based on behavioral profiles of process models. *IEEE Transactions on Software Engineering*, 2011, vol. 37, pp. 410-429.
- [7] Sergey Smirnov, Matthias Weidlich. Business Process Model Abstraction based on Synthesis from Well-structured Behavioral Profiles. *International Journal of Cooperative Information Systems*, 2012, vol. 21, pp. 55-83.
- [8] Matthias Weidlich, Artem Polyvyanyy, Nirmal Desai, Jan Mendling. *Process Compliance Measurement based on Behavioural Profiles*. 22nd International Conference on Advanced Information Systems Engineering Hammamet, TUNISIA 2010.
- [9] Xianwen Fang, Junzhi Wu, Xiangwei Liu. An Optimized Method of Business Process Mining Based on the Behavior Profile of Petri Net. *Information Technology Journal*. 2014, vol. 13, pp. 86-93.