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 be 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 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)[6] Let $PM_1$ and $PM_2$ be two Petri net models, $T_1$ and $T_2$ be the set of transitions respectively and $\sim \in T_1 \times T_2$ be the correspondence relation between them. Thus the matched transitions set of $T_1$ is $T'_1 = \{ t_i \in T_1 | \exists t_i \in T_2, t_i \sim t_i \}$, and the matched transitions set of $T_2$ is $T'_2 = \{ t_i \in T_2 | \exists t_i \in T_1, t_i \sim t_i \}$.

Definition 2 (Maximum matched regions) Let $mmr(PM_1)$ and $mmr(PM_2)$ be the part of $PM_1$ and $PM_2$ respectively, $T'_1 \subseteq T_1$ and $T'_2 \subseteq T_2$ 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 $MMRS(PM_1) = \{ mmr(PM_1) | \exists mmr(PM_2), mmr(PM_1) \sim mmr(PM_2) \}$, and the maximum matched regions set of
PM₂ is \( \text{MMRS}(PM₂) = \{ \text{mmr}(PM₂) \mid \exists \text{mmr}(PM_i), \text{mmr}(PM_i) \sim \text{mmr}(PM₂) \} \).

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[7] (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 \( x \succeq y \) and \( y \not\prec x \);
- interleaving order relation \( x \parallel y \), if \( x \succeq y \) and \( y \succeq x \);
- Exclusiveness relation \( x \wedge y \), if \( x \not\succeq y \) and \( y \not\succeq x \).

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

Definition 4[8] (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 \( x \succeq y \) and \( y \not\prec x \);
- interleaving order relation \( x \parallel y \), if \( x \succeq y \) and \( y \succeq x \);

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

Definition 5 (Region behavioral profiles) Let \( \text{mmr}_i, \text{mmr}_j \in \text{MMRS}(PM) \). A pair of the maximum matched regions \((\text{mmr}_i, \text{mmr}_j) \in \text{MMRS}(PM) \times \text{MMRS}(PM)\) is in one of the following three relations:

- strict order relation \( \text{mmr}_i \rightarrow \text{mmr}_j \), if \( \text{mmr}_i \succeq \text{mmr}_j \) and \( \text{mmr}_j \not\succeq \text{mmr}_i \);
- interleaving order relation \( \text{mmr}_i \parallel \text{mmr}_j \), if \( \text{mmr}_i \succeq \text{mmr}_j \) and \( \text{mmr}_j \not\succeq \text{mmr}_i \);
- Exclusiveness relation \( \text{mmr}_i \wedge \text{mmr}_j \), if \( \text{mmr}_i \not\succeq \text{mmr}_j \) and \( \text{mmr}_j \not\succeq \text{mmr}_i \).

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 mmr1 and mmr2 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 \((\text{mmr}_1, \text{mmr}_2)\) is strict order relation, that is \( \text{mmr}_1 \rightarrow \text{mmr}_2 \). In figure 1(b), the pair of maximum matched regions \((\text{mmr}_1, \text{mmr}_2)\) is exclusiveness relation, that is \( \text{mmr}_1 \wedge \text{mmr}_2 \). In figure 1(c), (d), the pair of maximum matched regions \((\text{mmr}_1, \text{mmr}_2)\) is interleaving order relation, that is \( \text{mmr}_1 \parallel \text{mmr}_2 \). 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 PM₁, PM₂, PM₃ and so on are established respectively. These sub-models constitute the model set S(PM)

Step 3: For arbitrary PM₁, PM₂ ∈ S(PM), finding out their maximum matched regions sets MMMR(PM₁) and MMMR(PM₂). Then merging the maximum matched regions according to the region behavioral profiles. The merged model is called as mpm(PM₁, PM₂)

Step 4: Calculating the behavior relation of transition pair (a, b) ∈ Tₛ × Tₛ, where Tₛ = (T₁ − T₁) ∪ (T₂ − T₂) ∪ MMMR(PM₁) ∪ MMMR(PM₂) * MMMR(PM₁)* is the source transitions set and MMMR(PM₂)* is the 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(PM₁, PM₂), The merged model is called as MPM(PM₁, PM₂)

Step 5: Then S(PM) = (S(PM) − {PM₁, PM₂}) ∪ {MPM(PM₁, PM₂)}, 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>
<thead>
<tr>
<th>TABLE</th>
<th>EVENT LOG L₁</th>
<th>EVENT LOG L₂</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ABVCDFIMU</td>
<td>AVBCDFIMOQU</td>
</tr>
<tr>
<td></td>
<td>AVBCDEGJMU</td>
<td>AVBCDFIMOSU</td>
</tr>
<tr>
<td></td>
<td>BVCDE</td>
<td>AVBCDEHKMOQU</td>
</tr>
</tbody>
</table>

Firstly, the log sequence BVCDE is removed because it is incomplete. Then the behavioral profiles of event log L₁ is calculated, and the key behavior relations are listed as follows: A → B, B → V, V → C, C → D, D → E, D → F, E + F, F → G, G → J, F → I, J → M, I → M, M → U. According to the behavioral profiles of L₁, constructing sub-model PM₁ as shown in Figure 2. Similarly, constructing sub-model PM₂ as shown in Figure 2 according to the behavioral profiles of L₂.

![Fig. 2. the process model of L₁ and L₂](image)

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 PM₁ from left to right denoted as mmr₁, mmr₁, mmr₃, mmr₄, mmr₅, and in PM₂, denoted as mmr₆, mmr₇, mmr₈, mmr₉, mmr₁₀, mmr₁₁. Thus we can get the region behavioral profiles as follows: mmr₁ → mmr₂, mmr₁ → mmr₃, mmr₁ → mmr₄, mmr₁ → mmr₅, mmr₂ → mmr₆, mmr₃ → mmr₇, mmr₄ → mmr₈, mmr₅ → mmr₉. The merged model mpm(PM₁, PM₂) 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 $(a, b) \in T \times T$, where $T = \{A, B, V, C, E, M, O, Q, S, U\}$. The result is, $A \rightarrow B \parallel V, B \rightarrow C, V \rightarrow C, C \rightarrow E, E \rightarrow G, E \rightarrow H, G \rightarrow H, G \rightarrow J, H \rightarrow K, J \rightarrow M, K \rightarrow M, O \rightarrow Q, O \rightarrow S, Q \rightarrow S, S \rightarrow U, Q \rightarrow U$. According to the behavioral profiles, inserting the rest transitions that do not included in any maximum matched region into $mpm(PM_1, PM_2)$. 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