

Mining the Model with the Hidden Transitions Based on Petri Net Probabilistic Behavioral Relations

Xiaoyue Wang^{1,a,*}, Xianwen Fang^{1,b} and Ruihao Cao^{1,c}

¹ School of Science, Anhui University of Science and Technology, Huainan 232001, China

^a827000951@qq.com, ^b280060673@qq.com, ^c601766120@qq.com

Keywords: Process mining, Hidden transition, Process model, Probabilistic behavioral relations, Module

Abstract. In process mining, one of the major challenges is to find a process model in the best possible way through the describing observed behavior. The hidden transitions exist in the process models, but not found in the log sequence. This paper is based on the probabilistic behavioral relations from the incomplete log of the business process to mining the model with hidden transitions. First of all, according to the probabilistic behavioral relations obtained the relation of the modules, then calculate the probability of the sub module obtained the relation between the transitions, Based on this, constructing the process model with hidden transitions .Finally, analyzing the default transitions in the log sequence to determine the location of the hidden transitions, the method can effectively and quickly find the location of the hidden transitions.

Introduction

In recent years, with the development of computer, the companies are looking for a way not only to satisfy the requirements of customers but also to leverage cost-saving through improve the business level. Hence puts forward the technology of business process mining, but mining process of business process models will often appear in the process model, but there is no transition in the log sequence. Such transitions are a large number of models which exist in reality. From the event log mining hidden transitions can be restore model, improve the operation efficiency and subsequently achieved high efficiency of production and service.

In the log mining invisible tasks in the beginning is put forward by the W.M.P. van der Aalst et al. He analyzed the cause of invisible tasks. The related concepts of Petri are given in [1]. The related concepts of log trace and event log are given by [2]. The analysis method of judging the consistency between the log and the model is proposed in [3]. Foreign scholars van der Aalst has done a lot of articles on the invisible task mining. [5, 6] tells the method of process mining. In this paper, we mainly mining the hidden transitions in the event log by using the probabilistic behavioral relations. First of all, according to the relation between the modules and then step by step calculate the probabilistic obtained the relation between the transitions. So that, we can get the source model. Finally, analyzing the default transitions in the log sequence to determine the location of the hidden transitions, the method can effectively and quickly find the location of the hidden transitions.

Motivation case

In the introduction, we introduce the importance of mining the model with the hidden transitions from the incomplete logs. In this section will use a simple example to prove the importance of mining hidden transitions from the incomplete logs. Below give a logL= [<A,B,E,F>,<A,B,E>,<C,D,G> , <C,D,E,F>,<A,B,E,F,E,F,E>,<C,D,E,F,E>,<B,E,F> ,<C,G>,<C,E,F,E>](This situation does not consider the number of the log sequence) according to the previous mining calculation method to mining the model as shown in Fig. 1.

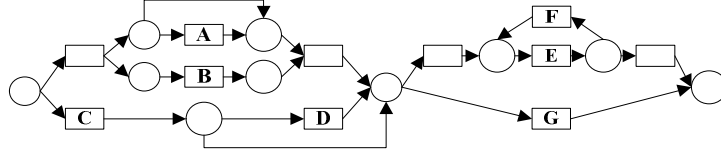


Fig. 1 The previous mining

Fig. 1 show that the fitness of the two models is low. The log sequences generated by these two graphs do not conform to the rule of process mining.

Preliminaries

Definition1^[2](Event Logs 、Traces) Let T is a collection of activities, $\sigma \in T^*$ is an event trace, $L \subseteq T^*$ is an execution log.

Definition2^[3](Fitness and Behavioral Appropriateness)

$$fitness = \frac{1}{2} \left(1 - \frac{\sum_{i=1}^k n_i m_i}{\sum_{i=1}^k n_i c_i} \right) + \frac{1}{2} \left(1 - \frac{\sum_{i=1}^k n_i r_i}{\sum_{i=1}^k n_i p_i} \right) \quad aB = 1 - \frac{\sum_{i=1}^k n_i (x_i - 1)}{(m-1) \sum_{i=1}^k n_i}$$

Note that, k be the number of different traces from the aggregated log. For each log trace $i (1 \leq i \leq k)$ n_i is the number of process instances combined into the current trace. m_i is the number of missing tokens, and p_i is the number of produced tokens during log replay of the current trace. m is the number of labeled tasks, x_i is the mean number of enabled transitions during log replay of the current trace.

Mining the Model with the Hidden Transitions Based on Probabilistic Behavioral Relations

In the introduction of this paper, we present the importance of mining the hidden transitions in the incomplete log.

Definition1(Hidden transition) T' is the transitions of the process model, l is the definition domain of tag function, the transition $t \in T'$ is known as the hidden transition if and only if $t \notin \text{dom}(l)$.

Definition2^[4](A block-structured workflow net) M_E is a block-structured workflow net, it can be divided recursively into workflow nets. An example is shown in Fig. 2.

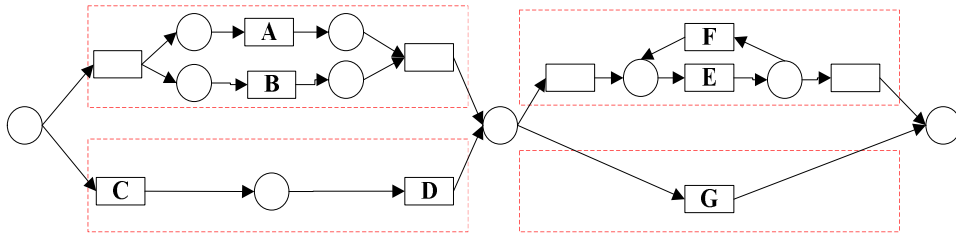


Fig. 2 filled regions denote the block-structured $(\rightarrow (\times (\wedge (A,B), \rightarrow (C,D)), \times (\Theta (E,F), G)))$

Definition3^[4](Process trees) A process tree is an abstract hierarchical representation of a block-structured workflow net. The leaves of the tree are activities, representing transitions. The nodes of the tree, operators, describe how their children are combined. This paper uses four operators: $\oplus \in (\times, \rightarrow, \wedge, \Theta)$ The \times operator describes the exclusive choice between its children, \rightarrow the sequential composition and \wedge the parallel composition. The first child of a Θ tree is the loop body, the non-first children are redo parts.

Definition4^[4](Partitions and Cuts) A partition is a distribution of an activity set Σ into disjoint non-empty subsets $\Sigma_1, \Sigma_2, \dots, \Sigma_n$, with $\Sigma = \Sigma_1 \cup \Sigma_2 \cup \dots \cup \Sigma_n$. A pair of activities (a, b) is partitioned by a partition $\Sigma_1, \Sigma_2, \dots, \Sigma_n$ if

a and b are not both in $n > 1$ the same Σ_i . A cut is a partition together with a process tree operator, for example $(\rightarrow, \{A\}, \{B, C, D, E, F\})$.

Definition 5^[4] (Transitive Closure \rightarrow^+ and Directly-Follows Relation \rightarrow) For two activities a and b , the relation $a \rightarrow^+ b$ if there exists a \rightarrow -path from a to b . If b can follow a directly in $M, \langle \dots, a, b, \dots \rangle \in L(M)$, then $a \rightarrow_M b$.

Definition 6^[4] (accumulated probability for \times, \rightarrow and \wedge) Let $c = (\oplus, \Sigma_1, \Sigma_2)$ be a cut, $\oplus \in \{\times, \rightarrow, \wedge\}$. Then $P_{\oplus}(\Sigma_1, \Sigma_2)$ denotes the accumulated probability of c : $P_{\oplus}(\Sigma_1, \Sigma_2) = \frac{\sum_{a \in \Sigma_1, b \in \Sigma_2} p_{\oplus}(a, b)}{|\Sigma_1| \cdot |\Sigma_2|}$

(P_{\oplus} denotes an artificially estimated probability) $z = \frac{(|a| + |b|)}{2}$ $P_{\oplus}(a, b) = 1 - \frac{(|a| + |b|)}{2}$ ($|a|$ denote the number of occurrences of a and b) Table 1 is shown our proposal for probabilistic activity relations for activities a and b

Table 1 Our proposal for probabilistic activity relations for activities a and b							
	$P_{\times}(a, b)$	$P_{\rightarrow}(a, b)$	$P_{\rightarrow}(b, a)$	$P_{\Theta_i}(a, b)$	$P_{\Theta_s}(a, b)$	$P_{\Theta_s}(b, a)$	$P_{\wedge}(a, b)$
nothing	$1 - \frac{1}{z+1}$	$\frac{1}{6} \cdot \frac{1}{z+1}$	$\frac{1}{6} \cdot \frac{1}{z+1}$	$\frac{1}{6} \cdot \frac{1}{z+1}$	$\frac{1}{6} \cdot \frac{1}{z+1}$	$\frac{1}{6} \cdot \frac{1}{z+1}$	$\frac{1}{6} \cdot \frac{1}{z+1}$
$a \rightarrow^+ b$	0	$1 - \frac{1}{z+1}$	0	$\frac{1}{4} \cdot \frac{1}{z+1}$	$\frac{1}{4} \cdot \frac{1}{z+1}$	$\frac{1}{4} \cdot \frac{1}{z+1}$	$\frac{1}{4} \cdot \frac{1}{z+1}$
$b \rightarrow^+ a$	0	0	$1 - \frac{1}{z+1}$	$\frac{1}{4} \cdot \frac{1}{z+1}$	$\frac{1}{4} \cdot \frac{1}{z+1}$	$\frac{1}{4} \cdot \frac{1}{z+1}$	$\frac{1}{4} \cdot \frac{1}{z+1}$
$a \rightarrow^+ b \wedge b \rightarrow^+ a$	0	0	0	$1 - \frac{1}{z+1}$	$\frac{1}{3} \cdot \frac{1}{z+1}$	$\frac{1}{3} \cdot \frac{1}{z+1}$	$\frac{1}{3} \cdot \frac{1}{z+1}$
$a \rightarrow b$	0	$1 - \frac{1}{z+1}$	0	0	$\frac{1}{2} \cdot \frac{1}{z+1}$	0	$\frac{1}{2} \cdot \frac{1}{z+1}$
$a \rightarrow b \wedge b \rightarrow^+ a$	0	0	0	0	$1 - \frac{1}{z+1}$	0	$\frac{1}{z+1}$
$b \rightarrow a$	0	0	$1 - \frac{1}{z+1}$	0	0	$\frac{1}{2} \cdot \frac{1}{z+1}$	$\frac{1}{2} \cdot \frac{1}{z+1}$
$b \rightarrow a \wedge a \rightarrow^+ b$	0	0	0	0	0	$1 - \frac{1}{z+1}$	$\frac{1}{z+1}$
$a \rightarrow b \wedge b \rightarrow a$	0	0	0	0	0	0	1

Definition 7^[4] (accumulated probability for Θ) Let $c = (\Theta, \Sigma_1, \Sigma_2)$ be a cut, L be a log, S_2 , E_2 be sets of activities. We aggregate over three parts: start of a redo part, end of a redo part and everything else:

$$redo_{start} = \sum_{(a,b) \in End(L) \times S_2} P_{\Theta_s}(a, b) \quad redo_{end} = \sum_{(a,b) \in E_2 \times Start(L)} P_{\Theta_s}(a, b)$$

$$indirect = \sum_{\substack{a \in \Sigma_1, b \in \Sigma_2 \\ (a,b) \notin (End(L) \times S_2) \cup (E_2 \times Start(L))}} P_{\Theta_i}(a, b) \quad P_{\Theta}(\Sigma_1, \Sigma_2, S_2, E_2) = \frac{redo_{start} + redo_{end} + indirect}{|\Sigma_1| \cdot |\Sigma_2|}$$

In the loop, the Directly-Follows Relation denotes by Θ_s , others denote by Θ_i .

Algorithm 1: Using the probability behavioral relations mining hidden transition

Input: Processed execution logs

Output: Petri net model

Step 1: To divide the first step from the record of the log sequence, and basing the Equations calculating the probability relations between the modules ,choosing the biggest.

Step 2: Then calculating the probability behavioral relations between the sub modules, choosing the biggest. So the relations between the transitions, the source model are derived.

Step 3: Comparing the log sequence to find the default location. And thus the corresponding to find the hidden transitions.

Step 4: According to the execution log to calculate the model's fitness.

Step 5: In the case of $f \rightarrow 1$, and then consider the behavioral appropriateness, at this time to find the hidden transition as a normal transition in the calculation, if $f_0 < f_1, aB_0 < aB_1$, then turn to step 6. the contrary to step 4. If $f_0 < f_1, aB_0 > aB_1$ or $f_0 > f_1, aB_0 < aB_1$ they need to be calculated by the formula $Q = 1 - \frac{mf + naB}{m + n}$ (m, n which represents the weight of them). The smaller the variance of the data, the smaller the variance, the smaller the fluctuation, the more stable, the better.

Step 6: All the log replay is completed, M_I is the model of the mining.

Experimental Result

In this section we will use a simple example to demonstrate the feasibility of the above method. Table 2 is the trace and frequency of each log.

Table2 the table of execution log

Execution Log		Execution Log	
Number of Instances	Log Trace	Number of Instances	Log Trace
245	ABEF EFE	1012	ABG
3241	BAEF EFE	268	BAG
220	ABE	284	CDEF EFE
1028	BAE	120	CDE
1250	CDG	224	BG
22	BEFEFE	34	CG
15	BE		CEFEFE
50	CE		

We can get the process trees as the Fig. 3:

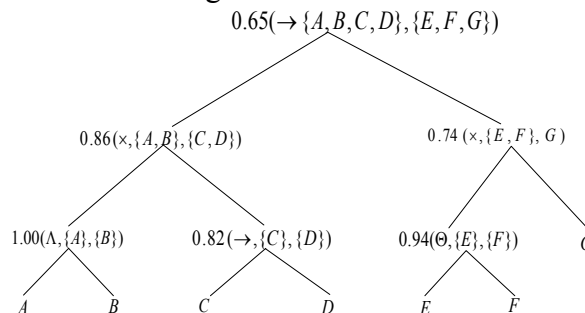


Fig. 3 Process Trees

According to the relationship of A, B, C, D, E, F, G , the source model M_0 is obtained, which is shown in Fig. 4.

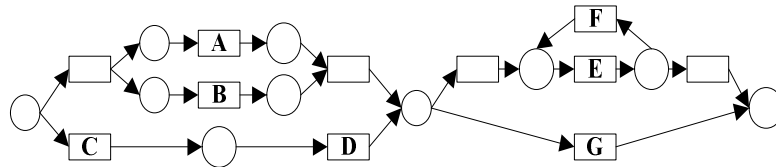


Fig. 4 The Source Model M_0

According to $L=\{BEFEFE, BE, BG, CG, CEF EFE, CE\}$ can be obtained with the hidden transition from the default transitions A and D . Draw Fig. 5, that is the optimized model M_1 .

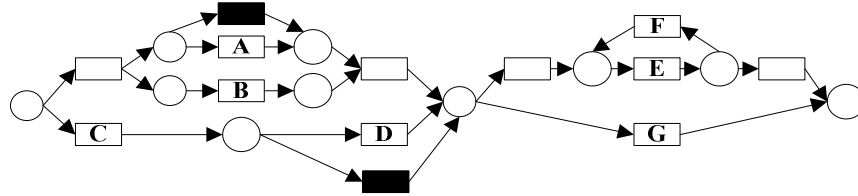


Fig. 5 The Optimized Model M_1

According to the model of the probability behavioral relations ,basing on the equations,we get $f = 1, aB = 0.89$. The completeness of the model is greatly improved.

Conclusions

This paper puts forward a method of mining hidden transitions in the event log based on probability behavioral relations. Firstly, the related concepts of process model and log are introduced. Then the consistency between the log and the model is described. After it, the core part of this paper is to mining the hidden transitions in the event log based on the Petri net probability behavioral relations, because the relationship of the probability is not very sensitive for the hidden transition. According to the relationship between the probability behavioral of the model, the relationship between the modules is obtained, and calculating the relationship between the transitions. At last, analyzing the default transitions in the log sequence to determine the location of the hidden transitions.

In this paper, we analyze the hidden transition in the log, but there are some other abnormal transitions in the actual process, such as block transitions. So in the future work, it should be more abnormal transitions to be analyzed; making process mining technology can be more perfect.

Acknowledgements

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 Soft Science Foundation (12020503031), the Natural Science Foundation of Educational Government of Anhui Province of China (KJ2012A073, KJ2014A067), Anhui Provincial Natural Science Foundation (1508085MF111), the youth academic backbone foundation of AUST, the Academic and Technology Leader Foundation of Anhui Province.

References

- [1] Matthias W, et al. Process Compliance Measurement based on Behavioral Profiles [J]. Computers in industry, 2004, 53(3):321~343
- [2] In Chinese Wen Li-jie.Study on process mining algorithm based on WF net[D].BeiJing: Tsinghua University.2007.
- [3] Rozinat A, van der Aalst W M P. Conformance Testing: Measuring the Fit and Appropriateness of Event Logs and Process Models [J].Computer Science,2005,3812 LNCS:163~176
- [4]Sander J.J. Leemans, Dirk Fahland, and Wil M.P. van der Aalst.Discovering Block-Structured Process Models from Incomplete Event Logs[J].Computer Science,Volume8489,2014,pp91-110

- [5] van der Aalst W M P. Process Mining: Discovery, Conformance and Enhancement of Business Processes[M]. 1st Edition. Berlin: Springer, 2011: 352
- [6] van der Aalst W M P et al. Work-flow mining: Discovering process models from event logs [J]. Knowledge and Data Engineering ,2004, 16(9): 1128~1142