A Method to Identify the Difference between Two Process Models

Min-Hsun Kuo
Department of Industrial Engineering and Management, Yuan Ze University, Taoyuan, Taiwan.
s909510@mail.yzu.edu.tw

Yun-Shiow Chen
Department of Industrial Engineering and Management, Yuan Ze University, Taoyuan, Taiwan.
ieyschen@saturn.yzu.edu.tw

Abstract—**Process mining is getting much attentions and interests in the development of the web-based information technique (IT) field. Process conformance, one of process mining techniques, may be the most common method used to find out the (dis)similar working process(es). The distinguishing processes found are greatly instrumental in making business decisions, such as conduction strategies, service tactics, manufacturing process, …, etc., and further to advance their working efficiency. Several process conformance approaches have been developed and discussed over the past decade, but those discussions involving in the investigation to seek for reasons why the working processes are (un)conformable are rare. To advance the function of the process conformance, this paper introduces the two parameters, Support and Confidence, used for newly defining the tinguishability among various processes; meanwhile, they are also taken to identify the roots resulting to the process distinguishing. "Support" parameter functions as the evaluation of the process similarity based on the working activity sequences (or called "from-to" workflows) and on the relationships among the various processes evaluated; "Confidence" as the measure of the process relationships defined by a ratio of the identical activities within two processes to the total activities of each individual process. Moreover, the two proposed parameters had also been applied to a real case, in which the nursing processes worked in the pediatrics department of a hospital were measured and improved. To the best knowledge of the authors, there does not have an exact technique that, so far, is intact of considering the whole situation where all of the process conformance factors are involved; even the presentation of this paper cannot be avoidable. Nevertheless, this paper truly provides a way to find a certain degree of a lower bound of the process distinguishability through the two proposed conformance parameters.**

Index Terms—***process*** **conformance, process mining, Petri nets, distinguishability, process similarity.**

I. INTRODUCTION

Facing to rapid changes in market environment, such as uncertain demands requiring from various customers, a business may have a trend to perform necessary strategic alliance by sufficiently aggregating its suppliers, partners, market, products and service, thus for making its conduction costs down. For such a complex allied organizations, a web-based IT system is often needed to aid and improve those running business process management (BPM) systems by means of performing certain process mining techniques. Since these techniques are able to define what and how business processes are working, thereby the BPM problems, such as process creation, conformance and oversight can be solved, and the efficiency of the process execution is then advanced [1]. Furthermore, the unique business intelligence (BI) can be gradually enriched, thus to gain the enhancement of the business core competence and the advancement of the global competitiveness.

The factors considered in designing BPM processes are pretty complicate and related with one another. They include cost, detention, quality, security, maintenance, and so on [2]. In the past, designing processes was a laborious job. It not only consumed a vast amount of time and resources, but also its quality designed and the relative avails strongly depended on the process developers' experience. To train a novice to become an experienced BPM designer quite needs extra time as well.

By contrast, nowadays, a BPM model can be built by the support of a web-based IT system consisted of the knowledge and experience of process designers. A planning process model can be newly formed, or an existing one can be improved, through mining the repository of the processes in the form of the empirical event (or activity) logs recorded in a web-based BPM system. Thereby, those time and costs spent on the business decisions making decrease with the performance increments. This taking advantage of preserving experienced knowledge related with the business running behavior into a knowledge base, and further being repeatedly used, is essential to a contemporary business for its sustainable running [2]. Van der Aalst et al. [3][4][5], who are the initiators for developing process mining techniques, described that process mining must enable a businesses to understand its real running situations via the provisions of the information mined, since the event

doi:10.4304/jcp.7.4.998-1005

behaviors taken place within those running processes and the expected processes have been distinguished each other.

Even a process mining is one of the best techniques used to deal with the BPM problems, but the mined quality always accounts on whether the parameters used in the mining processes are well defined. No matter with redesigning processes, detecting the exceptions (distinctions) in the processes, or making up the oversight of the process execution, those parameters should be capable of finding out the interesting working activity behavior (i.e., workflows) by means of conforming the event logs of various processes. Hereby, one matter in a process conformance has to be mentioned: what roots causing the process distinguishability interest a business process miner is much more than how much the distinguishability does. The reasons why this interest happens are as follows.

For a business, its new processes may neither be always stable, nor follow certain expected processing paths. When a working process is changed with the occurrence of new running factors, it may issue in the oversight of process variances. The supervision of the process variances is foremost for BPM; or, the business processes may fail because the variances will cause to the process's running away from the expected (normal) one. From the viewpoints of the previous BPM works, to build up a process model either by modifying the current ones, or by newly developing a process from scratch, measuring the conformance between the running and expected process models is definitely necessary for the accomplishment of the BPM improvement. Several methods proposed to perform the process conformance will be reviewed in the following section. Before that literature review, the structure of this paper is briefly described below.

The destinations of this paper are not only to present the parameters used for defining the distinguishability between the process activities, but also to describe the key factors resulting to the process dissimilarity. The rest of this paper is organized as follows. In Section 2, the works related to the destinations of this paper are reviewed. Section 3 presents the preliminary of the proposed process conformance parameters; subsequently, based on these defined parameters, the complete algorithm for performing the process conformance will be presented in Section 4. To explain the proposed algorithm, a manual numerical example is illustrated in Section 5. For demonstrating the power of the proposed algorithm, Section 6 studies a real case about the nursing tasks process done in a hospital. Finally, Section 7 concludes the findings and weaknesses out of this paper's investigation.

II. Literature Review

A well-defined process model should clearly involve what its working activity behavior is, how its various process behavior destinations are achieved, and what the treatment is taken to handle possible encounter problems during its executing. Wang et al. [6] presented that improving BPM processes can be a source of increasing competitive advantage, especially if the product development time, manufacturing quality and/or productivity management are given by the considering constraints; furthermore, they also pointed out that some possible encounter problems, such as activity interactions, have to be taken into account in mining process. If the oversight of the activity interactions happens, it may cause the design of the critical path of working activities (events) in a BPM process to be more difficult.

Based on Grammar inference, Cook and Wolf [7] developed an algorithm in which a similarity was defined and used to compare process models formed by a series of symbols. Their paper concluded that the process equivalence notions always result in a binary answer, equivalent or not, which can be helpful for a process manager to his modifying the process.

Additionally, Van der Aalst et al. [8] also presented their own similarity used to identify the conformance between process models based on the typical activity working behaviors. They concluded the similarity level plays an important role in determining how high or low is for the discrimination of the process models. Besides, they also found that each segment of a process model should have its own characteristic that may not be used to do the conformance checking. The reason for this infeasible checking is that not all of activity traces can be logged with the same series of the frequency of the event occurrences.

Zha el al.[9] proposed that many BPM tasks, such as process retrieval, process mining, and process integration, need firstly to determine the similarity between two processes. Their idea of the process similarity was inspired from Van der Aalst et al. [5]. They proposed two parameters to define the process similarity: the "reference similarity" based on process activity sequences and the "relationship similarity of the activity transition adjacency" based on the level of the event relations.

Kleiner [10] applied Delta analysis to evaluate the process conformance. The analysis utilized the event logs to perform the process similarity/difference. The results of his Delta analysis showed that it is more accuracy and less cost than those of the survey methods like interview and observation. However, Van der Aalst et al. [8] criticized that the Delta analysis has two drawbacks, one is that the event logs used to build up the process model may not stand for the real processes, and the other one the Delta analysis cannot precisely define the similarity between processes. Therefore, they proposed an alternative method, the conformance testing, to overcome the drawbacks of the Delta analysis.

Bae et al. [11][12][13] defined both working activity precedence graph and its corresponding workflow block structure as the two key mechanisms operating processes conformance. The block tree has binary structure that holds an inferable logic using to perform the yes/no process conformance. They proposed an algorithm to inspect the process similarity in the block structures for the achievement of the process conformance.

To sum up the literature review, almost previous discussions on the process conformance have focused on using workflow matrices or event traces to evaluate the difference between process models; yet, this type of process conformance is only able to identify whether the process activity sequences are equal or unequal with each other. Although this degree of the process difference or indifference may inform a process manager certain message about process modification, he more desires to understand what roots causing to the dissimilar processes. Although Wang et al. [14] and Yang et al. [15] have suggested that using the relationships between activity flows (sequences) can establish up the structure used to look into the causes contributed to the process differences, so far the works on introducing such a process relationships into the process conformance are still rare.

The goal of this paper is exactly to pave a way for advancing the efficiency of the previously proposed process conformance methods. This paper not only defines the similarity between processes, but also involves the investigation of the factors causing the process differences. By doing so, the proposed "complementary" conformance inspection procedure will provide BPM managers useful information to improve their business processes so that the business running effectiveness increases.

III. Preliminary

The mathematic preliminary needed in the process conformance proposed in this paper is based on Petri net theory. A number of literature published by Van der Aalst et al. [3][4][5] and others elsewhere have detailed the theory.

A Petri net is formed by a kind of the semantic description that can be used to describe the processes in diagramming. Such a graph-based process can be intuitively modeled and visualized by people, but it is difficult to be recognized by an IT system. How to store a process into a data repository, for example, is the first difficulty; the stored processes can be subsequently recognized by computer is another one. Others are like the process extraction, comparison, modification, and so forth.

In this section, both a Transitions-Places (T-P) table and a Places-Transitions (P-T) table are introduced to represent the arrowed (i.e., from-to) directions of diagramming a Petri net. The T-P table records a connection arrowed from transitions to places, and the P-T table shows a reversed connection, namely from places to transitions. Following up these two tables, the relations between process activity sequences can be readable and the structure of the process model is inferable.

Fig.1 pictures five kinds of the elemental activity relationship, i.e., one kind of activity sequence, two kinds of concurrence (AND-split and AND-join) and two kinds of alternative (OR-split and OR-join). The T-P table and the P-T table of these five relationships are organized in Table I and Table II, respectively. In principle, if an arrowed (i.e., from-to) activity relation is existed in a process model, the entry located at the intersection of the both kinds of the from-to tables is filled with 1; otherwise, zero. These entry numbers 0 or 1 can be taken to compute the relationships among the comparing processes.

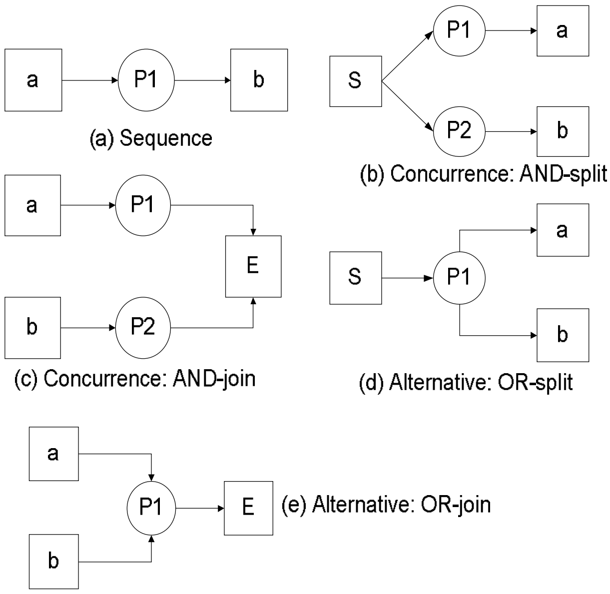


Figure 1. Five elemental activity relations in the form of Petri net

TABLE I
T-P DIRECTION FOR THE FIVE ACTIVITY RELATIONS

From\To	(a)		(b)		(c)		(d)		(e)	
	P_1	P_2	P_1	P_2	P_1	P_2	P_1	P_2	P_1	P_2
X	1	0	1	1	1	0	1	0	1	0
Y	0	0	0	0	0	0	0	1	1	0
Z	0	0	0	0	0	0	0	0	0	0

TABLE II
P-T DIRECTION FOR THE FIVE ACTIVITY RELATIONS

	From\To	X	Y	Z
(a)	P_1	1	0	0
	P_2	0	0	0
(b)	P_1	0	1	0
	P_2	0	0	1
(c)	P_1	0	1	1
	P_2	0	0	0
(d)	P_1	0	0	1
	P_2	0	0	1
(e)	P_1	0	0	1
	P_2	0	0	0

To compute the graphical relations in Petri net, there still needs five basic symbolized operators to respectively stand for the corresponding activity relations depicted in Fig. 1.

- Sequence: $a \to b$
- Concurrence: $S \to [a][b]$ for AND-split,
  $[a][b] \to E$ for AND-joint
- Alternative: $S \to [a,b]$ for OR-split,
  $[a,b] \to E$ for OR-joint.

A pair of brackets [ ] indicate a place. $S \to [a][b]$ means a connection of activity S is divided into two subsequent concurrent connections towards activity a and b through the individual place P1 and P2, respectively. $[a][b] \to E$ means two activities a and b are joined towards activity E through their respective two places P1 and P2 in parallel. $S \to [a,b]$ represents activity S separates its connection at one place P1 to respectively arrow towards either activity a or b. $[a,b] \to E$ represents connecting individual activity a with b together forwards to activity E via one intersection place P1.

IV. Proposed Method

As aforementioned, using sufficient event log traces of those activity sequences (flows) saved in the process repository can perform the process mining tasks, but the result mined may be inefficient if the mining procedure is indiscreet. Inspired by Bae et al. [11][12], since their binary answer (conformed or inconsistent) to the process comparison cannot provide a satisfactory solution for finding out the reasons why the comparing processes are different, this paper considers how to acquire the roots causing the process difference together with a difference level. By this kind of the complementary inspection, the capability of distinguishing two processes can be more informative, and the improvement of the processes thus can be adequately done well.

In real applications, according to the real numerical experiments experienced from the nursing process of the hospital studied in the latter section, the process conformance and improvement will be a hard work if the way to watch the distinguishing situations between the processes is unknown. In this section, the idea of the process relationships described in the previous section is detailed to make up the weakness of the Bae's approach.

In the proposed approach, two parameters, Support and Confidence, are used to complementarily measure the distinguishability expressed by the similarity and the relationships between two process models. "Support" is used to quantify the relational resemblances between the activities of two process models; "Confidence" computes the ratios of the identical sequential activities existing in the two comparing models with respective to the total activities of an individual model. Based on these two parameters, more information of the relationships between process models can be traced, and the distinguishability can be more clearly explained.

Since the size of the T-P table and the P-T table are dominated by both the numbers of activities (or so-called transition in Petri nets) and the "from-to" places in a process model, different sizes of T-P tables (or P-T tables) for different process models may cause obstacle to engaging the process model conformance. To overcome this kind of the problem, extracting the from-to relations appeared in T-P and P-T tables, rather than simply comparing the frequency of the same individual activity of the two processes, to quantify the similarity and the relationships among the various process activity sequences (flows) will be performed. The proposed algorithm with iterative heuristic for process conformance is distinguished with the following four steps:

Step 1. Tabulating both T-P and P-T relations with the "from-to" flows of each process activity on the basis of Petri net diagrammed from the process event log. The entry values of the T-P and the P-T tables are defined by the numbers of the both sequential activity occurrences and places diagrammed in a process model. The criteria used to finish the tabulation of the from-to relations of T-P and P-T was described in the previous section.

Step 2. Modifying those two kinds of tables established in Step 1 for refining the relations between process activities. The from-to relations refinement is treated until the activity direction logic meets the requirements of operations of Petri net. Based on the Petri-net technical operations, because a sequence-typed net is like a concurrence one (two activity sequences in parallel), the mathematical operations of these three elemental types are similar. Besides, a sequence net also can be regarded as a partial sequence of an alternative net, therefore, the alternative net can be adjusted to become the sequence one. This adjustment can make the computation of the various process relationships be more unified and time saving.

Step 3. Building a new table in which all relationships between the comparing process models are summarized. The table summarizes the number of the existing relationships of each process by two mathematical operating steps, i.e., Modeling and Comparison. In the Modeling step, the numbers of those process relationships occurred in the comparing processes are written in the last row of the summarized table (representing by "Count," in Table V shown in the next section). The Comparison step, on the other hand, achieves the establishment of the arrowed relationships of process models and the computation of discrepancy between the two processes. This discrepancy, or so-called difference, is indicated with 0 or 1. A slash in the summarized table (e.g., Table V) is used to discriminate the meaning of the relationships and the process identity (ID). If the corresponding relationship can be found with none of the comparing process models, the right side of the slash is with 0; otherwise, with 1 if the same relationships can be matched in both process models. The left side of the slash represents the ID of a process model if the process relationship computed is existed in the same ID model. When any of distinguishing "from-to" activity flows between the two process models occurs, the code of model ID will be 1; otherwise, 0.

Step 4. Computing the two distinguishability parameters (i.e., Support and Confidence) for determining the level of the process model conformance. Equations (1) and (2) are used to compute these two parameters.

$$Support(a,b) = \frac{A_{a\cap b}}{A_{a\cup b}} \times \frac{N_{a\cap b}}{N_{a\cup b}} \qquad (1)$$

$$Confidence(b|a) = \frac{N_{a\cap b}}{N_a} \qquad (2)$$

Where

$Support(a,b)$: similarity of process models *a* and *b*.

$Confidence(b|a)$: ratio of the similar (identical) relationships of processes *a* and *b* with respect to process a.

$A_{a\cap b}$: number of the total similar (identical) activities of processes *a* and *b*.

$A_{a\cup b}$: number of the activities of processes *a* and *b*.

$N_{a\cap b}$: number of the total similar (identical) relationships of processes *a* and *b*.

$N_{a\cup b}$: number of the relationships of processes *a* and *b*.

N_a : number of the relationships of process *a*.

Equation (1) is to compute the numbers of sequential activities and relationships. The computation is executed firstly by selecting an expecting process model as a normalized one, and then using it to conform whether the other comparing processes meet its normalized process requirements.

Equation (2) is used to evaluate the percentages of the similar (identical) activity (sub)sequences between two process models. This percentage is treated as an assistance of the Support similarity, so as to understand how the resemblance between two processes is. The usage of the frequencies of sequential activity occurrences is to conform whether they are consistent to the activity sequences of the resulting Petri net.

V. An Illustration Example

In this section, an example, shown in Fig. 2, is illustrated for the demonstration of the proposed algorithm used to perform the conformation of the activity flowing requirements existing in the four process models. In the figure, Model (I) is considered as a normal process used to draw a parallel among the four process models. Their corresponding T-P and P-T tables are transformed and tabulated in Table III. In this table, the relations between the activity sequences of the four process models are the data input to the proposed algorithm. The six tables, from Table III to Table VIII, respectively show the result of the four algorithmic operation steps depicted in the previous section.

In Table V, the relationships that cause the difference between two process models can be observed. For example, the relationships between Models (I) and (Ⅱ), namely, the activity relations $A \to B$, $B \to D$, $A \to [B][C]$ and $[B][C] \to D$ show the two models in difference. The first two relations are observed in Model (I), and the last two are just as those in Model (II). As to Models (I) and (III), $A \to C$ and $C \to D$ are two sequential activity relations appeared only in Model (III). Therefore, it is reasonable to regard Model (I) as a sub-model of Model (III).

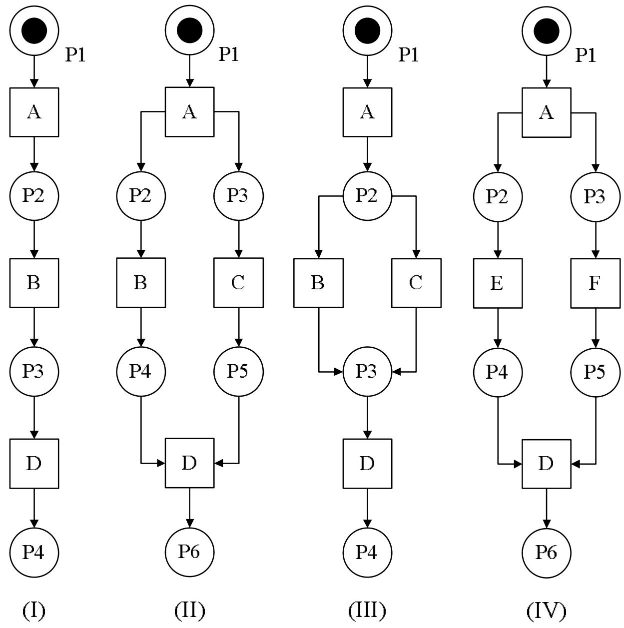


Figure 2. Illustrated process models for demonstrating the proposed process conformance algorithm

TABLE III
(STEP 1) RELATIONSHIPS OF SEQUENTIAL CONSEQUENCES OF THE FOUR PROCESS MODELS

Process Model	Model (I)	Model (Ⅱ)	Model (Ⅲ)	Model (Ⅵ)
Relationship	$Start \to A$	$Start \to A$	$Start \to A$	$Start \to A$
	$A \to B$	$A \to [B][C]$	$A \to [BC]$	$A \to [E][F]$
	$B \to D$	$[B][C] \to D$	$[BC] \to D$	$[E][F] \to D$
	$D \to End$	$D \to End$	$D \to End$	$D \to End$

TABLE IV
(STEP 2) RESULT OF REFINING RELATIONSHIPS OF THE ACTIVITY SEQUENCES IN TABLE III

Process Model	Model (I)	Model (II)	Model (III)	Model (VI)
Relationship	$Start \to A$	$Start \to A$	$Start \to A$	$Start \to A$
	$A \to B$	$A \to [B][C]$	$A \to B$	$A \to [E][F]$
	$B \to D$	$[B][C] \to D$	$A \to C$	$[E][F] \to D$
	$D \to End$	$D \to End$	$B \to D$	$D \to End$
			$C \to D$	
			$D \to End$	

TABLE V
(STEP 3-1) RESULT OF THE COMPARING PROCESS MODELS BASED ON THE NORMALIZED MODEL (I)

Relationship\ Model	Modeling Function				Comparison		
	Model(I)	Model(II)	Model(III)	Model(VI)	(I)(II)	(I)(III)	(I)(VI)
$Start \to A$	1	1	1	1	0/1	0/1	0/1
$A \to B$	1	0	1	0	I/1	0/1	I/1
$B \to D$	1	0	1	0	I/1	0/1	I/1
$A \to [B][C]$	0	1	0	0	II/1	0/0	0/0
$[B][C] \to D$	0	1	0	0	II/1	0/0	0/0
$A \to C$	0	0	1	0	0/0	III/1	0/0
$C \to D$	0	0	1	0	0/0	III/1	0/0
$A \to [E][F]$	0	0	0	1	0/0	0/0	VI/1
$[E][F] \to D$	0	0	0	1	0/0	0/0	VI/1
$D \to End$	1	1	1	1	0/1	0/1	0/1
Count	4	4	6	4	2/6	4/6	2/6

TABLE VI
(STEP 3-2) RESULT OF THE DIRECTED RELATIONSHIPS OBSERVED FROM THE FOUR MODELS

Activity\model	Model(I)	Model(II)	Model(III)	Model(VI)
A	1	1	1	1
B	1	1	1	0
C	0	1	1	0
D	1	1	1	1
E	0	0	0	1
F	0	0	0	1
Count	3	4	4	4

From Table VII, it can be observed that Model (I) is closer to Model (III) than to other models. Besides, from reviewing the rows of Table VIII, Model (I) can be thought of as the one completely contained by Model (III), because two third of the sequential activity occurrences of Model (III) are the same as those of Model (I). Moreover, Model (I) and Model (VI) are higher dissimilar each other. From Table VII or Table VIII, it observes that Model (I) and Model (III) are apparently similar each other.

TABLE VII
(STEP 4-1) SIMILARITY BETWEEN THE FOUR PROCESS MODELS ON THE BASIS OF MODEL (I)

Support (Equation I)		Comparing models			
		I	II	III	VI
Normal Model	I	-	1/4	1/2	2/15

TABLE VIII
(STEP 4-2) PERCENTAGE OF SIMILARITY ACTIVITY OF THE FOUR PROCESS MODELS

Confidence (Equation I)		Conforming models			
		I	II	III	VI
Relative Models	I	-	1/2	1	1/2
	II	1/2	-	-	-
	III	2/3	-	-	-
	VI	1/2	-	-	-

VI. REAL CASE STUDY

There two process models (a) and (b) are pictured in Fig. 3. Both of them had been mined from the event logs of nursing process proceeded at the pediatrics department of the studied hospital.

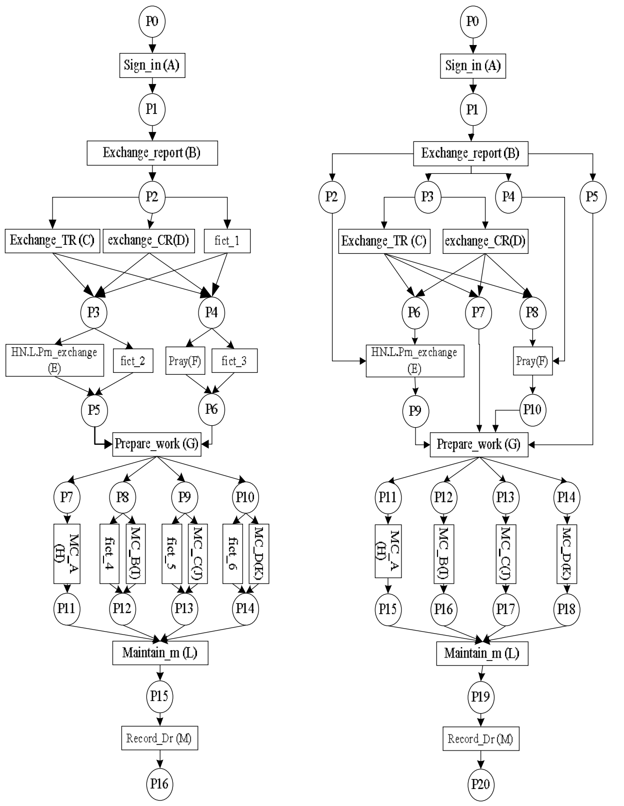
Table IX shows the compared result and the darken rows are the factors causing the difference between the nursing process models of the studied hospital. Their two process conformance parameters proposed in this paper are computed as follows.

$$Support\ (a,b) = \frac{13}{19} \times \frac{7}{43} \times 100\% \cong 11.14\% \tag{3}$$

$$Confidence\ \ (b|a) = \frac{7}{39} \times 100\% \cong 17.95\% \tag{4}$$

$$Confidence\ \ (a|b) = \frac{7}{11} \times 100\% \cong 63.64\% \tag{5}$$

From Table IX, it can be seen that Model (a) has much more connections and activities than Model (b). This information implies that those connections have higher percentage of the factors that cause the two studying process models to be different; besides, they all are Petri-net activity relations involving the alternative and concurrence types defined in Fig.1. In the Table, for the four activity relations $[C] \to [E][F]$, $[C] \to [fict\_2][F]$, $[C] \to [E][fict\_3]$ and $[C] \to [fict\_2][fict\_3]$, two activities should be selected out of the four activity candidates, E, F, fict_2 and fict_3, to perform the activity sequence conformances after activity C. On the other hand, the arrowed connections acquired only from Model (b) involve the relations among activities, B, C, D, E, F and G, and moreover an additional "non-free choice" structure [16] is inserted into them.



Model (a): The process model mined by mined the proposal method

Model (b): The process model by α/α++-algorithm

Figure 3. The nursing process models of a pediatrics department of the hospital studied

TABLE IX
RESULT OF THE COMPARING PROCESS MODELS

Relationship\Model	a	b	comparison	Relationship\Model	a	b	comparison
$Start \rightarrow A$	1	1	0/1	$[fict\_2][F] \rightarrow G$	1	0	A/1
$A \rightarrow B$	1	1	0/1	$[E][fict\_3] \rightarrow G$	1	0	A/1
$B \rightarrow [C]$	1	0	A/1	$[fict\_2][fict\_3] \rightarrow G$	1	0	A/1
$B \rightarrow [D]$	1	0	A/1	$G \rightarrow [H][I][J][K]$	1	1	0/1
$B \rightarrow [fict\_1]$	1	0	A/1	$G \rightarrow [H][fict\_4][J][K]$	1	0	A/1
$B \rightarrow [E][C][F][G]$	0	1	B/1	$G \rightarrow [H][I][fict\_5][K]$	1	0	A/1
$B \rightarrow [E][D][F][G]$	0	1	B/1	$G \rightarrow [H][I][J][fict\_6]$	1	0	A/1
$[C] \rightarrow [E][F]$	1	0	A/1	$G \rightarrow [H][fict\_4][fict\_5][K]$	1	0	A/1
$[C] \rightarrow [fict\_2][F]$	1	0	A/1	$G \rightarrow [H][fict\_4][J][fict\_6]$	1	0	A/1
$[C] \rightarrow [E][fict\_3]$	1	0	A/1	$G \rightarrow [H][I][fict\_5][fict\_6]$	1	0	A/1
$[C] \rightarrow [fict\_2][fict\_3]$	1	0	A/1	$G \rightarrow [H][fict\_4][fict\_5][fict\_6]$	1	0	A/1
$[C] \rightarrow [E][F][G]$	0	1	B/1	$[H][I][J][K] \rightarrow L$	1	1	0/1
$[D] \rightarrow [E][F][G]$	0	1	B/1	$[H][fict\_4][J][K] \rightarrow L$	1	0	A/1
$[D] \rightarrow [E][F]$	1	0	A/1	$[H][I][fict\_5][K] \rightarrow L$	1	0	A/1
$[D] \rightarrow [fict\_2][F]$	1	0	A/1	$[H][I][J][fict\_6] \rightarrow L$	1	0	A/1
$[D] \rightarrow [E][fict\_3]$	1	0	A/1	$[H][fict\_4][fict\_5][K] \rightarrow L$	1	0	A/1
$[D] \rightarrow [fict\_2][fict\_3]$	1	0	A/1	$[H][fict\_4][J][fict\_6] \rightarrow L$	1	0	A/1
$[fict\_1] \rightarrow [E][F]$	1	0	A/1	$[H][I][fict\_5][fict\_6] \rightarrow L$	1	0	A/1
$[fict\_1] \rightarrow [fict\_2][F]$	1	0	A/1	$[H][fict\_4][fict\_5][fict\_6] \rightarrow L$	1	0	A/1
$[fict\_1] \rightarrow [E][fict\_3]$	1	0	A/1	$L \rightarrow M$	1	1	0/1
$[fict\_1] \rightarrow [fict\_2][fict\_3]$	1	0	A/1	$M \rightarrow End$	1	1	0/1
$[E][F] \rightarrow G$	1	1	0/1	**Count**	**39**	**11**	**7/43**

To sum up, based on the numerical analysis to the two proposed process conformance parameters, as in (1) and (2), the conformance degree of these two real process models is 11.14%, and the ratio of identical activity sequence to Model (b) is much higher than that to Model (a). It is reasonable to infer that to adjust Model (a) by modifying Model (b) will be easier than the reversed modification, i.e., adjusting Model (b) according to Model (a) is harder.

VII. Discussion And Future Work

In process mining, process model conformance is an important topic and widely applied to various business decisions making. In this paper, the way to overcome the drawbacks occurred in the conventional Delta analysis was inspired by the ideas of Bae et al. [11][12]. Their proposing two terms, i.e., the connection of activities and the relations between activities, were considered as the basis of the improved process conformance examination. By this inspired consideration, the improved solution proposed in this paper can clarify the activities that cause to the discrepancy between two process models, and at the same time provides useful process information on the processes maintenance with the guidelines found from the roots causing the process differences.

There many methods have worked to discuss the issue of process model conformance on different bases such as the behaviors in processes or the process structure. Generally, handling model conformance from the process behavior should be more accurate; however, it could not be made sometimes. As shown in Fig. 4 below, for example, when the dual activities occur in the structure of a process model, although carrying on the process model conformance based on process structure can deal with the process working behaviors, it cannot provide the information about the roots causing the difference. Oppositely, this paper provided a solution to such the process distinguishability with the investigation of the roots influencing on the process working behaviors, and this solution exactly is its contribution.

Nevertheless, this paper still cannot involve the consideration of dynamically identifying the distinguishing behaviors with time, and neither performing the process difference evaluation between the multi-dimensional event logs occurred in a process and their sophisticate relationships with other processes. Much more efforts used to handle these problems should be paid in the future.

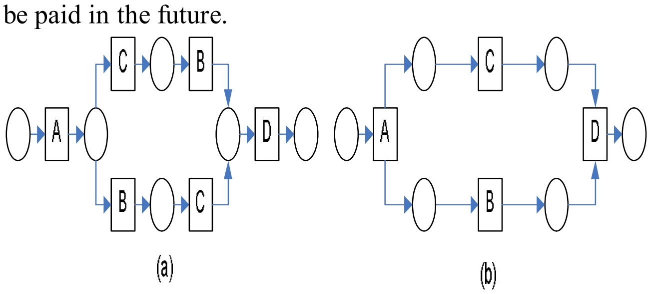


Fig. 4. Example of process models with dual activities.

REFERENCES

[1] Clarence Ellis, Karim Keddara, and Grzegorz Rozenberg, "Dynamic change within workflow systems," ACM Proceedings of Conference on Organizational Computing Systems, pp.10-21, 1995.

[2] Etienne Cocquebert, Damien Trentesaux, and Christian Tahon, "WISDOM: A website design method based on reusing design and software solutions," Information and Software Technology, vol.52, pp.1272-1285, 2010.

[3] W.M.P. van Aalst, , and A. J. M. M. Weijters, "Process mining: a research agenda, Computers in Industry," vol.53, pp.231-244, 2004.

[4] W.M.P. van Aalst, B.F. van Dongen, J. Herbst, L. Maruster, G. Schimm, and A. J. M. M. Weijters,"Workflow mining: A survey of issues and approaches," Data & Knowledge Engineering, vol.47, pp.237-267, 2003.

[5] W.M.P. van der Aalst, H.A. Reijers, A.J.M.M. Weijters, B.F. Van Dongen, A.K. Alves de Medeiros, M. Song, and H.M.W. Verbeek, "Business process mining: an industrial application," http://is.tm.tue.nl/staff/wvdaalst/publications/z7.pdf.

[6] Juite(Ray) Wang, David W. He, and Chang-Xue Feng, "A structured approach for analysis of design processes," IEEE Transactions on computers, packaging, and manufacturing technology-part A, vol.18, no.3, 1995.

[7] Jonathan E. Cook, and Alexander L. Wolf, "Event-based detection of concurrency," ACM SIGSOFT Software Engineering Notes , Proceedings of the 6th ACM SIGSOFT international symposium on Foundations of software engineering SIGSOFT '98/FSE-6, vol.23 , no.6, pp.35 – 45, 1998.

[8] W.M.P. van der Aalst, , A.K. Alves de Medeiros, and A.J.M.M. Weijters, "Process equivalence: comparing two process models based on observed behavior," in:S. Dustdar, J.L.Fiadeiro, and A. Sheth (Eds.): BPM 2006, LNCS 4102, Springer-Verlag Berlin Heidelberg, pp.129-144, 2006.

[9] Haiping Zha, Jianmin Wang, Lijie Wen, Chaokun Wang, and Jiaguang Sun, "A workflow net similarity measure based on transition adjacency relations, Computers in Industry, vol.61, pp.463-471, 2010.

[10] Nikolaus Kleiner, "Delta analysis with workflow logs: aligning business process prescriptions and their reality," Requirements Eng. 10: pp.212-222, 2005.

[11] Joonoo Bae, James Caverlee, Ling Liu, and Hua Yan, "Process mining by measuring process block similarity," in:J. Eder, S. Dustdar (Eds.): BPM 2006 Workshops, LNCS 4103, Springer-Verlag Berlin Heidelberg, pp.141-152, 2006.

[12] Joonoo Bae, Ling Liu, James Caverlee, and William B. Rouse, "Process mining, discovery, and integration using distance measures," IEEE International Conference on Web Services (ICWS'06) , pp.479-488, 2006.

[13] Joonsoo Bae, Hyerim Bae, Suk-Ho Kang, and Yeongho Kim, "Automatic control of workflow processes using ECA Rules," IEEE Transactions on knowledge and data engineering, vol.16, no.8, pp.1010-1023, 2004.

[14] Juite(Ray) Wang, David W. He, and Chang-Xue Feng, "A structured approach for analysis of design processes," IEEE Transactions on computers, packaging, and manufacturing technology-part A, vol.18, no.3, 1995.

[15] Rui Yang, Panos Kalnis, and Anthony K.H. Tung, "Similarity evaluation on tree-structrued data," SIGMOD ACM, pp.14-16, 2005.

[16] W.M.P. van der Aalst, "Business alignment: Using process as a tool for Delta analysis and conformance testing," Requirements Eng 10: pp.198-211, 2005.

Min-Hsun Kao a Ph.D. student studying industry engineering at the Yuan Ze university, Taiwan. During her research studies, she participated in several research projects in the areas of process modeling and data mining based on statistical methods. Currently she is working at the department for an instructor to teach engineering statistics and an assistant of statistical quality and experiments laboratory of the department.

Yun-Shiow Chen a full professor and the dean of industrial engineering department of Yuan Ze university Taiwan. She got her Ph.D. degree from the Iowa State University in USA. She is interested in the statistical learning methods, experiment design and quality control. She directs the laboratory of statistical quality and experiments laboratory, is leading a team of research project called Using Statistical Methods to Process Mining.