Indian Journal of Science and Technology, *Vol 8(S8), 620–626, April 2015*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645
DOI: 10.17485/ijst/2015/v8iS8/71707

Process Mining for Contextual Situations Adaptation using Goal-Heuristic Algorithm

Sungho Sim[1] and Sujin Baek[2*]

[1]Department of Liberal Education, Semyung University, Korea; shshim@semyung.ac.kr
[2]Division of Information and Communication, Baekseok University, Korea; croso@bu.ac.kr

Abstract

For surviving in their complex and dynamic environments, organizations should be adaptable and interoperable with these environments. However, there is a limit for distinguishing the range of occurrence and the level of severity in regard to the new external problems occurring in a complicated environment. In addition, they lack system servitization for dynamic business model processing and system integration. Therefore, a study upon the handling of service requirements by users, which could occur dynamically, needs a new concept of process mining to redesign a process to fit adaptation measure after analyzing the scope of problem recognition to change. The existing studies of process mining have been aimed to extract the information of redesigning business process at the level of simple field of database; examine dependency relation; and create a complete process model. Therefore, they lack in ways to analyze and evaluate business process of new requirements for external situations and solve the problems of dynamic reconfiguration. In this respect, the present study is aimed to analyze goal scenario so as to provide the service requirements of new scenarios that occurs dynamically and propose the process mining method that uses goal heuristic algorithm to analyze and evaluate process variable information. When a new requirement comes out, business process changes by goal on the basis of context information and searches a fit process model from existing goal scenarios and new goal scenario by goal heuristic algorithm. And then the similarity, importance, association of the process models are evaluated and analyzed for the purpose. The information analyzed in this way is capable of finding out problems resulting from change in requirements when necessary, determining the needs of adaptation and reconfiguring to keep providing services.

Keywords: Contextual Situation Adaptation, Process Mining, Reconfiguration, Security Policy Models

1. Introduction

Studies regarding process management system have currently been conducted on the flexibility of dynamic business model processing, system integration and situational business model processing[1,2]. However, the existing studies propose only partial solutions to new requirements and still lack in context awareness of the flexibility of business process and system integration. In addition, they lack the system servitization for dynamic business model processing and system integration[3,4]. To solve the problem, therefore, the concept of process mining that evaluates currently running business process and reconfigures previous business process or helps create a new business process on the basis of the information acquired through user-defined business process.

Existing studies on process mining include the research of Agrawal et al.[5] that proposes the business modeling process based on logs of workflow management system and that of Cook and Wolf[6] that analyzes process data by using neural network pure algorithm and Markovian method and improves process model in the field of software engineering science. In addition, Aalst et al.[7] carried out a study to re-discover the process model that considers the process of selective execution by using α-algorithm and Alast et al.[8] and Medeiros et al.[9] used genetic algorithm to develop and find the business model that considers the processes of both selective execution

**Author for correspondence*

and simultaneous execution. However, it was hard for them to extract process when there were no previously structured process models because they already analyzed on the basis of structured process models. Furthermore, such process mining methods are costly and time-consuming because they use the ignition rules of Petri-net in to evaluating the adaptation of derived solutions.

The existing studies of external situation evaluation for dynamic business process processing do not support abstraction technology necessary for problem recognition but define specific problem contexts by using mainly the conditions of performance environment resources. Furthermore, because it is difficult to expand the domain of evaluable problem with them, it is impossible to evaluate unpredicted situations that take place during execution. The existing studies that support dynamic changes also have several limitations, so that they are much limited to alternations. Moreover, as it is difficult for them to explore solutions precisely, they can hardly explore the flow relation between works of all tasks[10]. Therefore, those models have difficulty extracting information of new requirements unpredicted during designing process and dynamically reconfigure it.

In the analysis of external situations in the present study, new requirements are analyzed on the basis of goal scenario. The similarity, importance, association are measured for the analysis of goal process model and process variable information on the basis of existing process model information. In addition, the present uses goal heuristic algorithm to find a suitable model and analyzes changes. And it provides the expressions of the structure and reconfiguration behaviors through goal process modeling and changed information on the basis of goal scenario for new requirements.

2. Existing Process Mining

To optimize or reconfigure business process according to new requirements, the capability of workflow is essential to dynamically adjust to market change or when an exceptional situation brings out. Therefore, some companies consider as competitive edge how quickly process can be changed, react to external situations every changing and deal with the change.

Most of process management systems are those that execute in run-time by using business process defined in Build-time. But when a system cannot execute anymore due to the discrepancy or disagreement of information in two stages, it needs to provide flexibility to tackle such a matter. In addition, it is required to have dynamic function to change the attributes of running process and the activities involved in the concerned processing. On one hand, it finds all the possible paths by analyzing the business process designed in Build-time. On the other hand, it finds the information of actually executed path by using instance information. The two kinds of information obtained in this way are compared and analyzed to find useful information for a new execution path. Process mining is aimed to find useful information from work processing records that occur in business process and the results can be used to renovate the business process of a company[11]. Currently, the studies on process mining are mainly focused on searching structured process models and some pay attention to analyzing the results of fundamental performance or something from organizational perspective.

Agrawal et al.[5] proposed the modeling method of business process on the basis of logs of workflow management system. Cook and Wolf[6] analyzed process data by using neural network pure algorithm and Markovian method and improves process model in the field of software engineering science. In the meantime, Aalst et al.[7] carried out a study to re-discover the process model that considers the process of selective execution by using α-algorithm and Alast et al.[8] and Medeiros et al.[9] used genetic algorithm to develop and find the business model that considers the processes of both selective execution and simultaneous execution. However, it was hard for them to extract process without structured process models because initial solutions for causal matrix had already been created, based on previously structured process models. Furthermore, such process mining methods are cost and time-consuming because they use the ignition rules of Petri-net to evaluate the adaptation of derived solutions. In addition, as the volume of data increase, the effectiveness of algorithm decreases and it is hard to search solutions precisely. As a result, it was hard to explore detailed relation of process flow of all works and between works[12]. For the reason, such mining models have difficulty extracting data suitable for new requirements for dynamic reconfiguration.

3. Analysis using Goal-Heuristic Algorithm

3.1 Searching for the Similar Candidate Process Model

Analogous candidate model analysis does not analyze the whole of existing processes on the basis of goal process model but instead searches a candidate model, depending on the existence of activity through analogous analysis. It analyzes and changes only the necessary information of the candidate model. Analogous analysis computes by using vector space model similarity to find a similar model to goal process model in existing process models.

However, it has two restrictions in finding an appropriate analogous candidate model from existing system. First, all the traces of process logs should have a clear beginning and ending point. Second, the name of a task in process should not duplicated or overlapped but unique. If the two restrictions are lifted, correct process mining can be secured through the proposed analytic method.

Here, N is the set of activity; S is the process model for goal requirements; a is the activity that comprises goal process; and P is the set of all possible process models (S1 …Sn). And the model of goal requirements should be $S \in P$ and activity, $Aj \in N$, and $Ai \neq Aj$.

GA determines the existence of activity on goal requirements process and computes similarity. When activity is searched from existing process, it computes it as 1, otherwise it as 0.

$$GA : S_i = \begin{cases} 1 & \left(when\ a_i\ \ from\ S\ is\ detected\ within\ S_i\right) \\ 0 & \left(otherwise\right) \end{cases}$$

$$Sim\left(S, S_i\right) = \sum_{i=1}^{n} \left(GA : S_i\right) \tag{1}$$

The number of goal-based process activity is NS. Equation (3) calculates the similarity of activity (ASD) with the values of existing process activity obtained from Equation (2).

$$\text{ASD} = \frac{Sim\left(S, S_i\right)}{NS} \tag{2}$$

The values of ASD are used to find a fit process. Higher value means prior in order. In case that it is '0', it means no proper model suitable for goal requirements is searched out from existing process. It means a complete disagreement, so a new process is created.

Analogous model models are searched in the high order of ASD. And then weight values are assigned and activity relation is analyzed accordingly.

Although ASD is high, it does not satisfy the condition of being a fit process model. Therefore, contextual relation before and after all the activities of processes should be understood in finding a more suitable process model by using order matrix.

With the analogous candidate process model Figure 1, the similarity, importance and association of a process model should be analyzed. Such process models are composed only with completely analogous models that include the activities of goal process model and they are used in analyzing following relations. (S1 …Sn) means similar process models and the number of instance that is executed along with them was presented. S1~S3 in Figure 1 are the results from searching analogous candidate process models. When it is supposed that the total number of similar process models is n, 1/n is the weight of each process instances executed by them.

3.2 Activity Relational Analysis through the Order Matrix

The order matrix of similar process models comes by in order of importance and association on the basis of the similarities of the process models for goal requirements. The information of goal scenario modeling is expressed in order matrix and stored in tree shape and searched out for analysis.

First of all, order matrix is made to find the activities of similar process model with reference to data information archive where the information of process model and detailed activities are saved.

When order matrix is made, it is expressed with 4 types of control relations (0, 1, +, –). Here, N is the set of activities.

- The relation between Ai and Aj has value '1' if Ai comes before Aj.

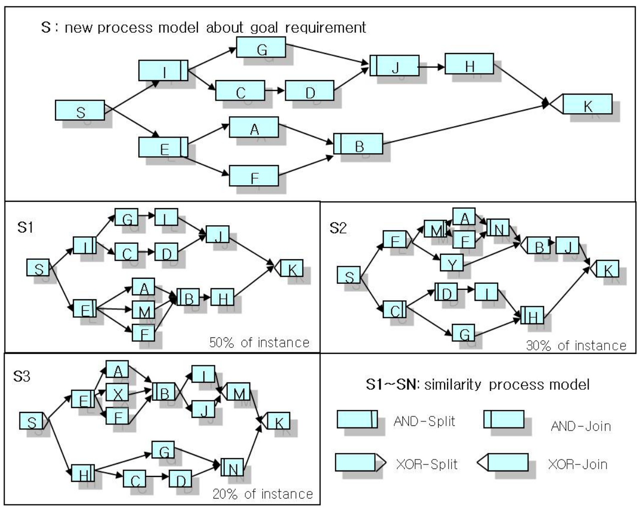

Figure 1. Similar candidate process model

- The relation between Ai and Aj has value '0' if Ai comes after Aj.
- The relation between Ai and Aj has value '*' when at least one trace of Ai exists before and after Aj and is included in another parallel (AND-block).
- The relation between Ai and Aj has value '–'when Ai does not have any trace before and after Aj and is included in another parallel (XOR-block).

In this way, order matrix is made on the process model for goal requirements by using those 4 types of control relations.

4 types of control relations of order matrix are used to analyze the relations with the activities between goal process model and similar process model. Importance (AID) of all the analogous candidate process models is measured. When 4 types of control relations of order matrix have the same values for the activities of goal process model and similar process model process, it is considered as important activities and expressed in the value of importance (AID). When they are used in the entire analogous candidate process models, importance (AID) is computed into '1'. This importance (AID) is very useful in finding important activities that are not yet found in the process of producing goal process model during the analysis of external situation. In addition, it is also used in the process of improving goal process model while heuristically searching analogous candidate process model.

Next, association (ARD) is measured to know if analogous candidate process model and goal process model, which are chosen on the basis of importance (AID) information, are structurally agreed structurally. It is to find if each activity value of candidate process models is same as the value of order matrix of goal process model.

$$\mathrm{ARD}\left(S_c\right)=\frac{\sum_{i=1,j=1}^{m}\sum_{j\neq 1}^{m}\left(f\left(V_{a_ia_j},V_{a_ia_j}^{c}\right)\right)}{m\times\left(m-1\right)}\times AID_{a_ia_j} \qquad (3)$$

"AID" means the importance of the activities and V is order matrix. Sc is the candidate process model and Vc is order matix of Sc. m = |Nc | means the number of the activities in SC. The activities are shown as ai, aj ∈ Nc, i ≠ j. The value is ARD (SC)∈[0,1].

If the activities of goal scenario should be all included and association (ARD) with existing processes is more than 50%, it should be adjusted and then adopted. If it is less than 50%, it should be reconfigured. Figure 2 shows the results.

Figure 2 shows the examples of association analysis with analogous candidate process model using order matrix. Association (ARD) in the example was computed with order matrix, focusing on the similar process model extracted on the basis of similarity. S1 with the highest similarity has 84% of association with goal process model is evaluated as the most similar model in terms of structural order. S2 has 68% of association and S3 has 56% of association. Therefore, S2 and S3 are evaluated to be models with lower similarity in structural order than S1.

3.3 The Heuristic Search Algorithm in Accordance with the Goal

Goal process model for new requirements first calculates and extracts similarity (ASD), importance (AID) and association (ARD) of activity in extracting a similar process model. To find a suitable similar process model on the basis of such information, meta-heuristic concept and search algorithm as suggested in this study is used see Figure 3(a).

In Figure 3(a), S is a model for goal requirements and S' is the fittest similar process model. S_B can be obtained through a single alternation work, showing other paths (sibling) of the model. S_K means child and it is searched out in a tree structure. All the adjacent process models analyzed from S on the basis of similarity are searched. If Si in a candidate model is found to be a better model, goal process model is changed to S'. AS includes all the

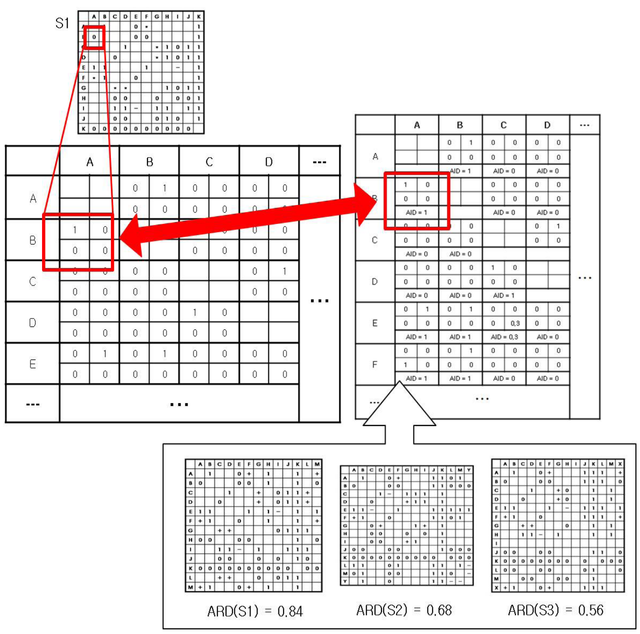

Figure 2. ARD by the order matrix.

changeable activities and comes out in one Si, at least. When aj is changed in Sc, $AS = \bigcup_{i=1}^{n} N_i$ each activity a_i $\in AS$ is determined to correspond with S_ K^{i}, the fittest in Sc. Here, the change of aj was considered for insertion, depletion and moving. Therefore, if S_Kj has better value than Sc, it will be included as AS of S', or aj will be deleted from AS. These processes repeat until a better model and candidate process model are searched out by phase and S' corresponds as Si reference model that is found in the last phase.

(a)

```
BEGIN
   AS = U_{i=1}^{n} N_i
   t=1
   while |AS|>0 and t<x
    do
      S_B = S
      Define a_k as the selected activity;
      foreach a_i ∈ AS
         do
            S_K = FindASD(S_i);
              if FindAID(S_K) > FindAID(S) then
                 if FindAID(S_K) > FindAID(S_B) then
                     S_B = S_K;
                     a_k = a_i;
                 end
              else
                 AS = AS / {a_i} ;
              end-if
       end
       if FindAID(S_B) > FindAID(S) then
           S= S_B;
           AS = AS / {a_k};
       else
           break;
       end-if
      t = t+1;
    end
end
```

(b)

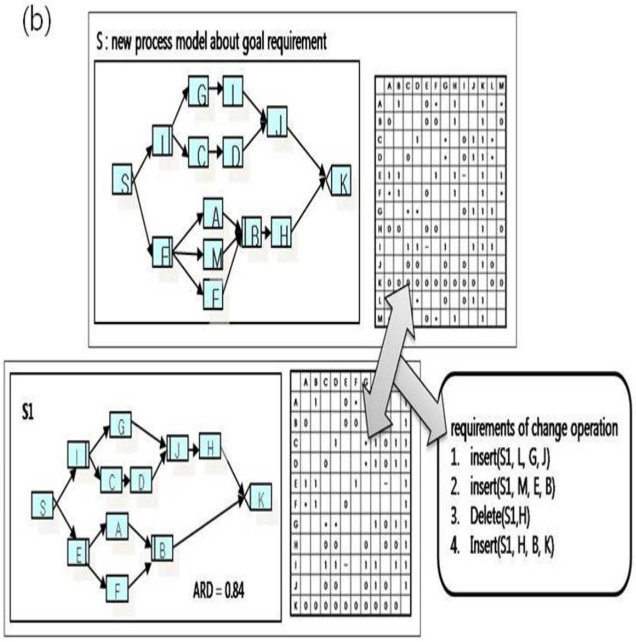

Figure 3. (a) Heuristic search algorithm in accordance with the goal. (b) ARD Analysis of Process Model Variants by the order matrix.

Figure 3(b) uses the correlation order matrix to show satisfactory results for the process-evaluating index through the comparison with the current goal process. It shows the necessary works for the solution of problems. Such works can be extracted through the order matrix. The process with a high level of similarity requires less works for adjustment compared to the one with a low level of similarity.

4. Evaluation

Changes in each similar process model should be known to change a suitable goal model through goal heuristic algorithm. Therefore, we compared existing process model and goal process model by using their similarity, importance and association information and could find work necessary to solve a problem.

Figure 4 demonstrates that s1, which is the most similar model to goal scenario-based process model, has the least number of works to change. It is also shown that the requirements according to the selection of process model to change with are clearly found because the similarity of order relations is found through association. In this case, the order of change can be different by user.

In Figure 5, it shows the necessary partial works for the solution of problems. As a S1 example, consider the graph shown in Figure 5 and assume that a new task, M shall be inserted between the AND-split E and its corresponding AND-join B. Note that it is possible to add M as a new branch between E and B, although the successors of the AND-split E, the nodes A and F, have already been completed respectively started.

Furthermore, by looking at the WF graph from Figure 5(a), a new task M may not be inserted between the nodes E and B. In order to insert a new task between E and F, first of all, the execution of F should be aborted by the user.

Therefore, it is possible to make a change for the demands related to an unexpected situation by using the current process. Also, it is possible to identify the works which need to be changed according to the order of priorities. Such a result is sent to the monitoring process in process in order to provide the necessary information for constant services.

5. Conclusion

The present study proposed a process mining method that analyzes business model on the basis goal scenario so that new requirements for external situations can be consistently

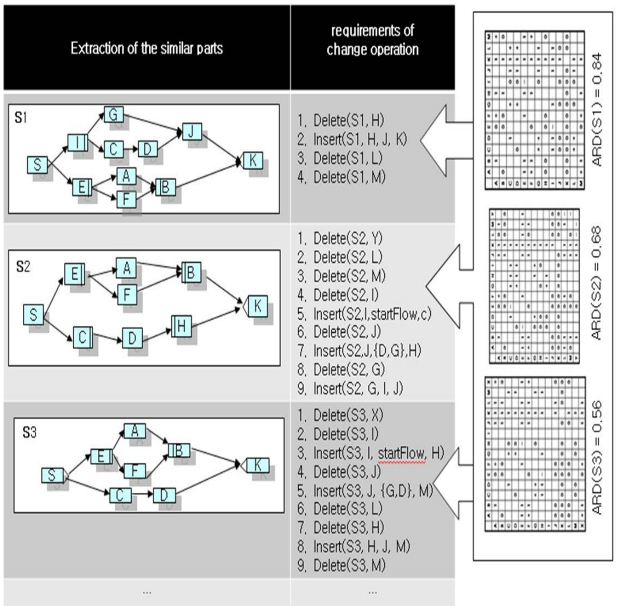

Figure 4. Analysis assessment for the extraction of the similar parts and the goal process model.

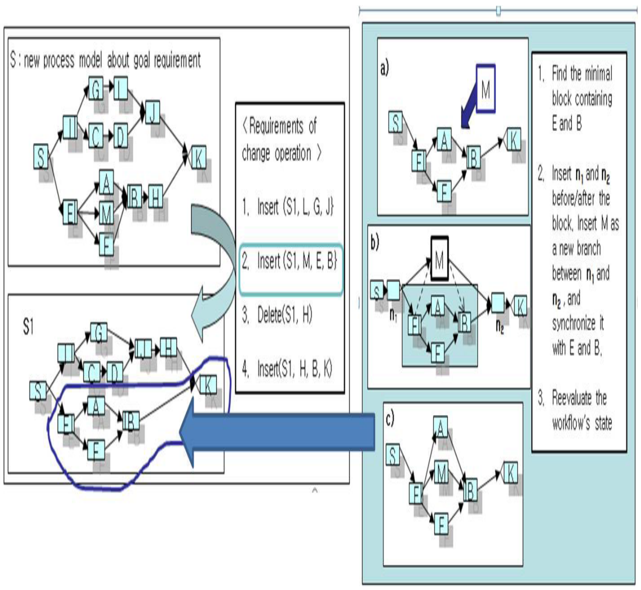

Figure 5. Adding a new task M between the AND-split E and its corresponding AND-join B in a part of S1.

provided and uses goal heuristic algorithm to analyze and evaluate process variable information. Analysis evaluation by goal does not analyze entire existing process models but only necessary information by using similarity (ASD) for the existence of activity. In addition, the proposed model uses 4 types of order matrix to measure and analyze the importance (AID) of activity and also measures association (ARD) between candidate process models. Based on the results, it extracts a fit candidate process model and searches suitable similar process models. Similar process models are searched to analyze process variable information by goal-heuristic algorithm and order matrix for changes.

Using such information can help quickly understand possible problems accompanied with change when necessary and forecast adaptation, so user can apply a quick and efficient decision on unpredicted situation to an early scenario. In addition, it can help service go on in other fields.

As a following study, the author place in consideration extraction of change effect and calculation of optimal scope for requirements for change. Thinking that this study has insufficiency in ways to verify extracted requirements through the proposed evaluation, it proposes a study to improve this. Therefore, the following study will be aimed for a strategy for adaptation measure on the basis of the goal scenario-based process evaluation information proposed in this study and a way to dynamically reconfigure business process model in a structure fit to changed requirements. And it is expected that such studies keep carrying out to provide users with a variety of dynamic services.

6. References

1. Baek SJ, Song YJ. Process evaluation model based on goal-scenario for business activity. International Journal of Maritime Information and Communication Sciences. 2011; 9(4):379–84.
2. Jaroucheh Z, Liu X, Smith S. Apto: a MDD-based generic framework for context-aware deeply adaptive service-based processes. ICWS, the IEEE 8th international conference on web services. IEEE Computer Society. 2010; 219–26.
3. Jehad Sarkar A, Lee Y-K, Lee S. A smoothed Naive Bayesbased classifier for activity recognition. IETE Tech Rev. 2010; 27(2):107–19.
4. Jaroucheh Z, Liu X, Smith S. Recognize contextual situation in pervasive environments using process mining techniques. Journal of Ambient intelligence and Humanized Computing. 2011; 2(1):53–69.
5. Agrawal R, Gunopulos D, Leymann F. Mining process models from work-flow logs. In 6th International Conference on Extending Database Technology. 1998; 469–83.
6. Cook JE, Wolf AL. Discovering models of software processes from event based data. ACM Transactions on Software Engineering and Methodology. 1998; 7(3):215–49.
7. van der Aalst WMP, Weijters AJMM, Maruster L. Workflow mining: discovering process models from event logs. IEEE Transcations on Knowledge and Data Engineering. 2004; 16(9) :1128–42.

8. van der Aalst WMP, de Medeiros AKA, Weijters AJMM. Genetic process mining. Lecture notes in computer science. 2005; 3536:48–69.
9. de Medeiros AKA, Weijters AJMM, van der Aalst WMP. Genetic process mining: a basic approach and its challenges. Lecture Notes In Computer Science. 2006; 3812:203–15.
10. Muller R, Greiner U, Rahm E. AGENTWORK : a workflow system supporting rule-based workflow adaptation. Data and Knowledge Engineering. 2004.
11. van der Aalst WMP, Reijers HA, Weijters AJMM, van Dongen BF, Alves de Medeiros AK, Song MS, Verbeek HMW. Business process mining: An industrial application. Information system. 2007; 32(5):713–32.
12. Chung SY, Kwon ST. A process mining using association rule and sequence pattern. Journal of the Society of Korea Industrial and Systems Engineering. 2008 June; 31(2):104–11.