Decision Mining for Multi Choice Workflow Patterns

Riyanarto Sarno[1], Putu Linda Indita Sari[2], Hari Ginardi[3], Dwi Sunaryono[4], Imam Mukhlash[5]
Informatics Department and Mathematics Department
Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[1]riyanarto@if.its.ac.id, [2]linda.indita10@mhs.if.its.ac.id, [3]hari@its.ac.id, [4]dwi@its-sby.edu, [5]imamm@matematika.its.ac.id

***Abstract*— Decision mining is combination of process mining and machine learning technique to retrieve information about how an attribute in a business process affects a case's route choice. It identifies decision point by looking for XOR-splits in petri-net workflow model and analyzing rules for each choice based on available attributes using decision tree. Problem emerges when decision mining technique is used on a workflow that does not use either XOR or AND splits, for example OR-split gateway logic. OR-split does not have explicit representation in petri nets and it makes decision mining algorithm cannot find its decision point. Workflow pattern that uses OR-split as its splitting logic is multi choice. Multi choice does not have its own explicit representation in form of petri net and it is problematic to apply decision mining to this workflow pattern. To make multi choice can be analyzed by decision miner some modification needs to be applied to the petri net representation of this pattern. This paper proposes modification of OR-split gateway representation in petri net. The new representation of OR-split uses combination the existing XOR-split and AND-split to make the model easier to be analyzed using decision miner. The proposed modification do not affect the conformance of event log and process model, but will allow each choice branch to be checked by decision miner.**

Keywords*— *decision mining, decision tree, multi choice, process mining, workflow pattern

I. Introduction

Decision mining is a technique to retrieve rules for choices in a business process. It combines process mining and machine learning technique to analyze a Process Aware Information System's event log and discover how the attributes in the process affects is being routed (Rozinat and van der Aalst 2006). The decision mining algorithm has been applied as plugin in ProM framework as a plugin named Decision Point Analysis. It will first discover decision points in a process model by identifying its XOR-splits. Only XOR-splits, place in Petri-net which has more than one outgoing arcs, will be identified as decision point (Subramaniam et al. 2007). Other split type, the AND-split, is a split which indicates all choices are taken. Therefore it cannot be identified as decision point.

Decision mining depends in process model, since it needs to identify the decision point from XOR-splits presented in model. The model itself must be presented in the form of petri-net. That is why, decision mining very related to petri-net model of a business process, especially its XOR-split to determine decision points for further analysis. But, in cases where its XOR-splits already found in Petri-net, decision mining can be a powerful analysis tool to identify rules of choices. Petri-net is the most common net used to model any workflow in process mining. In some literatures, petri-net's conditional routing was build using exclusive OR logic. That means in each case only one of all branches is executed. In petri-net, exclusive OR-split modeled using one place and more than one outgoing arcs from it.

Multi choice is a type of workflow which has ability to represent parallel branches which can be executed based on the input condition or attributes of the workflow (Wohed et al. 2005). It uses OR-split gateway logic to represent the split property. This pattern is different from petri net's exclusive OR choice, where only one of the choices can be executed. The fact that multi choice workflow pattern is different from XOR-splits in petri-net makes a process with multi choice pattern quite hard to be discovered. The OR-split gateway does not have explicit representation in petri net. Some available algorithms discover multi choice as AND-split, while others discover it as XOR-split. If multi choice model is discovered by common discovery algorithms as XOR-split, it can be analyzed using decision miner. But the analysis result will be invalid. It is due to different logical property between XOR-split and multi choice. XOR-splits only allow exactly one branch to be executed, while multi choice can allow from one to all branch of choices to be executed in a case.

This paper will propose two methods to represent multi choice workflow pattern in petri-net. The new representation will not change the property of business process. The petri net will be modified to let decision miner finding decision point for each branch of decision. Thus each branch of a decision will be represented by one decision point. Each branch of decision will be analyzed using available data. Then, rules for each choice can be found. This paper will first give review of already available methods and algorithms related to this work. Then new proposed representation of multi-choice will be given as well as an example case. Lastly, this paper will give experimental implementation of the proposed model for multi choice pattern.

II. Related Work

A. Decision Mining

Decision mining is a method to retrieve how the value of attributes in a business process can affect routing of a case. Decision mining uses combination of process mining to deal with Process Aware Information System's event log and machine learning to discover the rules. Decision mining method used in this paper is the same as presented by Rozinat and van der Aalst (2007). Decision mining generally consists

of two phases. First, discovery of decision points from process model. And second, construction of decision tree.

We will need event log or event log with petri-net model in decision mining. Log used in decision miner needs to contain attributes, i.e. variables with value. These values needed for construction of decision tree later. If the log does not come with petri net model, any appropriate discovery algorithm from process mining can discover the process model. The model which is used for this analysis must be in petri net. So, algorithms which results in other workflow net must be converted into petri net before processing.

In decision mining, the decision points are discovered by identifying conditional routing in petri net model. Therefore proper model needs to be found before doing decision mining. If the model does not show conditional routing as wanted, the result of decision mining might be unsatisfying. After all decision points identified, decision tree can be built for each decision point. Decision tree will be built using values of attributes in event log presented before.

B. Petri Net

Petri-net is the common model used in representing a process in process mining. Joins and splits in Petri-net consists of two kinds, AND-type and exclusive OR-type. AND-split and AND-join used in parallel routing. Parallel routing means that all branches will be executed. Exclusive OR split and join are used in conditional routing. Conditional routing means only one branch of choices will be executed, while the others will not be passed. Exclusive OR sometimes called XOR, the other literature also call it OR-split. But, logically, the properties of these branching are the XOR gateway logic, where only one of all choices will be executed. XOR-split modeled as more than one outgoing arc from one place, while XOR-join modeled as place with more than one incoming arc. Fig.1 shows split and join in petri net.

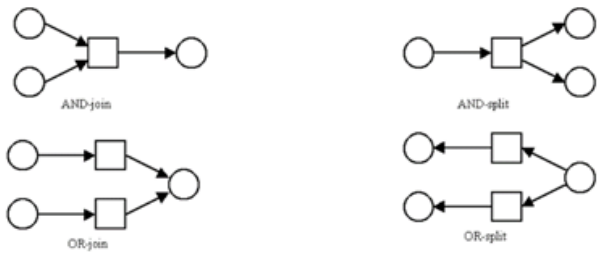

Fig. 1. Type of split and join in petri net

C. Decision Tree

Machine learning is a method to make a machine can identify provided data records and learn the rules to be used in future conditions. One of machine learning method is decision tree. Decision tree will identify the most affecting variables for classification of data records. The most affecting variable will become root of the tree, and continued by other variables as nodes until reaching leaves where the classification concluded. Each variable on decision tree's nodes are chosen based on information gain. The variable with largest gain will be chosen. Information gain can be computed using formula below.

$$Gain(S, A) = Entropy(S) - \sum_{v \epsilon Values(A)} \frac{|S_v|}{|S|} Entropy\ S_v \quad (1)$$

$$S_v = \{ s \epsilon S \mid A(s) = v \} \quad (2)$$

$$Entropy(S) = -p_{\oplus} log_2 p_{\oplus} - p_{\ominus} log_2 p_{\ominus} \quad (3)$$

Where S_v is the subset of S, with attribute A having value of v. Entropy of data partition computed by giving cost relative to size of S. In entropy formula, $p_{\oplus}$ is proportion of positive samples from S, while $p_{\ominus}$ is proportion of negative samples from S. In Decision Mining decision tree are built based on values of attributes in previous log. Rules for each decision point is learned trough these values. We can extract information about the logical rule for choices in each decision point based on the tree's form.

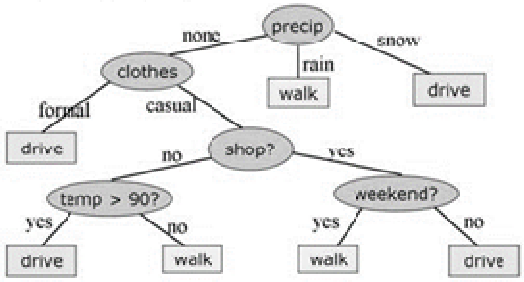

Fig. 2. Example of decision tree

D. Drawbacks of Decision Mining

Although decision mining has combines both process mining and machine learning, it still has some drawbacks. Some drawbacks of decision mining are stated by Leoni and van der Aalst (2013). Decision mining can only partially handle process models with invisible activity and duplicate task. Decision mining can only identify choices with invisible task and duplicate task if there are other visible transition after the invisible or duplicate task. Other drawbacks also come from the model. Decision mining is very dependent to process models. Some workflow patterns might cannot be handled using decision mining because it does not have explicit XOR-split representation in petri-net, thus the model resulted from discovery might not be fit. Wrong model can result in wrong analysis. Because the first step of decision mining is to identify decision points from petri-net model. When the petri-net cannot give the right model to identify decision points, further analysis result might be unsatisfying.

III. Choice Workflow Pattern

Decision mining needs to identify decision points before processing into construction of decision tree. Identification of decision points very related to process model. Inappropriate model to represent a business process will also results in wrong identification of decision point, which can yields to unsatisfying result.

There are many types of workflow that can be identified in real life processes. Russel et al. (2006) already compiled workflow patterns that has been grouped based on their flow similarity. Table I shows some workflow patterns. Since

decision mining only focused on decision points, the workflow patterns in table only show the choice patterns.

TABLE I. CHOICE WORKFLOW PATTERNS

NO	Workflow Pattern	Split Gateway Type	Description
1	Exclusive Choice	XOR-split	Branching of a process model into two or more branches such that it can only select one of the outgoing branches.
2	Multi Choice	OR-split (different from exclusive OR)	Branching of process model into two or more branches, where there are one or more branch of choice which is enabled based on the condition of the incoming attributes.
3	Deferred Choice	XOR-split	Point in a process where one branch of choice is chosen based on the interaction with the environment. In other word, there is a kind of race among the branches.

In Table I above, several workflow patterns are listed. There are three kinds of choice listed on the table, they are exclusive choice, multi choice and deferred choice. Exclusive choice is a decision point in a workflow where exactly one branch is chosen (Wohed et al. 2005). In this workflow pattern, exactly one branch is chosen at a case. The pattern can be expressed using XOR-split gateway in BPMN. In form of petri net, this pattern can easily be expressed by a place with more than one outgoing arc.

Multi choice is a workflow pattern which has the ability to select more than one branch based on the condition of the workflow (value of attributes) (Wohed et al. 2005). This pattern can be expressed in BPMN using OR-gateway. The modeler can make one condition as the default choice or no default choice for this workflow. It can also be modeled without a gateway, but the condition expressions are not exclusive. Lastly, it can be expressed using complex gateway which shows set of branches that will be chosen.

The other type of choice workflow pattern is deferred choice. In this workflow, one of the branches should be chosen. The decision on which branch will be activated is deferred to the latest moment based on which branch satisfies the workflow first (Wohed et al. 2005).

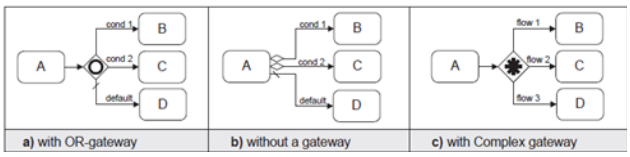

Fig. 3. BPMN expression for multi-choice

Only one pattern cannot be represented in petri-net using either XOR-split or AND-split, the multi-choice pattern. It needs an OR gateway to represent it. But, Petri-nets only have representation AND gateway and XOR gateway, even though XOR gateway sometimes called OR. Multi-choice pattern let one or all branch of choices executed during a case. Meanwhile AND-split will execute all branches and XOR-split will execute only one branch.

IV. Multi Choice Difficulty in Petri Net

Multi choice needs to be represented using OR gateway. While mining a log from multi choice workflow, some available discovery algorithm still cannot identify the correct representation of the workflow. Some algorithms identify the model as AND-split, while some other identify the model as XOR-split.

Result of multi choice process model discovery using several available algorithms is shown in Fig.4. Both alpha and alpha++ discover the log as AND-split. The AND-split actually means that both activity B and C are executed. In the event log, only two cases represent this property, case 3 and case 4. While other cases only execute B or only execute C. Heuristic algorithm discovers the log as XOR-split. It actually means that only B or C is executed during a case. The genetic algorithm is the only algorithm that gives fitness value 1.0 when conformance checking is done. But, the resulted process model contains too much invisible task. This model cannot be checked using decision mining since it cannot yet handle invisible task.

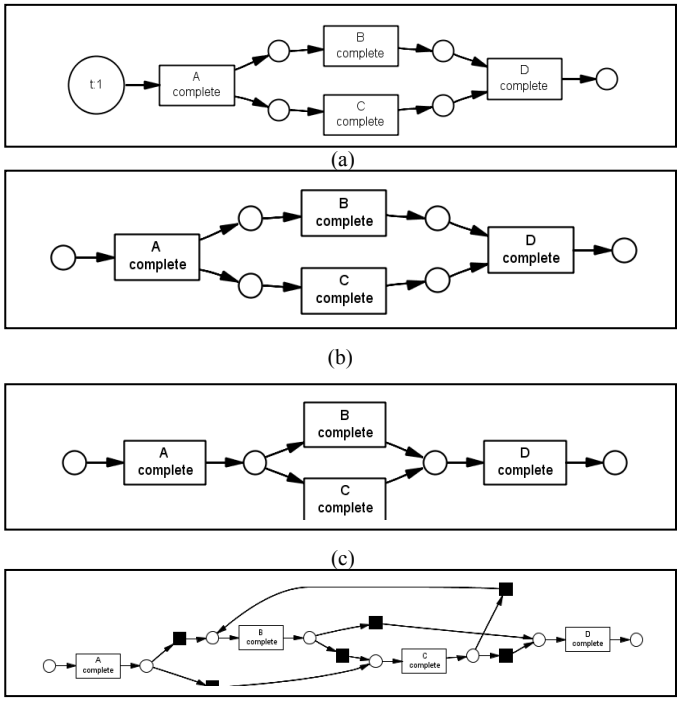

Fig. 4. Result of several discovery algorithm (a) Alpha algoritm (b) Alpha++ algorithm (c) Heuristic algorithm (d) Genetic algorithm

The available discovery algorithm cannot give satisfying representation of multi choice workflow pattern. The process models needs to be modified in order for this pattern can be identified using decision miner.

v. Modification of Multi Choice Model

Workflow patterns analyzed in decision mining might come from any level in business process, from operational to strategic planning. This time, we will use an example from business management process that use multi-choice pattern.

EXAMPLE 1. A bicycle production company, Polygon Cycle, is company which produces and sells their own bicycle. One of the bicycle types is XTRADA bike. Then, this company receives special order from another company to make 100 units of modified XTRADA bikes. These bikes will be given new brand Mountain Cycle. The client company offer price of Rp.1.790.000 for each unit of Mountain Cycle bike (1 US$ ≈ Rp.10.000). Meanwhile, the normal price of XTRADA bike is Rp.2.490.000 per unit with production cost of Rp.1.820.000.

In order to decide or accept these special orders, the company management needs to do analysis. The normal flow of analysis in general are: collecting XTRADA accounting data, collecting XTRADA and Mountain Cycle relevant costs, counting production incremental cost for both bikes, analyzing other opportunities, analyzing market share due to new brand, analyzing company image due to new brand, and analyzing gap positioning.

But, the management might not do all the analysis. Based on the management decision, some steps might be skipped. Flow of analysis is shown on Fig.5.

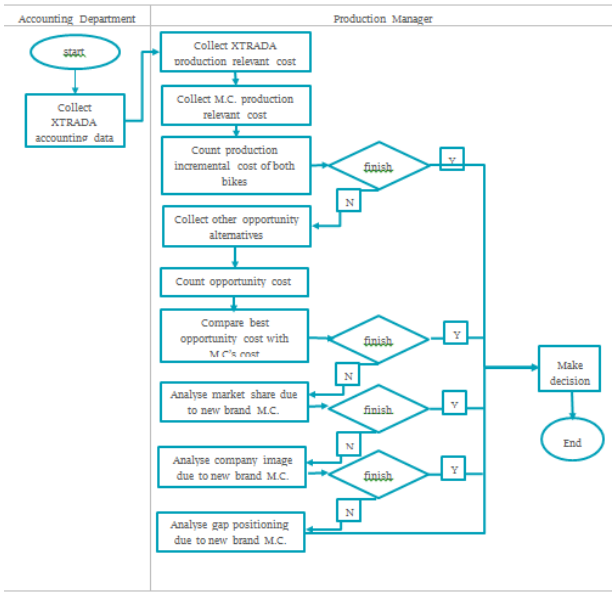

Fig. 5. Flow of analysis in Polygon Cycle special order case

The example case above shows a multi-choice workflow pattern. There are four block of choices. The first, identifying other opportunity, consists of collecting other opportunity alternatives, count opportunity cost, comparing best opportunity cost with mountain cycle's cost. Second choice consists of analyzing market share. Third choice consists of analyzing company image. And the fourth consists of analyzing gap position.

There are two ways proposed to modify a multi-choice pattern so that each choice can be analyzed by decision mining. Both representation will use XOR-split for each decision branch. Thus, identification to find why each step of analysis (branch of choice) is chosen becomes possible.

A. *Tricky OR-Split*

The first way to represent a multi-choice is using both AND-split and XOR-split. This representation also needs one extra dummy activity for each choice. AND-split will be placed after "Count Incremental Cost" as the last activity before the four choices. Then, XOR-split placed for each choice, where one branch pointing to the activity of the respective choice, and the other branch pointing to a dummy activity. Dummy activity will be executed if the choice is not chosen.

Tricky OR-split from example 1 is shown in Fig.6. There is one AND-split after "Count Incremental Cost" followed by four XOR-splits for each choices. In each choice, there is one dummy activity to represent that the choice is not chosen. This model cannot be mined by some discovery algorithms. That is why, separate petri-net model needed when decision mining needs to be done.

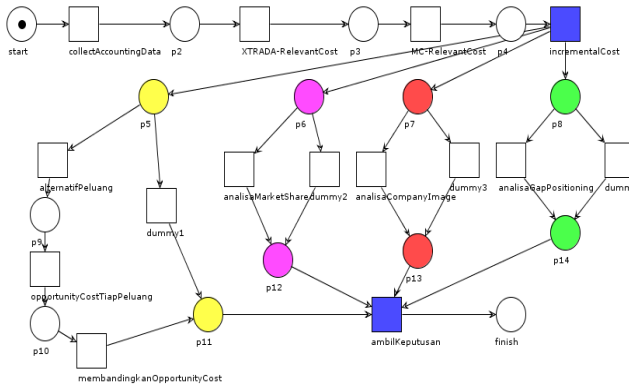

Fig. 6 The first representation of multi choice in WoPeD

B. *Sequential XOR-Split*

Sequential XOR-split can only be used for multi-choice that has "sequential" property, such as example 1. Sequential here means that even though there might be some activity skipped, but the order of execution are the same. For example, in example 1, "analyze company image" always comes after "analyze market share". The second representation retains the sequential property of the complete case, by inserting XOR-splits in each choice. Dummy activity used to represent that the corresponding choice is not chosen. Sequential XOR-split is shown in Fig.7.

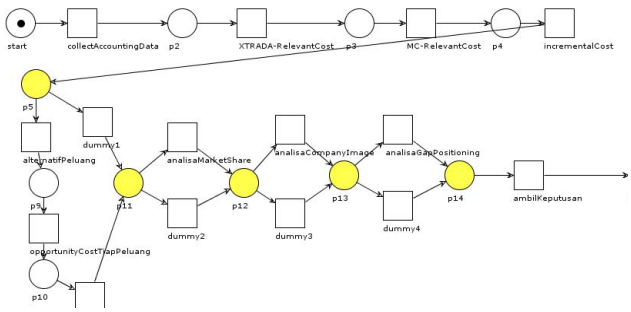

Fig. 7 The second representation of multi choice in WoPeD

The second representation can be applied if and only if the business process execution comes in a flow. For example activity B needs to be executed after activity A, activity C executed after B, etc. If the execution is not in sequence, then representation 2 cannot be applied.

VI. Implementation and Experiment

A. Scenario of Experiment

The proposed representation of multi-choice workflow pattern will be done using Example 1, special order case in Polygon Cycle. Several attributes needs to be assigned. From the "Count incremental cost" activity, there are several variables as shown in Table II.

TABLE II. EXAMPLE OF INCREMENTAL COST VALUE

	Each unit (in thousands)	Total 100 bikes (in thousands)
Selling increase	Rp 1.790	Rp 179.000
Incremental cost:		
Variable cost:		
Direct Materials	Rp 860	Rp 86.000
Direct Labor	Rp 450	Rp 45.000
***Overhead* variable**	Rp 60	Rp 6.000
Special Modification	Rp 170	Rp 17.000
Total variable cost	Rp 1.540	Rp 154.000
Fixed cost		
Design purchase		Rp 12.000

From data above, we can assign some variables/attributes to activity "Count incremental cost".

- Direct material (*Bahan Baku langsung / BBL)*
- Direct labor (*Tenaga Kerja Langsung/ TKL)*
- Overhead Variable (*OH*)
- Special Modification (*Modifikasi khusus/ MK*)
- Design Purchase (*Pembelian Desain / PD*)
- Total incremental cost (*Total peningkatan biaya / Total*)
- Incremental net profit (*Peningkatan laba bersih / Laba*)

Attributes above will contain values for each case. The values will be included in event log. Thus, decision tree will be built based on those values to retrieve rules why certain choice of analysis is taken.

Both proposed representation of multi-choice in petri-net for decision mining will be implemented in ProM 5.2. Special for tricky OR-split, separate model needs to be created using WoPeD. Then, for decision mining process, Decision Point Analysis plugin in ProM is used. Four blocks of choices are used in this event log. Choice 1 about alternative opportunity, choice 2 about analyzing market share, choice 3 about analyzing company image, and choice 4 about analyzing gap position. Dummy activity executed if the corresponding choices is not taken. Event log created to represent all choices.

B. Experimental Result

Experiment done for both proposed models to represent multi choice in example 1. Both proposed models using the same artificial event log. As any other ProM experiment, event log in .xml or .mxml file needs to be imported to ProM 5.2. Tricky OR-split needs separate models to be imported to ProM, so model in .pnml file should be prepared. Imported model has to be aligned with event log. Prepared model can be made in WoPeD, software to design a Petri-net workflow model.

Different from tricky OR-split, the sequential XOR-split model to represent multi-choice can be discovered by commonly available discovery algorithms in ProM 5.2. In this experiment alpha algorithm is used to mine the model.

Petri-net can be analyzed using decision point analysis plug-in in ProM 5.2. From the experiment, both models have same result in Decision Point Analysis plugin. The overall rules for each choice are as follows.

1) Analysis on alternate opportunity only done if profit (*laba*) <= 19 (in million Rupiahs)
2) Analysis on market share only done if direct material (*Bahan Baku Langsung/BBL*) > 69 (in million Rupiahs)
3) Analysis on company image only done if direct labor (*Tenaga Kerja Langsung/TKL*) <=59 (in million Rupiahs)
4) Analysis on gap positioning only done if special modification (*Modifikasi khusus/MK*) > 6 (in million Rupiahs)

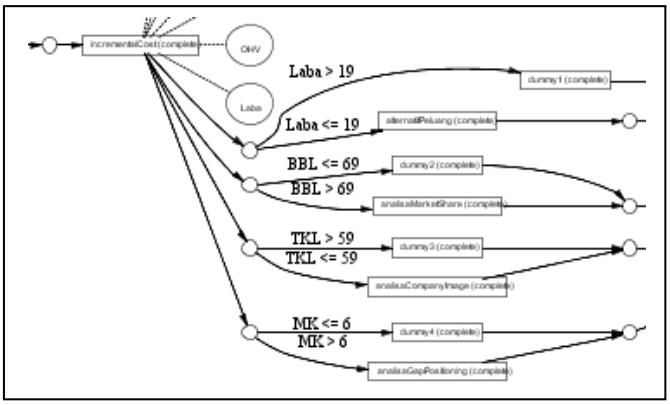

Fig. 10 Overall rule result for the workflow

VII. Conclusion and Future Works

Proposed models for multi choice using combination of AND-split and XOR-split and usage of dummy activity can be done in order to modify multi-choice representation in petri-net. These modifications retain the flow of case and make the decision points can be identified using decision point analysis plug-in in ProM framework.

The experiment shows that each choice can be analyzed separately using decision miner. So that each condition for multi choice/OR-split branch can be determined. The modification which lets decision miner to analyze multi-choice workflow pattern can help improving application of decision mining in real life business processes where multi-choice are often used.

Further works needs to be done to solve decision mining problems in another workflow patterns. Since decision mining has a very limited applicability to choices. Decision mining application in invisible activity, duplicate activity, and non-free choice cases are still limited. The current decision mining technique is limited to workflow that has appropriate representation in petri net.

Acknowledgment

We would like to thank the Higher Education Directorate of the Education Ministry of Indonesia, Japan International Cooperation Agency (JICA) and the Indonesian Agency for Agricultural Research and Development for supporting the research.

References

[1] Apte, C. and Weiss, S., Data mining with decision trees and decision rules, *Future Generation of Computer Systems,* vol. 13, pp. 197-210. 1997.

[2] Chandra, B. and Varghese, P.P., Moving towards efficient decision tree construction, *Information Sciences 179(8)*, pp.1059-1069. 2009.

[3] De Leoni, M. and van der Aalst, W.M.P., Data Aware Process Mining: Discovering Decision in Processes Using Alignments, *Proceedings the 28th Annual ACM Symposium on Applied Computing*, pp.1454-1461. 2013.

[4] Feng, Q., Miao, D., Cheng. Y., Hierarchical decision rules mining, *Expert Systems with Applications*, vol.37, pp.2081-2091. 2010.

[5] Grigorova, K., Process Modeling Using Petri Nets, *Proceedings International Conference in Computer Systems and Technologies*, 2003.

[6] Mehrez, A., Muzumdar, M., Acar, W., Weinroth, G., A Petri Net model view of decision making: an operational management analysis, Omega 23(1), pp. 63-78, 1995.

[7] Otero, F.E.B., Freitas, A.A., Johnson, C.G., Inducing decision trees with an ant colony optimization algorithm, *Applied Soft Computing*, vol. 12, pp.3615-3626. 2012.

[8] Rozinat, A. and van der Aalst, W.M.P., Decision Mining in Business Processes, BPM Center Report BPM-06-10, BPMcenter.org, 2006.

[9] Rozinat, A. and van der Aalst, W.M.P., Decision Mining in ProM, Lecture Notes in Computer Science, vol.4102, pp.420-425. 2006.

[10] Russell, N., ter Hofstede, A.H.M., van der Aalst, W.M.P., and Mulyar, N., Workflow Control-Flow Patterns: A Revised View. *BPM Center Report BPM-06-22*, BPMcenter.org, 2006.

[11] Salimifard, K. and Wright, M., Petri net—based modeling of workflow systems: An overview, *European Journal of Operational Research,* vol. 134, pp.664-676. 2001.

[12] Sarno, R., Sanjoyo, B.A., Mukhlash, I., Astuti, H.M., Petri net model of ERP business process variation for small and medium enterprises, *Journal of Theoretical and Applied Information Technology*, vol.54(1), pp.31-38. 2013.

[13] Seng, J.L. and Chen, T.C., An analytic approach to select data mining for business decision, *Expert System with Applications,* vol.37, pp.8042-8057. 2010.

[14] Subramaniam, S., Kalogeraki, V., Gunopulos, D., Casati, F., Castellanos, M., Dayal, U., Sayal, M., Improving process models by discovering decision points, Information Systems 32(7), pp. 1037-1055. 2007.

[15] Van der Aalst, W.M.P., The Application of Petri Nets to Workflow Management, *Journal of Circuits, Systems, and Computers 8(1)*, pp.21-66. 1998.

[16] Wohed, P., van der Aalst, W.M.P., Dumas, M., ter Hofstede, A.H.M., and Russell, N., Pattern-based analysis of BPMN - an extensive evaluation of the control-fow, the data and the resource perspectives, *Technical Report BPM Center Report BPM-05-26*, BPMcenter.org, 2005.