A Combination of The Evolutionary Tree Miner and Simulated Annealing

Afina Lina Nurlaili
Department of Informatiscs
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
afina.lina12@gmail.com

Riyanarto Sarno
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
riyanarto@if.its.ac.id

Abstract— In recent years, process mining is important to discover process model from event logs; however the existing methods have not achieved good in overall fitness. In this context, this paper proposes a combination of the Evolutionary Tree Miner (ETM) and Simulated Annealing (SA). The ETM aims to reduce randomness of population so that it can improved the quality of individuals. SA aims to increase overall fitness in the population. The results of the proposed method which was compared to other approaches show that the proposes method had better in overall fitness and better quality of individuals.

Keywords—The Evolutionary Tree Miner (ETM); Simulated Annealing (SA); process discovery; process tree; event logs .

I. INTRODUCTION

Information system is used by organization to support making decision and to organized process business. It executed activities from business process. Activities that was executed on information system will be recorded by information system. Activities that was recorded by information system usually called event log. Process mining was used to compare between event log and process model [1].

Process mining was distinguished into three part: (1) process discovery, (2) conformance checking, and (3) model enhancement [2]. Process discovery focus on make a model from an event log to obtain high quality model (replay fitness, precision, simplicity, and generalization), while conformance checking focus on detecting anomaly between model and event log [2]. After event log and process model were detected, knowledge from event log can be used to increase quality of process model [3].

Business process can be modelled using process model notations, for example Petri nets, Workflow, and Process Tree [1]. Process discovery created a process model using event log [3]. It became the most important research topics in process mining because it can be the initial point for activities in process mining [6].

There are many algorithms of process mining that has successfully to analyze event log, such as genetic algorithm, heuristic miner, and fuzzy miner [4]. Because of the aim of this paper is to present the optimal fitness in fewer generation, we combine the evolutionary tree miner with simulated annealing.

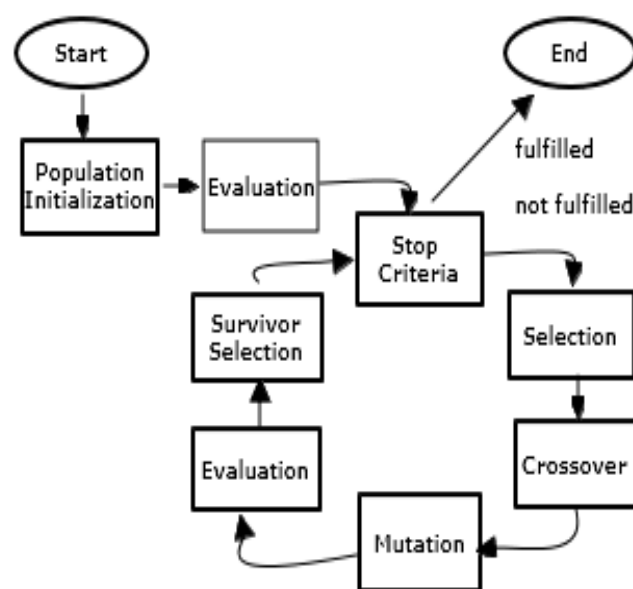The evolutionary tree miner is an algorithm from the family of

Fig. 1 Steps of Genetic Algorithm

Genetic Algorithms for mining business processes [5]. The evolutionary tree miner was used as the main process while simulated annealing was used for determining parent or child after apply crossover operator.

Genetic algorithm only focused on the quality dimensions of fitness. However there are four quality dimension of process mining: replay fitness (RF), precision (PE), generalization(Gv) and simplicity(Su) [5]. Steps of genetic algorithm is shown in Fig. 1. Based on Fig.1, the first step of GA is create population from random event log (you can create as many as possible generation to obtain the best result but do not forget genetic algorithm has high computation). Second, calculate the fitness each individual. Then, selecting process, crossover and mutation to get higher fitness. Finally, higher fitness is the best solution. The step will stop if GA produce maximum generation or maximum fitness 1.

The process of ETM is same with GA in nature. However in ETM use process tree to model the process of event log [9]. In this paper, by evaluating fewer models, we improve ETM. The goal is to get higher quality model in simplicity by

decreasing randomness of population. To reach the goal, we will explain two methods to improve ETM. First, create an initial populations of process trees [7]. Then we combine the Evolutionary Tree Miner with simulated annealing to improve the overall fitness of ETM [8]. Simulated annealing algorithm is a random search algorithm which use temperature and starts with a random initialization solution.

The ETM algorithm use process tree to represent process model. The Evolutionary Tree Miner algorithm use process tree to represent process model because it supports process that contain silent steps and duplicate activities [1]. Fig 2. show process tree with example event log. Fragment of event log can be seen in Table 1.

Table I Event log

| Case ID	Timestamp	Activity
1	30/12/10 11:02 AM	register request
1	31/12/10 10:06 AM	examine thoroughly
1	5/1/11 3:12 PM	check ticket
1	6/1/11 11:18 AM	decide
1	7/1/11 2:24 PM	reject request
2	30/12/10 11:32 AM	register request
2	30/12/10 12:12 PM	check ticket
2	30/12/10 14:16 PM	examine casually
2	5/1/11 11:22 AM	decide
2	8/1/11 12:05 PM	pay compensation
3	30/12/10 14:32 PM	register request
3	30/12/10 15:06 PM	examine casually
3	30/12/10 16:34 PM	check ticket
3	6/1/11 9:18 AM	decide
3	6/1/11 12:18 PM	reinitiate request
3	6/1/11 1:06 PM	examine thoroughly
3	8/1/11 11:43 AM	check ticket
3	9/1/11 9:55 AM	decide
3	15/01/11 10:45 AM	pay compensation
4	6/1/11 3:02 PM	register request
4	7/1/11 12:06 PM	check ticket
4	8/1/11 2:43 PM	examine thoroughly
4	9/1/11 12:02 PM	decide
4	12/1/11 3:44 PM	reject request
5	6/1/11 9:02 AM	register request
5	7/1/11 10:16 AM	examine casually
5	8/1/11 11:22 AM	check ticket
5	10/1/11 1:28 PM	decide
5	11/1/11 4:18 PM	reinitiate request
5	14/01/11 14:33 PM	check ticket
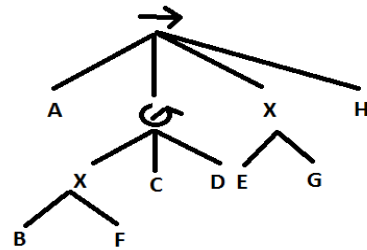
Fig. 2 A process tree from an example event log

II. RELATED WORK

In the related work section, we review previous work that relate to our proposed method.

Many existing algorithm to discover process model from event log for example α, α+, α++. They can discover all relation from all activities of event log. Firstly, an event log was taken as input in the algorithm. Then a workflow net was produced from mining an event log. The algorithm was modelled by Petri net.

Genetic Algorithm was used by Alves in business process mining especially in process discovery. Genetic algorithm was used by her to discover petri nets. Petri nets was used to modelling the event log. A two dimensional fitness measure was used by Alves. However there was a problem with another quality dimension such as preciseness, simplicity, and generalization.

The same method as Alve's was used by Tsai et al., the differences phase located on a time interval analysis between events in the log.

Buijs et al. explain that process tree better than petri net because it always sound process model while petri net can be unsound process model.

Amin et al. proposed Genetic Algorithm (GA) with Simulated Annealing (SA). He combined GA and SA to discover new process model with process tree and increase fitness.

III. PRELIMINARIES

3.1 Process Trees

Process tree was applied to describe our process model [7]. The operators of process tree are: sequential operator (→),non-exclusive choice operator (V), exclusive choice operator (X), parallel operator (Λ) and loop execution operator (⟲). The operator node of process tree that have children can be operator of process tree again or they can be leaf nodes. The children of process tree was read from left to right [12]. Children of process tree can have flexible number, except repeated execution. Repeated execution must have three children are: the left node is 'do', the middle node is 'redo' and the last node of the repeated execution is 'exit'. Unobservable activities of process tree indicated with a τ.

Based on Fig. 3, we can generate operator of process tree: the sequential operator and the its relation to BPMN (a), the parallel operator and the its relation to BPMN (b), the

exclusive choice operator and the its relation to BPMN (c), the non-exclusive operator and the its relation to BPMN (d), the repeated operator and the its relation to BPMN (e).
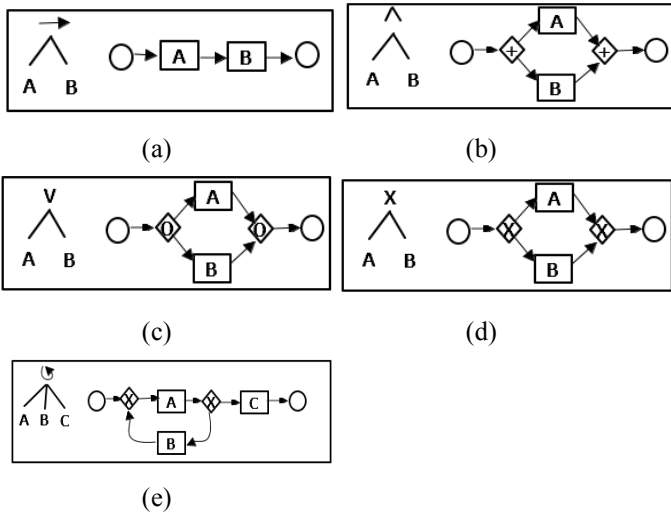

(a)

(b)

(c)

(d)

(e)

Fig. 3 Right is operator of process tree and left is translation to BPMN

3.2 The Evolutionary Tree Miner (ETM)

ETM had four main steps [1]. First, the ETM create initial population. Then calculated overall fitness of each individual. After that, checked the stopping criteria. The algorithm will be stopped if reach the maximum generation. And then copied an activities of elite individuals to the next generations. Elite rate determined the number of elite individual. The last step is to apply crossover and mutation operator. Stop the steps if new population is convergent. Alternatively, calculate fitness to repeat the process.

3.3 Simulated Annealing (SA)

Simulated annealing algorithm is a random search algorithm which use temperature and starts with a random initialization solution as present solution (xp) [3]. Then, choose the next neighbouring solution as the new solution (xn) and calculate $\Delta f = f(xn) - f(xp)$. If $\Delta f > 0$, calculate the probability $P = 1/1+\exp(\Delta f/t)$. Condition acceptance if $P > rn$. rn is random fix number, the value is between 0 and 1.

3.4 Quality of Process Model

Quality Dimensions of Fitness

The replay fitness was used to measure accuracy of the model with activity in the event logs done by calculating the value of fitness [11][12].

$$Replay = \frac{\sum_{i=1}^{n} parse(trace_i)}{\sum_{i=1}^{n} length(trace_i)} \qquad (1)$$

IV. PROPOSED METHODS

We explain how this method will be implemented is discussed in this section. We proposed our algorithm in Fig.4 in seven steps:

1. Create an Initial Population
   Create initial population by randomly selecting single trace from the event log by using process tree.
2. Calculate Fitness
   Calculate fitness of each individual based on trace.
3. Apply Crossover Operator
   Apply crossover operator of each individual.
4. Determine new individual using probability from Simulated Annealing
   Determine new individual from crossover operator by using probability of Simulated Annealing
5. Apply Mutation Operator
   Apply mutation operator of each individual. New generation be the new population.
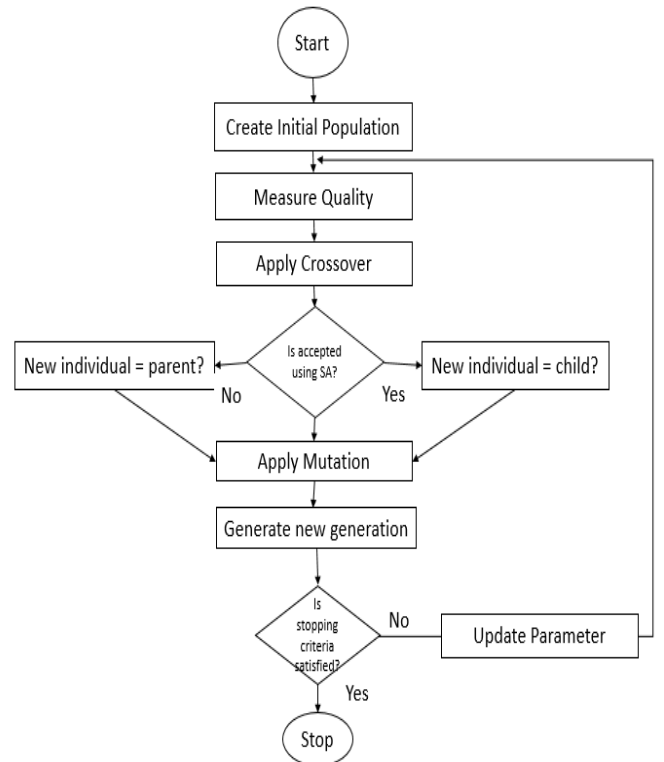6. Stop the steps if new population is convergent. Alternatively, go to step 2 to repeat the process.

Fig. 4 Flowchart of proposed method

Initial Population

Initial population was created at the first step of the evolutionary tree miner [10]. We explain how this initial population of process trees will be implemented is to create one of the process tree [13]. This initial population will increase high quality of process model. Fig. 5 show each trace of event log (i) A, F, C, D, H, C, F, D, H, F, C, D, E (ii) A, F, C, D, H, B, C, D, G (iii) A, B, C, D, E (iv) A, C, B, D, E (v) A, C, F, D, G (vi) A, F, C, D, G.
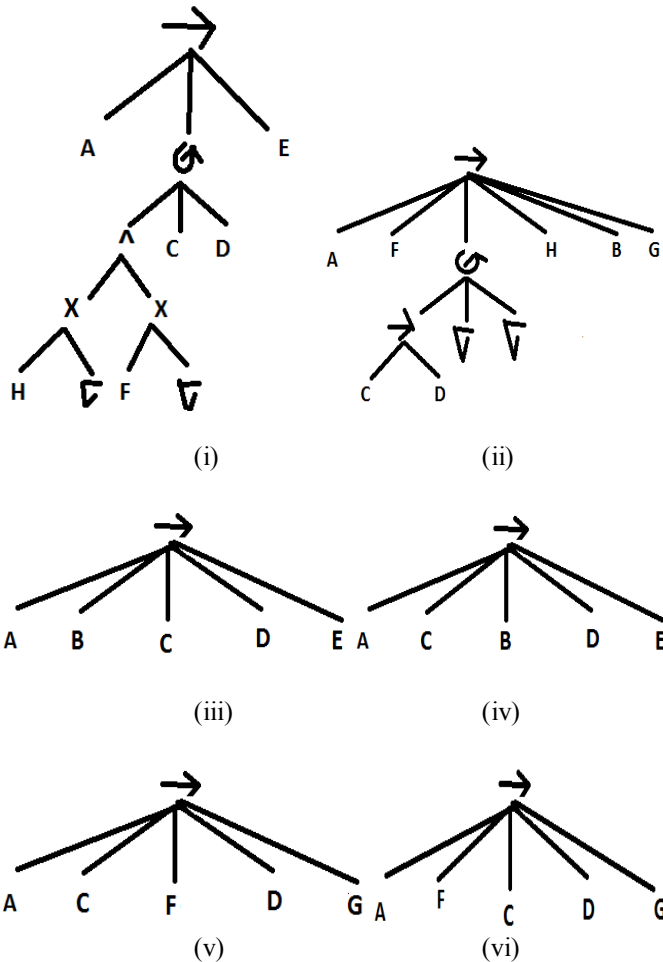
Fig. 5 Process tree each trace of event log

Crossover

The evolutionary tree miner has many method of crossover, there are single point, multi-point, and uniform crossover.

Mutation

Mutation is a little modification to the new offspring to prevent the population stagnating in their local optimal solution.

Elitism

Elistism(nElit) allowing n of the fittest parent to be n offspring of the next generation. Elitism is created to maintain the quality between two generation, so there will be no degradation of quality during succession.

Stopping Criteria

The maximum number of generation (nMax) is selected as the stopping critera. The best offspring then selected as final solution [14]. There is no rule how many nMax is required.

V. EXPERIMENTAL RESULTS

In this section, our experimental results were presented using event log. We used a set of example event log of handling a compensation request in an airline to test the evolutionary tree miner and simulated annealing. This experiments have performed with 20 initial population and stop at generation 400. We selected the best generation from 400 generation. An overall fitness weighted average of RF (x10), Pe (x5), Su (x1) and Gv (x0.1). Parameter used for running proposed method shown in Table II. Overall fitness measurement using ETM shown in Table III. The bold number in table is the highest quality.

Table II Parameter was used for run proposed method

Population size	Elite count	Maximum Generation	Crossover	Random tree creation	Temperature
20	5	400	0.25	0.25	20

Table III Fitness using ETM

Overall fitness	Replay fitness	Precision	Simplicity	Generalisation
0.830	0.764	**1.0**	**1.0**	**0.469**
0.824	0.764	1.0	0.933	0.437
0.820	**0.784**	0.933	1.0	0.445
0.811	0.767	0.961	0.894	0.408
0.809	0.751	0.974	0.941	0.426
0.807	0.767	0.916	1.0	0.463
0.799	0.761	0.904	1.0	0.456
0.793	0.715	0.974	1.0	0.459
0.784	0.687	1.0	1.0	0.465
0.775	0.673	1.0	1.0	0.456
0.767	0.660	1.0	**1.0**	0.438

The final experiments in TABLE IV give comparative analysis of average fitness of each algorithm. We compare a combination of the ETM with SA and GA. We can analyze that the proposed method has the best results in terms of average fitness.

Table IV Fitness using Genetic Algorithm vs Evolutionary Tree Miner

Number of Generation	ETM	GA
1	**0.830**	**0.744**
2	0.824	0.743
3	0.820	0.697
4	0.811	0.669
5	0.809	0.648

Fig. 7 shows the overall fitness values each individual of ETM and proposed method. It shows that the proposed

method had better overall fitness individuals in its populations each generation. So using SA can improve the algorithm's fitness.
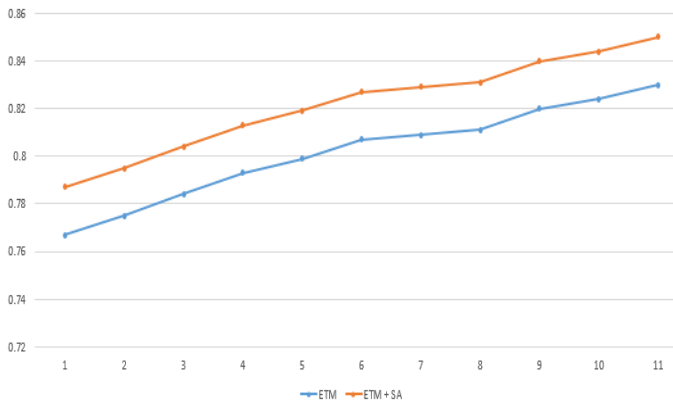

Fig. 6 Fitness values of the individuals throughout the generation

VI. CONCLUSION AND FUTURE WORK

We propose a combination of the Evolutionary Tree Miner and Simulated Annealing for business process mining. In this algorithm, for the first step, we have to define initial population in order to apply the Evolutionary Tree Miner algorithm. Using the Evolutionary Tree Miner in the context of process discovery has a lot of advantages. When using the Evolutionary Tree Miner, it is important to have a correct fitness quality. Nevertheless, the quality of process model is not only fitness but other qualities are important, such as precision, simpicity, and generalisation. Using the Evolutionary Tree Miner can balance four quality of model. We use Simulated Annealing (SA) for finding a good solution to an optimization problem, which uses probabilistic technique. Furthermore using combination the evolutionary tree miner and simulated annealing can increase the quality of fitness of individuals in the population.

In future work we will try to implemented using real-life event log. The crossover and mutation operator will also be point in ETM.

REFERENCES

[1] M.L. van Eck, J.C.A.M. Buijs and B.F. van Dongen, "Genetic Process Mining: Alignment-based Process Model Mutation," 2013. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-15895-2_25

[2] R. Sarno and K. R Sungkono, "Coupled Hidden Markov Model for Process Mining of Invisible Prime Tasks," International Journal of Software Engineering and Its Applications, vol. 11, 2016. [Online]. Available: 10.15866/irecos.v11i6.9555

[3] R. Sarno and Y. A. Effendi, "Modified Time-Based Heuristics Miner," International Review on Computers and Software (IRECOS), vol. 11, 2016. [Online]. Available: 10.1109/ICTS.2015.7379899

[4] I. G. Anugrah, R. Sarno and R. N. E. Anggraini, "Decomposition Using Refined Process Structure Tree (RPST) and Control Flow Complexity Metrics," in ICTS 2015 co-sponsored by IEEE, Surabaya, 2015. [Online]. Available: 10.1109/ICTS.2015.7379899

[5] K. R. Sungkono and R. Sarno, "Hidden Markov Model for Process Mining of Parallel Business Processes," International Review on Computers and Software (IRECOS), pp. 290-230, 2016. [Online]. Available: https://doi.org/10.15866/irecos.v11i4.8700

[6] W. v. d. Aalst, "Process Mining: Discovery , Conformance and Enhancement of Business Process, Springer, 2011. [Online]. Available: 10.1007/978-3-642-19345-3

[7] A. Adriansyah, B.F. van Dongen and W.M.P. van der Aalst, "Conformance Checking using cost-based fitness analysis. In Enterprise Distibuted Obhect Computing Conference (EDOC)," 15th IEEE, pp. 55-64, 2011. [Online]. Available: 10.1109/EDOC.2011.12

[8] J.C.A.M. Buijs, B.F. van Dongen and W.M.P. van der Aalst, "On the Role Fitness, Precision, Generalization and Simplicity in Process Discovery," In On the Move to Meaningful Internet Systems:OTM2012, pp. 305-322, 2012. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-33606-5_19

[9] M.L. van Eck, Alignment-based process model repair and its application to the Evolutionary Tree Miner, Master's thesis, Technische Eindhoven, 2013. [Online]. Available: https://pure.tue.nl/ws/files/46956499/770544-1.pdf

[10] D. Fahland and W.M.P. van der Aallst, Model repair - aligning process models to reality, 2013. [Online]. Available: http://dx.doi.org/10.1016/j.is.2013.12.007

[11] S. Huda, R. Sarno and T. Ahmad, "Identification of Process-based Fraud Patterns in Credit Application," in 2nd International Conference on Information and Communication Technology (ICoICT), 2014. [Online]. Available: 10.1109/ICoICT.2014.6914045

[12] W. Song, S. Liu and Q. Liu, "Business Process Mining Based on Simulated Annealing," in The 9th International Conference for Young Computer Scientist, 2008. [Online]. Available: 10.1109/ICYCS.2008.279

[13] A. Vahedian Khezerlou and S. Alizadeh, "A new model for discovering process trees from event logs," Applied Intelligence, vol. 41, no. 3, p. 725–735, Jul. 2014. [Online]. Available: 10.1007/s10489-014-0564-7

[14] B. Vazquez-Barreiros, M. Mucientes and M. Lama, "A Genetic Algorithm for Process Discovery Guided by Completeness, Precision and Simplicity," Business Process Management, p. 118–133, 2014. [Online]. Available: 10.1007/978-3-319-10172-9_8