

A Temporal Work Transference Event Log Trace Classification Algorithm and Its Experimental Analysis

Dinh-Lam Pham*, Hyun Ahn*, Kwanghoon Kim*

*Collaboration Technology Research Lab. Division of Computer Science and Engineering,
KYONGGI UNIVERSITY, 154-42 Youngtong-Gu Suwon-Si Gyeonggi-Do, 16627, Republic of Korea

Corresponding Author: Kwanghoon Pio Kim

phamdinhlam@kgu.ac.kr, hahn@kgu.ac.kr, kwang@kgu.ac.kr

Abstract—In the field of process mining, a lot of information about what happened inside the information system has been exploited and has yielded significant results. However, information related to the relationship between performers and performers is only utilized and evaluated in certain aspects. In this paper, we propose an algorithm to classify the temporal work transference from workflow enactment event log. This result may be used to reduce system memory, increase the computation speed. Furthermore, it can be used as one of the factors to evaluate the performer, active role of resources in the information system.

Keywords: Process mining, Temporal Work Transference Pattern, Trace Classification, Information System, Workflow Event Log

I. INTRODUCTION

In process mining, we take the data that exists in an organization information system and use that data to visualize what's really happening and how they are done in the process. Unlike some other resources such as financial resources, material resources, technology resources, etc. Human resources are an indispensable resource that determines organizational success. While building an organization's workflow model, we regularly pay close attention to how to effectively exploit human resources, combine resources to solve the problem in the organization's activities. However, that is the plan what we build the project, and we don't know what is going on in the organization's operations. To capture what happened, we have to consider, evaluate from the operating history and compare with the planning model.

Workflow management systems register the start and completion of activities, ERP systems log all transactions. These examples have in common that there is some kind of event log. Unfortunately, the information in these logs is rarely used to derive information about the process, the organization, and the social network. We can get information by observing the behavior from one individual to another [1]. Process mining techniques can be applied to analyze and improve any operational processes in a variety of domains. The final process model that has been extracted based on the real log history would us have a clear model that only includes the most relevant activities, and ignores the rest of activities that will not have a significant impact on the model quality [2]. During the operation of the system, one of the operational

resource relations formed was the Temporal Work Transference (TWT) Network [3][4]. This network tells us the role of resources in the system and the interaction between these resources. Through the network, we can see the connection of the performers together, the role of each performer in the whole process to complete the work. Temporal work Transference network is formed by combining the TWT of each trace from the log file of the information system. In other words, TWT is a significant component of this network. Specifically, we will see that TWT networks are made up of different TWT patterns. Some patterns are used many times, with high frequency, besides, the others are used less or only once in creating the network.

In this paper, we propose an approach to identify and classify the TWT pattern. This work will help:

- Reduced storage space during the organization of the data, extraction of information from the workflow event log: Instead of storing the same TWT sets many times; we only need to save an unique TWT pattern representation and store the number of times it acted in the system.
- Indicate the significant TWT patterns that are important in the information system. Significant pattern sets are components that have a high reuse index in the TWT network. From this perception, we see the critical performer of the system, the interaction of the performer with the remaining performer.

These results help the administrators, the operators of the organization have a comparative basis, evaluate the performance of the system compared to the original model. These will also be components that aim to analyze the role of the performer, the network of performer that is built into our process mining system.

The paper is structured as follows. Section 2 discusses background and motivation. Section 3 proposes the technique while Section 4 shows the experimental result. Finally, Section 5 concludes the paper and discusses future work.

II. BACKGROUND AND MOTIVATION

In this section, we discuss the concepts and related knowledge used in this paper:

A. Temporal Work Transference Model

During the operation of the system, the components of the workflow engine record events that occurred in a log file, the event data being recorded in the sequence of the events that occurred. This execution sequence of a workflow instance is forming a workflow instance event trace, from which we can extract a workflow instance activity trace and its formal representation is specified by a model of temporal work-case.

Definition of Temporal Work Transference Model [5]. A temporal work transference model is formally defined through 3-tuple $TWTM = (X, F_r^t, T_o^t)$ over a set P of performer trace-nodes $\forall \eta_p^{\tau[\emptyset]}$ on a temporal work transference $TWT(C)$ of a workflow instance (work-case), c , and a species $K = (\{e, u, o\})$ of the timestamp origins, where:

- F_r^t is a coordinator or a coordinator-group linked from an external temporal work transference model;
- T_o^t is a coordinator or a coordinator-group linked to an external temporal work transference model;
- $X = X_i \cup X_o$ on $\forall \eta_p^{\tau[\emptyset]} \in P$
 - $X_o : P \rightarrow \wp(P)$ is a single-valued mapping function of a performer trace-node $\eta_p^{\tau[\emptyset]} = we_p^\tau \wedge \emptyset \in K$, to its (immediate) successor in a temporal worktransference;
 - $X_i : P \rightarrow \wp(P)$ is a single-valued mapping function of a performer trace-node $\eta_p^{\tau[\emptyset]} = we_p^\tau \wedge \emptyset \in K$, to its (immediate) predecessors in a temporal worktransference;
- The species of temporal worktransference models: $TWTM\emptyset$
 - AssessedTime Temporal worktransference Model: $\emptyset = 'e'$ in $\forall \eta_p^{\tau[\emptyset]}$ of a temporal worktransference model
 - StartedTime Temporal worktransference Model: $\emptyset = 'u'$ in $\forall \eta_p^{\tau[\emptyset]}$ of a temporal worktransference model
 - CompletedTime Temporal worktransference Model: $\emptyset = 'o'$ in $\forall \eta_p^{\tau[\emptyset]}$ of a temporal worktransference model

The execution sequence of a workflow instance is also involved with a performing sequence of the performers who are in charge of the executions of the work items in the corresponding workflow instance. We can also extract a workflow instance performer trace from a workflow instance event trace and its formal representation is specified by

a model of temporal work transference.

B. Work Transference Network

We know that the procedural activities on a workflow procedure eventually trigger the work transferences among the performers who are involved in the workflow procedure. Analyzing work transferences among performers is one of the essential planning activities in human resource management and decision-making support management, in general. Especially, the work transference relationships in

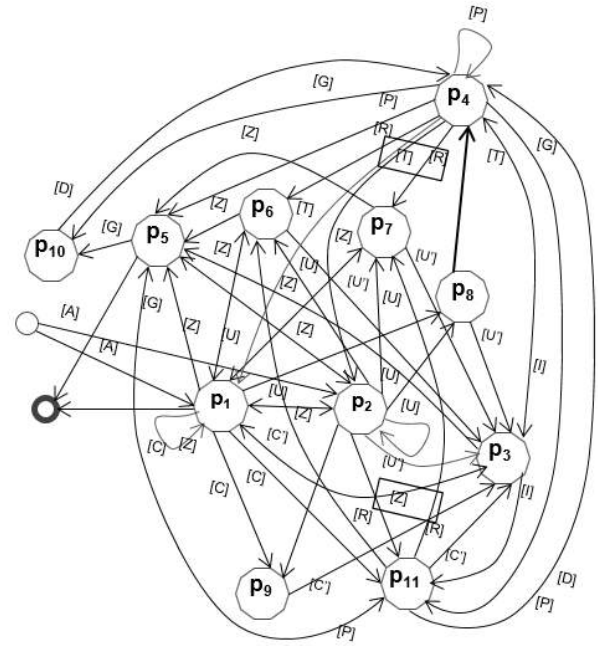


Fig. 1: A Work Transference Network of Buildtime From The Library Book Acquisition Workflow Model [5]

performing a workflow procedure turn also out the essential knowledge and criteria for evaluating and validating the human resource performance of the corresponding workflow procedure. A work transference network is a formal and graphical structure of a directed graph (or digraph) model to represent the relationship of work (activity) transferences and their associated works among performers who are involved in a corresponding workflow procedure. Each vertex represents a performer, and each ordered pair of vertices (or a directed edge) represents a work transference relationship [5][6].

We can judge that a work transference network is made up of a combination of temporal work transference from the history log. In this network, each TWT represents a path from the starting work

to the end work. More specifically, it demonstrates the order in which a work instance performs work with the participation of performers into the event in it.

III. TEMPORAL WORK TRANSFERENCE PATTERN CLASSIFICATION

A. Temporal Work Transference Pattern

In the history of an information system, many TWTs are generated as mentioned before. Among them, some TWTs are repeated numerous times. Conversely, the other being used only times or only one time. We call it a TWT pattern. Identifying TWT patterns for each group will help:

- Optimal storage of information on memory: When we dig up data from the history log, instead of saving all information of trace instances, we only need to keep the TWT pattern and the number of its representation for the whole group.
- Identify regular performers involved in implementation in the system. The number of times these performers participate in the system and the number of times they compete in different types of TWT.

One problem is that in the history log there are thousands of trace instances generated, how to identify the TWT pattern and organize them optimally. In the next session, we propose an approach and algorithm to classify the TWT patterns from the workflow event log.

B. Classifying The Temporal Work Transference Pattern

The job of identifying and classifying TWT patterns requires that all trace instances in the history log be visited at least once. We need to avoid revisiting trace instances because their numbers are numerous and this will affect the performance of the system. The main idea of the approach is: While reading a trace instance from the history log, we consider whether it meets a TWT pattern by comparison with the recorded TWT patterns. The TWT pattern is put in a "silo," and the "silo" label is the number of times that pattern has appeared in the history log to the current position.

Fig. 2 shows the idea of TWT pattern classification. Instead of having to store information about all trace instances, after classification, the system only

saves the silos containing information about the TWT pattern.

The Algorithm classify TWT pattern

Input: Workflow Enactment Log ($\omega\varepsilon$) file with XES standard.

Output: List of TWT pattern.

BEGIN

```

1: Open the enactment log file
2: Initialize box in listBox $\beta$  to  $\emptyset$ 
3: for ( $\forall_{trace[i]} \in (\omega\varepsilon)$ ) do
4:   currentTWT = GetTWT(trace[i]) (*)
5:   for ( $\forall_{box} \in listBox\beta$ ) do
6:     flag=false
7:     if (currentTWT  $\in$  currentBox) then
8:       flag=true
9:     end if
10:  end for
11:  if (flag=true) then
12:     $\beta$ .newBox.put(currentTWT)
13:  end if
14: end for
15: Close the enactment log file
16: Return listBox $\beta$ 

```

END

In this algorithm, we use the **GetTWT** function(*) which used to get the temporal work transference [3].

IV. EXPERIMENTAL RESULTS

To evaluate the effectiveness of the TWT pattern classification and the algorithm that was mentioned above, we performed with the data set provided by 4TU. Center for Research Data [7].

TABLE I: INFORMATION ABOUT THE LOGS

Log File	Number of Traces	Number of Individual	Number of Individual Performers
Hospital billing	100.000	18	1150
BPI Challenge 2012	13.087	24	68
BPI Challenge 2018	43.809	41	165

With the synthetic event logs data set, we select the following logs:

- BPI Challenge 2012: Event log of a loan application process.

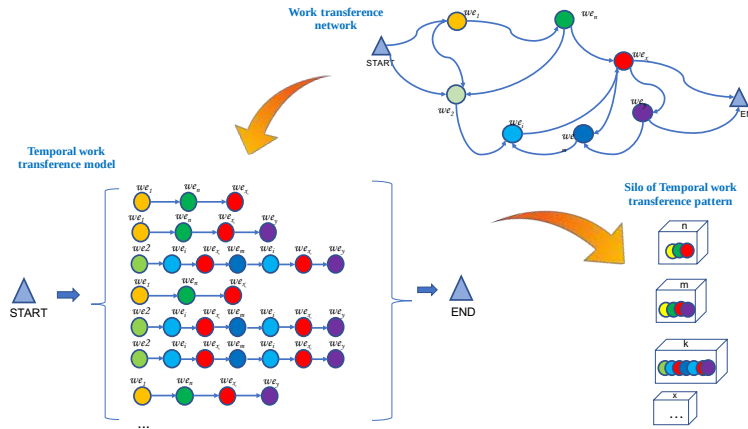


Fig. 2: The Idea of Classification Temporal Work Transference Patterns From Workflow Event Log

- BPI Challenge 2018: The process considered in this dataset covers the handling of applications for EU direct payments for German farmers from the European Agricultural Guarantee Fund.
- Hospital Billing: The 'Hospital Billing' event log was obtained from the financial modules of the ERP system of a regional hospital.

Table I shows general information about the logs used. Table II presents the results of the TWT classification algorithm.

Base on the result in Table II we can see, many different TWT pattern were built in an information system. The *Degree of Pattern Requirement* in the log 'hospital billing' is 19.67% and 80.33% degree of Reduction. Meanwhile, the log 'BPI Challenge 2018' has a higher *Degree of Pattern Requirement* with 68.62% and 31.38% of Reduction. We can see more detail in Fig 3.

The *Average of Using Traces Per Pattern* value of log 'Hospital billing' is highest with 5.08 times, the others are 1.82 and 1.46 times. In the field of store TWT data, the log 'Hospital billing' has 2.56 times data saving level compare to 'BPI Challenge 2018' log.

V. CONCLUSION

In this paper, we have devised an approach to identify temporal work transference patterns from the workflow enactment event log. We presented an algorithm and applied it to defining TWT patterns from real life logs [7]. By storing the profiles which represent for the group instead of the entire

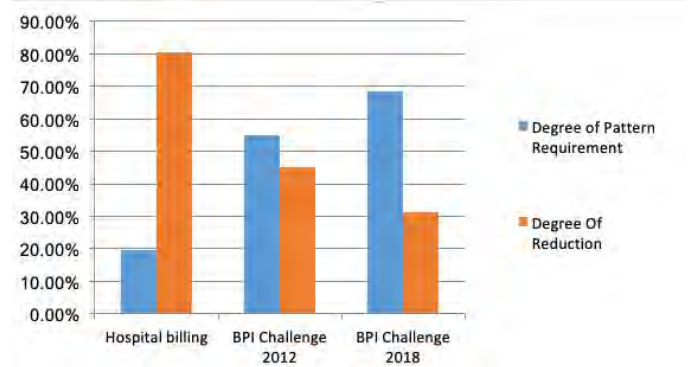


Fig. 3: The Degree of Pattern Requirement and Degree of Reduction

history log, we saved memory, as well as increased computing speed for process mining systems. Also by using these patterns, we determine the level of participation of performers on different traits as well as different temporal work transference. It gives a way of assessing the level, efficiency, and frequency of performers in the system. In the future work, we will continue to improve this algorithm, applying deep learning technology to dig up more information from the Information system history log.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT Future Planning (Grant No.2017R1A2B2010697).

TABLE II: THE RESULT OF CLASSIFICATION TWT PATTERN

Log File	Number of Traces	Number of TWT Pattern	Average of Using Traces Per Pattern	Degree of Pattern Requirement	Degree of Reduction
Hospital billing	100.000	19.670	5.08	19.67%	80.33%
BPI Challenge 2012	13.087	7.171	1.82	54.79%	45.21%
BPI Challenge 2018	43.809	30.061	1.46	68.62%	31.38%

REFERENCES

- [1] W. M. Van der Aalst and M. Song, "Mining social networks: Uncovering interaction patterns in business processes," in *International conference on business process management*. Springer, 2004, pp. 244–260.
- [2] A. Fares, J. Gama, and P. Campos, "Process mining for analyzing customer relationship management systems: A case study," in *Learning from Data Streams in Evolving Environments*. Springer, 2019, pp. 209–221.
- [3] D. Pham, H. Ahn, and K. P. Kim, "A temporal work transference discovery algorithm and experimental results on xes-formatted workflow logs," in *Procs. of the 13th Asia Pacific International Conference on Information Science and Technology (APIC-IST)*, pp. 41–46, 2018.
- [4] M. Park, H. Ahn, and K. P. Kim, "Workflow-supported social networks: Discovery, analyses, and system," *Journal of Network and Computer Applications*, pp. 355–373, 2016.
- [5] M. Jin, K. Kim, and K. P. Kim, "Discovering a work transference network from a workflow model of library book acquisition procedures," in *Procs. of the 9th International Conference on Internet*, pp. 221–225, 2017.
- [6] H. Ahn, M. Park, and K. P. Kim, "A graph-based data model for process-centric enterprise social networks," in *Procs. of the 12th Asia Pacific International Conference on Information Science and Technology*, pp. 210–212, 2017.
- [7] 4tu center for research data. [Online]. Available: <https://data.4tu.nl/repository/collection>



Dinh-Lam Pham Dinh-Lam Pham is a full-time Ph.D. student in Kyonggi University, South Korea since March, 2018. He got the MSc in computer science, Thai Nguyen University, Vietnam (2010). He worked in Thainguyen University from 2008 to 2016 as an IT engineering. From 2016, he works in the Information Technology Institute, Vietnam National University, Hanoi. His research interests include Process Mining, Workflow Management System, Information Control Net.



Hyun Ahn Hyun Ahn is an assistant professor of computer science and engineering department at Kyonggi University, South Korea. He received his B.S., M.S., and Ph.D. degrees in computer science from Kyonggi University in 2011, 2013, and 2017, respectively. His research interests include workflow systems, BPM, business process intelligence, and process mining.



Kwanghoon Pio Kim Kwanghoon Pio Kim is a full professor of computer science department and the founder and supervisor of the collaboration technology research laboratory at Kyonggi University, South Korea. He received B.S. degree in computer science from Kyonggi University in 1984. And he received M.S. degree in computer science from Chungang University in 1986. He also received his M.S. and Ph.D. degrees from the computer science department at University of Colorado Boulder, in 1994 and 1998, respectively. At Kyonggi University, he is Dean of the Computerization and Informatics Institute, and the director of the contents convergence software research center, as well. He had worked as researcher and developer at Aztek Engineering, American Educational Products Inc., and IBM in USA, as well as at Electronics and Telecommunications Research Institute (ETRI) in South Korea. In present, he is a vice-chair of the BPM Korea Forum. He has been in charge of a country-chair (Korea) and ERC vice-chair of the Workflow Management Coalition. He has also been on the editorial board of the journal of KSII, and the committee member of the several conferences and workshops. His research interests include groupware, workflow systems, BPM, CSCW, collaboration theory, Grid/P2P distributed systems, process warehousing and mining, workflow-supported social networks discovery and analysis, process-aware information systems, data intensive workflows, and process-aware Internet of Things.