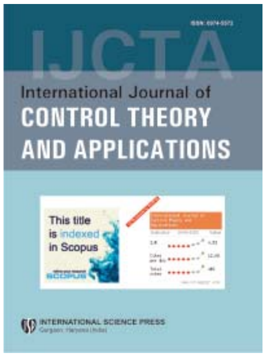


International Journal of Control Theory and Applications

ISSN : 0974–5572



Volume 9 • Number 42 • 2016

# 2-D Method of Assessment for Corrugated Industry using Process Mining Approach

**Aruna Devi .T [a] Kumudavalli M.V[b] and Sudhamani[c]**

*[a]Research Scholar, Rayalaseema University, Andhra Pradesh, India.*

*E-mail: arunadevi@dayanandasagar.edu*

*[b]Dept. of Computer Applications, Dayananda Sagar College, Bangalore, India.*

*E-mail: kumudavalli@dayanandasagar.edu*

*[c]Research Supervisor, Rayalaseema University, Andhra Pradesh, India.*

*E-mail: ewsudha@hotmail.com*

***Abstract:*** Process is the back bone of any industry. Now-a-days industry are keen on increasing their efforts to improve process as it plays an important role in boosting the work productivity and reducing the cost. One of the key factors involved in the productivity of the industry is its Data. Process mining is an exponentially growing area which plays a major role in the field of manufacturing industries in terms of helping the industries in understanding the status, check for compliance and plan for improving their process. Today's industries use information systems invaried dimensions and tailor made structures. The main aim of process mining is to extract knowledge from event logs, build a suitable model and use the same for analysis. It also gives the provision for enhancing the models by BPM/WFM systems. Previously process models were typically made by hand without using event data. However, activities executed by people, machines, and software leave trails in so-called event logs. In this paper we (authors) are giving a 2-Dimensional analysis to the data/ event-log created with respect to the corrugated industry using some of the important techniques of Process Mining, specially footprint to improve the process and hence to enhance their productivity.

***Keywords:*** *Process Mining, Event Logs, Corrugated Industry, Footprint.*

## 1. INTRODUCTION

In today's dynamic market, it is necessary for industries to gradually streamline their process so as to reduce costs and to improve performance. Process automation leads us in a new path for creating different forms of process analysis in comparison to traditional hand-made models. Digital event data has become an important factor in every sector and is continuing to grow exponentially. Presence of such data allows the industries to create and analyze different forms of process analysis, especially based on the observed facts. Process mining is a relatively recent research discipline ([1], [2]) that can be highlighted among the seminal articles. The main focus of process mining is to extract knowledge from the data collected in the information systems/databases in order to create custom made event logs [3]. An event log can be viewed as a collection of foot marks, each containing all the activities executed for a particular process instance. Specific examples of such applications include Enterprise Resource Planning systems (*e.g.*, SAP), Customer Relationship Management Systems (*e.g.*, Sales force) but event log data are not limited only to data from these existing applications, as many other systems can also provide useful data about process execution. Complex process may have data related to more than one source of information. Any activity performed in an industry by a worker the related data is stored in the system. Activities are recorded in event logs for support, control and further analysis. Process models are built so as to specify the order in which different workers are suppose to perform their task within a given process, or to analyze critically the process design. Process mining generally includes three main types: discovering process models from event logs, conformance checking and organizational mining.

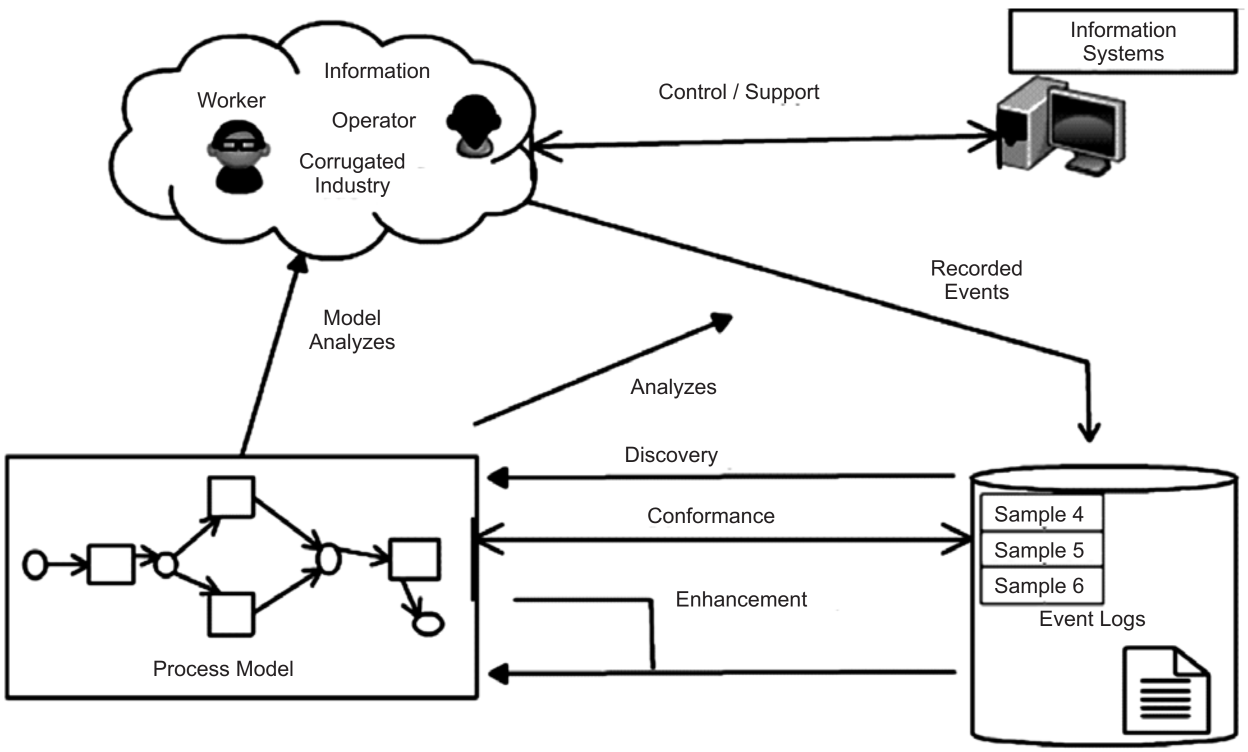


**Figure 1: Process Mining Overview of Corrugated Industry**

In reference [3], it is explained how automatic process discovery allows process models to be extracted from an event log; how conformance checking allows monitoring deviations by comparing a given model with the event log; and how enhancement allows extending or improving an existing process model using information about the actual process recorded in the event log. Using process mining techniques in corrugated manufacturing industry, process is not only ensures that such procedures can be thoroughly understood, but can also generate benefits associated with process efficiency. Process mining techniques can be of help to identify and understand the real behaviour of the process by analysing the performance of the process. In this paper an exhaustive study is made pertaining to corrugated industry and a part of the data is used to build the event logs, which are further used to build the relationship table and footprint matrix to give a 2-D analysis to the process.

In this paper we have adopted algorithmic approach to create the footprint matrix. The event log created for the corrugated industry is analysed and a part of alpha algorithm is used to create 2D- matrix. The further study of this matrix would be the implementation of the ProM Tool to create pictorial representation of the data, as it is more appropriate and understandable.

This paper is organized as follows: Section II Related work, Section III Introduces process mining, Section IV explains about Event Logs, Section V Discusses about footprint in process mining, Section VI deals with Process Discovery. Section VII gives the methodology followed for the analysis; the outputs of the study are presented as results in Section VIII. A brief conclusion is discussed in section IX.

## 2. RELATED WORK

Process mining has growing interest in healthcare which has resulted in analysis of existing logs for representing the process model ([11], [12]). The process mining work focuses on discovery of resource queues where the ability to determine the number of cases waiting for an activity to discover the queue lengths [13]. The study of process mining has stretched its roots into animation world because of its huge volume of data that can be visualized easily [14]. The creation of footprint leaves the researchers with challenging problems to optimize the matrix so as to improve the results/production. Not much work is done with regards to footprint analysis related to corrugated industry which motivated us to do the current study.

## 3. PROCESS MINING

The purpose of Process Mining is to discover, monitor and improve real process by extracting knowledge from event logs which are available in the information systems. Event logs are the stepping stone for process mining. Each event in a log is referred to an activity which is related to a particular case. The events belonging to a case are ordered and can be seen as one "run" of the process. Event logs may store additional information about events in the information system. The additional information stored in the system are the resource (*i.e.*, person or device), executing or initiating the activity, the timestamp of the event or data elements recorded with the event (*e.g.*, the size of a corrugated box). Event logs can be used to conduct three types of process mining[4] (*i*) discovery, (*ii*) conformance, and (*iii*) enhancement. Different types of techniques are used to build a process model.

## 4. EVENT LOGS

Event logs are the footprints left in the system by the execution of processes. Event logs are the key objects for any process mining technique to work. We rely on event logs to obtain different process models. An event log can be viewed as a multi set of traces [5]. The life-cycle of a particular case is traced in terms of activities executed. Process mining techniques use event logs to discover, monitor and improve real-life processes [5]. Event logs are the input for process mining techniques. At least three properties should be found in each event data: (*i*) data should have timestamps, (*ii*) activity labels should be present and (*iii*) case id of each record should be specified (case *id* is the *id* of each process instance). Therefore, there is a need to standardize the logging format of event logs. Three main process mining tasks are: Learning a process model from the behavior recorded in an event log is process discovery, for detecting and diagnosing the observed behavior and modelled behavior by aligning the event log and process model for conformance checking. Replaying the observed behavior on process models for identifying the bottlenecks, delays and inefficiencies in a process is called performance analysis The data should be a high quality event log to build an accurate model. The quality of the resulting model depends on the quality and volume of the behavior that has been observed.

**Event Log :** The starting point for any process mining technique is an event log, which is formally defined as [6]:

**(Event, Trace, and Event Log):** Let $E_L$ be the set of events occurring in the event log. A trace is a sequence $\sigma \in E^*_L$ of events. An event log $L \subseteq E^*_L$ is a collection of traces. Each trace corresponds to an execution of a process, *i.e.*, a case or process instance.

**Business Transactions*:*** A business transaction is in the context of analyzing system behavior where we can recognize a special type of trace [7]. It consists of a sequence of related events. Event logs are characterized by the relationships between activities. Casual footprint can be expressed using the dependencies between activities in an event log.

An event log is a sequence of events. Each event is associated with a set of attributes. The list of most common attributes in an event logs for the practice of process mining analysis [8] are:

**Case:** Process instance *id* of the event.

**Activity:** Name of the action performed in the event.

**Timestamp:** Moment of the event execution, establishing an order between the events.

**Resource:** Name of the resource initiating the event.

**Data:** Data attribute related to the event.

## 5. CONFORMANCE CHECKING USING FOOTPRINT

Conformance checking is a part of process mining analysis which is used to analyze the abstraction of the behavior in the log and an abstraction of the behavior allowed by the model. A footprint is a 2-D matrix showing causal dependencies between the activities. For example, the footprint of an event log may show that '*r*' is sometimes followed by '*s*' but never the other way around. If the footprint of the corresponding model shows that *r* is never followed by *s* or that *s* is sometimes followed by *r*, then the footprints of event log and model disagree on the ordering relation of *r* and *s*. Conformance checking is used for deviation detection, prediction, decision making and recommendation systems. In conformance checking, an event log is compared with its existing corresponding process model and it reveals that if process model conforms to reality and vice versa.

## 6. PROCESS DISCOVERY

Process discovery is one of the most challenging process mining tasks. A process discovery algorithm is a function that maps an event Log L into a process model that represents it.

### 6.1. Quality measures in process mining

The quality measures in process mining should be considered for construction and analysis of any model. The challenges for process discovery are classified into four competing forces [9] as follows:

1. **Fitness:** It is the ability to explain observed behaviour as how much a discovered process model is in accordance with a corresponding event log. A basic approach for calculating fitness is replaying all traces on the process model and calculating the number of cases that trace cannot be replayed on the process model.
2. **Simplicity:** More the discovered model is simple, the more desirable it would be. Various approaches of computing simplicity based on the number of activities and relations are as explained in[10].
3. **Precision:** It is to avoid under fitting. A model is precise if it does not allow too much unobserved behaviours in the event logs.
4. **Generalization:** It is to avoid overfitting problem. A model should have a minimum generalization and not be restricted to the behaviour that is seen in the logs. Because we should always consider that the event logs might be incomplete.

In many cases, these challenges are competing with each other and increasing one may cause decrease another (*e.g.*, a model that has high fitness might have low generalization). So, process mining algorithms should balance between these quality dimensions.

The idea of the **Alpha** algorithm is simple and used by many other process mining algorithms. The alpha algorithm scans the event log for particular patterns. For example, if in the event logs, activity *r* is always followed by *s* but *s* is never followed by *r*, then it is assumed that there is a causal dependency between *r* and *s* [2].

### 6.2. Basic ordering relationship for footprint

**Direct Succession:** $r > s$ iff for some case $r$ is directly followed by $s$

**Causality:** $r \rightarrow s$ *iff* $r > s$ and not $s > r$

**Parallel:** $r \parallel s$ *iff* $r > s$ and $s > r$

**Choice:** $r \ \# \ s$ *iff* not $r > s$ and not $s > r$

These relations are used to know the pattern in a process. Event logs should be scanned for extracting these relations. The result can be depicted as a 2-D matrix known as footprint matrix as shown in Fig 2.

## 7. METHODOLOGY

The corrugated industry in Chennai was visited and a thorough study on the working system was done. Related to the industry was collected and pre-processing of data is done to suit the standard process mining tool for process analysis. Based on the events handled by the industry a log is created and the Footprint for the log $L_1$ is derived from the basic ordering relationship. Suppose that $L_1$ is a log describing the history of six cases:

$$L_1 = [< A, B, C, D, F, G >^2, < A, B, D, F, G >^2, < A, B, E, F, G >^1, < A, B, C, E, F, G >^1]$$

The following event log $L_1$ is based on the sequence of industry process.

**Table 1**
**Corrugated Industry Event Log**

| *Job Identifier* | *Process Identifier* |
|---|---|
| Case 1 | Task A |
| Case 2 | Task A |
| Case 3 | Task A |
| Case 3 | Task B |
| Case 1 | Task B |
| Case 1 | Task C |
| Case 2 | Task B |
| Case 4 | Task A |
| Case 2 | Task C |
| Case 2 | Task D |
| Case 5 | Task A |
| Case 4 | Task B |
| Case 1 | Task D |
| Case 3 | Task D |
| Case 3 | Task F |
| Case 4 | Task D |
| Case 1 | Task F |
| Case 5 | Task B |
| Case 6 | Task A |
| Case 1 | Task G |

| *Job Identifier* | *Process Identifier* |
|---|---|
| Case 6 | Task B |
| Case 6 | Task C |
| Case 2 | Task F |
| Case 2 | Task G |
| Case 5 | Task E |
| Case 4 | Task F |
| Case 6 | Task E |
| Case 3 | Task G |
| Case 5 | Task F |
| Case 4 | Task G |
| Case 5 | Task G |
| Case 6 | Task F |
| Case 1 | Task G |

The event log in Table 1 show that all case begins with the execution of task A and ends with the execution of task G. The log contains information about six different cases which is about the process in the corrugated industry. Cases are 1, 2, 3, 4, 5, 6 and tasks are A, B, C, D, E, F, and G which has been executed. Case 1 and 2 has executed the task A, B, C, D, F, and G. Case 3 and 4 has executed the task A, B, D, F and G. Then case 5 has executed the task initially A, then B, E, F, and G. Finally the case 6 has executed the task A, B, C, E, F, and G. In some cases we can find that the tasks are following the same execution compared to other tasks. The industry manufactures corrugated boxes in various sizes and quantity. The event logs which are stored in the industry systems give us the details about the process. The main aim of process mining is to extract knowledge from event logs, build a suitable model and use the same for analysis. The different cases the tasks are A, B, C, D, E, F, and G. These tasks are referred to different process in the industry. Task A is to load paper, Task B is corrugation, Task C to Print, Task D is gluing, Task E is stitching, Task F is bundling, and Task G is Dispatch these are the different process which is performed in the industry. From the log $L_1$ we can derive the ordering of relationships in Table 2 to know the pattern of a process. The event log in Table 1 is scanned for extracting the different relations. These results can be shown as footprint matrix as shown in Fig 2.

## 8. ANALYSIS AND RESULTS

After following the methodology described in section VI, we have obtained the relationship table and the 2-D footprint depicting the work process of the corrugated manufacturing industry. The results obtained for the Log $L_1$ is generated based on the ordering relationships are given in the Table 2 below:

Table 2 describes the four different ordering relationships of event log $L_1$ which is called the footprint of the process. The table is divided into different cells based on the events and has any one of the following four relationships $\rightarrow$ , $\leftarrow$ , #**,** || either has a causality on one direction or causality in the other direction. The two activities never follow one another or they sometimes follow each other in one direction and sometimes in another direction.

**Table 2**
**Derived Relationship for $L_1$**

| A > B | A → B | ‖ | A # C |
|---|---|---|---|
| B > C | B → C | | A # D |
| B > D | C → D | | A # E |
| B > E | D → F | | A # F |
| C > D | E → F | | A # G |
| C > E | F → G | | B # F |
| D > F | | | B # G |
| D > F | | | C # A |
| F > G | | | C # F |
| | | | C # G |
| | | | D # A |
| | | | D # E |
| | | | E # A |
| | | | E # D |
| | | | E # G |
| | | | F # A |
| | | | F # B |
| | | | F # C |
| | | | G # A |
| | | | G # B |
| | | | G # C |
| | | | G # D |
| | | | G # E |

Using the footprint matrix particular patterns can be discovered easily. The concept of footprint matrix which is used in the alpha algorithm to construct a petri net model to analyze the process is being implemented to the corrugated industry data. Scanning the event log $L_1$ we obtained Fig 2 which explains the footprint matrix pertaining to the subject of study. The data being collected and analyzed based on the process mining techniques to help the corrugated manufacturing industry in improvising their productivity which is given as a 2-Dimensional matrix known as footprint. From Fig 2 we can observe that task A is always followed by task B and not the other tasks. If we consider task B either follows task C or task D ortask E. Then the case C follows the task D or task E never follows F. Task E and D always follows task F.

| $L_1$ | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| A | # | → | # | # | # | # | # |
| B | ← | # | → | → | → | # | # |
| C | # | ← | # | → | → | # | # |
| D | # | ← | ← | # | # | → | # |
| E | # | ← | ← | # | # | → | # |
| F | # | # | # | ← | ← | # | → |
| G | # | # | # | # | # | ← | # |

**Figure 2: Footprint for $L_1$**

## 9. CONCLUSION

Data from nontrivial process of the corrugated manufacturing industry is pooled up into the process. The event log is created and used for the analysis. The process mining techniques enable the researchers to obtain the results in the form of footprint to analyze the process. Unlike existing analysis approaches, process mining is process-centric (and not data-centric), truly intelligent (learning from historic data), and fact-based (based on event data rather than opinions). The main idea of this study is to use the discovered footprint as an objective start point to deploy systems that support the execution of corrugated industry processes or as a feed-back mechanism to check if the prescribed step wise process fit the executed ones.

## 10. ACKNOWLEDGMENT

One of the authors Mrs. Aruna Devi T., acknowledges Dayananda Sagar Institutions, Bangalore, Karnataka and Rayalaseema University, Andhra Pradesh, India for providing the facilities for carrying out the research work.

## REFERENCES

[1] W.M.P. van der Aalst, A.J.M.M. Weijters, and L. Maruster, "Workflow Mining: Discovering Process Models from Event Logs", IEEE Transactions on Knowledge and Data Engineering, 16(9):1128–1142, (2004).

[2] W.M.P. van der Aalst, M., Song, "Mining Social Networks: Uncovering interaction patterns in business processes", In: Diesel, J. Pernici, B, Weske, M. (eds.) BPM 2004. LNCS, vol. 3080, pp. 244–260, Springer, Heidelberg (2004).

[3] W.M.P. van der 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 and Knowledge Engineering, 47(2):237–267, 2003.

[4] W. M. P. van der Aalst, and A. J. M. M. Weijters, "Process mining: a research agenda, Computers in Industry", Volume 53, Issue 3, Process / Workflow Mining, April 2004, pp. 231244, ISSN 0166-3615, DOI: 10.1016/j.compind.2003.10.001.

[5] W. M. P. van der Aalst, "Process Mining: Discovery, Conformance and Enhancement of Business Processes", Springer-Verlag, Berlin, 2011.

[6] W. M. P. van der Aalst, K. A., V. Rubin, and H. M. W. Verbeek, "Process discovery using localized events," 2015, to appear in Petri nets in 2015.

[7] Agrawal R., Gunopulos D., Leymann F., "Mining process models from workflow logs", In: Sixth International Conference on Extending Database Technology, pp. 469–483 (1998).

[8] Van der Aalst, W.M.P., Van Hee, K.M, "Workflow Management: Models, Methods, and Systems", MIT Press, Cambridge, 2002.

[9] IEEE Task Force on Process Mining, "Process Mining Manifesto", In F. Daniel, K. Barkaoui, and S. Dustdar, editors, Business Process Management Workshop, volume 99 of Lecture Notes in Business Information Processing, pages 169–194. Springer, Berlin, (2012).

[10] W. M. P. van der Aalst, "Process Mining - Data Science in Action", Second Edition. Springer 2016, ISBN 978-3-662-49850-7, (2016), pp. 3-452.

[11] Felix Mannhardt, Massimiliano de Leoni, Hajo A. Reijers, Wil M.P. van der Aalst, Pieter J. Toussaint, "From Low-Level Events to Activities A Pattern-based Approach", In La Rosa, Marcello and Loos, Peter and Pastor, Oscar (editors), Business Process Management: 4th International Conference, BPM 2016, Rio de Janeiro, Brazil, September 18-22, 2016. Proceedings, 2016, pages 124-141.

[12] Caron F, Vanthienen J, Vanhaecht K, Limbergen E.V, de Weerdt J, Baesens B, "Monitoring care processes in the gynecologic oncology department", Comput Biol Med, 2014, 44:88–96, doi: 10.1016/j.compbiomed.2013.10.015.

[13] Arik Senderovich, Sander J.J. Leemans, Shahar Harel, Avigdor Gal, Avishai Mandelbaum, Wil M.P. van der Aalst. Discovering Queues from Event Logs with Varying Levels of Information. In *Business Process Intelligence* 2015, Innsbruck, Austria.

[14] de Leoni, M, Suriadi, S, Arthur H. M. ter Hofstede, Aalst van der W.M.P, "Turning Event Logs into Process Movies: Animating What Has Really Happened", 28 September 2014 , DOI: 10.1007/s10270-014-0432-2.