A Process Mining Based Approach to Complex Manufacturing Process Flow Analysis: A Case Study

Sook-young Son

Department of Management Engineering

Graduate School of UNIST

2014

A Process Mining Based Approach to Complex Manufacturing Process Flow Analysis: A Case Study

A thesis

Submitted to the Graduate School of UNIST

In partial fulfillment of the

Requirements for the degree of

Master of Science

Sook-young Son

06.23.2014

Approved by

Major Advisor

Minseok Song

A Process Mining Based Approach to Complex Manufacturing Process Flow Analysis: A Case Study

Sook-young Son

This certifies that the thesis of Sook-young Son is approved.

06.23.2014

Signature

Thesis supervisor: Minseok Song

Signature

Duck-Young Kim

Signature

Daeil Kwon

ABSTRACT

With recent advances in IT infrastructure in manufacturing environments, a large amount of manufacturing data are collected and stored in a database at various stages of production. These data may include valuable information for manufacturing companies to improve their manufacturing processes. The method of manufacturing data analysis is crucial for understanding the manufacturing data. However, traditional manufacturing data analysis methods such as data mining, simulation, etc. have limitations for this purpose since those are difficult to provide overall process-level information. Therefore, in this thesis, a process mining based approach for analyzing complex manufacturing processes is proposed. Process mining is a useful tool for process-related knowledge acquisition since it enables users to derive not only manufacturing process models, but also several performance measures related to processes, resources, and tasks. This thesis suggests a framework for the manufacturing process analysis. To do this, it applies process mining techniques to perform four types of analysis, which are visualization of production flows, machine-to-machine inter-relationship analysis, machine utilization, and monitoring & diagnosis of task performance regarding yield rate and lead time. Furthermore, a case study is conducted to support the proposed framework with an event log of an electronic components manufacturing process.

Keywords: Process Mining, Manufacturing Process Analysis, MES, Case study, Framework

Contents

List of Figures

List of Tables

I. Introduction

1.1. Motivation and Background

Manufacturing processes are defined as all the steps of transforming raw materials, components, or parts into final product that meet a customer's expectations [1]. With recent advances in IT infrastructures in manufacturing environments (e.g., RFID and sensor technologies), a large amount of data related to tasks, products, machines, and workers are collected and stored in database at various stages of production. The huge amount of data may include valuable knowledge for manufacturing companies to improve their manufacturing processes. For example, detecting defects (e.g., bottleneck point) of manufacturing processes or finding a problem in resource management can be a typical example of the knowledge. Thus, it is important for the companies to analyze the large amount of data to extract, capture, understand, and use the knowledge.

Such analysis for the manufacturing processes improvement is called manufacturing process analysis that includes modeling and analysis of manufacturing processes. It usually establishes a process model to understand the structure of manufacturing, and reveals relationships among machines, tasks, and materials. (Lin et al. 2009). According to the several review papers in manufacturing process analysis, the most commonly used approaches are data mining (e.g., association, classification, prediction, and clustering) and simulation. However, data mining have limitation to provide overall process level analysis results, and the results are sometimes too complex to understand. Moreover, simulation has limitation that it takes too much time to build a complex manufacturing process model. Therefore, there is a need for intelligent manufacturing data analysis methodology that may be useful to discover process-level knowledge faster and easier. This thesis will suggest a methodology by applying process mining that is useful to discover a manufacturing process model from the event logs. The methodology is supported by MES (Manufacturing Execution System) for the data extraction and data analysis.

---

[1] http://www.chegg.com/homework-help/definitions/manufacturing-processes-5

Process mining is able to extract process-oriented knowledge from event logs (Van der Aalst et al. 2004; Van der Aalst et al. 2007) extracted through the MES. Process mining provides a lot of analysis techniques, and some of them can be useful to manufacturing process analysis. Firstly, process mining has various mining tools that discover a manufacturing process model providing an insight of actual manufacturing processes. Moreover, it is able to perform further performance analysis for the discovered model e.g. bottleneck analysis and conformance checking. Furthermore it is useful to show the utilization of machines or workers.

1.2. Objective

This thesis aims to propose a framework for manufacturing process analysis to extract valuable knowledge for manufacturing process monitoring and defect diagnosis to improve current manufacturing environment. The framework consists of four phases: *data preparation*, *data preprocessing*, *manufacturing process mining analysis*, and *Interpretation and Evaluation*. In *manufacturing process mining and analysis* step, we applied process mining techniques to perform the following types of analysis.

1. *Manufacturing process visualization*. Hong & Soon (2001) defined *visualization* as a method that presents a manufacturing process model to give domain experts visible information, e.g., how products are transformed and where is the bottleneck point. Manufacturing process visualization is essential to generate efficient manufacturing processes through improvement of production defects such as bottleneck and abnormal flows. Simulation has been common in use to visualize manufacturing processes, which allows users to examine the complex processes in real-time and from different aspects (Zhong and Shirinzadeh, 2008). However, it requires a lot of time and cost to discover a complex process model, thus this thesis will suggest a way of discovering an actual process model from the event logs by applying process mining techniques. Moreover, we will suggest a way of enhancement of manufacturing processes by comparing an existing model with actual processes.

2. *Machine to machine interrelationship analysis*. To improve manufacturing processes, it is important to visualize logistics flows among the resources (e.g., machines and human resources). It is important to find an essential resource (leading machines) for the overall logistic processes. This thesis applied social network analysis to generate machine network model that shows relationship among the resources.

3. *Machine utilization*. This analysis includes structuring the relationship between resources and tasks regarding their working frequency, and analyzing machine idle time & working time. Machine allocation analysis will use organizational miner to structure "who or which" resource worked for an activity, and to examine resource allocation in respect of working frequency. It will be used to improve resource allocation for the better manufacturing processes. In addition, we used several performance analysis techniques to monitor the current resource operations about their idle time.

4. *Monitoring and Diagnosis of task performance*. Various literature researches have suggested many different performance indicators such as cycle time, through put, yield rate, and lost rate. This thesis aims to monitor and diagnosis the production ability for each task about their yield rate and lead time. Production yield rate is important to reduce their lost rate that related to production cost. Moreover, production lead time is a major determinant of production speed. Thus, it is essential to examine their yield rate, working, and waiting/transformation time for each task.

1.3. Outline of the thesis

The remainder of this thesis is organized as follows. We provide a brief introduction of the state-of-the-art of manufacturing process management in regards of data analysis, process mining, and MES in section 2. Section 3 lists the questions we would like to answer and suggests the way of measurements of analysis results and presents the main idea of process mining framework for manufacturing process analysis. Section 4 describes the case study with event logs of electronic components manufacturing processes. As a result, section 5 concludes the paper with limitations of the research and future works.

II. Related Works

2.1. Manufacturing Process Management with Data Analysis

2.1.1. Simulation for Manufacturing Data Analysis

Because of the complexity and dynamic behavior of manufacturing systems, simulation becomes popular methods of facilitating their designed and assessing operation strategies (Rajes et al. 2008). According to the Heilala, J. (1999), simulation in manufacturing has several advantages such as diagnose problems, identify constrains, visualize the plan, and explore possibilities (Heilala, 1999). With a lot of benefits of simulation in manufacturing process management, simulation has been successfully adopted in manufacturing system design and operation (Ashkan and Jeffrey, 2014). Ashkan & Jeffrey (2014) provided 290 review papers on simulation applications in manufacturing systems. For example, Chan et al (2002) used simulation for Flexible Manufacturing System (FMS) scheduling (Chan et al. 2002) and Allahverdi et al (2008) applied simulation for scheduling problems with setup times or costs (Allahverdi et al. 2008).

However, simulation modeling and analysis can be time consuming and expensive since building a model requires specific training (Heilala, 1999). Additionally, it is hard to set a standard input data for the simulation model, thus it is possible to generate incorrect analysis results depending on the input data. Process Mining discovers a process model based on the event logs collected in information systems. Thus it is easy to extract manufacturing process model quickly and correctly.

2.1.2. Data mining for Manufacturing Data Analysis

Due to the large amount of collected manufacturing data, extracting knowledge from the data becomes important for optimization purpose (Wang et al. 2007). Therefore, recently, more and more manufacturing enterprises have trying to use data mining for problem solving and enhancing their capability (Harding et al. 2005). Several researches already conducted literature review for data mining in manufacturing, and they addressed some fields of application with data mining techniques (Wang et al. 2007; Harding et al. 2005; Polczynsk and Kochansk, 2010; Wang, 2006). Dan Braha addressed several domains of data mining in manufacturing data analysis, which are fault diagnosis, preventive machine maintenance, manufacturing knowledge acquisition, operational manufacturing control, quality and process control, and so on (Braha, 2002). Additionally, Harding (2006) reviewed applications of data mining in manufacturing, and it summarized some applications in operations, fault detection, maintenance, decision support, and so on. For example, Chen et al (2004) used association rules for defect detection by determining the association between different machines and their combination with defect (Chen et al. 2004).

According to the researches on review of application of data mining in manufacturing, data mining techniques are helpful to enhance the capability of manufacturing enterprises by providing useful knowledge for better decision. However, there are some limitations on data mining in manufacturing data analysis. J.A. Harding (2006) addressed that the discovered knowledge using data mining sometimes is too complex to understand, so they need to effort to enhance the expressiveness of the knowledge. (Harding et al. 2006). To solve the limitation, this thesis is going to apply process mining techniques to analyze manufacturing data analysis since process mining is useful to visualization.

2.2. Process Mining

2.2.1. Overview of Process Mining

Process mining provides process-related useful information by analyzing event log data stored in information systems (Van der Aalst et al. 2004; Van der Aalst et al. 2007). Process Mining is on the intersection between Business Process Intelligence (BPI) and Business Activity Monitoring (BPM) (Song and Van der Aalst. 2008; Jochen et al. 2013). BPI supports business by providing analysis, prediction, monitoring, control, and optimization features (Grigori et al. 2004). And BPM can be defined that it supports business processes using methods, techniques, and software (Weske et al. 2004). Unlike BI and BPM, process mining aims to provide insight of processes, for example, “where is bottleneck point?” or “how the lots are flowed?”

Process mining consists of three types, which are *discovery, conformance*, and *enhancement*. A *discovery* technique takes an event log and produces a model without using any information. Alpha miner, Heuristic miner, Fuzzy miner, and Comp miner are the example of discovery techniques (Van der Aalst et al. 2004; Weijters et al. 2003; Günther et al. 2007). Next *conformance* compares an a-priori model with the observed behavior as recorded in the logs (Song and Van der Aalst. 2008). It can discover potential cases of fraud by scanning the event log (Van der Aalst. 2011). Lastly, the idea of *enhancement* is to extend or improve an existing process model using information about the actual process recorded in some event log (Van der Aalst. 2011).

Process mining has four kinds of perspectives: (1) Control-flow, (2) Organizational, (3) Case, and (4) Time. Control-flow perspective focuses on the control-flow with aiming of finding a good characterization of all possible paths, e.g., expressed in terms of a Petri-net (Reisig and Rozenberg. 1998; Van der Aalst. 2011). Organizational perspective focuses on information about resources such as workers, machines, and department and how they are related. And it has a goal of structuring the organization by classifying resources in terms of roles and organizational units or to show the social network (Scott, 1988; Wasserman and Fust, 1994, Van der Aalst et al. 2005). Case data perspective focuses on the features of cases. Cases can be characterized by their path in the process or by the originators working on a case. (Van der Aalst, 2011). Time perspective focuses on the timing and frequency of events. It can be used to discover bottlenecks, measure service level, monitor the utilization of resources, and predict the remaining processing time of running cases (Van der Aalst, 2011).

2.2.2. ProM framework

Process mining has several useful analysis tools such as ProM and Disco. ProM supports a lot of process mining techniques in the form of plug-in, and use MXML file as input format. Fig. 1 shows a screen shot of ProM. The ProM framework has developed as a plug-in environment, so it enables rapid development of new algorithms and techniques as plug-ins (Van Dongen et al. 2005; Song et al. 2008).

The architecture of ProM consists of five types of plugins, which are mining, export, import, analysis, and conversion. Mining plug-ins has some mining algorithms such as heuristic and fuzzy mining. Export plug-ins has function as “save as” for some objects. Import plug-ins has a function of “open”. Analysis plug-ins has some property of analysis, such as conformance checking and LTL checker. Lastly, Conversion plug-ins converts different data format, for example it converts format from EPCs to Petri nets (Van Dongen et al. 2005). The plug-ins support three kinds of analysis perspective (e.g., process, resource and task) as mentioned in the previous section.

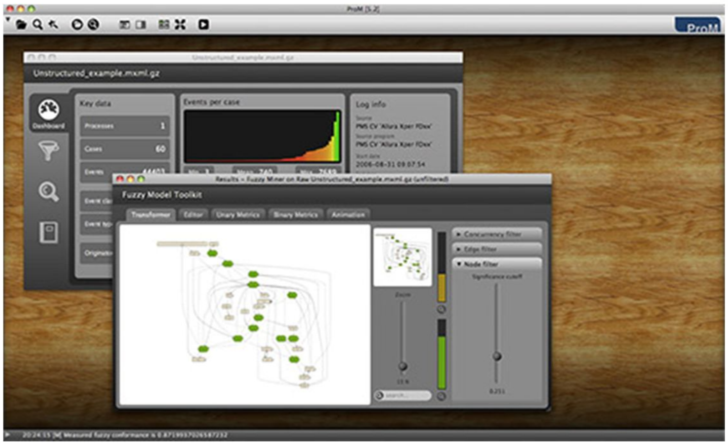


Figure 1: Screenshot of the ProM framework[2]

2.2.3. *Application of process mining*

Traditionally, process mining has been focusing on control-flow perspective (Van der Aalst et al. 2004; Wen et al. 2006; Weijters and Van der Aalst, 2001). However, process mining techniques is not limited for control-flow perspective, e.g., organizational and case perspective. There has been an increase in the application of process mining in real-environment industries regarding the various perspectives. Table 1 shows challenges of process mining in industry application. Jochen De Weerdt (2013) suggested a methodology framework of process mining in financial service organization. Alvaro Reburge (2012) introduced a methodology for the application of process mining techniques in healthcare environments. Additionally, Ying Wang (2014) presented a comprehensive methodology for the bulk port. There have been some studies for service industry, but only few research works have paid attention to manufacturing process analysis by applying process mining techniques, e.g., Lee (2013) suggested a methodology for the application of process mining in after-assembly block manufacturing process in the shipbuilding industry and Rozinat (2009) applied process mining techniques to discover a manufacturing process model of wafer scanner production processes.

The starting idea of process mining application in the manufacturing processes is to mine process models from the event logs. Process discovery has numerous results, i.e. discovering process models based on observed events, and aggregating process models based on frequency (Weijters et al. 2006). Not only discovering process model, but process mining approach has other functionalities to show the relationships of resources (i.e. machine, human), and build the organizational view of an enterprise (Van der Aalst et al, 2005). Moreover, in order to enhance the discovered model, it is

---

[2] Available in http://fluxicon.com

necessary to conduct further analysis about manufacturing processes efficiency indicators such as bottleneck point and defective production. In the section 5, there will be detail process mining application results.

Table 1: Challenges for process mining in manufacturing organizations

Year	Authors	Perspective	Application	Description
2009	**Rozinat et al.**	Control-flow	Manufacturing	In this paper, we investigate the applicability of process mining to less structured processes. And We report on a case study on the test processes of ASML (the leading manufacturer of wafer scanners in the world)
2012	**Stuit and Wortmann**		e-mail	Methodology of business process modelling language to visualize the discovered e-mail-driven business process
2012	**Rebug and Ferreira**		Healthcare	Methodology of leading to the identification of regular behaviour, process variants, and exceptional medical cases.
2012	**Damer et al.**		Finance	Methodology which first clusters the event log into homogeneous groups of event traces and then computes the compliance degree for each cluster separately.
2013	**Caron et al.**	Overall	ERM	Methodology for the context of the eight components of the COSO Enterprise Risk Management Framework.
2013	**De Weerdt et al.**	Overall	Finance	Methodology for a multi-faceted analysis of real-life event logs based on Process Mining.
2013	**Lee et al.**	Control-flow	transportation logs	Methodology for automatically extracting the most frequent task flows from transport usage histories.
2013	**Mans et al.**		Healthcare	Methodology for evaluating the impact of IT on a busines s process.
2013	**Jans et al.**	Overall	Auditing	Making a case for why internal and external auditors should leverage the capabilities process mining offers to rethink how auditing is carried out.
2013	**Fernandex-Gallego et al.**	Control-flow	3D educational vitual world	Methodology for 3D educational virtual worlds that focus on discovering learning flows and checking its conformance through process mining techniques
2014	**Wang et al.**	Control-flow	Port	Methodology for applying process mining in logistics, covering the event log extraction and pre-processing as well as the execution of exploratory, performance and conformance analyses.

2.3. MES (Manufacturing Execution System)

Manufacturing Execution Systems (MESs) are defined by MESA (Manufacturing Execution Systems Association) international as followed: MESs deliver information that enables the optimization of production activities from order launch to finished goods (MESA. 1997). The MES bridges the gap between the planning system and the controlling system using on-line information to manage the current application of manufacturing resources: people, equipment and inventory (McClellan, 2001).

The MES has evolved in useful software applications as computing technologies, and they have been advanced since the in the mid-1990s (Saenz de Ugarte et al. 2009). It helps manufacturing experts to make a better decision by providing real-time feedback to reduce the gap between the production plan and actual performance (Qiu and Zhou, 2004). According to MESA's survey, MESs provided benefits to manufacturing enterprises such as an average 45% reduction in manufacturing cycle time (MESA. 1997). Additionally, Douglas Scott introduced a survey of the MES benefits conducted by MESA International, and it shows that MES has benefits such as reducing manufacturing cycle times, reducing or eliminating data entry time, and so on (MESA. 1997; Douglas Scott et al, 1996; Michael McClellan, 2001; Jurgen Kletti, 2007). Thanks to those a lot of benefits, several manufacturing enterprises have adopted MESs as the solution to manufacturing cost competitiveness (David, et al. 1995)

The MESA International has identified 11 principal MES functions, which are Operation/Detail Scheduling, Resource Allocation and Status, Dispatching production units, Document control, Production tracking and Genealogy, Performance analysis, Labor Management, Maintenance Management, Process Management, Quality Management, and Data collection/Acquisition (MESA. 1997). Traditionally, production data collection systems using either databases or spreadsheets are commonly developed on the shop floor to monitor and control real-time and variable execution processes, but maintenance and data consolidation is obviously complex in such an environment as the number and the structure of these small applications vary over time (Saenz de Ugarte et al. 2009). Now, data collection function of MES can solve the difficulty. This thesis mainly focuses on the performance analysis function of the MES. We are going to suggest manufacturing process analysis framework based on the event data gathered through the MES. As well as the performance analysis function, we focus on the data collection/acquisition function.

III. A Process Mining Framework for Manufacturing Process Analysis

3.1. Problem statements

Manufacturing process analysis provides modeling and performance analysis of a production process, which is essential for manufacturing companies to improve their market competition (Lin et al. 2009). Several methodologies, models, and tools have been developed for the manufacturing process analysis to generate data modeling, simulation, decision-making support, expert systems, and standard model (Hernandez-Matias et al. 2006). However, those established manufacturing process analysis methodologies, tools, and models have limited about what is actually happening. It is important to capture the actual manufacturing processes, since it is possible to be a significant gap between what is supposed to happen and what actually happened (Song et al. 2008). Process mining can extract a manufacturing process model from real event log data with aim of "what is really going on" (Reijers et al. 2007; Tiwari et al. 2008; Van der Aalst et al. 2003). Additionally, process mining provides a lot of analysis tools as well as the mining algorithms. Therefore, this thesis is going to suggest a framework of manufacturing process analysis by applying process mining techniques.

Generally, domain experts in manufacturing industry have several questions about the manufacturing process analysis. Among the questions, this thesis is going to answer the following four questions.

- Question 1: *What are the flows of lots (part flows) in the current manufacturing processes?* For the manufacturing enterprises, it is critical to capture the current statue of the manufacturing processes. In the regard to this, domain experts want to make sure what is going on the processes. Process mining can generate a process model from event logs, and it is useful to visualize actual flows including reworks and unexpected processes. Additionally, it is possible to visualize all possible production patterns. Conformance of standard model will compare the standard model (or reference model) with event logs. Moreover, we can find bottleneck point based on the generated model, thus the experts can make a decision of reducing the through put time.

- Question 2: *What is the interrelationship between resources (e.g., machines) about the flows of lots?* As lots are moved through the manufacturing processes, the event logs are automatically collected on the MES. The event logs contain precedence relationship among the resources. Hence, it is possible to generate social network model and structure their relationships about the handover the works.

- Question 3: *What is the performance of each resource?* Resources (e.g., machines, human resources) are important elements of manufacturing processes. Hence, it is essential to analyze the performance of each resource. Firstly, we will structure the relationship between resources and tasks to reduce failure and improve resource operation. Second, performance analysis of each resource with respect of idle time is required to know how long resources waited.

- Question 4: *What is the performance of each resource with respect of its lead time and production lost rate?* Manufacturing enterprises want to reduce their production cost to increase their profit. Thus, it is important to reduce the production cost by reducing loss rate and total lead time. Lead time is time interval between start and end time. And production lost rate is ratio of output on input quantity. Performance analysis of each task can provide information about the production lead time and lost rate.

Table 2 shows relationship between the four questions with manufacturing process analysis, which is organized according to the three kinds of perspectives on process mining.

Table 2: Process mining techniques for manufacturing process analysis

Q	Perspective	MPA	Analysis	Goal
Q1	Control-flow (Process)	Manufacturing Process Visualization	Actual flow discovery	To discover a manufacturing process model and visualize rework flows.
			Conformance of standard model	To compare the standard model with event logs by calculating fitness.
			Visualization of bottleneck point	To visualize bottleneck point on the process model and compare the bottleneck points among different production patterns.
Q2	Organizational (Resource)	MTM Inter-relationship Analysis	Machine network analysis	To examine inter-relationship between machines
Q3	Organizational (Resource)	Machine Utilization	Machine allocation analysis	To structure relationship between machine and task and compare working frequency of machines for each task
			Machine idle time analysis	To conduct performance analysis for each machine
Q4	Task (Task)	Monitoring and Diagnosis of task performance	Lead time monitoring	To monitor performance with respect of lead time
			Yield monitoring	To monitor performance with respect of yield rate

3.2. A Process Mining Framework

This section explains a framework for the manufacturing process analysis by applying process mining techniques. Fig. 2 shows the framework consisting of four major steps: *data preparation*, *data preprocessing, manufacturing process mining and analysis*, and *evaluation and interpretation*. In the data preparation step, manufacturing event logs are collected using information systems (e.g., MES). Next, data pruning and filtering should be done in the data preprocessing step. After that, the refined data is converted into a standard data form for the analysis step, i.e. MXML or XES. Then, several process mining techniques are applied according to the three perspectives: *process*, *resource*, and *task*. In the process perspective, we can discover a manufacturing process model, conformance the standard model, and visualize bottleneck point on the discovered model. Those analyses will support manufacturing process visualization. And resource perspective analysis consists of machine-to-machine inter-relationship analysis and machine utilization that includes machine allocation analysis and machine idle time analysis. Lastly, task perspective focuses on monitoring and diagnosis of task performance by monitoring production time and production yield rate. As a result, the analysis results should be evaluated and interpreted by domain experts. Then, they can improve the existing processes based on the interpreted results. The detailed explanations of the framework are described in the followed.

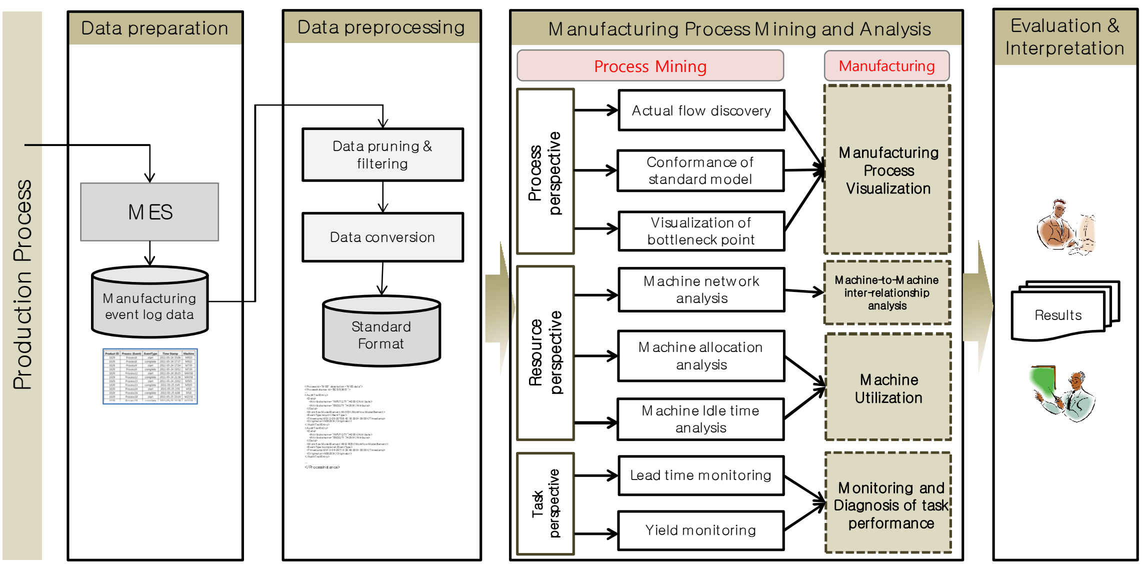


Figure 2: A Process mining framework

3.2.1. Data preparation

A manufacturing process consists of the sequence of activities to make final products from raw materials. The activities are performed by manufacturing resources including human resources, machines, and software. For process mining analysis, event log containing a set of cases (i.e. process instances) is required. Each case consists of a sequence of events with various attributes such as activity name, event type, timestamp, resource name, event data, and so on. Table 3 shows a fragment of manufacturing processes with six cases and six activities including resource attributes.

Table 3: Example of traces with activity and resource information

Case	Traces (activity, resource)
1	(A, M1), (B, M3), (C, M5), (D, M6), (E, M8), (F, M9)
2	(A, M2), (C, M4), (D, M7), (E, M8), (F, M10)
3	(A, M1), (C, M5), (D, M7), (F, M9)
4	(A, M1), (D, M6), (E, M8), (F, M10)
5	(A, M2), (B, M3), (D, M6), (E, M8), (F, M10)
6	(A, M2), (D, M6), (F, M10)

Manufacturing process has specific process ID and process name (we refer to all processes in the manufacturing as "activities" in the later part of this paper). In the manufacturing processes, it is possible to record the event type that identifies the transactional information at a given point in time. For example, starting to do Task_A would be a start event. Finishing the Task_A is a complete event. The time (i.e. timestamp) of the event is to show the order of the events in a case. The event also contains information about name of resources and lot number. Resource is an entity (e.g. machine or performers) that performed an activity. Other data attributes e.g. a set of product quantity (input or output quantity) can provide additional analysis for managers. Table 4 shows the essential information of event log for process mining with an example of manufacturing process.

Table 4: Essential information of event log for process mining

Attribute name in process mining	Attribute name in manufacturing	Data example
Case	Lot ID	BC4052849
Activity	Process ID	ABG0490
	Process Name	Task_A
Event Type	Start	Start
	Complete	Complete
Timestamp	Start Time	2012-05-02 05:16:17
	Complete Time	2012-05-02 05:16:19
Resource	Machine ID	N08032
	Performer ID	J0253B
Data	Input Quantity	6710
	Output Quantity	6710

By looking at the manufacturing event log data in Table 4, we can formalize the event log as follow.

Definition 1. Event Log

An event log, denoted as L, consists of a set of process instances or cases, and each case is described by a sequence of events. The sequence of events contains behaviors of activities to indicate the flow of activities from beginning until the end. Thus, an event log is a tuple of <E, C> which is defined as follows:

- **Event.** $E = A \times Y \times R \times T \times D$ is a set of events, where A is a set of activities, Y is a set of event type (start and complete), R is a set of resources, T is a set of timestamps, and D is a set of quantity data.
- **Case.** C is a set of event for its one instance where a collection of all cases is an event log (L). Cases always have a trace, denoted as $\sigma \in E^*$.$C=E^*$ is the set of possible event sequences (traces describing a case). Then, $L \in \mathcal{B}(C)$ is an event log. Note that $\mathcal{B}(C)$ is the set of all bags (multi-set) over C. Each element of L denotes a case.
-

For any event e E, $\pi(e)$ is the value of an attribute for event e. If event e does not have a designated attribute named, then $\pi(e) = \perp$ (null value). Moreover, each attribute has each mapping function. For example, $\pi A(e)$ is the activity associated to event e, $\pi Y(e)$ is the transaction type associated to event e, $\pi R(e)$ is the resources associated to event e, $\pi T(e)$ is the timestamp associated to event e, $\pi C(c)$ is the trace associated to case c, and $\pi D(e)$ is the quantity data associated to event e. For example, if e = (Task_A, start, N08032, 2012-05-02 05:16:17) then $\pi A(e)$='Task_A', $\pi Y(e)$='start', $\pi R(e)$= 'N08032', and $\pi T(e)$='2012-05-0205:16:17'.

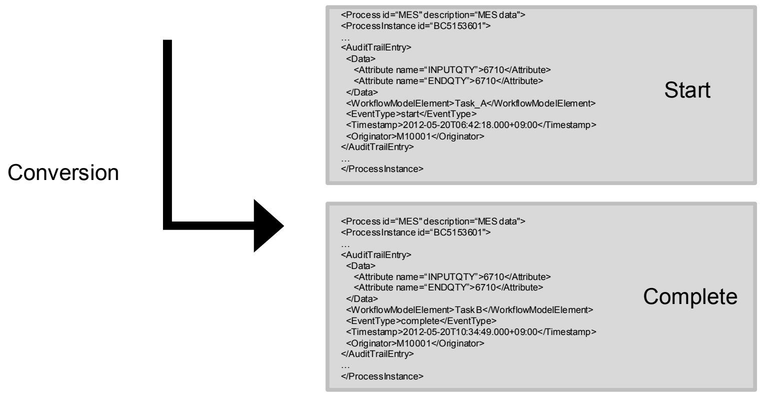
3.2.2. Data preprocessing

Since the raw data may be incomplete and inconsistent, data preprocessing step is required to improve the data quality. In this thesis, data preprocessing consists of two steps: data cleansing and data conversion. The first step, data cleansing, involves removing redundant data, revising errors, and smoothing out noise. Before the data cleansing, it is essential to set a rule about the noise data since one of our objectives is visualize actual process from the event logs, some noise data may be important to provide insights. For example, a manufacturing process should have one start activity and one end activity or *imaginary activity* should be excluded from the analysis.

After data cleansing step, data conversion step transforms the data into appropriate form for process mining such as MXML (Mining eXensible Markup Language) and XES. Since process mining tool, ProM, requires the MXML as a format of input file, data conversion is necessary. *ProMimport* is a tool supporting data conversion of different data sources (e.g., CSV, Text) to MXML or MEX. The MXML has a standard notion for storing case, activity, resource, timestamp, and data attribute. Fig. 3 (a) shows an example fragment of the manufacturing event log. Each line corresponds to the execution of one activity. The number at the beginning of the line identifies the LotID (i.e. a product lot) that is executed. Afterward, the process ID, the process name, the machine ID, the performer ID, the start time, the complete time, the input quantity and the output quantity for the executed product are recorded. Depending on the kind of analysis, we are able to convert the log into XES log fragment (see Fig. 3(b)) for the highlighted event in Fig. 3(a). A separate audit trail entry is created for each event in which each activity name has specific event type.

Case	Activity	Resource		Timestamp		Data	
LOT_ID	TASK_ID	MACHINE	WORKER	START_TIME	END_TIME	INPUTQTY	ENDQYT
BC5153601	Task_A	M10001	W10001	2012-05-02 5:16	2012-05-02 5:23	6710	6710

(a) Example fragment of manufacturing event log



(b) Example fragment of XES file

Figure 3: A fragment of a log in XES format

3.3. Manufacturing Process Mining and Analysis

This section describes the application of process mining techniques to analyze manufacturing process according to the three perspectives: process, resource, and task. As explained in the framework, the analysis results will be interpreted (or used) as followed in the manufacturing industry: manufacturing visualization, machine-to-machine interrelationship analysis, machine utilization, and monitoring & diagnosis of task performance.

The process perspective focuses on the control-flow, i.e., the ordering of activities. The objective of the process perspective is visualization of manufacturing processes expressed in terms of specified format, e.g. graph. It also attempts to compare a reference model (a-priori model) with the observed behavior as recorded in the log. Next, the resource perspective focuses on the resource entity, i.e. machine-to-machine interrelationship analysis that finds how resources are related and machine utilization that exams individual performance of each machine as well as mining the organization structure of resources in terms of organizational units. Lastly, the task perspective focuses on properties of activities, especially, take performance analysis. A specified activity can be characterized by the time spent on a case and product quantity as the execution result. Process mining has a lot of useful techniques for the manufacturing process analysis. Among them, following table 5 shows summary of process mining techniques can be applied for the three kinds of perspectives.

Table 5: Summary of process mining techniques for manufacturing process analysis

Perspective	**Analysis**	**PM techniques**										
		HM	FM	CC	PAWP	PA	SNA	OM	RbT	DC	BPA	EDAV
Process perspective	Actual flow discovery	●	●									
	Conformance of standard model			●		●				●		
	Visualization of bottleneck point				●							
Resource perspective	Machine network analysis						●					
	Machine allocation analysis							●	●			
	Machine idle time analysis									●	●	
Task perspective	Lead time monitoring									●	●	
	Yield monitoring	●										●

HM: Heuristic Miner, FM: Fuzzy Miner, CC: Conformance Checker, PAWP: Performance Analysis With Petri-net, PA: Pattern Analysis, SNA: Social Network Analysis, OM: Organizational Miner, RbT: Resource by Tasks, DC: Dotted Chart analysis, BPA: Basic Performance Analysis, EDAV: Event Data Attribute Visualizer

3.3.1. Process perspective: Manufacturing Process Visualization

This section aims at visualization of manufacturing processes, as a part of the process perspective analysis. For the purpose, we are going to discover an actual manufacturing process model, derive some patterns based on the event log. It aims construct a manufacturing process model that reflects current situation from enactment logs. Additionally, we will check a conformance of standard model with event logs by calculating a fitness value. To calculate the conformance, we will apply conformance checking of ProM plug-in, which focuses on finding the discrepancies between the process model and the event log. Lastly, we will visualize bottleneck points leading longer production time on the discovered process model and compare the bottleneck points among the different production patterns.

Actual flow discovery

Actual flow discovery is considered as of the most challenging process mining tasks (Van der Aalst, 2011). A manufacturing process model is constructed based on an event log, thus it captures the actual behaviors seen in the event log. Fig. 4 (b) shows a discovered process model (e.g. Petri-net) from a log traces in Fig. 4 (a) using a naive α-algorithm based on Petri net (Van der Aalst et al. 2004).

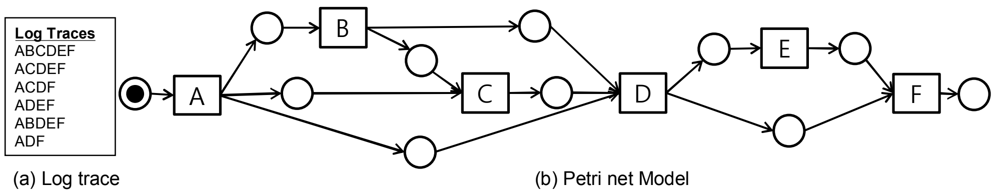


Figure 4: (a). Log traces, (b). Process model using Petri Net based on the log traces

There are a number of algorithms concerned with actual process discovery. This thesis chooses heuristic mining and fuzzy mining to construct the process model. Heuristic mining takes frequencies of events and sequences into account when constructing a process model. Moreover, it includes the discovery of control flow behaviors (i.e. sequence, AND and OR) which comes from the causal dependency of the events. In the case of fuzzy mining, it has advantages on finding abstraction process model when the process is complicated (i.e. there are a large number of activities). In addition, fuzzy mining can highlight significant information by visual means, e.g. animation.

First, we discuss the mechanism of heuristic mining on process discovery. The starting point of the heuristics mining is the construction of dependency graph. To obtain the dependency graph, it is required to consider a trace that contains the ordering of the events. If an event is always followed by another event, it is likely that there is a dependency relation between both events. To analyze the dependency relations among events, we introduce the following notations (Weijters et al. 2006).

Let L be an event log over A and a1, a2∈A. a1>L a2 indicates that there is a sequence of activities in the log L such that the relation of a1 and a2 is direct succession. a1>>L a2 denotes as a sequence of activities such that the relation of a1 and a2 is direct succession and there is a sequence from a2 to a1. It is considered as a loop with length two (Weijters et al. 2006). A frequency based metric is used to indicate how we certain that there is a dependency relation between two activities a1 and a2 (notation a1⇒L a2). The calculated ⇒L values between the activities are used in a heuristic search for retrieving the dependency relations (eq. 1). This thesis uses |a1>La1| as the number of times a1>La1 takes place in L (represent as a loop of length one) (eq. 2) and | a1>>La2 | as the number of times a1>>La2 occurs in L (represent as a loop of length two) (eq.3). For further detail, reader can refer to the original publication (Weijters et al. 2006).

$$a_1 \Rightarrow_L a_2 = \left(\frac{|a_1 >_L a_2| - |a_2 >_L a_1|}{|a_1 >_L a_2| + |a_2 >_L a_1| + 1} \right) if\, (a_1 \neq a_2) \tag{1}$$

$$a_1 \Rightarrow_L a_1 = \left(\frac{|a_1 >_L a_1|}{|a_1 >_L a_1| + 1} \right) \tag{2}$$

$$a_1 \Rightarrow_{2L} a_2 = \left(\frac{|a_1 >>_L a_2| - |a_2 >>_L a_1|}{|a_1 >>_L a_2| + |a_2 >>_L a_1| + 1} \right) if\, (a_1 \neq a_2) \tag{3}$$

Next, we discuss how the concept of fuzzy mining used in the manufacturing process. Fuzzy mining is designed for mining a less-structured process (Gunther and Van der Aalst. 2007). The less-structured process, or usually called as complex process, exhibits a large amount of unstructured and conflicting behavior. Fuzzy mining introduces the concept of aggregation, abstraction, emphasis and customization. The aggregation is to limit the number of information displayed items to show coherent clusters of low-level detailed information in an aggregated manner. Abstraction is to show higher-level information and omit the lower-level information which is insignificant in the chosen context. Emphasis is to highlight the significant information by using visual means such as color, contrast, saturation and size. Finally, customization defines local context which has a specific level of details and a dedicated purpose.

In addition, we will discover all production patterns with aims of finding a representative production flow of underlying manufacturing process. A pattern consists of a set of sequences of activities. The time information should be available for each pattern as well as the number of cases and frequency of patterns. For each pattern, it is possible to extract the number of frequency and production time.

The detail characteristics of each pattern can be visualized using dotted chart analysis. Dotted chart that is similar to a Gannt chart shows a spread of events of an event log over time (Song and Van der Aalst. 2008). It plots dots (i.e. events) according to the time and component types. The component types refer to case, task and resources information from the event log. For example, if the instance (e.g. lot ID) is used as a component type, it is easy to identify which product takes longer process and which activities cause the high completion time.

Fig. 5 shows an example of spread of events in the log according to the actual time and it has been sorted in descending order based on the duration. We can see that case 5 has the longest processing time among cases. To see the detail problem, we select only case 5 and conduct a new dotted chart with the activity and resource component. Fig. 6 shows that there is a big time interval between activity D and E. In other word, there might be a delay after finishing activity D before starting activity E (see Fig. 6). To find the resources which performed those activities, we can provide a dotted chart as shown in Fig. 7. It represents the resources used in case 5 and shows there is a big time interval after the product is executed by M6 moving into M8. It shows that the use of dotted chart can clearly describe the real world situation based on the information in the log.

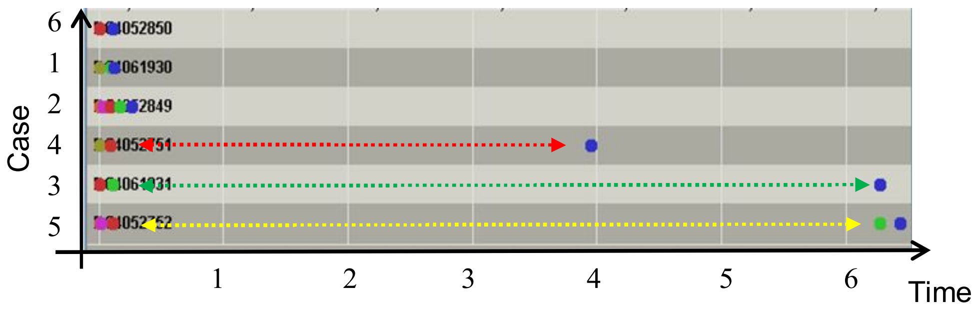


Figure 5: Example of pattern analysis using dotted chart analysis to visualize the big gap of the events

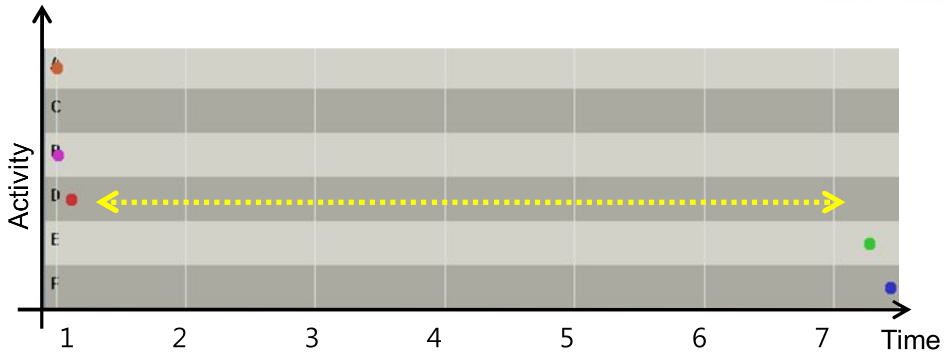


Figure 6: Dotted chart with activity component information for specific case (e.g. case 5 in Figure 5)

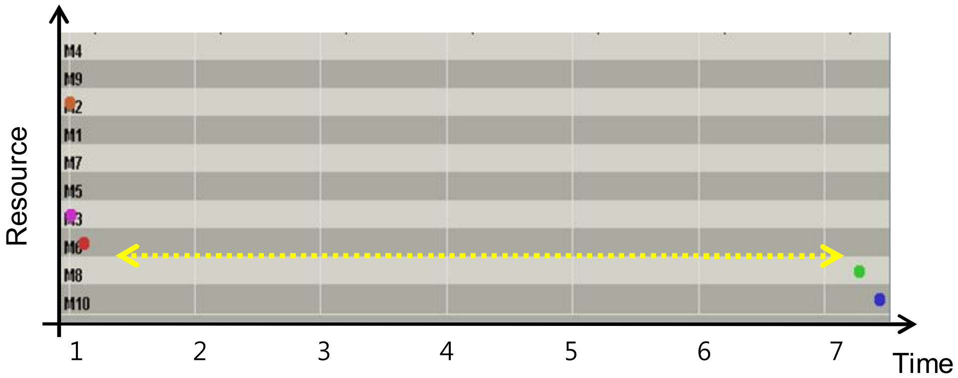


Figure 7: Dotted chart with resource component information for specific case (e.g. case 5 in Figure 5)

Conformance of standard model

It aims to check conformance of standard model with discovered model with event logs. We will apply conformance checking of ProM plug-in. Conformance checking is a technique to check whether a reality conforms to the model or not. It is also important for compliance checking, auditing, certification, and run time monitoring (Rozinat and van der Aalst. 2006). Moreover, it can be used to judge the quality of discovered models. As a verification method of discovered process model, there are two perspectives of conformance. First, it does not capture the real behavior (i.e. "the model is wrong"). Second, it shows that the reality deviates from the desired model (i.e. "the event log is wrong"). Typically, four quality dimensions for comparing model and log are considered; such as fitness, simplicity, precision and generalization (Rozinat and van der Aalst. 2006). These dimensions can be derived using Petri net.

A Petri net is a directed bipartite graph, in which the nodes represent transitions (i.e. events that may occur, signified by bar) and places (i.e. conditions, signified by circles). It also has arcs which run from a place to transition or vice versa, never between places or between transitions (Van der Aalst et al. 2011). Graphically, places in a Petri net may contain a discrete number of marks called tokens. In abstract sense relating to a Petri net diagram, if there are sufficient tokens in all of input places, when the transition enable, it consumes the required input tokens and produces tokens in its output places.

We apply fitness measure to verify the quality of process models over the log. The fitness measure is a mismatch value as a result of replaying the log in the model. The replay of a log starts with marking the initial place in the model and then the transitions that belong to the logged events in the trace are enabled one after another (Rozinat and van der Aalst. 2006). We count the number of tokens that had to be created artificially. For example, the transition belonging to the logged event was not enabled and therefore could not be successfully executed (missing tokens). It also counts the number of tokens that had been left in the model, which indicates the process does not properly complete (remaining tokens). The fitness measure (f) (eq. 4) is formalized as follow.

$$f = \frac{1}{2}\left(1 - \frac{mt}{ct}\right) + \frac{1}{2}\left(1 - \frac{rt}{pt}\right) \tag{4}$$

Where *mt* = the number of missing tokens

rt = the number of remaining tokens

ct = the number of consumed tokens

pt = the number of produced tokens.

It should be noted that mt should be less or equal than ct and rt should be less or equal than pt, and therefore $0 < f < 1$. For example, let consider a trace ABDEF in Fig. 4 (a). By using the model in Fig. 4(d), the token produced on each transition can go to other transition from start to end. The token begins from ‘Start’ (pt=0, ct=0, rt=0, mt=0) and the transition A consume the token and produce a token to c1 (pt=1, ct=1, rt=0, mt=0). The token from c1 goes to B (B consumes the token) and B produces a token to c2 (pt=2, ct=2, rt=0, mt=0). The token continues and is consumed by D and D produces a token to c3 (pt=3, ct=3, rt=0, mt=0). It keeps continuing to follow the path of E, c4, F and End (pt=5, ct=5, rt=0, mt=0). When the tokens are perfectly consumed and produced from the start until the end, then the fitness value equals to 1. By applying the replay to trace ABDE, we can get fitness value as 0.42 with following information (pt=4, ct=3, mt=2, rt=2). After finishing all token reply, we can define the fitness of an event log L where L = event logs, N = model, $L(\sigma)$ = frequency of trace σ, $m_{N,\sigma}$ = number of missing tokens for a single instance σ, and $\sum_{\sigma\in L} L(\sigma) \times m_{N,\sigma}$ = is total number of missing tokens.

$$\mathrm{fitness}(L, N) = \frac{1}{2}\left(1 - \frac{\sum_{\sigma\in L} L(\sigma)\times m_{N,\sigma}}{\sum_{\sigma\in L} L(\sigma)\times c_{N,\sigma}}\right) + \frac{1}{2}\left(1 - \frac{\sum_{\sigma\in L} L(\sigma)\times r_{N,\sigma}}{\sum_{\sigma\in L} L(\sigma)\times p_{N,\sigma}}\right) \quad (5)$$

Bottleneck Analysis

Bottleneck analysis is associated with finding the problem of the long operation time. To find the bottleneck, we need to measure the duration of each event. Since each event has timestamp information, we can measure the time between two events with some indicators. Previous work classified the time-related performance indicator into four; sojourn time, working time, waiting time and synchronization time (Wang et al. 2009). Each of the performance indicators is explained in Table 6.

Table 6: Description of time-related performance indicator [Hornix, 2007]

Time-related performance indicator	Description
Sojourn time	The time spent after the completion of predecessor activity until the completion of that activity.
Working time	The time needed to complete an activity from the start
Waiting time	The time to wait a particular activity to start after the completion of predecessor activity
Synchronization time	Synchronization times are only greater than 0 when it executes AND-join.

In this study, we focus on working and waiting time for two reasons. First, the purpose of using time-related performance indicator is to find the bottleneck which can be extracted from working and waiting time. Note that the sojourn time is the summation of working and waiting time. Second, the manufacturing process has most probably a sequential flow which includes no parallelism in the production flow. Therefore, we can disregard the use synchronization time. The time-related performance indicators can also be used to analyze other perspectives, such as resource and task perspective. In terms of resource, we can find resource utilization based on the working time and waiting time indicator. In the case of activity, the working and waiting time of two activities can be identified. We will use these performance indicators and explain them in the later section.

3.3.2. Resource perspective Ⅰ: Machine-to-Machine interrelationship analysis

Resource perspective deals with all information related to the performance of resources e.g. individual performance of each resource and relation between resources. In this study, the resource perspective analysis refers to the work of machine-to-machine interrelationship analysis (e.g., social network analysis) and machine utilization (e.g., resource allocation and resource working/idle time analysis).

The first resource perspective analysis, machine-to-machine interrelationship analysis, is basis of resource perspective analysis. Machine-to-machine interrelationship analysis derives a social network that builds a network based on the handover of work from one resource to the next resource (Van der Aalst et al. 2005). Fig. 8 shows an example sociogram based on the transfers of work from one resource to another. The node represents the resources and the arc represents that there has been a transfer of work from one resource to another. The definition of "transfer of work from R1 to R2" means that there is an activity executed by R1 directly followed by an activity executed by R2 in the same case. For instance, there is a transfer of work from M1 to M3 in case 1, but the sociogram does not show frequencies. However, the frequencies can be added for analysis purposes. For further detail, reader can refer to (Van der Aalst et al. 2005)

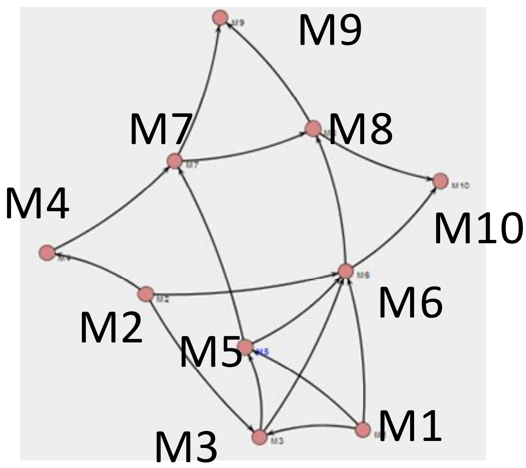


Figure 8: Example of the sociogram of machine network based on the log example

As illustrated in Fig. 8, we can formalize the handover of work as follows (Van der Aalst et al. 2005).

Definition 2. (⊳, ⊵) Let L be a log. Assume that ⟶ denotes some causality relation derived from the process model. For $a_1, a_2 \in A$, $r_1, r_2 \in R$, $c = (c_0, c_1, \ldots) \in L$ and n $\in \mathbb{N}$:

$$r_1 \rhd_c^n r_2 = \exists_{0 \le i < |c| - n} \pi_R(c_i) = r_1 \wedge \pi_R(c_{i+n}) = r_2 \quad (5)$$

$$| r_1 \rhd_c^n r_2 | = \sum_{0 \le i < |c| - n} \begin{cases} 1 & if\, \pi_R(c_i) = r_1 \wedge \pi_R(c_{i+n}) = r_2 \\ 0 & otherwise \end{cases} \quad (6)$$

$$r_1 \unrhd_c^n r_2 = \exists_{0 \le i < |c| - n} \pi_R(c_i) = r_1 \wedge \pi_R(c_{i+n}) = r_2 \wedge \pi_A(c_i) \to \pi_A(c_{i+n}) \quad (7)$$

$$| r_1 \unrhd_c^n r_2 | = \sum_{0 \le i < |c| - n} \begin{cases} 1 & if\, \pi_R(c_i) = r_1 \wedge \pi_R(c_{i+n}) = r_2 \wedge \pi_A(c_i) \to \pi_A(c_{i+n}) \\ 0 & otherwise \end{cases} \quad (8)$$

r1 ⊳nc r2 (eq. 5) denotes the function which return true if within the context of case c resources r1 and r2 both executed some activities such that the distance between these two activities is n. For example, for case 1 shown in fig.8, M1 ⊳1c M3 equals 1 (i.e., true) and M1 ⊳3c M3 equals 1 (i.e. true). In this definition, the value of n refers to the relation i.e. a value of 1 refers to direct succession and if the value is greater than 1, it refers to indirect succession. | r1 ⊳nc r2 | (eq. 6) denotes the function which returns the number of times r1 ⊳ nc r2 in the case c. In other words, it considers multiple transfers within one instance e.g. a rework occurs in an instance. r1 ⊵ nc r2 (eq. 7) and |r1 ⊵ nc r2| (eq. 8) are similar to r1 ⊳ nc r2 and |r1 ⊳ nc r2| but in addition they take into account whether there is a real causal dependency. The information on causal dependency (i.e. causal dependency graph) can be added if the process model is known. If necessary, this information can also be derived from the log by using for example the α-algorithm (Van der Aalst et al. 2004). Based on the above relations, we define handover of work metrics.

Definition 3. Handover of work metrics

$$r_1 \rhd_L r_2 = \left(\sum_{c \in L} | r_1 \rhd_c^1 r_2 |\right) / \left(\sum_{c \in L} | c | - 1\right) \quad (9)$$

$$r_1 \rhd^{\bullet}_L r_2 = \left(\sum_{c \in L \wedge r_1 \rhd_c^1 r_2} 1 \right) / | L | \quad (10)$$

r1 ⊳L r2 (eq. 9) means dividing the total number of direct successions from r1 to r2 in a process log by the maximum number of possible direct successions in the log. r1 ⊳•L r2 (eq. 10) ignores multiple transfers within one instance (i.e. case). For example, in figure 9, M1 ⊳L M3 equals 1/21 and M1 ⊳•L M3 equals 1/6. Note that metric ⊳L defines a weight function W, i.e. r1 ⊳L r2 = W_{r_1, r_2} is the weight of the link from r1 to r2 in the corresponding sociogram. A threshold may be used to remove links (e.g. finding a high relationship with high weight) from the sociogram.

3.3.3. Resource perspective Ⅱ: Machine Utilization

The second resource perspective analysis is machine utilization that aims at deriving machine allocations and analysis machine working/idle time using process mining techniques. For the machine allocation analysis, we used “Organizational mining” and “Resource by Task Matrix” plug-ins of ProM. Organizational model describes the organizational knowledge of machines such as organizational structures based on executed activities and it enables managers to understand the group of machines for each activity. Machine by task assignment explains the frequency of a machine performed in an activity. And for the machine working/idle time analysis, we used “Basic performance analysis” and “Dotted chart” to delineate other performance indicators such as time and equipment efficiencies in the log.

Machine allocation

The machine organization applies organizational mining (Song and Van der Aalst. 2008). There are several principles that define the relationship characteristic of machines using organizational mining. They generally consist of *doing similar task*, w*orking together* and d*efault mining*. *Doing similar task* focuses on the machines assigned similar tasks with similar knowledge. *Working together* is related to the project groups in which machines usually progress different skills and work together at the same cases. We can infer the working together through transfer of work concept as explained previously.

Default mining is simply to show the matching between activities and machines (Song and Van der Aalst. 2008). The development of organizational mining is highly relevant to processes that are not completely controlled by systems, e.g. humans play a dominant role in a test process. However, the application of organizational mining can also assist domain experts to understand and improve machine allocation in manufacturing process i.e. it can show the machines that performed in some different activities. Fig.9 shows an example of organizational structure. For example, M7 and M6 worked for task D, and M8 assigned for task E.

Resource by task matrix shows the relationship between machine and activity as shown in the fig. 10. Each machine (M) has a profile (i.e. activity-resource relationship matrix) based on how frequently they conduct specific activity (A). If a machine executed an activity, the activity is assigned to the machine. Let consider the example in Fig.9. For example, we can deduce that activity A is executed by either M1 or M2, activity B is executed by M3, C is executed by M4 or M5, D is executed by M6 or M7, E is executed by M8 and F is executed by M9 or M10. This example can show a typical method to find the organizational structure according to executed activity. For larger data set, it is possible to find the resource which performed in different activity (e.g. a machine can be used in different activity)

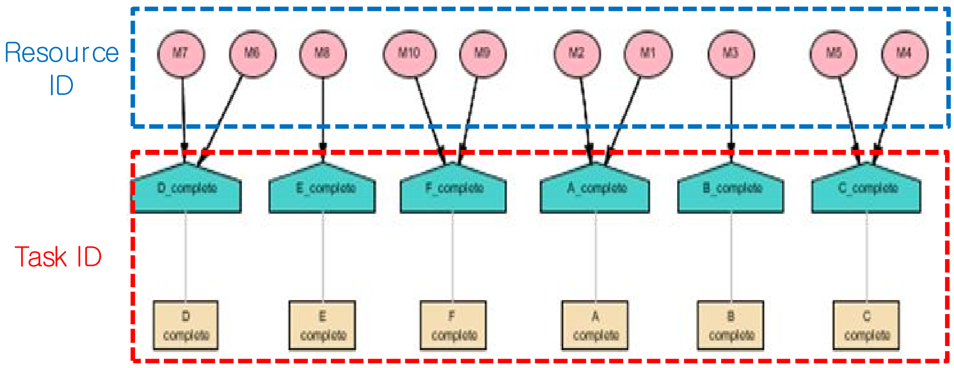


Figure 9: Organizational structure of resources according to the activity

Resource ID

Task ID	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
A	10	2								
B			5							
C				10	10					
D						20	1			
E								10		
F									5	15

Figure 10: Example of resource by task matrix

3.3.4. Task perspective: monitoring & diagnosis of task performance

In this section, we explain the task analysis in regard to production lead time and production lost rate. Monitoring & diagnosis of task performance aims at analyzing the performance of each task, and the ability focuses on their production lead time and production lost rate. Manufacturing enterprises wants to increase their profit by reducing the production cost that is affected by lead time and lost rate. Therefore, it is essential for the manufacturing enterprises to find a task that has high lead time or high production lost rate. Production lead time refers to the Sojourn time. For example, the production lead time of task B is time interval between completed time of Task A and complete time of *Task B* when a lot moved though from task A to task B.

For the production lost rate analysis, we used the data product quantity from transaction logs of the MES. Since we discovered the process model, we can see the specific position where the product quantity is decreasing. The decrement of product quantity can have two meaning. First, the product quantity in a lot is decreased because of assembly process. Second, it is decreased because of defective reason. Since the intention of product quantity analysis is not about finding the reason of decrement, we will show the result of product quantity analysis in the way of task perspective with additional information on product quantity.

Production lead time analysis

Using timestamp, we can find several kinds of analysis such as lead time and bottleneck analysis. Bottleneck analysis is associated with finding the problem of the long operation time. To find the bottleneck, we need to measure the duration of each event. Since each event has different event type and a timestamp attached to it, we can measure the time between events with some indicators. Previous work classified the time-related performance indicator into four; sojourn time, working time, waiting time and synchronization time (Wang et al. 2009). Each of the performance indicators is explained in Table 4.

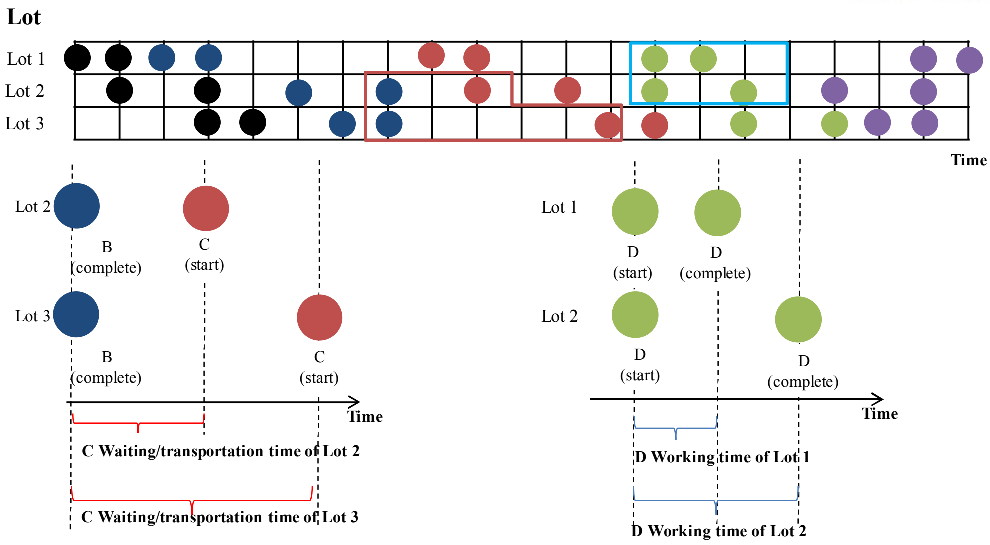


Figure 11: Example of time-related performance indicator based on the spread of events visualization

In this study, we focus on working and waiting/transportation time for two reasons. First, the purpose of using time-related performance indicator is to find the bottleneck which can be extracted from working and waiting/transportation time. Second, the manufacturing process using MES has a sequential flow. Most probably, there is no AND-join in the production flow. Fig.11 illustrates the calculation of working and waiting/transportation time of two consecutive events with different event type. It should be noted that time-related performance indicator can be used to analyze other perspectives, such as resource and task perspective. In terms of resource, we can find resource utilization based on the working time and waiting/transportation time indicator. In the case of task, the working and waiting/transportation time of two activities can be identified. We will use these performance indicators and explain them in the later section.

Incorporating with pattern analysis in the previous section, process mining tool can provide working time and waiting/transportation time analysis between particular activities. The notation *t(e1, e2)* is denoted as the working time and *w(e1, e2)* is denoted as the waiting time between event e1 and e2. Those working time and waiting time can be derived using eq. (11) and eq. (12), respectively.

$$t(e_1, e_2) = \pi_T(e_2) - \pi_T(e_1) \qquad (11)$$

where $e1 < e2 \wedge \pi A(e1) = \pi A(e1) \wedge \pi Y(e1)=\text{"start"} \wedge \pi Y(e2)=\text{"complete"}$

$$w(e_1, e_2) = \pi_T(e_2) - \pi_T(e_1) \qquad (12)$$

where $e1 < e2 \wedge \pi A(e1) \neq \pi A(e1) \wedge \pi Y(e2)=\text{"start"} \wedge \pi Y(e1)=\text{"complete"}$

Production yield analysis

In order to improve the yield rate of manufacturing process, it is necessary to take into account the product quantity in the analysis. Since the manufacturing data includes the input and output of product quantity, process mining tool can detect the activity that causes the decrement of product quantity as the effect of defect. Let *diffqty(e1,e2)* be the number of quantity difference between event e1 and event e2 with the formula as follow.

$$diff_{qty}(e_1, e_2) = \pi_{qty}(e_2) - \pi_{qty}(e_1) \qquad (13)$$

If the *diffqty(e1,e2)* is greater than 0, then there is a decrement on the number of product quantity. If *πA(e2)* equals to *πA(e1)* then we can probably say that the decrement is because of the operation failure. But, if *πA(e2)* is not equal to *πA(e1)*, we can say that the cause of decrement is the loss during transporting the product from one activity to another activity. It is important to keep in mind that there are a lot of other factors about the quantity decrement. Thus, domain experts' opinions should be considered to find the rational reasons of the quantity decrement.

IV. Case Study: An electronic components manufacturing process

This section discusses the obtained results by applying the proposed framework to the real manufacturing process analysis. The case study is based on an event log collected in the electronic component manufacturing processes at a manufacturing company in the South Korea. In the case study, we focus on showing how process mining techniques can be applied in a real manufacturing process analysis. This section consists of three parts. First, we explain the context of the case study. Then, we propose a result of case study according to the three perspectives. Lastly, the discussion is described.

4.1. Contexts

The manufacturing company produces various kinds of electronic components such as MLCC. We obtained event logs using a MES of the company. In the manufacturing industry, a basic unit is a lot (or a batch) and we used a lot number as a case ID. Case ID is an identification number assigned to a particular quantity or lot of material. In the manufacturing process, a lot pass through several activities. In each activity, a start event and a compete event are recorded. Each event contains some information such as lot ID, task ID, resource ID, and event time.

The machines are categorized into three types: manual, semi-automatic, and automatic. For activities with a manual or a semi-automatic machine, a worker involved in the execution of a task is specified. An activity with an automatic machine has no information about the worker. In addition, an event contains information about the quantity in a lot. Quantity in a start event refers to the input quantity of an activity and one in the end event means the output quantity of an activity. In the pre-processing step, we include the cases that start with Input task and finish with Package task. The event logs include 11,226 lots passed along the production process. The number of total working events is 990,542 and the process consists of 361 activities. It also included 1,217 machines performed for the activities in 5 months.

4.2. Summary of manufacturing process mining and analysis

4.2.1. Manufacturing process visualization of the process perspective

As explained in the framework, analysis in process perspective are going to visualize manufacturing processes, compare the existing model with discovered model, and find bottleneck points on the discovered model to improve the current manufacturing processes. First, we discover actual processes by using process modeling algorithms i.e. Heuristic mining and Fuzzy mining. Second, we conducted two analyses to enhance the standard model. Conformance checking compares discovered process model with event logs or with standard model. Additionally, performance analysis with petri-net offers all possible manufacturing flows as well as statistical information, e.g. average lead-time and number of tasks. Finally, a bottleneck analysis visualizes the bottleneck points on the discovered model.

We describe results of actual process discovery of manufacturing processes of electronic components in the process perspective to identify actual flows of lots. Among the several mining plug-ins, we applied heuristic mining and fuzzy mining. Fig. 12 shows the fragment of manufacturing process model. Fig. 12 (a) displays the overall manufacturing process model derived by Heuristic miner. All the processes start with the *input* task and finish with the *packaging* task. The model is useful to provide general manufacturing flows as well as number of lots passing between activities. Additionally, it visualizes the rework flows and how many lots are moved through the flows. For instance, four lots made rework flow from task C to task B. Moreover, some lots moved through ETC task being considered as unexpected task for the experts. Compared to Heuristic mining results, fuzzy mining shows a simple process model in the Fig. 12 (b). The red circles are Input and Package, respectively. The model represents the main manufacturing process model by aggregating some activities according to the parameter in fuzzy miner. Blue triangle activities are in the main production flows.

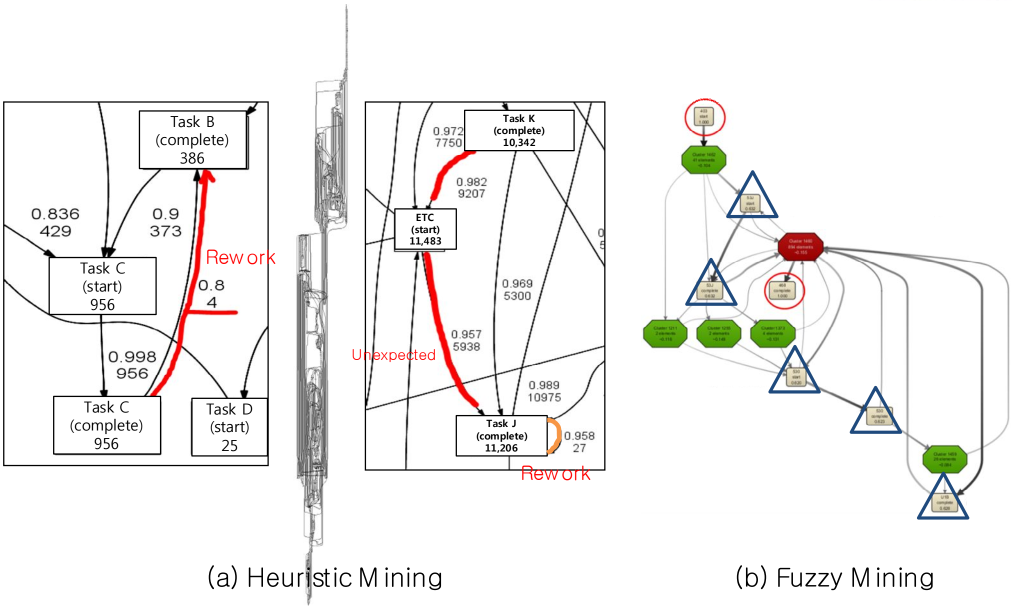


Figure 12: Manufacturing process models

Moreover, to compare the actual process patterns with existing model, we discovered all possible manufacturing flows using "Performance analysis with petri-net", which is used to find the best practice processes. To see the event-level information for each pattern, dotted chart analysis was applied. For example, there are 176 patterns in the electronic-components production process. A dotted chart provides the event-level information of a pattern. Fig. 13 and 14 show the dotted chart results of two patterns. As explained in Section 3, the x-axis represents time, in minute, and the y-axis represents cases having same pattern. A dot is colored based on the name of the activity and the cases are sorted by case duration. The gap between two dots describes the time interval of the two events, and the longer distance affect to total lead-time. Fig. 13 has the shortest average production time while right-hand pattern has the longest average production time among the 176 patterns. In the fig. 14, durations are lower than 20 hours, but the other one have a long duration time. The duration takes longer than 3 months (about 104 days), and it affects to increase production time to average 111 days 23 hours with 81 activities.

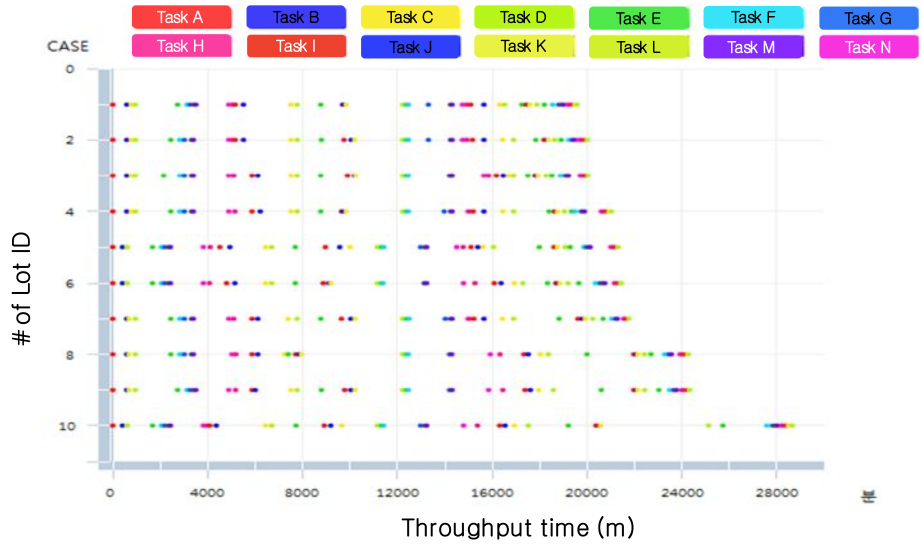


Figure 13: Lead time analysis of a pattern having the lowest average entire lead-time with dotted chart

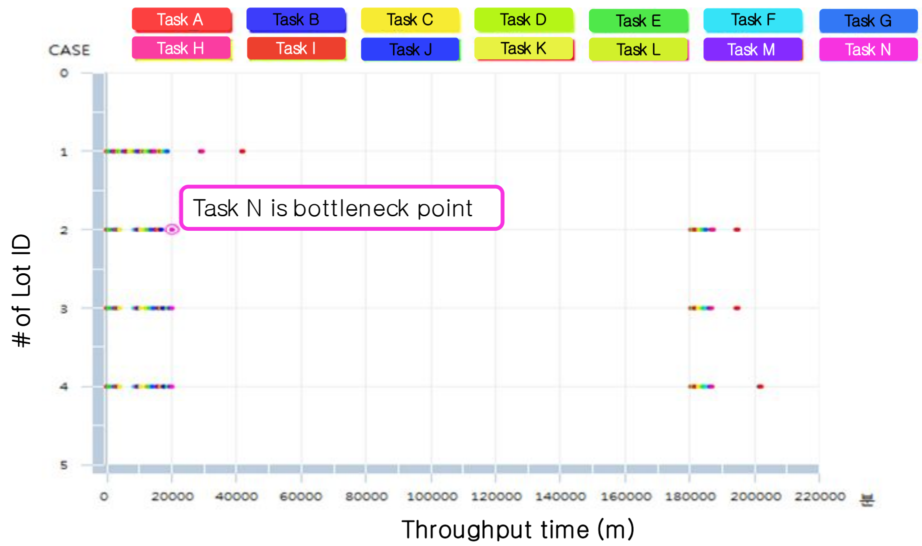
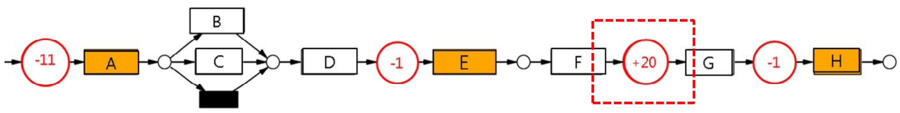


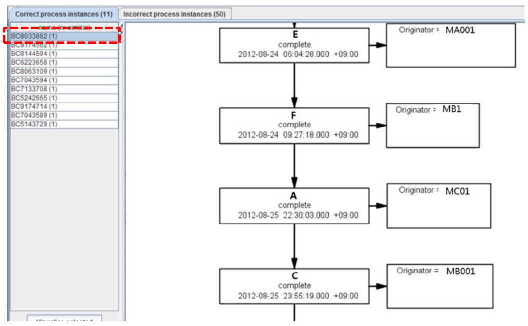
Figure 14: Lead time analysis of a pattern having the highest average entire lead-time with dotted chart

Manufacturing company usually has standard manufacturing process model, which is expected flows. But, the actual process model may be different with the existing model. To check the difference and to enhance the existing model, we applied two process mining techniques. Firstly, conformance checking provides a fitness value of the discovered process model from event logs by comparing with the event log. For this analysis, it is necessary to convert a heuristic process model to a Petri-net model. The detailed explanation about Petri-net is in (Van Dongen et al. 2009). The existing conversion technique in process mining is utilized to convert the heuristic model into the Petri-net model. Additionally, it is possible to compare the discovered model with existing model.

Fig. 15 shows a fragment of the conformance checking result. The final fitness value of overall process is 0.9961815. It indicates that almost all of cases in the log (99%) conform to the mined model. The token replay function was applied to get the fitness value. For example, the place before activity package has a red color with a value of +20. It means that there are 20 remaining tokens before the activity start. For investigating the 20 missing token, we generated linear temporal logic (LTL) checker, an existing technique in process mining used to check the event log. The Fig. 15 (a) shows 11 lots of 20 remaining tokens went back to task A before the task G.



(a) Conformance checking result



(b) LTL Checker result

Figure 15: Conformance checking results

The bottleneck analysis visualizes bottleneck points where an activity has long lead time on the discovered process model. Fig. 16 shows the fragment of the analysis result. The bottleneck points are represented by red, while the yellow shows medium waiting time and the blue represents low waiting time. For instance, it is easy to find the bottleneck point from A to D and E to F or G. Additionally, the analysis result also shows the time in-between two tasks. For example, on average it takes 98 hours between task A to task H, which represented by red box.

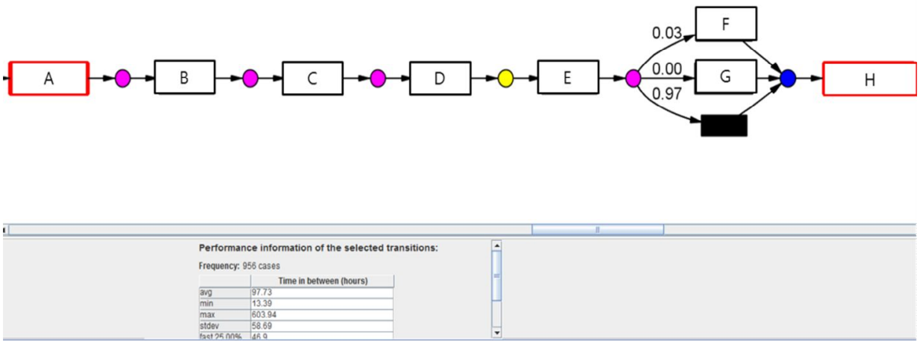


Figure 16: Bottleneck analysis results

4.2.2. MTM interrelationship analysis of the Resource Perspective

For the analysis in the resource perspective, we discovered machine network model and analyze machine utilization. First, we derived machine network model that show interrelationship between machines in respect of logistic flows using the ProM plug-in “Social Network Analysis”. Fig. 17 shows the results obtained for machine-to-machine interrelationship analysis, and Table 7 shows top 5 machines having high betweenness that is degree of mediator role in the network. A social network shows how lots are transferred among machines, and the table shows the machine has the most frequent relationship between other two machines. According to the result, the machine M0052 has the highest value of betweenness, and the machine is in the important location of the network. Since a trouble on the machine may influence the entire manufacturing processes, manufacturing managers need to pay more attention to the machine.

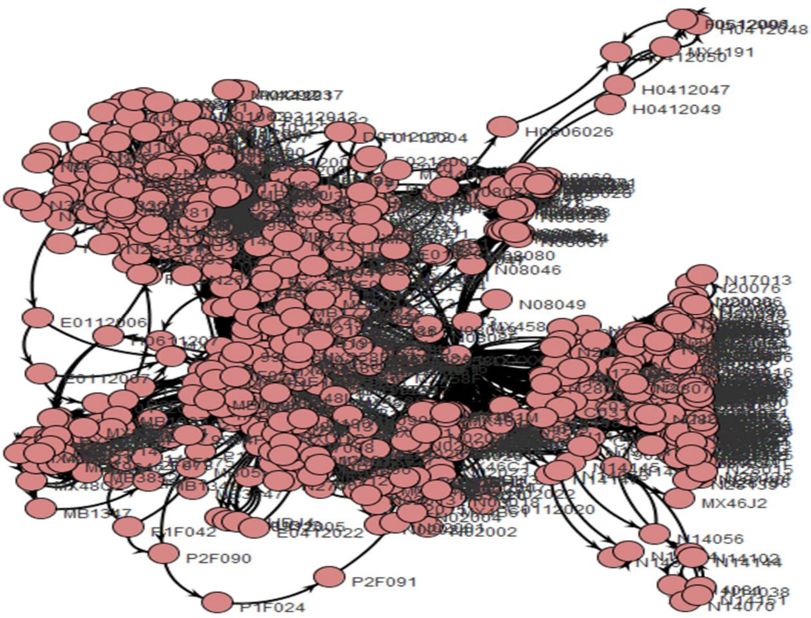


Figure 17: Machine-to-Machine interrelationship analysis result

Table 7: Top 5 machines having high betweenness

Machine code	Machine name	Betweenness
M0052	**Task A**	131,045.07
M0014	**Task B**	125,163.11
M0075	**Task C**	90,114.86
M0018	**Task D**	77,860.14
M0086	**Task E**	77,454.90
…	…	…

4.2.3. Machine utilization of the Resource Perspective

The second analysis in resource perspective is machine utilization with aims of managing the machines on manufacturing processes in a balanced way. For the machine allocation analysis, we used ProM plug-in "Organizational mining" and "Resource by Task matrix", and we used "Basic performance analysis" and "Dotted chart" to analyze machine idle time.

Fig. 18 shows one fragment of organizational mining results, which are useful to visualize the relationship between tasks and machines. Task A is performed by seven machines called M0001, M0002, M0003, M0004, M0005, M0006, and M0007. Among them two machines are also working for task B. With the result, managers can check whether machines were used for desired works by comparing the model with the standard machine assignment guideline in a company. Moreover, the machine by task assignment matrix shows the frequencies of machines used for tasks. It provides the number of events that each machine worked for the activity. Table 8 shows example of machines for Task A. In the table, M0001 machine worked 5,612 times for package for 5 months while M0005 machine only did 2 times. The frequencies are quite different each other. It means that all machines are not operating at the similar capacity.

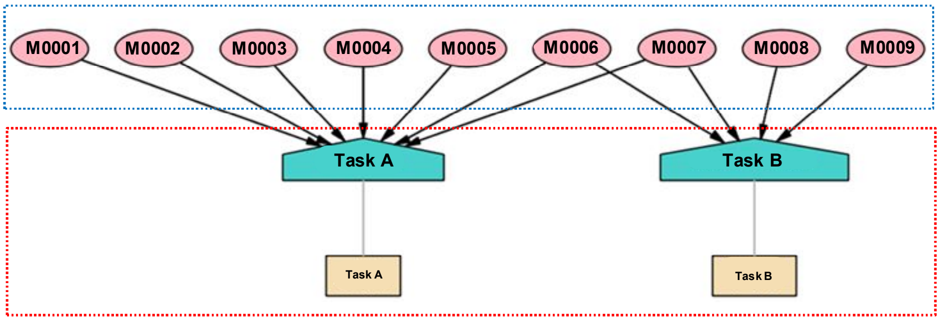


Figure 18: Organizational model of task A and task B

Table 8: Machines for task matrix of 'task A'

Machine	Task A
M0001	5,612
M0002	2,965
M0003	2,143
M0004	500
M0005	2
M0006	2
M0007	2

In addition, we also analyzed machine working time and idle time. Machine idle time is a period of time during a regular work cycle when a worker is not active because of waiting for materials or instruction, it also known as waiting/transportation time. Fig. 19 shows the working time of the seven machines of task A. It provides working frequency as well as average and median of working time. The dashed line indicates the average working time of the task, i.e. 62 minutes. M0003 machine has the highest value of working time than other machines. Fig. 20 shows the dotted chart for the M0003 machine to check the detailed working duration. The figure shows that about 90% of cases finished within 5 hours. However, there are few cases which take more than 1 day.

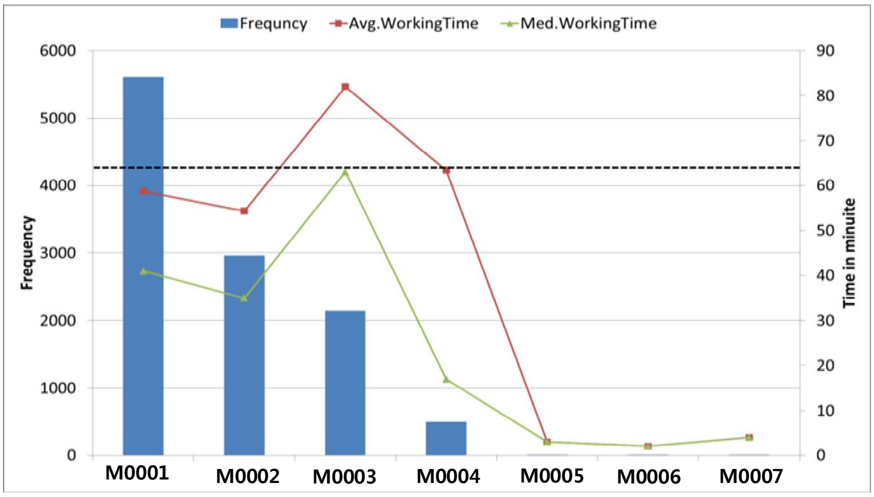


Figure 19: Bar chart for performance analysis result of task A

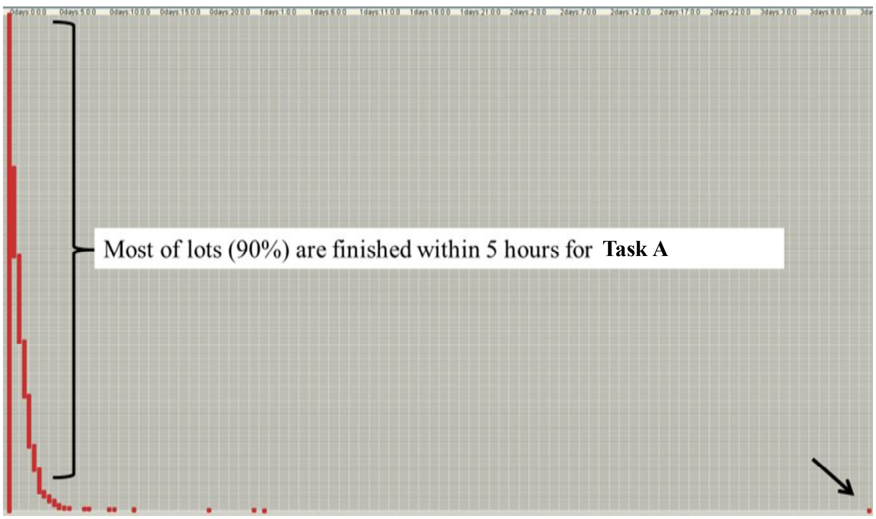


Figure 20: Time analysis of task A (Dotted Chart for M0003 machine of task A.)

4.2.4. *Monitoring & Diagnosis of task performance*

In this part, we conduct several analysis related to task perspective. Task perspective provides information of the task that reduces manufacturing efficiency. One of the main purposes of manufacturing management is increasing production efficiency by reducing total cost and lead time. For this reason, root-cause analysis is necessary to find the activity having high fraction defective and lead time. Production time analysis focuses on high lead-time while production quantity loss does on fraction defectives. Actually, we already introduced another root-cause analysis, bottleneck analysis detecting high lead time tasks in the process perspective. The production time analysis has difference on the results are used statistical information not process model.

The proposed bottleneck analysis finds a bottleneck point having high lead time on the manufacturing processes. On the other hands, a production time analysis use statistical data such as working frequency, average working, waiting /moving time rather than process model. This analysis is performed by BPA (Basic Performance Analysis) of ProM. The BPA have several positive points to show the statistical time information.

Production yield analysis generates a graph showing the change of product quantity along the manufacturing process. Fig. 21 shows one example, which shows the change of input and output product quantity. The output quantity is smaller than input quantity at 'Horizon 5 copper plating', It means that there might be some problems during the working.

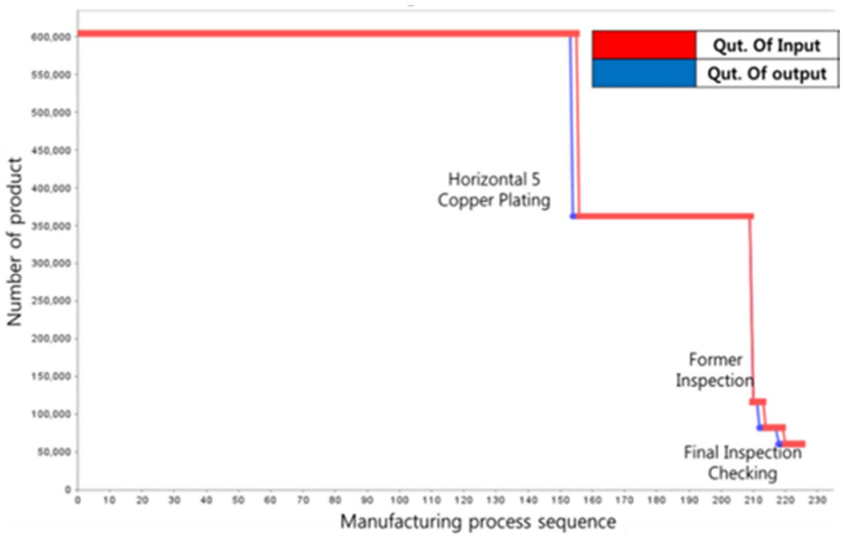


Figure 21: Change of product quantity along the time sequence

V.Conclusion

In this thesis, we suggested a framework for manufacturing process analysis based on process mining techniques. MES records production activities when they take place and store as event logs that can be valuable source for the manufacturing process analysis according to the three perspectives. The proposed framework applied several process mining techniques for the four kinds of analysis: *Manufacturing process visualization, Machine-to-Machine interrelationship analysis, Machine utilization,* and *Monitoring & diagnosis of task performance.*

For the *manufacturing process visualization*, firstly, we discovered manufacturing process models using mining techniques such as heuristic and fuzzy mining. The discovered models showed all production flows including the main flows, rework flow, and abnormal flows. Additionally, pattern analysis examines all possible sequences of manufacturing processes using performance analysis with petri-net plug-in. Moreover, we also conducted conformance checking that compares the derived process model with event logs and standard model with event logs. Finally, we also visualized bottleneck point on the discovered process model.

Second, this thesis conducted two analyses in the resource perspective: *machine-to-machine interrelationship analysis* and *machine utilization*. The machine-to-machine inter-relationship analysis used social network analysis that shows "how" resources are related to each other. Next, machine utilization conducted two analyses: machine allocation analysis and machine idle time analysis. Machine allocation analysis used organizational mining that represents "which" machine worked for each activity and machine by task matrix that represents frequency of working events for each tasks. Next, for the machine idle time analysis, dotted chart analysis and basic performance analysis examined the detailed machine performance over time. Based on the results, manufacturing manager can make a better decision about machine allocation and operation.

Third, task perspective conducted *monitoring & diagnosis of task performance* that analyzes the performance of each tasks regarding their production yield rate and production lead time. Production yield rate analysis is important to reduce their lost rate that related to production cost, and production lead time is a major determinant of production speed. For the analysis, we made a formula, and evaluate their capability.

We applied the proposed framework to real-manufacturing data in electronic components manufacturing processes. After have a discussion about the results, the experts agreed that the proposed framework is useful to visualize manufacturing processes and machine interrelationship, conduct machine utilization, and monitor task performance. Therefore, the framework helps users to obtain useful information that is used to improve existing processes. It has the following major contributions:

- It suggested a framework that support manufacturing process analysis with process mining techniques. It gives several important opportunities for applying process mining in manufacturing process analysis: manufacturing process visualization, machine-to-machine interrelationship analysis, machine utilization, and monitoring & diagnosis of task performance.
- It demonstrated application of the framework using real manufacturing event logs.
- It demonstrated the ability of process mining to provide insight from manufacturing data.

However this thesis has several limitations that lead us to future works. First, it took a lot of time for data preparation including data collection, preprocessing, and conversion since the availability and quality of event logs are key factors for the framework. Second, the framework did not consider a big data. With the advanced technology, manufacturing data are getting to bigger than ever. Accordingly, it needs to develop and improve the current framework to deal with the big data. In the future works, we plan to improve the framework by verifying the effectiveness of the framework and applying the framework to other manufacturing processes.

Reference

1. Adler, D.J., Herkamp, J., Henricks, D., & Moss, R. (1995). "Does a manufacturing execution system reduce the cost of production for bulk pharmaceuticals?" *Isa Transactions,* 34(4): 343-347.

2. Agrawal, R., Gunopulos, D., & Leymann, F. (1998). "Mining process models from workflow logs," *Advances in Database Technology — EDBT'98,* Springer Berlin Heidelberg. 1377: 467-483.

3. Allahverdi, A., Ng, C.T., Cheng, T.E., & Kovalyov, M.Y. (2008). "A survey of scheduling problems with setup times or costs," *European Journal of Operational Research,* 187(3): 985-1032.

4. Braha, D. (2001). "Data mining for design and manufacturing: methods and applications," Kluwer Academic Publishers; USA.

5. Caron, F., Vanthienen, J., & Baesens, B. (2013). "A comprehensive investigation of the applicability of process mining techniques for enterprise risk management," *Computers in Industry,* 64(4): 464-475.

6. Chan, F.T.S., Chan, H.K., & Lau, H.C.W. (2002). "The State of the Art in Simulation Study on FMS Scheduling: A Comprehensive Survey," *The International Journal of Advanced Manufacturing Technology*. 19(11): 830-849.

7. Chen, W.C., Tseng, S.S., & Wang, C.Y. (2004). "A Novel Manufacturing Defect Detection Method Using Data Mining Approach," *Innovations in Applied Artificial Intelligence*, Springer Berlin Heidelberg. 3029: 77–86.

8. Chien, C.F., Wang, W.C., & Cheng, J.C. (2007). "Data mining for yield enhancement in semiconductor manufacturing and an empirical study," *Expert Systems with Applications*, 33(1):198-198.

9. Damer, N., Jans, M.J., Depaire, B., & Vanhoof, K. (2012). "Making Compliance Measures Actionable: A New Compliance Analysis Approach," In Proceedings of international conference *Business Process Management Workshops*, Springer Berlin Heidelberg. LNBP 99: 159-164.

10. De Weerdt, J., Schupp, A., Vanderloock, A., & Baesens, B. (2013). "Process mining for the multi-faceted analysis of business processes—A case study in a financial services organization," *Computers in Industry,* 64(1): 57-67.

11. De Medeiros, A.A., Weijters, A.J.M.M., & van der Aalst, W.M.P. (2006). "Genetic process mining: a basic approach and its challenges," In Proceedings of the 3th international conference *Business Process Management Workshop*, Springer Berlin Heidelberg. LNCS 3812: 203-215.

12. Elovici, Y. and D. Braha (2003). "A decision-theoretic approach to data mining Systems," *Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions* on 33(1): 42-51.

13. Fernández-Gallego, B., Lama, M., Vidal, J.C., & Mucientes, M. (2013). "Learning Analytics Framework for Educational Virtual Worlds," *Procedia Computer Science,* 25: 443-447.

14. Günther, C.W., & Van der Aalst, W.M.P (2007). "Fuzzy mining: adaptive process simplification

based on multi-perspective metrics," In Proceedings of the 5th international conference on *Business process management*, Springer Berlin Heidelberg. LNCS 4714: 328-343.

15. Giachetti, R. (1999). "A standard manufacturing information model to support design for manufacturing in virtual enterprises," *Journal of Intelligent Manufacturing,* 10(1): 49-60.
16. Griori, D., Casati, F., Castellanos, M., Dayal, U., Sayal, M., & Shan, M. (2004). "Business Process Intelligence," *Computers in Industry*, 55(3): 321-343
17. Gröger, C., Niedermann, F., & Mitschang, B. (2012). “Data mining-driven manufacturing process optimization,” In Proceedings of international conference on *the World Congress on Engineering (WCE 2012),* 3:1475-1481
18. Harding, J.A., Kusiak, A., & Shahbaz, M. (2005). "Data Mining in Manufacturing: A Review," *Journal of Manufacturing Science and Engineering,* 128(4): 969-976.
19. Hernandez-Matias, J.C., Vizan, A., Hidalgo, A., & Rios, J. (2006). "Evaluation of techniques for manufacturing process analysis," *Journal of Intelligent Manufacturing,* 17(5): 571-583.
20. Heilala, J. (1999). "Use of simulation in manufacturing and logistics systems planning," In Proceedings of *the VTT Manufacturing Technology.* Finland.
21. Hong, Z., Soon, T. H., & Sivakumar, A. I. (2001). “Digital models for manufacturing process visualization,” In Proceedings of International Conference on *Integrated Logistics,* Singapore, 113-122.
22. Jans, M., Alles, M., & Vasarhelyi, M. (2013). "The case for process mining in auditing: Sources of value added and areas of application," *International Journal of Accounting Information Systems*, 14(1): 1-20.
23. Kletti, J. (2007). "Manufacturing Execution Systems-MES," Springer; Berlin.
24. Kusiak, A. (2006). “Data mining: manufacturing and service applications,” *International Journal of Production Research*, 44(18-19): 4175-4191.
25. Lee, S. K., Kim, B., Huh, M., Cho, S., Park, S., & Lee, D. (2013). "Mining transportation logs for understanding the after-assembly block manufacturing process in the shipbuilding industry," *Expert Systems with Applications,* 40(1): 83-95.
26. Lin, H., Fan, Y., & Newman, S. T. (2009). "Manufacturing process analysis with support of workflow modelling and simulation," *International Journal of Production Research,* 47(7): 1773-1790.
27. Mans, R., Reijers, H., Wismeijer, D., & Van Genuchten, M. (2013). "A process-oriented methodology for evaluating the impact of IT: A proposal and an application in healthcare," *Information Systems,* 38(8): 1097-1115.
28. Mans, R. S., Schonenberg, M. H., Song, M., van der Aalst, W. M., & Bakker, P. J. (2009). "Application of process mining in healthcare–a case study in a dutch hospital," In proceedings of International Joint Conference on *Biomedical Engineering Systems and Technologies*, Springer Berlin Heidelberg, CCIS 25: 425-438.

29. MESA, (1997). "MES functionalities & MRP to MES data flow possibilities, " *MESA International White Paper* 2, Available: http://www.mesa.org

30. MESA, (1997). "MES explained: A high level vision," *MESA International White Paper* 6, Available: http://www.mesa.org

31. McClellan, M. (1997). "Applying manufacturing execution systems," CRC Press; USA.

32. McClellan, M. (2000). "Introduction to manufacturing execution systems," *MES Conference & Exposition,* Phoenix; USA. 3-11.

33. Polczynsk, M. and Kochanski, A. (2010). "Knowledge Discovery and Analysis in Manufacturing," *Quality Engineering*, 22(3): 169-181

34. Negahban, A. and Smith, J. S. (2014). "Simulation for manufacturing system design and operation: Literature review and analysis," *Journal of Manufacturing Systems,* 33(2): 241-261.

35. Qiu, R. G. and Zhou, M. (2004). "Mighty MESs; state-of-the-art and future manufacturing execution systems," *Robotics & Automation Magazine, IEEE,* 11(1): 19-25, 40.

36. Rajesh, V., Ashu, G., & Kawaljeet S. (2008). "Simulation Software Evaluation and Selection: A Comprehensive Framework," *Journal of Automation & Systems Engineering*, 1:221-234.

37. Rebuge, Á. and Ferreira, D. R. (2012). "Business process analysis in healthcare environments: A methodology based on process mining," *Information Systems*, 37(2): 99-116.

38. Reijers, H. A., Song, M., & Jeong, B. (2007). "On the Performance of Workflow Processes with Distributed Actors: Does Place Matter?" In proceedings of 5th international conference on *Business Process Management*. Springer Berlin Heidelberg. LNCS 4714: 32-47.

39. Reisig, W. & Rozenberg, G. (1998). "Lectures on Petri Nets I: Basic Models, Advances in Petri Nets," LNCS 1491, Springer Berlin Heidelberg.

40. Rozinat, A. and Van der Aalst, W.M.P. (2006). "Conformance Testing: Measuring the Fit and Appropriateness of Event Logs and Process Models," *Business Process Management Workshops*, Springer Berlin Heidelberg, LNCS 3812: 163-176.

41. Saenz de Ugarte, B., Artiba, A., & Pellerin, R. (2009). "Manufacturing execution system – a literature review," *Production Planning & Control,* 20(6): 525-539.

42. Scott, D. (1996). "Comparative advantage through manufacturing execution systems," In proceedings of *Advanced Semiconductor Manufacturing Conference and Workshop, ASMC 96*, IEEE Cambridge, 179-184.

43. Scott, J. (1988). "Social network analysis," *Sociology,* 22(1): 109-127.

44. Song, M. and Van der Aalst, W.M.P. (2008). "Towards comprehensive support for organizational mining," *Decision support systems,* 46(1): 300-317.

45. Stuit, M. and H. Wortmann (2012). "Discovery and analysis of e-mail-driven business processes," *Information Systems,* 37(2): 142-168.

46. Tiwari, A., Turner, C.J., & Majeed, B. (2008). "A review of business process mining: state-of-the-art and future trends," *Business Process Management Journal,* 14(1): 5-22.

47. Van der Aalst, W.M.P. (2011). "Discovery, Conformance and Enhancement of Business Processes," Springer; Heidelberg.

48. Van der Aalst, W.M.P., Weijters, T., & Maruster, L. (2004). "Workflow mining: discovering process models from event logs," *Knowledge and Data Engineering, IEEE Transactions* on 16(9): 1128-1142.

49. Van der Aalst, W.M.P., Reijers, H. A., & Song, M. (2005). "Discovering social networks from event logs," *Computer Supported Cooperative Work (CSCW)*, 14(6): 549-593.

50. Van der Aalst, W.M.P., & Weijters, A.J.M.M. (2004). "Process mining: a research agenda," *Computers in industry*, 53(3): 231-244.

51. Van der Aalst, W.M.P., Reijers, H.A., Weijters, A.J., van Dongen, B.F., Alves de Medeiros, A.K., Song,M., & Verbeek, H.M.W. (2007). "Business process mining: An industrial application," *Information Systems,* 32(5): 713-732.

52. Van der Aalst, W.M.P., Schonenberg, M. H., & Song, M. (2011). "Time prediction based on process mining," *Information Systems,* 36(2): 450-475.

53. Van der Aalst, W.M.P., Lohmann, N., & La Rosa, M. (2012). "Ensuring Correctness During Process Configuration Via Partner Synthesis," *Information Systems,* 37(6): 574-592.

54. Van Dongen, B.F., De Medeiros, A.A., & Wen, L. (2009). "Process Mining: Overview and Outlook of Petri Net Discovery Algorithms," In Proceedings on 26^{th} international conference on *Transactions on Petri Nets and Other Models of Concurrency II*, Springer Berlin Heidelberg. LNCS 5460: 225-242.

55. Van Dongen, B.F., de Medeiros, A.K.A., Verbeek, H.M.W., Weijters, A.J.M.M., & Van Der Aalst, W.M. (2005). "The ProM Framework: A New Era in Process Mining Tool Support," *Applications and Theory of Petri Nets 2005*, Springer Berlin Heidelberg. 3536: 444-454.

56. Verma, R., Gupta, A., & Singh, K. (2008). "Simulation software evaluation and selection: a comprehensive framework," *Journal of Automation and System Engineering* (in-press).

57. Wang, K. (2006). "Data Mining in Manufacturing: The nature and implications," *Knowledge Enterprise: intelligent strategies in product design, manufacturing, management.* Springer New York, 1-10

58. Wang, K., Tong, S., & Eynard, B. (2007). "Review on Application Data Mining in Product Design and Manufacturing," in Proceedings of the Fourth International Conference on *Fuzzy Systems and Knowledge Discovery* - FSKD 2007, IEEE Los Alamitos, 4: 613-618.

59. Wang, Y., Caron, F., Vanthienen, J., Huang, L., & Guo, Y. (2014). "Acquiring logistics process intelligence: Methodology and an application for a Chinese bulk port," *Expert Systems with Applications,* 41(1): 195-209.

60. Wang, Z., Du, P., & Yu, Y. (2009). "An intelligent modeling and analysis method of

manufacturing process using the first-order predicate logic," *Computers & Industrial Engineering*, 56(4): 1559-1565.

61. Wasserman, S. (1994). "Social network analysis: Methods and applications," Cambridge university press; USA.

62. Weber, P., Bordbar, B., & Tino, P. (2013). "A Framework for the Analysis of Process Mining Algorithms," *Systems, Man, and Cybernetics: Systems, IEEE Transactions* on 43(2): 303-317.

63. Weijters, A.J.M.M., van der Aalst, W.M.P., & De Medeiros, A.A. (2006). "Process mining with the heuristics miner-algorithm," *Technische Universiteit Eindhoven, Technical Report*. WP, 166.

64. Weijters, A.J.M.M. and van der Aalst, W.M.P. (2003). "Rediscovering workflow models from event-based data using little thumb," *Integrated Computer-Aided Engineering*, 10(2): 151-162.

65. Weijters, A.J.M.M. & van der Aalst, W.M.P. (2001). "Process mining: discovering workflow models from event-based data," in Proceedings of the 13th Belgium-Netherlands Conference on *Artificial Intelligence (BNAIC 2001),* 283-290.

66. Wen, L., Wang, J., & Sun, J. (2006). "Detecting Implicit Dependencies Between Tasks from Event Logs," In proceedings of 8^{th} international conference on *Frontiers of WWW Research and Development - APWeb 2006*, Springer Berlin Heidelberg, LNCS 3841: 591-603.

67. Weske, M., van der Aalst, W.M.P., & Verbeek, H.M.W. (2004). "Advances in business process management," *Data Knowledge Engineering*, 50(1): 1-8.

68. Witsch, M. and B. Vogel-Heuser (2012). "Towards a Formal Specification Framework for Manufacturing Execution Systems," *Industrial Informatics, IEEE Transactions* on 8(2): 311-320.

69. Zhong, Y., & Shirinzadeh, B. (2008). "Virtual factory for manufacturing process visualization," *Complexity International*, 12: 1-22.

Acknowledgement

Foremost, I would like to show my greatest appreciation to my advisor Prof. Minseok Song for the continuous support and enthusiasm. His guidance helped me in all the time of research and writing of this thesis, and it was very valuable experience for me. I could not have imagined having a better advisor. Besides my advisor, I would like to express my gratitude to my thesis committees: Prof. Duck-Young Kim and Prof. Daeil Kwon, for their encouragement, insightful comments, and hard questions.

Additionally, I thank my lovely lab-mates in BPI Lab: Hanna Yang, Dr. Seong-Cheol, Dr. Bernardo, Min-su, Ye-rim, Minjeong, HoJung, and Ik Jae, for the sleepless nights we were working together and for all the fun we have had in the last years. I would like to offer my special thanks to Hanna for her endless love. I also thank to all Management Engineering members, Su-bin, Jinseok, Jason, Jung-bum, Jiyu, and all other members in management school for giving enjoyable time and valuable friendship to me. In addition, my appreciation goes to my all other friends for the pleasant time together.

Finally, I wish to thank to my family.