ULSAN NATIONAL INSTITUTE OF SCIENCE AND TECHNOLOGY

Master's Thesis

Process Mining-driven Performance Analysis in Manufacturing Process: Cost and Quality Perspective

Tu Thi Bich Hong

Department of Management Engineering

Graduate School of UNIST

2016

Process Mining-driven Performance Analysis in Manufacturing Process: Cost and Quality Perspective

Tu Thi Bich Hong

Department of Management Engineering

Graduate School of UNIST

Process Mining-driven Performance Analysis in Manufacturing Process: Cost and Quality Perspective

A thesis
submitted to the Graduate School of UNIST
in partial fulfillment of the
requirements for the degree of
Master of Science

Tu Thi Bich Hong

06. 10. 2016

Approved by

Advisor
Changyong Lee

Process Mining-driven Performance Analysis in Manufacturing Process: Cost and Quality Perspective

Tu Thi Bich Hong

This certifies that the thesis of Tu Thi Bich Hong approved.

06/10/2016

Signature

Advisor: Changyong Lee

Signature

Committee Member #1: Minseok Song

Signature

Committee Member #2: Marco Commuzzi

Abstract

The dynamics of globalization and high expectation of customers make manufacturing enterprise move towards three primarily competitive factors, namely, time, cost, and quality. The desire for continuous performance improvement of manufacturing processes is as old as manufacturing itself. However, the latest industrial revolution Industry 4.0 and the digital revolution have opened up new avenues, intertwined information systems and the operation processes. As a result, enterprises face a challenge in extracting value from a massive amount of events recorded by today's information systems. Process mining is well recognized as a valuable tool for observing and diagnosing inefficiencies in business processes based on event data. It turns out that process mining is a viable solution to this challenge. Nevertheless, significantly less attention has been paid on investigating cost and quality perspective in process mining. In these respects, this thesis suggests a framework for performance analysis in manufacturing processes based on process mining. The proposed approach focuse on cost and quality perspective. Specifically, the contributions of this thesis are in four-fold (i) to suggest a method to extend event log of manufacturing process with manufacturing information, i.e. cost, quality; (ii) to analyze manufacturing information, i.e. cost, quality with process model; (iii) to utilize various existing process mining techniques and develop new approaches to analyze and predict manufacturing cost; (iv) and to enable quality report in manufacturing process.

Keywords: Process Mining, Manufacturing Process, Performance Analysis, Manufacturing Cost, Quality-related KPIs.

Contents

List of Figures

List of Tables

List of Algorithms

Chapter 1

Introduction

This chapter introduces the motivation of the thesis. The problem statement and objectives are discussed in Section 1.2 and Section 1.3. The outline of the thesis is explained in Section 1.4.

1.1. Motivation

Manufacturing process concerns all efforts of an organization to add values to the inputs (raw materials, semi-finished products, know-how, etc.) and transform them into the outputs (finished goods) that meet an expectation of customers [1]. Nowadays, manufacturing environment is becoming more and more sophisticated [2]. Furthermore, increasing the level of competition in the global market has a significant impact on enterprises to stay competitive to survive. Leading manufacturers have seen the continuous improvement of process quality and the reduction of manufacturing cost as strategic weapons in the market to compete against peers. Quality is being known as one of the essential dimensions of business process performance [3-6]. Efficiency in managing of manufacturing cost is considered as a critical factor in obtaining competitive advantages. Also, making production decision is based on cost [7]. Therefore, getting a clear understanding of manufacturing process regarding cost and quality constitutes the most important prerequisite in this respect.

The paradigms of information and sensor technology development have enabled large-scale data collection when monitoring manufacturing processes. The knowledge learned from those data may gain a potential value for performance improvement of the process. However, the discovery of knowledge hidden in the data without proper tools is a challenge due to a significant amount of data. Commonly used approaches in manufacturing process performance analysis have a vast amount of literatures, for example statistical method [8], data mining techniques, i.e. association rule, sequence data mining [9, 10], and simulation method [11]. An enterprise has applied various quality control techniques to improve the process quality by reducing its unevenness. These include Six Sigma [12], Statistical Process Control (SPC) tools, Total Quality Management (TQM) [13], Business Process Re-engineering (BPR) [14, 15], and Business Process Improvement (BPI) [16]. However, related methods have not provided an adequate solution to this issue yet. For instance, statistical and data mining results are sometimes too complicated to understand [17]. Furthermore, simulation has the limitation that it takes too much time and cost to build acomplex manufacturing process model [17].

Process mining is well recognized as a valuable tool for analyzing an operational process and tracking down its problems or inefficiencies using event logs [18]. These event logs can be obtained from Process Awareness Information Systems (PAISs) such as Manufacturing Execution System (MES), Enterprise Resource Planning (ERP), Customer Relationship Management (CRM), and so forth [19, 20]. Furthermore, process mining framework also allows us to discover models, analyze bottlenecks, conduct conformance checking and analyze process performance [21, 22].

Process mining has been applied in various fields such as finance, healthcare and port. In recent years, process mining has begun to be implemented in manufacturing areas. For instance, Son et al. [23] proposed a method to analyze an overall production process based process mining. Rozinat et al. [24] introduced a way to improve the test processes in ASML using process mining. Park et al. [25] suggested an approach to measure the performance of the ship block manufacturing processes. Hence, process mining is the most promising approach for performance analysis of manufacturing processes.

1.2. Problem Statement

In the manufacturing industry, it is essential that enterprises reduce their manufacturing cost and improve process quality continuously. Unfortunately, they encounter a challenge to analyze manufacturing cost due to the complexity of manufacturing processes. For instance, cost for a particular task or an individual resource is hard to be calculated. The questions regarding consuming cost can be only answered at the end of the process. The quality of a process refers to the ability of that process to ‘produce and deliver quality products’ [26], and to ‘conform to manufacturing specifications’ [27]. Interest in improving the quality of a process has grown throughout the manufacturing community. Unfortunately, one of the difficulties in improving the quality of a process is figuring out where to start and deciding whether the implemented changes are beneficial. Therefore, quality aspects of a manufacturing process are often neglected or deferred due to the difficulty in measurement [28]. Considering these issues, there is a need of proper method allowing access to the cost and quality status of each stage or each activity in the process.

Although process mining literature has discussed a wide range of its application to solve these issues, not all of them have adequately supported these perspectives. Firstly, although significantly amounts of research have been conducted in process mining, they have focused on time and resource perspective rather than the perspective of cost and quality. Next, there exist few studies on cost mining in process mining [29]. However, these efforts have not considered manufacturing processes [30-32]. Furthermore, they have not adequately assisted cost mining of manufacturing process where cost relies on not only time but also production volume. Production volume means the total units of products coming into a process. Assume that there is no loss or omission of goods during the manufacturing process. Meanwhile total output equivalent to total input. The output may include good products and defect products. Lastly, the approach is taken by [23] enables a way to control the quality of manufacturing process through yield analysis by combining the input and output of each activity with event logs. It has become an emerging topic in the field of quality engineering using process mining. Nevertheless, this thesis has not entirely proposed a systematic and holistic method towards the quality perspective of manufacturing process yet. These drawbacks necessitate the development of a proper method to handle the entire cost and process quality mining in the manufacturing sector.

1.3. Objectives

To overcome such limitations, this thesis suggests a framework for performance analysis in manufacturing processes based on process mining and focuses on cost and quality perspective. The primary objectives of the study are as follows:

1) *To extend an event log of manufacturing process with manufacturing information, i.e. cost, quality*

2) *To analyze manufacturing information, i.e. cost, quality with a process model*

3) *To utilize various process mining techniques and develop new techniques to analyze and predict manufacturingcost*

4) *To develop a method to generate quality report in manufacturing process*

1.4. Outline

The rest of this study is organized as follows. Chapter 2 explains related works which are process mining, its application in the manufacturing sector, cost and quality mining based on process mining. An overview of performance analysis with cost-related KPIs and quality-related KPIs is presented in Chapter 3. Chapter 4 introduces a performance analysis in manufacturing process. An implementation is illustrated in Chapter 5. Finally, this thesis ends with conclusions and possible future works in Chapter6.

Firstly, in order to extend an event log with manufacturing cost, this study first formalizes and extracts manufacturing cost model. Furthermore, a cost database structure is presented to support the extraction cost model from various information sources. Later, this study formalizes event logs of manufacturing process. With the respect to enhancing the event log with quality and quality model with cost information is generated based on quality-related KPIs. Then, the event log is updated with quality components, and a quality index is directly computed in event log using the quality model.

Secondly, the manufacturing information would be easier to be interpreted and more significant and precise for decision makers if they are associated with the corresponding elements of the process model. Therefore, a process model extension with manufacturing information such as cost and quality are presented.

Thirdly, various visualization methods are used to present analysis results, such as cost analysis with a process model, visualization of a cost breakdown, and structure resource utilization cost. Cost analysis with process model is a systematic method that allows users to access the detailed cost of a particular activity or an individual resource involved in a manufacturing process. Visualization of cost breakdown identifies the distinct cost types including material cost, labor cost and overhead cost comprised of the total cost of a task during process execution. Resource utilization cost deals with all information related to the cost of resources e.g. individual use of each resource. Furthermore, this thesis develops a cost prediction method which relies on time prediction and production volume using progress of manufacturing processes.

Finally, this study suggests quality analysis in process mining. This goal is accomplished by the generation of accurate, relevant quality report.

Chapter 2

Related works

In this chapter, basic concepts of process mining and its applications in manufacturing are introduced. In particular, cost and quality mining are discussed in detail.

2.1. Process mining

The purpose of process mining is, on the one hand, to extract process-related information from transaction logs stored in information systems [33]. The starting point of process mining is an event log in which each event refers to a particular activity or task and is related to a case. Here, a particular activity or task is a well-defined step in some process, whereas a case is defined as a process instance. In addition, each event referes to an originator who (which) is a person (device) exercuting the task or timestamp of the event. Event logs for analysis can be derived from PAISs (Process Awareness Information Systems) such as Workflow Management, CRM (Customer Relationship Management), ERP (Enterprise Resources Planning), and so on [19, 20]. On the other hand, techniques in process mining can be applied to support monitoring phases in BPM lifecycles by analyzing processes as they are being executed. Process mining can also be understood in a broader context of BI and BAM [34], but with the aim of offering insights into processes (Where is the bottleneck?).

Process mining is categorized into three types: discovery, conformance, and extension. Traditionally, based on the process information stored in a log, process mining focuses on a discovery process model. One then performs a conformance check by comparing the priori model with the observed behavior [35]. Finally, an extension is carried out with the aim of improving the initial process model by taking other attributes such as time and cost into consideration [20]. Besides the three types mentioned above, other perspectives in process mining can be identified. These include *Control-flow perspective* (“How?”), *Organizational perspective* (“Who?”), *Performance perspective* (“What?”). A *control-flow perspective* focuses on control-flow, an ordered flow of activitives, aiming at finding good characterization of each possible path. A control-flow can be expressed either by a Petri net or [19] some other notations such as Event-driven Process Chain (EPC) [36]. An *organizational perspective* targets at resource fields, e.g. people, systems, roles, departments, etc. and their correlations. Its goal is to construct an organization by classifying people regarding their roles and organizational units [36-38]. Finally, a *performance perspective* has the focus on identifying bottenecks, measuring service levels, and monitoring resource utilization [38].

2.2. Application of process mining in manufacturing

Process mining has been successfully applied in many fields, for example, finance, healthcare, port, to name but a few. In the case of manufacturing areas, there have been studies in process mining conducted in various ways. For example, Rozinat et al. [24] presented the applicability of process mining to less structured processes of wafer steppers in ASML based on process mining. It was a pioneering study which applied process mining to analyze manufacturing processes. Son et al. [23, 39] conducted research on overall manufacturing process examination based on process mining. They performed four types of analysis, i.e. the machine-to-machine inter-relationship analysis, visualization of production flows, machine utilization, and monitoring & diagnosis of task performance based on process mining [17]. Park et al. conducted research on performance analysis based on a dataset of Korean ship block manufacturing processes [25]. In that study, they proposed a systematic approach to evaluating the performance based on actual work data that are stored in the database of manufacturing information systems and provided a guideline for the improvement of underperforming Business Process Management about the manufacturing process. Additionally, manufacturing performance analysis including workload and delay using event logs was suggested by Park et al. [39]. These researchers have shown that process mining is a valuable tool for tracking down problems with current situations on manufacturing processes. However, the research has not focused on cost perspective and quality perspective yet.

2.3. Cost and quality mining in process mining

Nauta [29] proposed an architecture to support cost awareness in process mining by integrating a management accounting field and business process management, and this was the first implementation in this regard. The research has laid a good foundation of cost mining for later studies Wynnet al. [30] made an explicit link between cost and processes in all phases of the lifecycle of the business process management. They also described a research agenda that considered a holistic approach to managing the cost of business operations in a structured manner. Additionally, Wynn et al. [31] first implemented the cost reporting and cost prediction functionalities in the process mining framework ProM. A short research paper [32] defined how Workflow Management Systems (WfMS) can provide support for strategic cost-informed operational decisions. The article [40] extended the earlier work by providing a detailed discussion of the realization of the cost-informed operational support within the well-known open-source WfMS system environment YAWL.

Eventhough there have been many researches on cost mining in process mining, cost mining of production processes where the cost relies on not only time but also production volume has not been adequately investigated. Moreover, it is crucial that a manufacturing cost model be developed based on the characteristics of manufacturing cost. Therefore, this study is conducted. Based on the initial idea of event log-annotated cost, this study proposes a method to enhance event logs of manufacturing with manufacturing information such as cost. A database structure is also created to support generating a manufacturing cost model. Manufacturing cost analysis and prediction then follow.

Son et al. [17] enabled a way to control the quality of production processes through yield analysis by combining the input and output of each activity with event logs. It has become an emerging topic in the field of quality engineering using process mining. Nevertheless, [17] has not entirely proposed a systematic and holistic method towards the quality perspective of manufacturing process yet.

Chapter 3

Performance Analysis

The concept "performance" of an organization can be seen as the degree that organization meets its objectives [41, 42]. Performance analysis related to supports an organization to measure and analyze the performance of their business processes. Supporting an organization means not only describing what has happened and diagnosing why it happened, but also guiding to set the performance target for future. The process performance can be measured through Key Performance Indicators (KPIs) [43]. Key Performance Indicators are considered as the best representations of the strategy of an organization. They reflect the essential success factors of an organization and support it to assess sustainable manufacturing performances towards its organizational goals. Typically, performance measurement covers three dimensions including time, quality and cost [44-49]. This study focuses on cost and quality. We have conducted literatures review in an attempt to determine which indicators are commonly used in manufacturing performance evaluation in these two dimensions.

Cost-Related KPIs: To improve a manufacturing process, it is vital to be aware the cost aspect. There are different perspectives on cost. Cost can be categorized into fixed cost (overhead cost) and variable cost such as material cost, labor cost, etc. The cost of each activity depends on its utilization, types of resource used, and the duration of activities. This paper focuses on both the costs associated with a task and the cost associated with a resource.

Quality-Related KPIs: The well-known quality-related KPIs are commonly used in manufacturing performance evaluation to give us a better understanding of the process quality. Quality process control needs to be based on KPIs that serve as guard rails to keep quality management on track to meet and exceed customer requirements first. The quality of a manufacturing process can be viewed from at least two different faces: the process participant's side (internal quality) and client's side (external quality) [43]. A literature review has been carried out in an attempt to determine quality-related indicators commonly used in manufacturing performance evaluation. The most popular quality-related KPIs using to measure process quality are defect rate [50, 51], repetition [43, 51, 52], first pass yield [51, 53].

3.1. Cost-related KPIs

The commonly used cost-related KPIs in manufacturing performance analysis include overhead cost [52], material cost [52][51] and labor cost [52][51]. Previous studies have shown that manufacturing costs are the essential ones for convertin inputs into products [54]. They also indicated that manufacturing cost is typically divided into three categories including direct labor cost (DL), direct material cost (DM), and manufacturing overhead cost (OH). A manufacturing process concerns all efforts of an organization to add values to the inputs and transform them into the outputs [7, 55]. Each activity includes attributes of actuality as well as plan data such as timestamps, working progress, etc. Thus, the total cost of a manufacturing process is the summation of all costs of individual activities in that process. Early organizations used management accounting techniques, e.g. activitity-based costing, to allocate cost per activity. However, these costing techniques are difficult (or expensive) in implementation and maintenance. In this study, we analyse them based on data inputs and attempt to design a conceptual cost database structure.

3.3.1. Activithy-Based Costing (ABC)

Activity-Based Costing (ABC) is that costing in which the overhead cost is assigned to a product based on activities required for the production [56, 57]. This first defines all overhead cost such as depreciation, indirect salaries, and utilities allocated to activity pools. An activity pool is a collection of the costs that relate to an activity. Activity in the ABC is an event that causes the consumption of overhead cost [56, 57]. Next, costs are assigned from activity pools to a cost object, such as good or service.

ABC is a valuable tool to improve the accuracy of products costs, helps managers to understand the nature of overhead costs. However, it has limitations. Firstly, the implementation and maintenance of an ABC system are costly [56, 57]. Specifically for short-term plans in which balancing between the implementation cost andits expected benefits could be formidably challenging. . Secondly, information gathering and updating for ABC are also expensive [58]. In addition, an ABC model could be not accurate due to a number of critical assumptions based on abstractions of reality [59]. Lastly, ABC does not take excess capacity of resources into account [60].

3.3.2. Manufacturing Cost

The basic concept for calculating manufacturing cost has been employed elsewhere [54]. This study aims to extend the original ideas mentioned above by extending a simplification step in computing the cost per activity (or job) of a manufacturing process. In particular, each of the three cost types of manufacturing cost is rigorously computed and presented in the below sections. Each formula consists of two main parts. The first part is constant which is the same value of an event representing an activity performed by the same resources. The second part is the so-called variable part which varies from event to event. The latter can be derived from an event log. For instance, production volume is used for material cost calculation and the duration time is used for calculation of labor cost and machinery cost, another overhead cost. The first part becomes part of a cost driver, whereas the second part is embedded in an event log. In Table 1, we include the cost notations used in Sections 3.1.1, 3.1.2, and 3.1.3.

Table 1. Cost Notation

Notation	Explanation
DM	Direct material cost
MR	Material rate
Con	Unit of net consumption
t	Ratio of waste materials allowance
$\Pr$	Standard price of material
Q	Production volume
k	A task in a manufacturing process
j	The order of material in material list $(1 \le j \le M)$
M	The number of materials involved in a task
DL	Direct labor cost
LR	Labor rate
T	Time for task completion
E	The number of different employees concerned task
a	The order of employee $(1 \le a \le E)$
IDM	Indirect material cost
IL	Indirect labor cost
MC	Machinery cost
T	Duration
w	Percentage of value added
Pur	Purchase cost of machine
Tr	Tradeoff price of machine
l	Expected life span of machine (hours)
K	The number of difference machines used for task
b	The order of machine $(1 \le b \le K)$
AOH	Other overhead cost
AQ	The actual quantity of the allocation base used by each job
M	The number of difference materials used for task
P	The predetermined manufacturing overhead rate
AQ	The actual quantity of the allocation based used by each job

3.3.1.1. Direct material cost

Direct material cost is the cost of materials which are parts of a finished product. The formula of the direct material cost is given in Equation (1). Here, two main parameters are taken into consideration: the material rate (MR_j) and the production volume ($Q_{j)}$. The former is derived from a cost model, whereas the latter is stored in the event logs. The material rate is derived from cost model while production volume is obtained in event logs. The material rate is calculated as the product of the unit of net consumption (Con_j), is multiplied bythe ratio of waste materials allowance (t_j),the standard price of material ($P_{j)}$.

$$DM_k = \sum_{j=1}^{M}\left(MR_{j,k} * Q_{j,k}\right) = \sum_{j=1}^{M}\left(Con_{j,k} * t_{j,k} * \Pr_{j,k} * Q_{j,k}\right) \tag{1}$$

3.3.1.2. Direct labor cost

Direct labor cost is the cost of workers who directly add value to a product. The direct labor cost is calculated as follows. The labor rate is taken from cost model while duration time is derived from event log.

$$DL_k = \sum_{a=1}^{E}\left(LR_{a,k} * T_{a,k}\right) \tag{2}$$

3.3.1.3. Overhead cost

Overhead cost includescost of all activities that support a manufacturing processbut often are not directly related to any particular product. Moreover, this study has separated OH cost of the indirect labor cost, indirect material cost, machinery cost, and other overhead costs for the aim of detailedextraction.

The overhead (OH) cost of Task k can be expressed as:

$$OH_k = IDM_k + IL_k + MC_k + AOH_k \tag{3}$$

Machinery cost is the cost of all machines used in a manufacturing process. The machinery cost is computed in Equation (3.1) below, which takes the percentage of value added (*w*), the purchase cost of all machine (*Pur*), the expected lifespan of machine (in hours) (*l*) into account. Similarity to direct labor cost, machinery rate is obtained from cost model and duration is taken from event log.

$$MC_k = \sum_{b=1}^{K} \left(w_{b,k} * \left(\frac{Pur_{b,k} - Tr_{b,k}}{l_{b,k}} \right) * T_{b,k} \right) \quad (3.1)$$

In order to compute other overhead costs, we adopt the formula in [61] as follows:

$$AOH_k = P_k * AQ_k \quad (3.2)$$

The predetermined manufacturing OH rate is calculated by the summation of estimated manufacturing OH cost multiplied with a summation of the estimated quantity of manufacturing OH allocation based. Traditionally, manufacturing enterprises use allocation bases, for example, direct labor hours (for labor – intensive production environments) or machine hours (for machine-intensive production environments) [62].

3.4. Quality-related KPIs

Various definitions of quality are given in different fields. Traditionally, industrial manufacturing literature emphasizes quality as conformance to manufacturing specifications [63]. literature in economics and marketing often take quality to a term of performance quality that is virtually synonymous with the class or position of the product [63]. Karmarkar (1991) separated process quality as conformance to manufacturing specifications, with product quality as conformance to customer expectation [64]. Quality of a process refers to the ability that the process to produce and deliver quality products [26]. Product quality relies on the quality of the final product while process quality emphasizes on steps used to produce the end product.

In this research, the term quality refers to process quality. The most popular quality-related KPIs used to measure process quality are defect rate [50, 51], repetition [43, 51, 52], and firstpass yield [51, 53].

3.4.1. Defect rate

A defect is defined as nonconformity to intended usage demands or is a deviation from the requirements of the specification [50]. There are three categories of defects, i.e. critical defect, major defect, and minor defect. Typically, a critical defect is likely to be hazardous for users when using the product. It may lead to the loss of a tactical function of the product. A major defect could result in failure or reduction in the usability of the unit of goods for its intended purpose. A minor defect is deemed to a discrepancy from the standards. However, it is not likely to affect the usability of the unit of product for its intended purpose. In [50], it is pointed out that the root of defection is an error. However, errors do not always lead to defects. Defects may or may not result in failure of products. Defects may happen at any stage of a production process. There is no other option except that a defected product is removed from a process output once detected [65].

Defect rate (Dfr) can be measured by counting the number of defects (df) per the total number of manufactured products (N) [50].

$$Dfr = \frac{\sum df}{\sum N} \tag{4}$$

Defect rate plays a crucial role in the improvement of yield and financial conditions of any enterprise [66]. Furthermore, the reputation of an organization will be ruined if defective products reach customers. Therefore, enterprises have attempted to to reduce the defect rate of their production as much as possible during the production procesess. Inspection of defects at each stage of a manufacturing process allows us to figure out at which stage more defects are likely to occur. The cause factor may be then detected to reduce the defection. Causes of defects are categorized into three types: material, human, and machinery.

3.4.2. First pass yield

First pass yield (FPY) is an important manufacturing metric for measuring quality and production performance, which is given as the ratio of the number of units completed from a production process to specification subtracted by the number of scraps and reworks (Y) compared to the total number of units coming into the process (X) [52].

$$FPY = \frac{X - Y}{X} \tag{5}$$

Clearly, reduction of rework should be the goal of every manufacturer for improving first pass yield to achieve the possible lowest product cost. This necessitates the design of an accurate method for measuring and tracking FPY throughout a production process.

It is vital to use a process model as a guide for evaluating how efficient each step involves in the process and how well the overall process is performing. all reworked products are assumed to take only one time to be qualified. Fig. 1 demonstrates how FPY is calculated in a process model. According to this example, 90 of the 100 pieces that entered step A went through Step A correctly the first time. Therefore Step A FPY= 90 / 100 = 90.0%. 80 of the 90 pieces went through Step B correctly the first time through. Therefore Step B FPY = 80 / 90= 88.9% and so on.

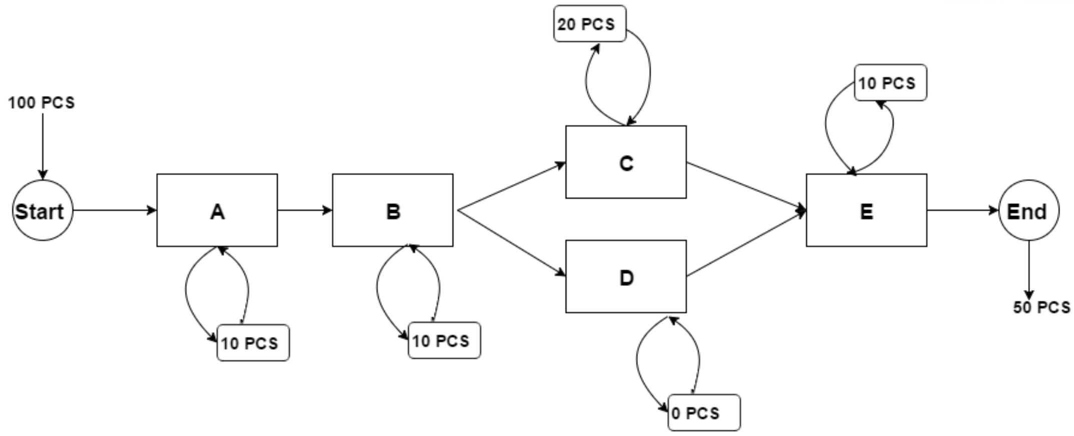


Figure 1. Example of first pass yield caculation

3.4.3. Repitition

Rework or repetition is an indicator of the level of production quality. In [3], rework is defined as the process by which a product is reformatted to conform to the standard requirement by correction. Reference [5] sees rework as thequality deviation, for example, a product rejected from a regular production and reprocessed into a finished product. Rework is often extremely time-consuming and labor-intensive. It involves nonproductive activities for which clients are not willing to pay for. Mishandling of rework may result in the loss of customers, degradation of enterprise reputation, and consequently the loss of profits. Reworks in the industry are popular works that hamper the smooth production process. By reacting quicker in minimization of reworks, productivity can be improved significantly.

It is essential to identify causes of manufacturing rework to amend the performance of products [68]. Mainly, rework results from errors, omissions, failures, damages, and change orders throughout the procurement and the production process [63, 64]. References [63,65, 66] categorized causes of rework into three types of rework factors, i.e. technical, quality, and human resource factors.The repetition can be found in sequences of a task and measured through execution logs[43]. We define the repetition in Equation (5) below, which is adapted from [43].

$$r = 1 - \frac{T}{CT} \tag{6}$$

Here, T denotes the time to execute Task k only once or the average execution time of k per instantiation, CT is the cycle time of task k and can be calculated as the average executnumber of times to perform the execution time of k per case, and r is the repetition probability from a series of event logs.

Chapter 4

Performance Analysis in Manufacturing Process

4.1. Overview

In this section, we explain the overall performance analysis in manufacturing processes, as well as give a brief explanation of each step at the same time. As described in Fig. 2, the proposed framework consists of two main steps: (1) an extension of event logs with manufacturing information and (2) a performance analysis of manufacturing processes. Step 1 emphasizes on the extension of event logs with manufacturing cost and process quality. First of all, this work formalizes the event logs of a manufacturing process for a plan with actual data. Most of the manufacturing industry has not only real information but also planned information. The planned data are optimized in consideration of cost, time, and so on. The actuality data represent results recorded in information systems in real time. This study uses both planned and actual data for cost prediction which is presented in section 4.4.2.3. With respect to the extension of the event logs with cost, a cost model is presented, and a cost database model is created for the cost model extraction. For the event logs extended with quality, we provide guidance for creating a quality model and computing quality-related KPIs. Finally, we form the event logs including manufacturing information which serve as the inputs for various visualization techniques for both cost and quality perspective discussed in Step 2.

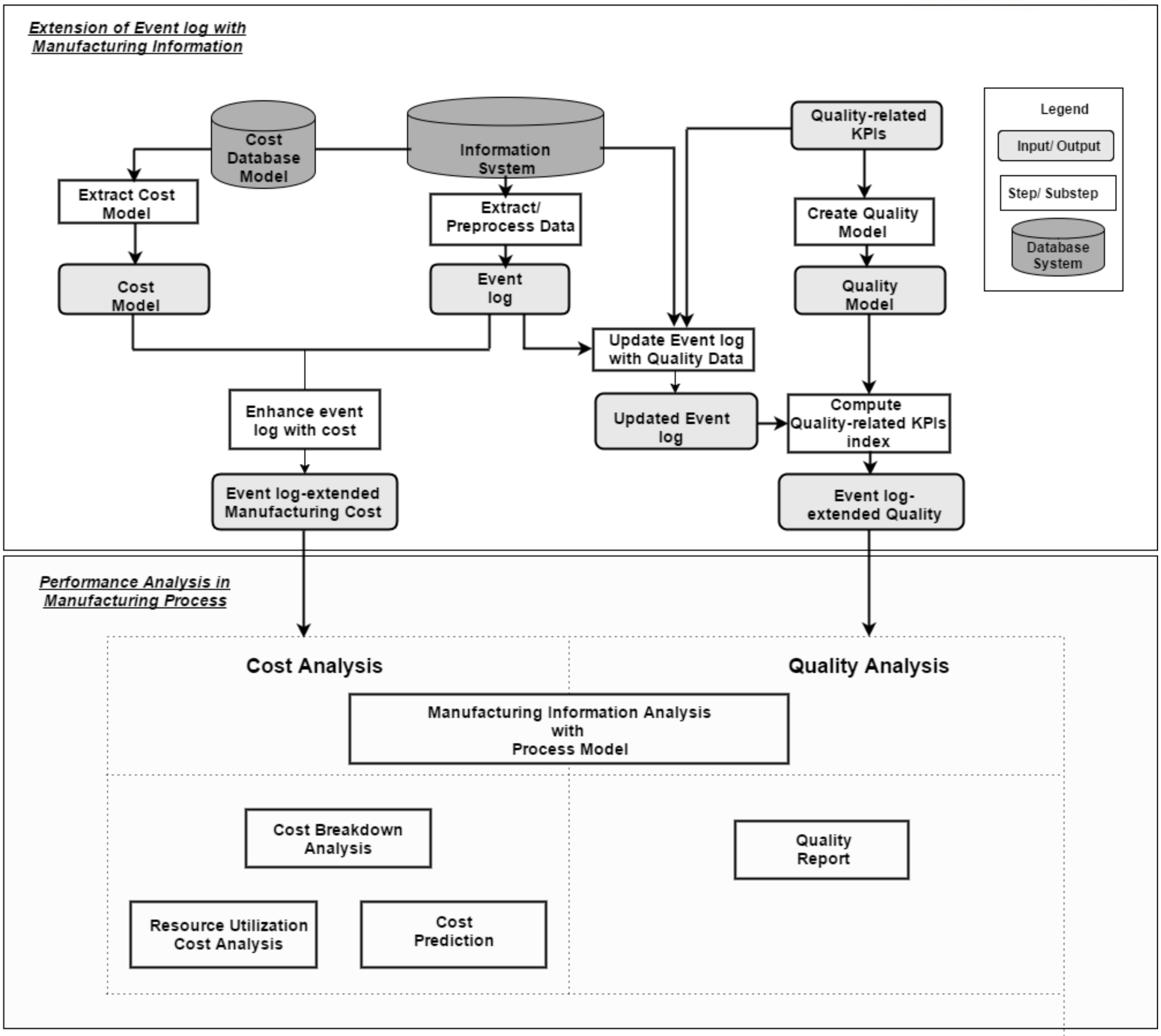


Figure 2. A methodology for performance analysis in manufacturing process

Definition 1. Event, Case, Trace of manufacturing processes

Let K be the event universe, i.e., the set of possible event identifiers of a manufacturing process. Denote by A the set of activities. Let the set attributes of an activity in the planned and actual data be denoted as follows: S and S' the sets of start timestamps, R and R' the sets of resources, Q and Q' the sets of production volumes, W and W' the sets of working progresses ($0 \leq W \leq 1$). *An event E is a tuple of (A, S, C, R, Q, W,* S', C', Q', W'*). For any event* $e \in E$, $\pi(e)$ *is the value of attribute n of event e.*

Case*: A case C is a set of events with attributes. Cases always have a trace, denoted by* $\sigma \in E^*$ *is a finite sequence of events such that each event appears only once, i.e., for* $1 \leq i \leq j \leq \ / j_\sigma \ / : \sigma(i) \neq \sigma(j)$ *and time is non-descending denoted* $\overline{\sigma(i)} \leq \overline{\sigma(j)}$ $(if\ i$ *occurs before j).*

Definition 2. Event log of manufacturing processes

Let $L \in \beta(C)$ *be an event log and* $\beta(C)$ *is the set of all bags (multi-set) over C.*

4.2. Running Example

Throughout this thesis, we implement a running example to demonstrate the applicability of our approach discussed in this chapter. This section discusses the example in detail. We notice that no real world examples of a manufacturing process together with event logs and cost or quality data are available in this work. Therefore, an artificial dataset of Jean's manufacturing process, its manufacturing cost, and quality data are used. The event logs include information about activities, resources, etc. These event logs can be created by executing a process simulation, and we assume that they do not contain noise or unexpected behaviors. We further suppose the followings: the total number of outputs equals those of inputs, there is no loss or omission of products during the manufacturing process, and the output may contain both good and bad products.

The example information is summarized as follows: 11 cases with 80 events in three months from March 14th, 2015 to May 20th, 2015 and there involves 8 activities (Cutting, Embroidery, Sewing, End line checking, Washing, Finishing, Checking finish garment and Packing out).

The Cutting process receives primary and sub fabric from the warehouse, does the shrinkage tests of material, and cuts them. Depending on customer requirements, Embroidery may be involved in the process. Otherwise, the semi-products will be transferred directly from the cutting stage to sewing stage. Next, quality assurance (QA) department checks on products at the production lines. The failure ones will be returned, and quality products will continue to generate in washing department. The Finishing process stacks the button and rivet. Then QA department checks the quality of goods again before delivering to customers. Finally, the Packing out process stitches hand tags, barcode stickers for the products, and put them in the poly bags and carton boxes ready for delivery.

Table 2. Fraction of event log

Case ID	Event Id	A	S	C	R	Q	W	S'	C'	R'	Q'	W'
1	MFP28981	Cutting	03.14.15	03.30.15	Cutting Shop	5565	100%	03.14.15	03.30.15	Cutting Shop	5565	100%
1	MFP28982	Sewing	03.30.15	04.15.15	Sewing shop 2	5565	100%	03.30.15	04.15.15	Sewing shop 2	5565	100%
1	MFP28983	Endline checking	04.10.15	04.17.15	QA dept	5565	100%	04.17.15	-	Washing Shop	4174	75%
1	MFP28984	Washing	04.14.15	04.23.15	Washing Shop	5565	100%	03.19.15	03.20.15	Cutting Shop	4174	100%
1	MFP28985	Finishing	04.13.15	05.03.15	Finishing Shop	5565	100%	03.29.15	04.03.15	Sewing shop 4	4174	100%
1	MFP28986	Checking finish garment	04.27.15	05.05.15	QA dept	5565	100%	04.03.15	-	Washing Shop	4174	35%
1	MFP28987	Packing out	05.05.15	05.11.15	Packing Shop	5565	100%	-	-	Finishing Shop	4174	0%
2	MFP29371	Cutting	03.19.15	03.20.15	Cutting Shop	500	100%	03.21.15	03.22.15	Cutting Shop	480	100%
2	MFP29372	Embroidery	03.22.15	03.27.15	Embroidery	500	100%	03.22.15	03.27.15	Embroidery	480	100%
2	MFP29373	Sewing	03.29.15	04.03.15	Sewing shop 4	500	100%	03.30.15	04.03.15	Sewing shop 4	480	100%
2	MFP29374	Endline checking	04.01.15	04.04.15	QA dept	500	100%	04.01.15	04.04.15	QA dept	450	100%
2	MFP29375	Washing	04.03.15	04.11.15	Washing Shop	500	100%	04.03.15	04.11.15	Washing Shop	450	100%
2	MFP29376	Finishing	04.07.15	04.14.15	Finishing Shop	500	100%	04.07.15	-	Finishing Shop	450	50%
2	MFP29377	Checking finish garment	04.12.15	04.15.15	QA dept	500	100%	-	-	QA dept	420	30%
2	MFP29378	Packing out	04.15.15	04.20.15	Packing Shop	500	100%	-	-	Packing Shop	420	5%

In Table 2, we show a fraction of the event logs which consist of cases, events, timestamps, working progress, resources, and production volume. For instance, the case with the case ID 1 consists of sequence of events and activities: MFP28981 (Cutting), MFP28982 (Sewing), MFP28983(Endline checking), MFP28984 (Washing), MFP28985 (Finishing), MFP28986 (Checking finish garment), and MFP28987(Packing out) ; the event MFP28981 (Cutting) has the attributes such as start timestamps for plan and actual (S = {03.10.15}, S' = {03.14.15}), complete timestamps (C={03.20.15}, C'= {03.30.15}), resources (R ={ Cutting Shop}, R' = { Cutting Shop}), working progress (W = {100%},W' = {100%}), production volume (Q ={5565}, Q'= {5565}).

4.3. Extension of event log with manufacturing information

Before proceeding to the performance analysis of manufacturing processes, we discuess in this section the concepts of extension of event logs with manufacturing information, i.e. cost-related KPIs, quality-related KPIs. These form the basis of our main research.

4.3.1. Extension of event log withcost-related KPIs

With respect to the event log extension with cost-related KPIs, a cost model and the creation of cost database model for manufacturing cost model extraction are defined in the following sub-sections.

4.3.1.1. Cost database model

Creation of a cost database model that supports the extraction of a cost model component, namel a cost driver, is the primary focus of this section. As mentioned earlier, manufacturing cost includes material cost, labor cost, and overhead cost. Acost model should comprise these cost types. Therefore, the core idea for the construction of a cost database model is that a cost driver connects to the cost types and event logs through a bridge over activities and resources. Furthermore, an organization may assign raw materials to each pair of activities and resources using a material requisition form. This form contains all information about activities, resources, material IDs, consumption, as well as material rates. Employees, machines, and other OHs are allocated to each pair of activities and resources through employee allocation, machine allocation, and other OH allocation. To formalize the cost database model, we need to define a cost model. To do so, we introduce and define a cost driver and cost function as follows.

Definition 3. Cost driver of manufacturing process

Let $L \in \beta(C)$ *be an event log. Denote by A a distinguishable set of activities of L, R' a set of actual resource of event log L,* $AR' \subseteq A \times R'$ *a set of non-ordered pairs of activity from A and resource from* $R^{'}$ *of event log L. Assume that Q is some universe of values. A cost driver of a manufacturing process is a tuple CD = (*A^{cd}*,* R^{cd}*, lr, ilr, mr, imr,mar,ohr) over K. Here,* A^{cd} *is a finite set of activities,* R^{cd} *is a finite set of resources,* AR^{cd} *is a finite set of pairs activity from* A^{cd} *and resource from* R^{cd}*,* $lr \in AR^{cd}$ *is a set of direct labor rates,* $ilr \in AR^{cd}$ *is a set of indirect labor rates,* $mr \in AR^{cd}$ *is a set of direct material rates,* $imr \in AR^{cd}$ *is a set of indirect material rates, mar* $\in AR^{cd}$ *is a set of machine rates, ohr* $\in AR^{cd}$ *is a set of other overhead rates. We required that:* $A^{cd} \subseteq A$*,* $A \in L$*,* $R^{cd} \subseteq R'$*,* $R' \in L$*,* $AR^{cd} \subseteq AR'$*,* $AR \in L$*,* $AR^{cd} = [a_i]_{i=1}^{M_{AR'} \leq N_{AR'}}$ *such that* $\begin{cases} a_i \in AR', \forall i = 1, M \\ \forall a_i, a_k \in AR^{cd}, l \neq k \end{cases}$

We assume in this study that each cost function is a formula for supporting coherence between a cost driver and anevent log. All cost functions are mutually distinct from one another in the cost model. Each of them has a unique identifier. We give the definition of a cost function below.

Definition 4. Cost function

Assume that I is some space of values (formulas, strings, etc). Let CF = (I^{cf}, N^{cf}, F, D) *be a cost function over I where* I^{cf} *is a set of function identifiers,* N^{cf}*is a set of function names, F is a set of function formulas ,and D is a set of descriptionssastifying that* $\forall i_l^{cf}, i_k^{cf} \in I^{cf} \left| i_l^{cf} \neq i_k^{cf} \right., \forall f_i, f_j \in F \left| f_i \neq f_j \right.$. *Given a cost model of manufacturing process, we define four basis cost functions*

- *Material cost, denoted by* $cf_1 = MR*Q$*, satisfying a material rate MR and production volume Q*
- *Labor cost, denoted by* $cf_2 = LR*T$*, satisfying a labor rate LR and processing time T*
- *Machinery cost, denoted by* $cf_3=MaR*T$*, sastisfying a machine rate MaR and T*
- *Other OH cost, denoted by* $cf_4=OHR*T$*, sastisfying other overhead rate OHR and T*

Definition 5. Cost model of manufacturing process

Let CM =(CD, CF) where CD is a cost driver defined in Definition 3, and CF is a cost function defined in Definition 4. We call CM a cost model if there exists a uniqueness CD in CM, i.e., $\exists!$ $CD \in CM$*, then there is exist* $\exists!\ CF \in CM$ *.HereCF and CM are disjoint,* $CD \cap CF = \{\phi\}$.

We call CM a cost model if there if there exists uniquely $\exists!\ CD \in CM$*, then there exists uniquely* $\exists!$ $CF \in CM$.

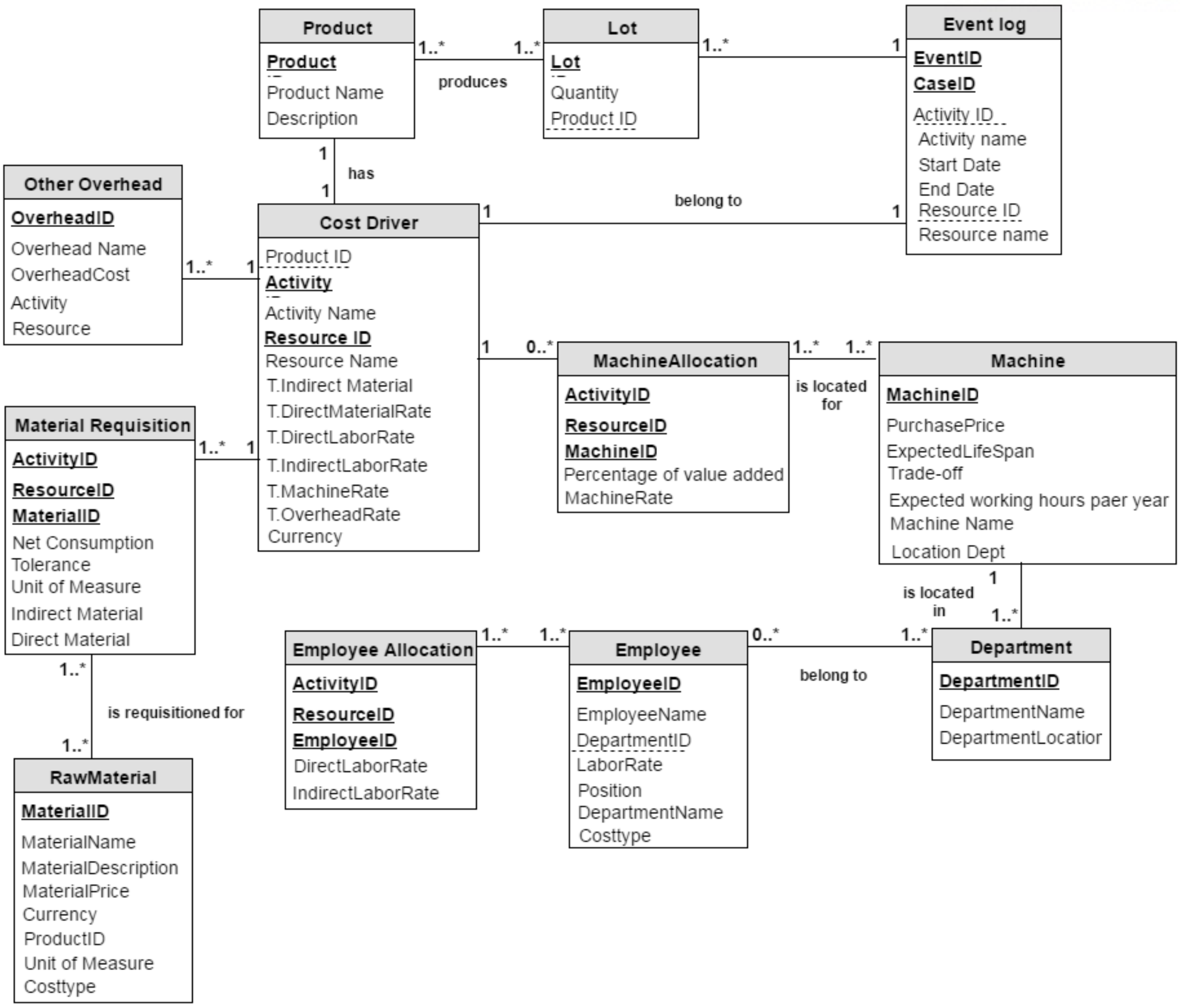


Figure 3. Cost database model for cost model extraction

In this work, we create a cost database structure with an example shown in Fig.3. Our database structure focuses on business rules as well as significant constraints. In the database, the entities are connected through their relationships which provide concrete notations for the data model structure used in this study. Entity-relationship provides concrete notation for data model structure used in this study [67]. Here, the solid lines indicate primary keys, whereas the dotted ones represent foreign keys. these pairs of keys, we can consider the attributes referred to them as the equivalence for the purposes, i.e. relating to the same event. We will do so in attempting to describe events involved in different tables but use different column names (attributes in the events) to store the same values. Therefore, attributes Activity ID and Resource ID in the cost driver entity and other entities of cost types such as other overhead, Material requisition, Machine allocation, Employee allocation are considered to be equivalent. In the same meaning, we have the following pairs of primary and foreign keys: Material ID (RawMaterial) and Material ID (Material Requisition), Machine ID (Machine Allocation) and Machine ID (Machine), Employee ID (Employee Allocation) and Employee ID

(Employee). The cardinality constraints in Fig.3 impose restrictions on object models. For example, a cost driver corresponds to one or more material requisition and each material requisition corresponds precisely one cost driver (see annotations “0..*” and “1” in solid line between cost driver and material requisition).

Definition 6. Cost database structure

Assuming V to be some space of values. A data structure is a tuple DS = (E^{ds}, A^{ds},R^{ds}, PK, FK) where E^{ds} *is a set of entities,* R^{ds} *is a set of relationships between two entities,* $A^{ds} \in E^{ds}$ *is a set of attribute names,* $PK \in E^{ds}$ *is a set of primary key names,* $FK \in E^{ds}$ *is a set of foreign key names, Such that:* $PK \neq \phi, \forall_{i,j} \in PK$ such that $i \neq j$, $FK \neq \phi$, PK *and* FK *are disjoint sets, that is* $PK \cap FK = \phi$, *there exists a unique entity in the database model, that is* $\exists! \, E^{ds} \in DS$.

4.3.1.2. Event log-extended manufacturing cost

Algorithm 1 presents an approach to extending an event log with manufacturing cost. A cost model and an event log form the input for event log-extended manufacturing cost.

Fig.4 shows a fraction of an event log-extended cost which consists of cases, events, timestamps, working progress, resources, and production volume. For instance, the case with case ID 1 consists of a sequence of events and activities: E01 (A01), E02 (A02), and E03 (A03); event E01 (A01) has the attributes such as start timestamps for plan and actual (S = {03.10.15}, $S^{'}$ = {03.14.15}), complete timestamps (C={03.20.15}, $C^{'}$= {03.30.15}), resources (R ={R01}, $R^{'}$ = {R01}), working progress (W = {100%},$W^{'}$ = {100%}), production volume (Q ={5565}, $Q^{'}$= {5565}), direct labor cost(lr ={0.71}), indirect labor cost (ilr ={3.49}), direct material (mr ={14.15}), indirect material (imr = {0.04}),machinery cost (ma ={0.5}), other overhead cost (oh = {0.02}). As shown in Fig.4 the material cost is calculated based on production volume. Labor cost, machinery cost, and other overhead cost are carried out based on processing time.

Algorithm 1. Extension of event log with manufacturing cost (L^{CM})

Input: an event log L, a cost model CM

Output: an event log-entended manufacturing cost L^{CM}

For(each $trace \in L$) **do**

  For(each $e \in L$) **do**

  For(each $cd \in CD$) **do**

    If($A^{cd} = A \cap R^{cd} = R$) **then**

      calculate cost from driver for event;

      If $(c = mr \cup c = imr)$ **then**

        using $cf_{1,}$ $L^{CM}.add(c \rightarrow e)$;

        else if $(c = lr \cup c = idr)$ **then**

        using cf_2, $L^{CM}.add(c \rightarrow e)$;

        else if $(c = mar)$ **then**

        using cf_3,$L^{CM}.add(c \rightarrow e)$;

        else

        using cf_4,$L^{CM}.add(c \rightarrow e)$;

      End

    End

  End

  End

End

Case ID	Event ID	A	S	C	R	Q	lr	mr	ilr	imr	ma	oh
1	E01	A01	03.10.2015	03.20.15	R01	5565	0.71	14.15	3.49	0.04	0.5	0.02
1	E02	A02	03.21.15	04.15.15	R02	5565	0.86	1.17	14.15	0.41	0.7	0.02
1	E03	A03	04.17.15	04.23.15	R03	5565	0.18	0	0.04	0.01	0.78	0.01
2	E11	A01	03.19.15	03.30.15	R01	500	0.25	0	1.17	0.5	1.04	0.03
2	E12	A02	03.29.15	04.03.15	R02	500	0.3	4.08	0	0.7	0.95	0.04
2	E13	A03	04.03.15	04.11.15	R03	500	0.27	0	0	0.78	0.36	0.01
2	E14	A04	04.07.15	04.14.15	R04	500	0.04	3.49	4.08	1.04	0.41	0.02

Figure 4. Event log-extended cost of manufacturing process

4.3.2. Extension of event log with quality-related KPIs

4.3.2.1. Quality-aware Event log

While the cost attributes of an event log are hooked up from a cost model satisfying the matching of activity and resource from both sides, we can directly compute the quality index in the event log by using its attributes in the response to the quality perspective. In this respect, the process quality model is formalized in Definition 9. The way to measure and aggregate them into an event log using a quality model is shown in Algorithm 2.

A set of appropriate quality-related KPIs is derived from a consideration of enterprise goals and observations in manufacturing processes. Quality-related KPIs can be measured either directly or indirectly. We first define their characteristics, components, and how to calculate them. For example, let us-consider a defect rate. The quality components of this KPI are defective products and total input products. Afterward, the related data are collected and updated for each event in an event log. Simultaneously, a quality model which supports functions to compute a quality index from the updated event log is generated based on the quality-related KPIs. Meanwhile, quality models are necessary for providing consistent terminology and guidance forthe quality computation and are the basis for the evaluation of any process. A quality model includes quality drivers and quality functions. The quality driversare associated with quality-related KPIs and their quality components used to calculate them. The quality functions contain all functions that support the computation of the quality index in the event log.

Definition 9. Quality Model

Let $QM = \{QD, QF\}$ *be a quality model where* QD *is a cost driver,* QF *is a cost function sastisfying*

- $QD = \{I^{qd}, N^{qd}, QC^{qd}\} \mid \forall i,k \in I^{qd}, i \neq k$ *. Here* I^{qd} *is a quality driver identifier,* N^{qd} *a quality driver name,* QC^{qd} *a quality component.*

- $QF = \{I'^{qd}, I^{qf}, F, D\} \mid \forall m,l \in I^{qf}, m \neq l; \forall n \in I^{qd}, n' \in I^{qd} \mid n \equiv n'$ *. Here* I'^{qd} *is a quality driver identifier,* I^{qf} *is a quality function identifier,* F *is an expression of formula,* D *is a description.*

- $\exists! QD \in QM$, $\exists! QF \in QM$ *. Here* QD *and* QF *are disjoint,* $QD \cap QF = \{\phi\}$.

This work only covers three popular quality-related KPIs including defect rate, first pass yield, and repetition which are usually used to evaluate the quality of manufacturing process. The following definitions formalize quality-related KPIs in an event.

Definition 10. Deriving Defect Rate in Event

Let $L \in \beta(C)$ *be an event log. Let* b_i *be a number of defect product of event* e_i $(1 < i \leq n)$, p_i *be number of input product of event* e_i. *A defect rate* dfr_i *of event* e_i *can be measured by counting the number of defect product per total input product denoted as* $dfr_i = \frac{b_i}{p_i}$.

Definition 11. Deriving First Pass Yield

Let $L \in \beta(C)$ *be an event log. Let* X_i *the number of units coming out of aevent* e_i $(1 < i \leq n)$, Y_i *be number of rework and scrap products ofevent* e_i. *A first pass yield* FPY_i *of event* e_i *can be measured by* $FPY_i = \frac{X_i - Y_i}{X_i}$.

Definition 12. DerivingRepitition Probability in Event log

Let $L \in \beta(C)$ *be an event log. Let* T_i *be time to execute the event* e_i *in event log L only once or the average execution time of* e_i *per instantiation* $(1 < i \leq n)$; CT_i *be the cycle time of event* e_i *and can* *be calculated as the average execute number of times to perform the execution time of event* e_i *per case. A repetition probability* r_i *of event* e_i *can be measured by* $r_i = 1 - \frac{T_i}{CT_i}$.

4.3.2.2.Computation of quality-related KPIs index

The input for the computation of quality-related KPIs index is is an updated event log which is well defined in the previous section, together with the quality model. The computation are well described in Algorithm 2 below.

Algorithm 2. Computation of quality-related KPIs index (L^{QM})

Input: an event log *L*, a quality model *QM*

Output: an event log-entendedqualityL^{QM}

For(each $trace \in L$) **do**

    For(each $e \in L$) **do**

    For(each $qd \in QD$) **do**

        calculate quality for event;

        Using $qf \in QF$) **sastisfying** $qd \equiv qd', qd \in QD \cup qd' \in QF$

        $L^{QM}.add(qn \to e)$;

    End

    End

End

4.4. Performance analysis

We discuss in this section the event log-extended cost and event log-extended quality. These are inputs for all the analyses below.

4.4.1. Manufacturing information analysis with process model

The cost or quality information would be easier to be interpreted, more significant and accurate for decision makers if they are associated with the corresponding elements of the business process model. In this regard, manufacturing information analysis with process model in the form of process model-extended manufacturing information, i.e. cost and quality is presented. Process model-extended manufacturing information is defined in the form of a frequent sequence graph-extendedmanufacturing information. A frequent sequence graph has been retrieved from event log using frequent sequence mining [68, 69]. The following definitions formalize a frequent sequence graph and frequent sequence graph-extendedmanufacturing information.

Definition 13. Frequent sequence graph

We denoted by G= $(N, a_s, a_e, E, L_N, L_E, f^n, f^e, l)$ *a frequent sequence graph with its attribute defined as below.*

- *N is a finite set of nodes represening an event* e_i*; thereforeall nodes have all attributes of the event they represent for and are denoted by* $\#_n(e_i) = \#_n(n_i), n_i \in N$,
- $a_s \in N$, *is the start node such that* $a_s \neq \phi$,
- $a_e \in N$, *is the start node such that* $a_e \neq \phi$,
- $E = \{(n_i, n_j) \mid (n_i, n_j) \in N \times N, and (n_i, n_j) \in E\}$ *is a set of edges represeningt the frequent sequence among nodes,*
- $L_N \in N$ *is a set of node labels,*
- $L_E \in N$ *is a set of edge labels,* $f^n \in N$ *is a set of frequency such that f:* $n_i \rightarrow n_j, denoted(n_{ij})$ *which represent the total flow occurring from* n_i *to* n_j .

Definition 14. Measurement function

A measurement function is a function that, given a bag of measurements, produces some cost values, e.g., the sum, average, min, max. Formally, measure $\in P(M) \rightarrow M$, i.e, *for some bag of* *measurements p,* $\cos tmeasure(p)$ *returns some cost value,* $qualitymeasure(p)$ *returns some quality value.*

Let us assume that $p = [p_i]_{i=1}^{n}$, *i.e., the* measurements taken as a sample. The sample total is defined as follows: $measuretotal(p)$ = sum $\{[p_i]_{i=1}^{n}\}$. Then the sample mean is defined as follow $\bar{p} = \frac{\sum_{i=1}^{n} p}{n}$. Other prediction functions can be used for the measurements, for example $measure\min(p)$ = min $\{[p_i]_{i=1}^{n}\}$ or $measure\max(p)$ = max $\{[p_i]_{i=1}^{n}\}$.

Definition 15. Frequent sequence graph-extendedmanufacturing information

Let $G(L) = (N, a_s, a_e, E, L_N, L_E, F, l)$ *be a frequent sequence graph of event log L. A frequent sequence graph –enhanced manufacturing information is a tuple* $C(L) = (G(L), O, P)$ *in which*

- *O: E* $\rightarrow$ *N is a function that associates an event [e]with cost or quality.Its event label occurs in the event log, and corresponds with the equivalence switched measurement, i.e. O([e]) = P([e]),*
- *P: is a set of aggregate functions to measure the cost or quality in a graph defined in Definition 4,* $\forall i, j \in P, i \neq j$.

Fig.5 illustrates the conceptual idea of the process model-extended cost. Each event in the graph is represented as a node, i.e., start event a_s = {DP01}, end event a_e = {DP08}. The graph shows the occurrence of nodes, e.g. N_{DP01}= 11, N_{DP04} =*3*, and flow between nodes, e.g. $N_{DP01,DP04}$ = 3. Based on this result, we can deduce the detailed cost of execution time of each activity or resources in the process. By comparing them, we can determine which the cheapest or the most expensive activity is. For instance, the total cost of DP02 is 2406.12 USD and the execution time is 85 days. Thus, cheapest one is Activity *DP04* with 384.64 USD, the most expensive one is Activity *DP05* with 6716.84 USD.

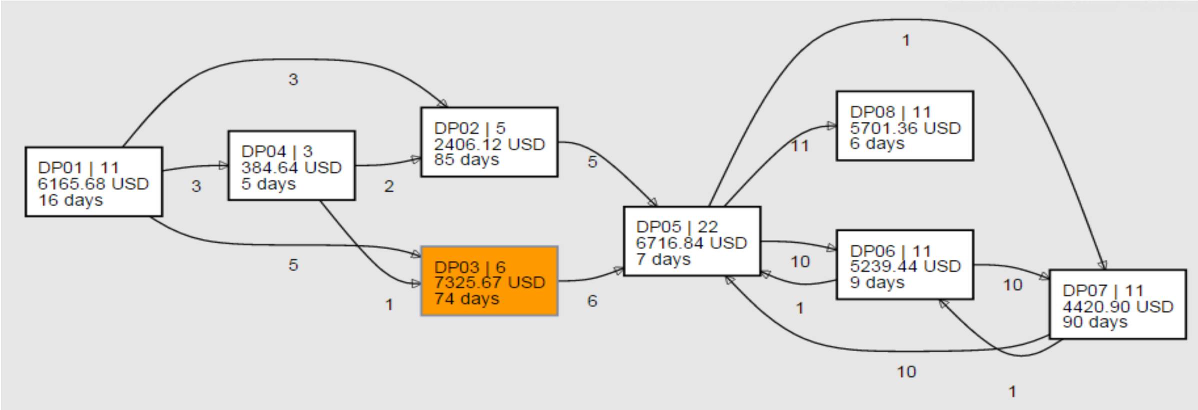


Figure 5. An example of cost analysis with process model

4.4.2. Cost Analysis

4.4.2.1. Visualization of cost breakdown

While a process model-extended cost only provides an overview of the cost of activities involved in a process, visualization of cost breakdown, on the one hand, shows a detailed view of specific cost components of each activity in the process. On the other hand, cost breakdown is the systematic technique of identifying the distinct cost types, i.e. material cost, labor cost, overhead cost comprised of the total cost of a task during process execution. It assigns a particular cost value to each cost category. Also, the value of the individual cost type is expressed as a percentage of the total cost. Furthermore, this plug-in offers a function that tracks back to cost value of tasks per each cost category. Moreover, by clicking each bar located on the bar chart, one can interpret information such as direct labor cost, indirect labor cost, direct material cost, indirect material cost, machinery cost, and other overhead expenses in the pie chart.

Definition 16. Visualization of Cost breakdown

We define CB = (B, P, T, R, r, v, M) *a visualization of Cost breakdown. Here the attributes are defined as below.*

- *B is a bar chart showing cost per activity a* $\mid \forall a_i, a_j \in A, i \neq j$,
- *P is a pie chart showing cost per cost type cp* $\mid \forall cp_k, cp_l \in P, k \neq l \cup cp \in C = \{lr, mr, ilr, mr, imr, ma, oh\}$,
- *T is a summary table such that* $a \in T \cup cp \in T$,
- *R is a set of relationship between B and P,*
- $r \in R \rightarrow Q(BP)$ *is a function mapping the selected activity in B onto a set of values of cost*

type in P,

- *v* $\in R \rightarrow Q(PB)$ *is a function describing the selected cost type in P onto a set of values of activity in B ,* $r \cap v = \phi$*,*
- *M is a finite set of measurement* $m \in P(M) \rightarrow M$ *|P(M)=(Total, Min, Max, AVG).*

4.4.2.2. *Resource utilization cost*

Resource utilization cost, on the one hand, takes an event log-extended cost as the only input and calculates the cost per resource in a given log. On the other hand, resource utilization cost deals with all information relating to the cost of resources, i.e. individual use of each labor utilization cost and machine utilization cost (e.g., the cost of machine allocation and cost of machine working/idle). Users can select one or more of the provided measurements: total, max, min, AVG. This plug-in also offers many chart types, for example,bar, line, pie, and column. Furthermore, users can quickly execute the resource cost distribution per timeline of each project e.g. month, year. Based on this result, the process management or board of directors may define which resource is the most expensive or which is the cheapest. Moreover, they can quickly interpret whether resource utilization cost is reasonable or not. As a result, they can seek solutions for timely adjustments. Understanding cost allocation rules of each activity in its project timeline, the users can effectively manage their project budget, as well as avoid a deficit. Based on this result, they can accurately predict the budget for the follow-up.

Definition 17. Resource utilization cost

A resource utilization cost is a tuple RC= (RM, CT, M, vis) *where*

- *RM is a set of resource cost types such that* $RM \in C = \{lr, ilr, ma\}$*,*
- *M be a finite set of measurements* $m \in RM(M) \rightarrow M$ *|RM(M)={Total, Min, Max, AVG}*
- *CT is a set of chart types such that* $CT \in$*{Column, Bar, Pie, Line}*
- *vis* $\in RM \cap M \cap CT$ $\rightarrow$ *RC (vis) is a function mapping each resource cost type and measurementonto its chart type.*

4.4.2.3. *Cost prediction*

Inspired by the demand to predict the cost of a manufacturing process, this research introduces a cost prediction that allows us not only to execute the cumulative cost at the present statebut also to predict the completion cost of some process instance. Also, the prediction can be applied for each category of cost types.

We assume that there is a close relationship between time and cost, between costand production volume. Therefore, we conjecture that cost prediction should be based on time prediction and production volume. Whenever an organization requires a prediction of the completion time of process instance, we can take a partial trace and consider its working progress. The working progress of an event denoted by "W" refers to the percentage completion of that event $(0 \le W \le 100\%)$. In case an event has started but not completed yet $(0 < W < 100\%)$, we can easily track back to the consumed time, as well as to calculate the remaining time using working progress. Otherwise, we can learn from its plan timestamps.Definition 18 below gives a formal definition of time prediction for an event in an event log.

Definition 18. Event, time prediction

Let σ *be the Trace in event log* $L \subseteq C$. G*iven event* e_m *in the trace such that* $e_m \in \sigma_i, 1 \le m \le n : n = length(\sigma_i)$, *the remaining time until completion event* e_m $d_{remaining(e_m)}$ is computed by

$$= \begin{cases} d_{plan(e_m)} & if \quad w_{uptodate}(e_m) = 0 \\ d_{uptodate} * \dfrac{(1 - w_{uptodate}(e_m))}{w_{uptodate}(e_m)} & if \quad 0 < w_{uptodate}(e_m) \le 1 \end{cases}$$

in which:

$d_{plan(e_m)} = E_{plan(e_m)} - S_{plan(e_m)}$ *: the budget execution time to complete event* e_m *,*
$d_{uptodate(e_m)} = E_{uptodate(e_m)} - S_{actual(e_m)}$ *: the actual execution time until reporting date.*

The partial trace has fully filled the timestamps according to Definition 18. Subsequently, we can calculate the cost of the events in the trace. Given a trace σ_i in an event log, a cumulative cost of an event e_m in the trace refers to a sum of cost of all event follow e_m from the start event, and e_m cost itself. Thus, the last event of the trace has the maximum cumulative cost. Chosen any event e_m in a trace, we suppose that the remaining cost refers to the abstraction of maximum cumulative cost and its cumulative cost.

Definition 19. Trace, Event, Cost measurement

Let σ *be a trace in event log* $L \subseteq C,$. *Given event* e_m *in the trace such that* $e_m \in \sigma_i$ *for m is the order of event in trace,* $1 \le m \le n : n = length(\sigma_i), 1 \le i \le /\sigma/,$ *produces some measurment* $ACost(e_m)$: *cumulative cost and* $RCost(e_m)$ *:remaining cost.*

$$ACost(e_m) = \begin{cases} Cost(e_m) & if \quad m=1 \\ ACost(e_{m-1}) + Cost(e_m) & if \quad m \neq 1 \end{cases}$$

Where:

$$ACost(e_{m-1}) = Cost(e_1) + Cost(e_2) + \ldots + Cost(e_{m-1})$$

$e_{m-1} > \sigma e_m$: *event* e_{m-1} *directly follows event* e_m

$$RCost(e_m) = \begin{cases} 0 & if \quad m=n \\ \max \sigma_i (ACost) - ACost(e_m) & if \quad m \neq n \end{cases}$$

Note that these are not real estimators for the whole event log, but only for the trace in a log. We suppose that the cost of an event in a log refers to a set of cost of that event in the traces denoted by $Cost(e_m)_L = \left[Cost(e_m)_{\sigma_i}\right]_{1 \le i = a \le n}^{n} = \left[Cost(e_m)_{\sigma_i}, Cost(e_m)_{\sigma_{i+1}}, \ldots\right]$ Therefore, we also have the cumulative cost of event in log = $ACost(e_m)_L = \left[ACost(e_m)_{\sigma_i}\right]_{1 \le i = a \le n}^{n}$ *and remaining cost of event in log =*

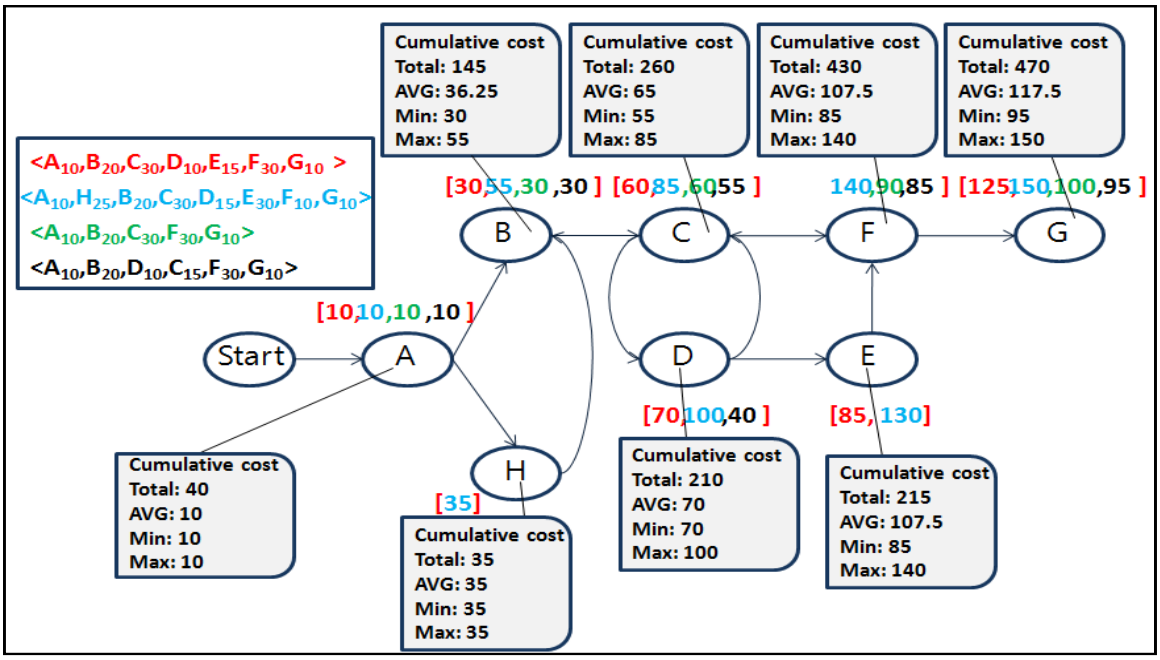


Figure 6. Cumulative cost showing per activity

$$RCost(e_m)_L = \left[RCost(e_m)_{\sigma_i}\right]^n_{1\le i=a\le n}$$

As illustrative example, we employ the event log and present it in Fig.6. Each row corresponds to a process instance, e.g., the first trace <A10, B20, C30, D10, E15, F30, G10 > refers to a process instance where task A has a cost of 10, task B has a cost of 20, task C has a cost of 30. Similarity, we cost of activity D, E, F, G are performed. The first trace starts from A and ends at G, so the cumulative cost is minimal at A and maximal at G. Now we are ready to calculate the cumulative cost of each task. A is the start task, so cumulative cost of A is itself 10. Therefore, 10 is added to state {A}. Next, B is directly followed by A. Therefore, the cumulative cost at B refers to the cost of B 20 plus the cumulative cost of A 10, thus equals to 30. Then 30 is supplemented to state {B}. Similarly, we insert 60 (=30+30) to {C}, 70 (=60+10) to {D}, 85 to {E}, 115 to {F}, 125 to {G}. Consider the second trace < A10, H25, B20, C30, E30, F10, G10>. As another example, we add 10 into {A}, 35 into {H}, 45 into {B}, 75 into {C}, 105 into {E}, 115 into {F}, 125into {G}. These steps are echoed for all other traces. Consequently, state{A} is annotated with a bag containing four elements:[10, 10, 10, 10]. State{B} is annotated with a bag containing four elements: [30, 55, 30, 30]. State{C} is also annotated with a bag containing four elements: [60, 85, 60, 55], so on. Now cost measurement functions are applied to calculate promptly the demanded cost. For example, cost of task B is defined as follow: total = sum (30, 55, 30, 30)=145, mean = 145g/4 = 36.25, max = max (30, 55, 30, 30) = 55, min = min (30, 55, 30, 30) =30.

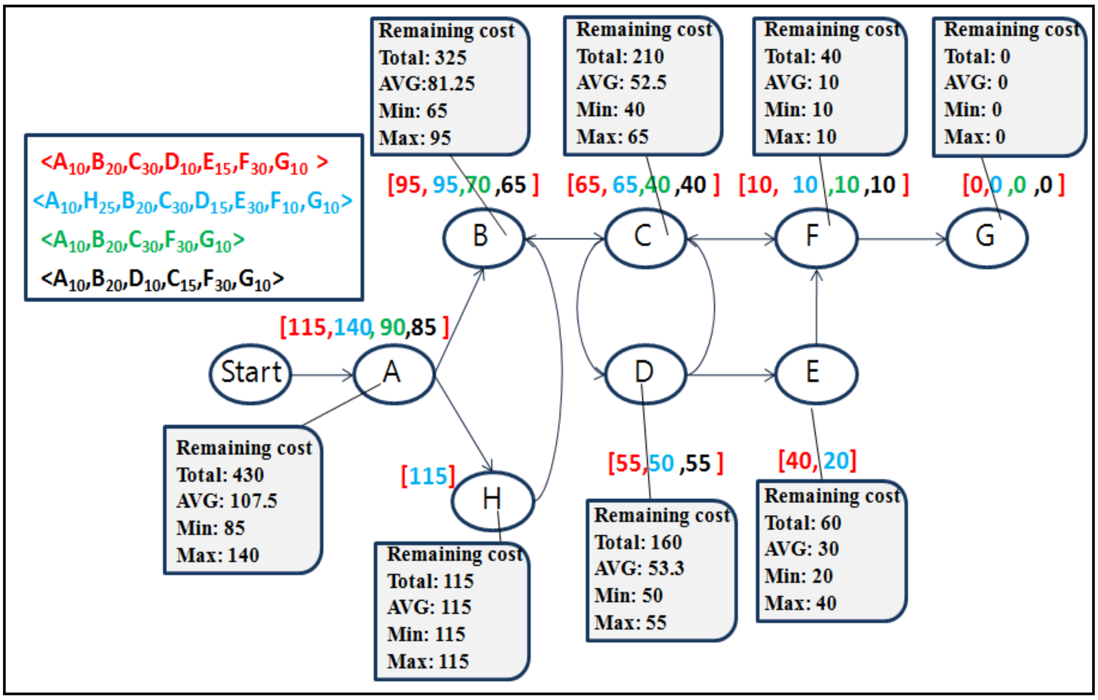


Figure 7. Remaining cost showing per activity

We now turn our attention to calculate the remaining cost. Returning back to the first trace, the maximum of cumulative value of this trace is 125. Using Definition 11, it is ready to knowabout the remaining cost of each task in a trace. For instance, task A has 10 in the cumulative cost . Thus, the remaining cost of A is115 (= 125-10). Then, 115 is added to state {A}. Task B has a cumulative cost 30. So the remaining of the cost of B is 95 (= 125-30). As a result, 95 is added into state {B}. Following the are same way, we have the remaining cost of C 65, D 55, E 40, F 10, G 0. Next, these steps are echoed for all other traces and are the same with the calculation of the cumulative cost. Consequently, we have the remaining cost shown per activity in Fig.7.

4.4.3. Quality Analysis

4.4.3.1. Quality report

Event log-extended quality-related KPIs can be used to generate quality reports aggregating quality by case, quality type, task, case data, or any other data available in completed task instances. Quality reports can provide different views (i.e., process, resourcehave been identified to support on-demand quality reporting. Quality functions have been defined in the quality model describing the content of a quality report. Dynamic reporting using quality functions is left as future work. Static reports are implemented taking in a quality annotated event log producing an overview of quality by case, quality type, or both, showing the potential for using quality annotated event logs for quality reporting.

Generating reports from different viewpoints is also an important consideration for this study. Quality reports that were generated from the process or resource point-of-view enable users to analyze and understand business operation quality from different perspectives. From the process's viewpoint, users can identify which task of the process consumes the highest quality or the most time. From the resource perspective, users can determine which resource was fully utilized (or underutilized) or the quality incurred in resource consumption. By illustrating these different aspects of quality reports, organizations can make better decisions based on the correlation of different perspectives in their business operations.

Chapter 5

Implementation

In this chapter, we present the implementation of our proposed approach using My SQL (), JavaScript, HTML (Hyper Text Markup Language) environment. Fig. 8 shows the system architecture of the proposed method. We first use My SQL to handle our database including the cost data, quality data, and the event logs. We also use this to extract an event log-enhanced cost. We remind that in our approach, the event log-extended cost and event log-extended quality are treated separately. We next employ PHP () to fetch data from the event log-enhanced cost, transform it into a JSON format, and save it as a PHP script file. This is later served as the input for cost plug-in's which are run under Javascript and HTML framework. These include cost analysis with a process model, cost breakdown analysis, cost prediction, resource cost utilization, quality analysis with a process model, and quality report. Depending on different plug-in's, PHP extracts various types of data such as process model information (nodes, link, frequency, duration,i.e.,) or cost, quality information only. Finally, the result is generated and visualized to users.

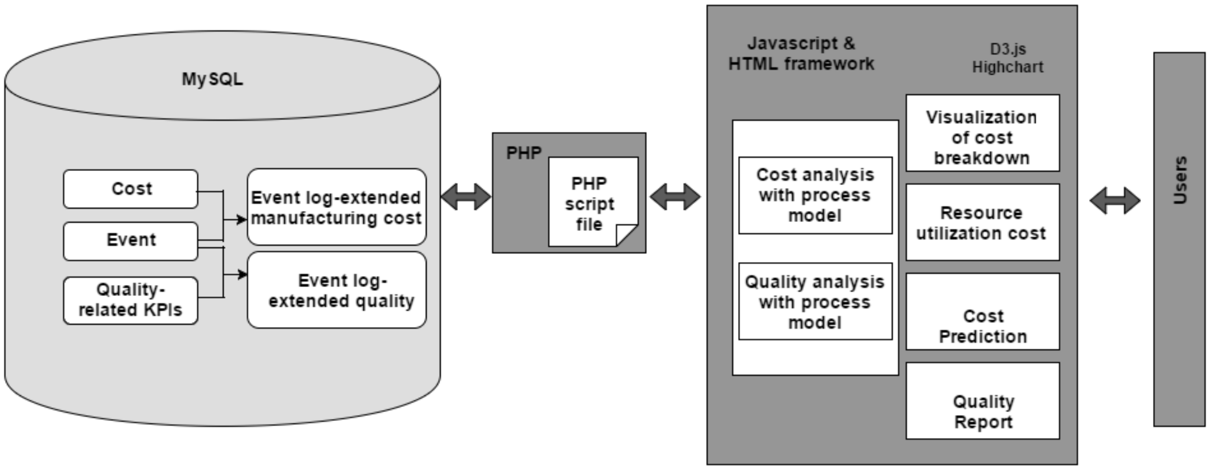


Figure 8. System architecture of performance analysis

5.1. Process model-enhanced manufacturing information

As shown in Fig.9, the inputs for the process model-enhanced manufacturing information are event log-extended manufacturing cost and event log-extended process quality. The proposed approach includes three main steps: (1) extraction of the process model information and extract manufacturing information such as cost or quality from event log- extended process quality; (2) generation of the process model based on frequent sequence mining, and enriching it with cost or quality information; and finally, (3) display of the process model along with the cost or quality information.

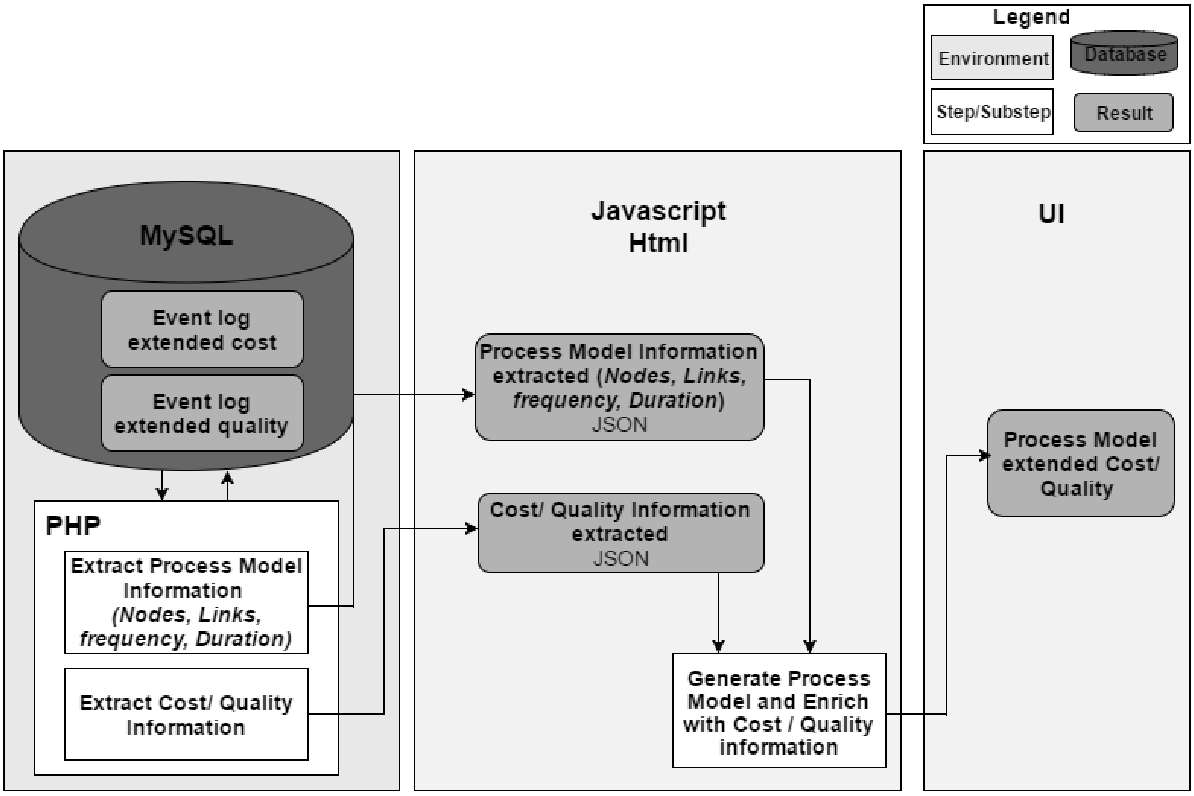


Figure 9. A system architecture for manufacturing information analysis with process model

We show in Fig. 10 the frequent sequent graph-enhanced cost. The Map view contains the following elements:

① *Canvas with process map*: The reserved main area for the process map visualization.

② *Zoom slider*: A zoom-in and zoom-out tool. Users can alternatively use the mouse wheel to make the process map larger and smaller.

③ *Measure*: This allows the user to choose the type of measurements for the visualization, including Activity and Resource cost with Total, Min, Max, or AVG view filtering for each cost type.

④ *Tittle:* Chart title displays.

Each node represents an event with its activity name, frequency, execution time, and cost. Arrows show links between two nodes and their frequency. The left panel displays the major measurements, i.e. “Activity” and “Resource” as well as the minor ones, i.e. Total, Max, Min, AVG, and Median. From this, users can choose an appropriate measurement for the visualization.

To illustrate, hereafter we limit the cost measurement to Activity on average. By checking the process model-enhanced cost of Jean's manufacturing process in Fig.10, users can obtain the information about the number of activities, its frequency, and the detailed cost of each activity included in that process. There are totally 8 activities A= {Cutting, Embroidery, Sewing, Washing, End line checking, Finishing, Checking finish garment, and Packing out}. The start event a_s = {Cutting}, end event a_e = {Packing out}.The graph shows the occurrence of nodes, e.g. $N_{Cutting}$= 11, $N_{Embroidery}$=3, and flow between nodes, e.g. $N_{Cutting,Sewing}$= 3, $N_{Cutting,Embroidery}$ = 8. There are reworks at "Sewing", "Finishing", and "Washing" stage. The manufacturing cost on average of each task are as follows: "Cutting" $45294.33, "Embroidery" $919.41, "Sewing" $1711.18, "End line checking" $78.48, "Finishing" $14476.41, "Washing" $350.10, "Checking finish garment" $668.08, "Packing out" $349.71. By comparison, users can determine which the cheapest or the most expensive activity is. That is, the cheapest one is "End line checking" activity with $78.48; the most expensive one is "Cutting" activity with $45294.33. The "Packing out" activity with $349.71 is cheaper than "Cutting" activity costs $45294.33, "Sewing" costs $1711.18, "Washing" $350.10, "Checking finish garment" $668.08, and "Embroidery" $919.41. Thus, equipped with this overview of the allocation of costs per each activity involved in the process. Based on these, process managers or boards of directors are able to prevent from risks, adjust unreasonable points, avoid waste, and optimize the process.

Moreover, this plug-in allows users to access the execution time of each activity involved in the process. For example, the execution time of "Cutting" activity is 17.4545 days, "Embroidery" activity is 4.3333 days, "Sewing" activity is 30.4545 days, e.g. "Sewing" activity has the longest execution time of 21 days, and "Embroidery" activity has the shortest execution time only 4.333 days.

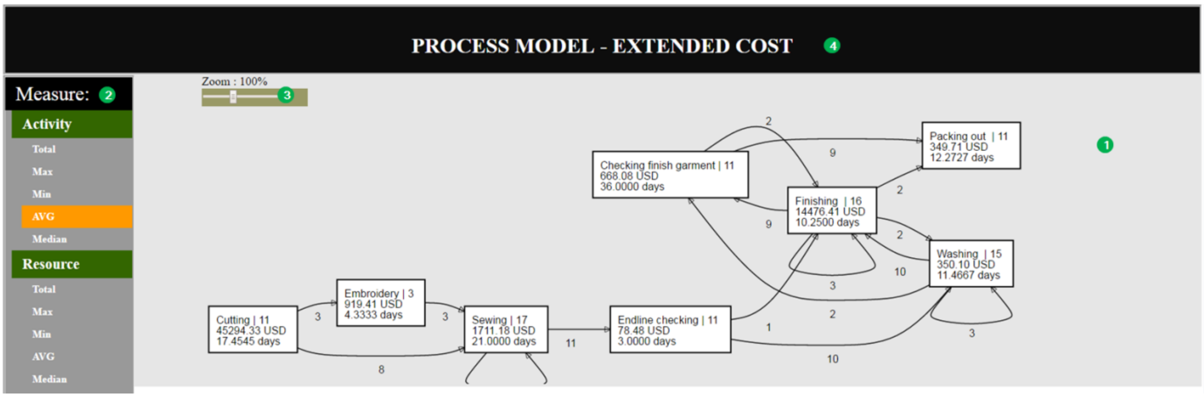


Figure 10. A screen shot of cost analysis with process model

5.2. Cost perspective

In order to support the construction of the cost database structute, information from Material requisition, allocation Labor, Machine and other OH costs is used. This information is stored in the MySQL server. Then, a cost drivers are created. Finally, this cost driver together with the event logging generates an enhanced cost log-event by some cost function which is used as the input for a later analysis process.

5.2.1. Visualization of cost breakdown

Fig. 11 visualizes an example of cost breakdown. The bar chart on the left indicates the total cost of the process distributed per activity in which each bar represents a particular activity. The pie chart and legend table on the right show the detailed cost components, i.e. direct labor cost (DL), direct material cost (DM), indirect material cost (IDM),

① *Chart area*: The main area displays the chart visualization.

② *Measure*: This allows the user to choose the type of measurement used for the visualization. For now, the measurement is only limited to the Activity cost. Other measurements such as Resouce cost are to be implemented in the future.

③ *Tittle:* Displays the title of the plug-in.

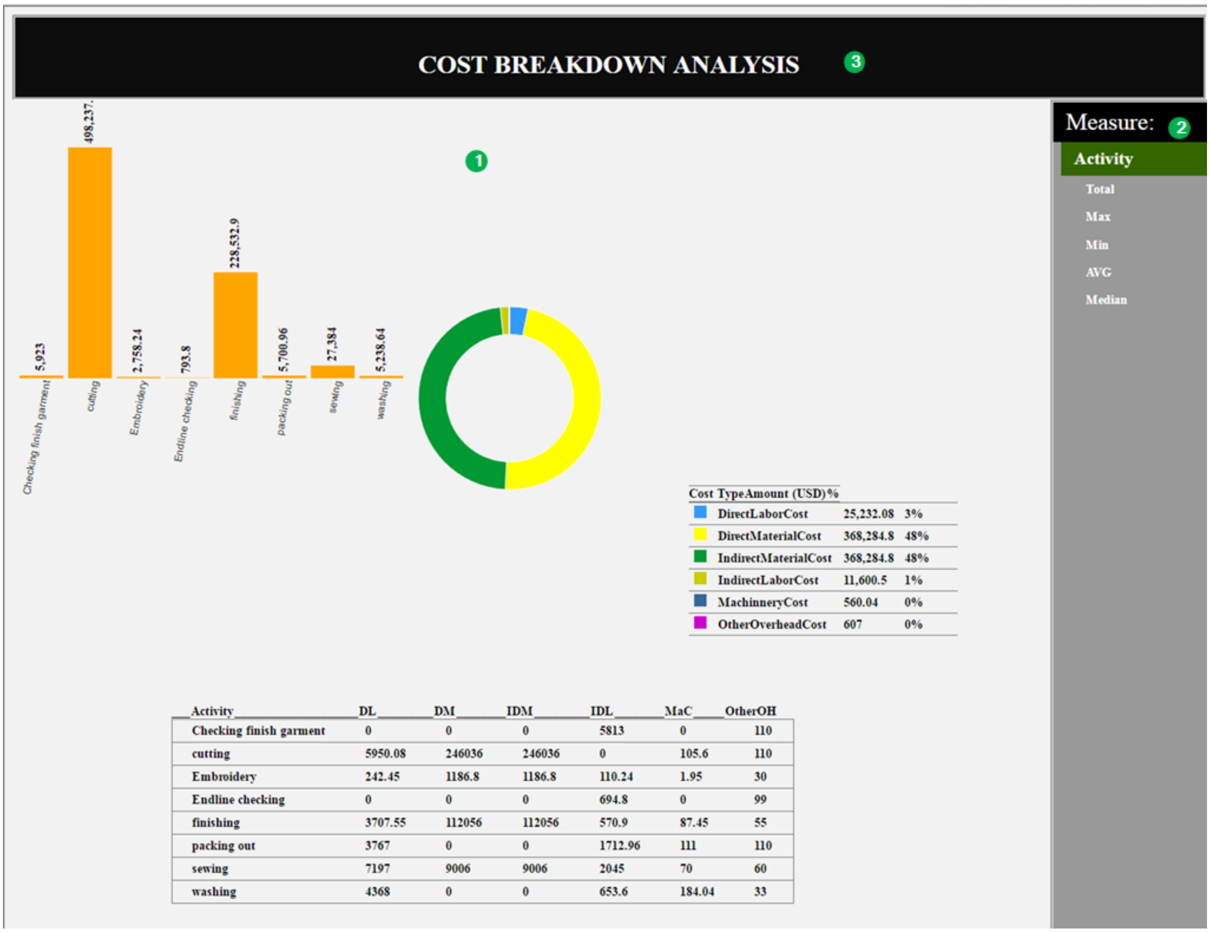
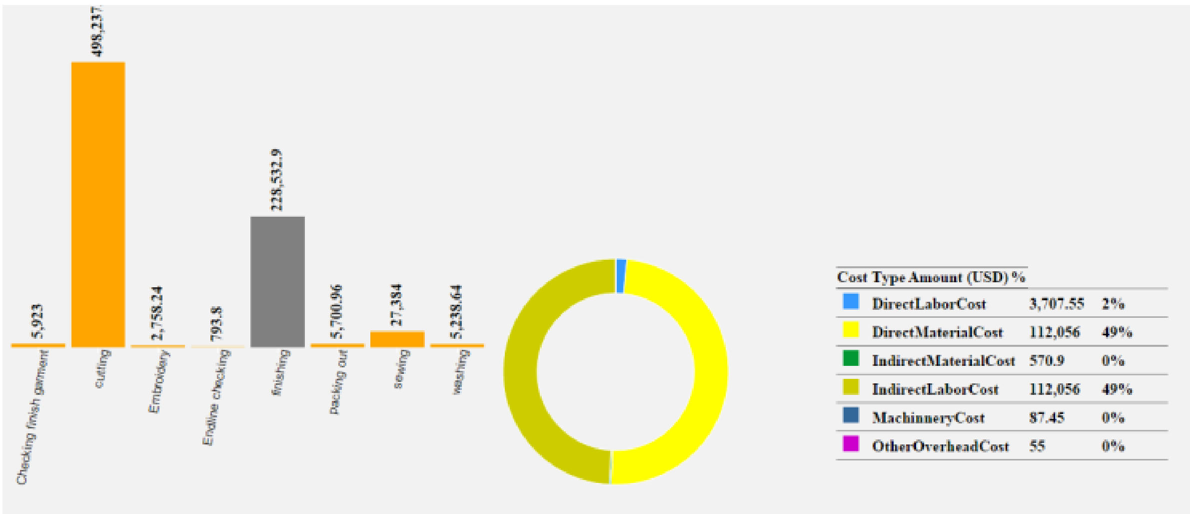


Figure 11. A screen shot of cost breakdown analysis



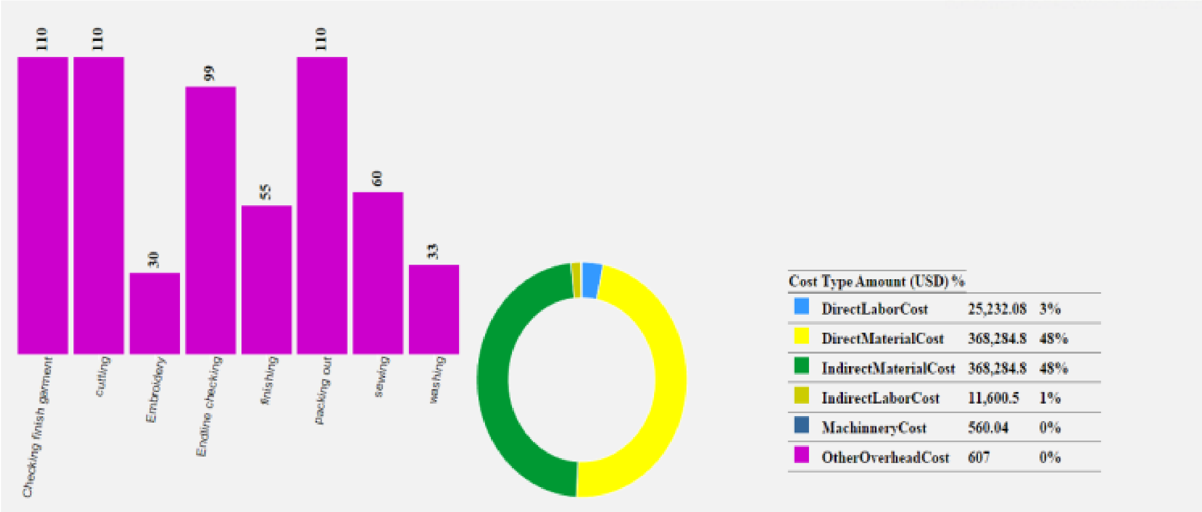


Figure 12. Selection of cost parameter in cost breakdown plug-in

machinery cost (MaC), other overhead cost (OtherOH) and the percentage of them attribute to the total cost. As we can see in Fig.11, the highest one are material cost; whereas machinery cost and other overhead cost are not significant. For instance, DM costs $368,284.8 account for 48%, IDM costs $368,284.8 makes up for 48%, DL costs $22,235.08 account for 3%, IDL costs $11,600.5 makes up for 1%, MaC costs $560.04 approximately constitutes 0%, and OtherOH costs 607 approximately constitutes 0%. Furthermore, the table in the bottom of Fig.11 shows a summary specific cost components which are distributed per activity. Those costs can be analyzed in terms of either cost type or activity. By comparing them, users can know the highest or the lowest cost of each activity or cost type. In terms of cost type, for example of IDM, “Cutting” activity is the highest one with a cost of $246,036 USD and greater than “Finishing” activity which costs $112,506. The “Embroidery” activity is cheaper with a cost of $1186.8, while the “Sewing” activity costs $9006. In term of activity, for example of “Checking finish garment” activity, IDL account for most of the cost value with $5813, OtherOH is lower with a cost of $110, and the lowest ones are DL, DM, IDM, MaC with a cost of $0.

By clicking each bar, users can see its detailed cost components visualized in the pie chart and legend table. The left side of Fig.12 shows that users has selected to “Finishing” activity which is in a gray color bar. The detailed cost components of this activity are as follows: DL costs $3,707.55 accounting for 2%, DM costs $112,056 accounting for 49%, IDM costs $570.9 approximately making up for 0%, IDL costs $112,056 making up for 49%, MaC with a cost of $87.45 roughly accounting for 0% and Other Overhead Cost with a cost of $55 nearly making up for 0%. In a similar way, users can quickly figure out the detailed cost of other activities.

On the contrary, by clicking at the pie chart, the results are visualized in the bar graph on the left side. In Fig. 12, users have selected “Other Overhead Cost” which is presented in the purple (on the right side). Overall, “Checking finish garment” activity, “Cutting” activity, and “Packing out” activity cost the same value of $110. The OtherOH of “ End line checking” activity of $60 is higher than “Sewing” activity $60, “Finishing” $55, “Washing” $33, “Embroidery” $30. By comparing them, the users can quickly determine that know the most expensive ones are “Checking finish garment”, “Cutting” and “ packing out”; the cheapest one is “Embroidery”.

Obviously, the cost information would be easier to be interpreted and more significant and accurate for decision makers if they are visualized using cost breakdown visualization plug-in.

5.2.2. Resource utilization cost

Fig.13 shows the visualization of resourcing of cost consumption.

① *Chart area*: Displays the chart visualization.

② *Measure:* This allows the user to choose the type of measurement used for the visualization. For now, the measurements are Labor cost and Machine cost.

③ *Chart type:*. This allows the user to choose the type of chart visualization. For example, it is possible to display the data using a Column, Bar, Pie, Line.

④ *Tittle:* Displays the title the chart visualization

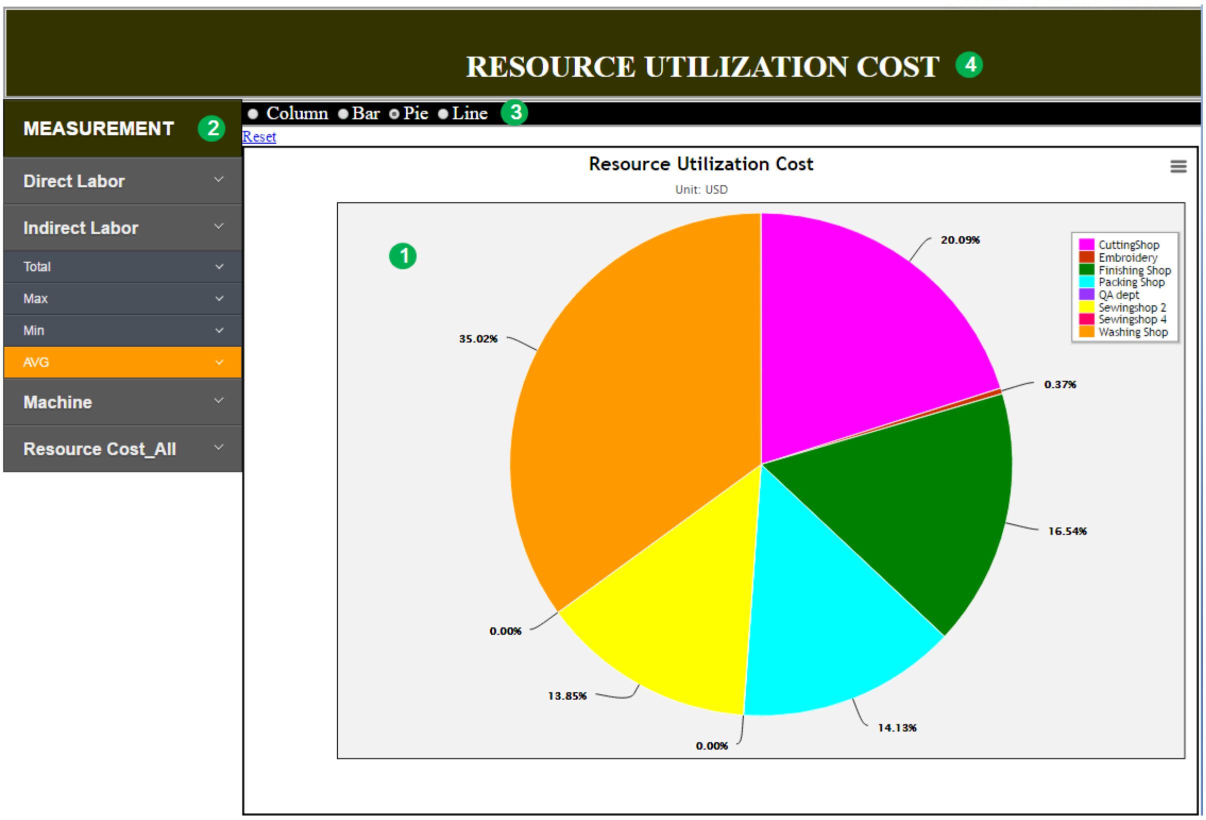


Figure 13. A screen shot of resource utilization cost plug-in

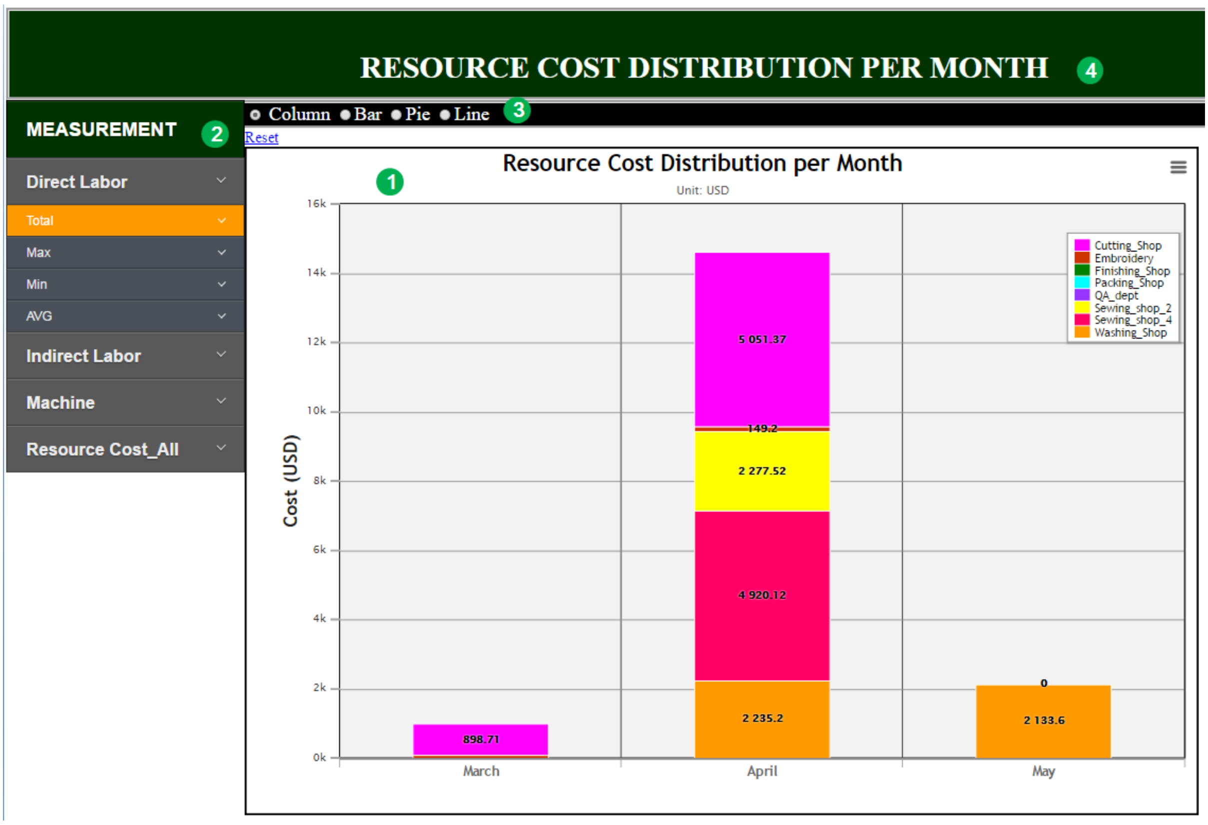


Figure 14. A screen shot of resource cost distribution per month plug-in

Here users have selected “Indirect Labor” cost type with AVG measurement and rendered with a pie chart. Interpreting these results, we can know the cost of resources on average are know as follows: “Washing shop” account for 35.02%, “Cutting Shop” account for 20.09%, “Packing Shop” makes up for 14.13%, “Finishing shop” produces 16.54%, “Sewing Shop 2” 13.85%, “Sewing Shop 4” 0.37%, “QA dept” and “Embroidery” approximately 0%. “Washing shop” is the most expensive resource. “QA dept” and “Embroidery” are the cheapest ones.

Fig.14 displays the resource cost distribution per month. Users have selected direct labor cost for total measurement. Examining Fig.14, we can figure out that direct material cost covers through three months March, Apr, and May. “Cutting shop” only locates on March with a cost of $898.71, Apr with a cost of $5051.37. Most of the resourcing of cost consumption locate on Apr. For instance, “Washing shop” costs $2235.2, “Sewing_shop_4” costs $4920.12, “Sewing_shop_2” costs $2,277.52, “ embroidery” costs $149.2. May only expenses for “Washing shop” with a cost of $2,133.6. Based on this acquired knowledge, users can efficiently adjust their process management.

5.2.3. Cost prediction

Process model-extendedcost, visualization of cost breakdown, resource utilization only analyze the data after the completion of an activity or a case. Cost prediction, on the other hand, looks for cost patterns and characteristics from event log-enhanced cost so that one can predict the possible cost consumption of current, as well as ongoing business processes based on historical data. In this work, we first predict the completion time through the stage of completion of work by using work progress of manufacturing process. Then, cost prediction is conducted. It is necessity to know exactly how much the consumed cost and remaining one are at the moment. Based on this, administrators are able to efficiently manage their manufacturing processes companies and avoid deficit budget or early precautions financial risks.

In Fig.15, we show a screenshot of cost prediction of the process. The cost prediction contains the following elements which are the same as those with process map-extended cost in Section 5.1:

① *Cost prediction area*: The main area is reserved to visualize the process map with accumulative cost and remaining cost

② *Measure:* Allows the user to choose the type of measurement used for the visualization. Users can select Activity or Resource cost. Furthermore, users also can view Total, Min, Max, or AVG of that cost.

③ *Zoom slider*: Gives an explicit control to make the process map larger and smaller. Alternatively, users can simply use mouse wheel to zoom in and out.

④ *Tittle:* Displays the tittle of the visualization.

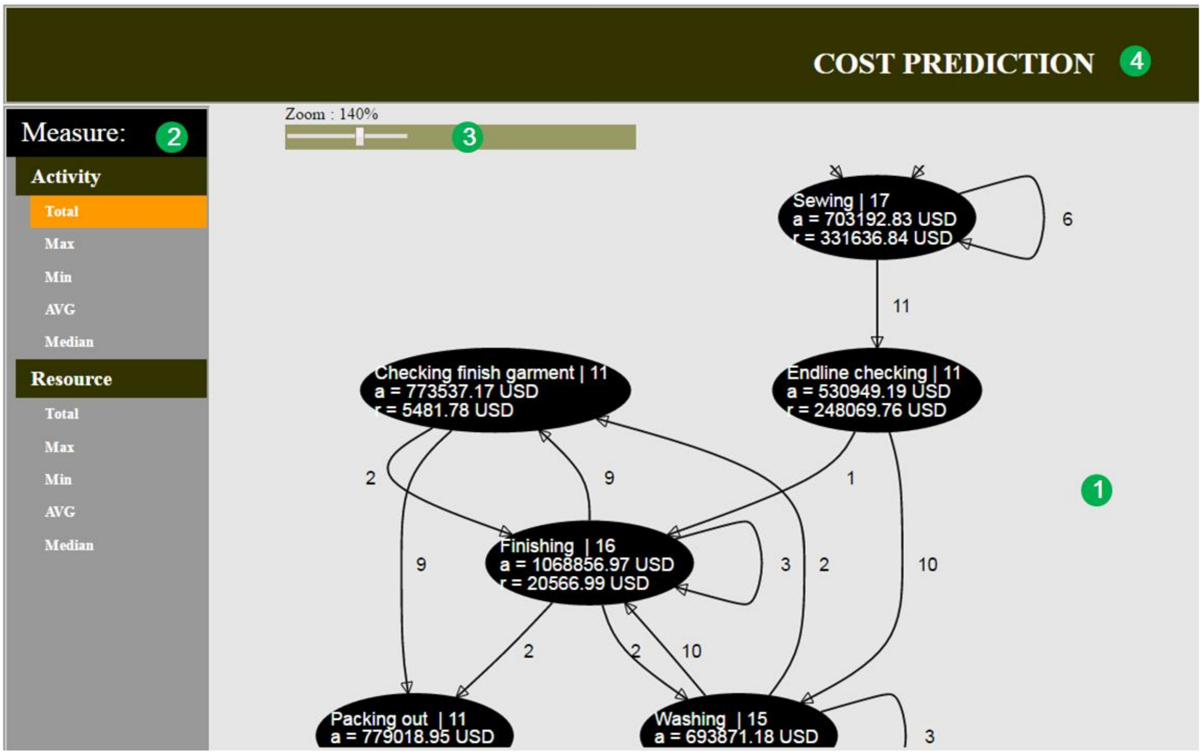


Figure 15. A Screenshot of Cost Prediction

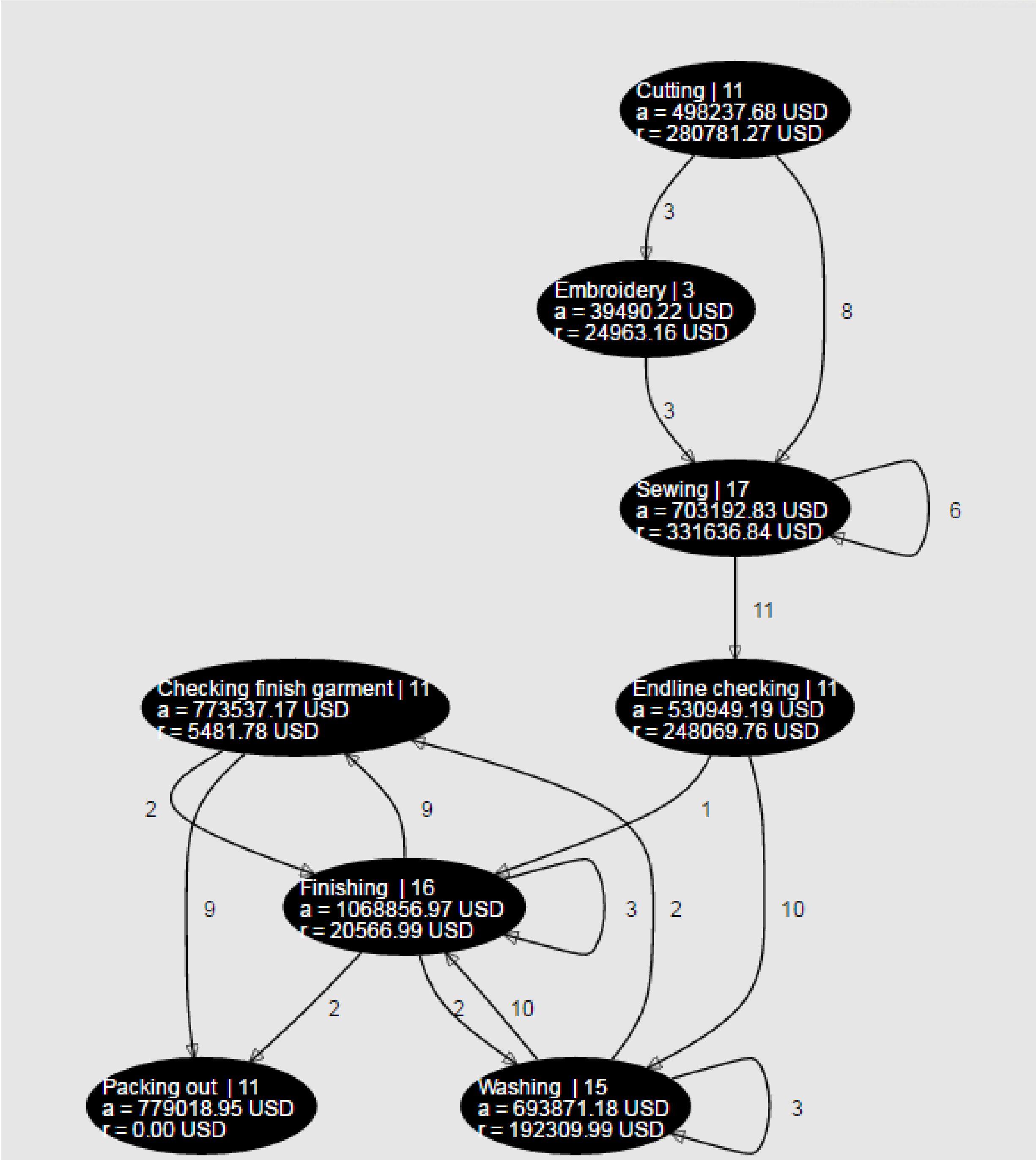


Figure 16. Cost Prediction

As shown in Fig. 16, an oval node presents an activity (each node here represents an activity). The accumulated cost is denoted by “a”, and the remaining cost is denoted by “r” of each activity or resource. The process includes eight tasks, i.e. Cutting, Embroidery, Sewing, etc. Cutting is the start task, and Packing out is the end task. By checking the graph, we can figure out the accumulated cost and the remaining cost of each activity comprised in the process. For instance, “Cutting” had consumed a= {498237.68 USD} and required r= {280781.27 USD} to complete the process;

"Embroidery" activity had consumed a= {39490.22 USD} and had required r= {24963.16 USD} to finish the process. Based on the result, we can know the highest remaining cost at the start event and zero remaining cost at the end event. Users also can use measurement such as total, max, min, AVG, median to customize their demands.

5.3. Quality report

The quality extended event log with data from the running example is used to generate a quality report with quality aggregated per case and quality-related KPIs, which is shown in Figure 17 (The report contains 11 cases and is large to be included as a whole.) Cases are identified by a Case ID in the cost report. Each row shows cost associated with a single case; every column illustrates the index for a quality-KPIs.

Quality per trace and Quality-related KPIs

10 25

External filter for "Numbers" column :

CaseID	ActivityName	Defect rate	First pass yield
1	Cutting	0.00	100.00
1	Sewing	0.00	100.00
1	Endline checking	0.00	100.00
1	Washing	0.00	100.00
1	Finishing	0.00	100.00
1	Checking finish garment	0.00	100.00
1	Packing out	0.00	100.00
2	Cutting	0.00	100.00
2	Embroidery	0.00	100.00
2	Sewing	0.00	100.00
2	Endline checking	0.00	100.00
2	Washing	0.00	100.00
2	Finishing	0.00	100.00
2	Checking finish garment	0.00	100.00
2	Packing out	0.00	100.00
3	Cutting	0.00	100.00
3	Sewing	16.67	83.33
3	Sewing	40.00	60.00
3	Sewing	0.00	100.00
3	Endline checking	0.00	100.00

Figure 17. Cost per Case

Chapter 6

Conclusions & Future Works

6.1. Summary of contributions

For manufacturing enterprises, it is vital that detailed and reliable insights be otained into their manufacturing processes regarding cost and quality. This thesis developed the approach to analyze performance in manufacturing processes. The propsed approach focused on cost and quality perspective. The main contributions of this thesis are: (i) a proposed method to extend event log of manufacturing process with manufacturing information, i.e. cost, quality; (ii) a detailed analysis on manufacturing information, i.e. cost, quality with process model; (iii) utilization of various existing process mining techniques and development of new approaches to analyze and predict manufacturing cost; (iv) and quality report generation in manufacturing proces.

The first contribution was well described in Section 4.3. With the respect to the event log extension with cost, a cost model and the creation of cost database model for cost model extraction was presented. Guidance for creating a quality model and computing quality-related KPIs index was introduced in the response to event log extension with quality as well.

For the second contribution, the process model-extended manufacturing cost and process quality was introduced in Section 4.4.1.

For the third contribution, various visualization methods were presented in Section 4.4.1 and Section 4.4.2. They are cost analysis with the process model, visualization of cost breakdown, and resource utilization cost. Cost analysis with a process model is a systematic method that allows users to access the detailed cost of a particular activity or individual resource involved in a manufacturing process. Visualization of a cost breakdown identifies the distinct cost types including material cost, labor cost and overhead cost comprised of the total cost of a task during process execution. Resource utilization cost deals with all information related to the cost of resources. Furthermore, this paper also developed a cost prediction method which relies on time prediction and production volume using working progress of manufacturing processes.

Finally, this study enabled quality analysis in process mining. This goal is accomplished in Section 4.4.3 by the generation of accurate, relevant quality report.

6.2. Future works

Despite various contributions, the current approach and implementation also suffer from some limitations that lead us to possible directions for future research. Firstly, the current costing techniques i.e. cost analysis with the process model, cost breakdown analysis, resource utilization cost and cost prediction consider only processing time. They neglect other time-related KPIs such as waiting time, throughput time, synchronization time.

Secondly, there exists a problem which we have not addressed before. In this thesis we limit the scope to discuss the area of quality reporting within a workflow management system environment only. That is, only the data which are relevant for quality reporting are consideredsuch event log-extended quality. Although static quality reporting has been implemented, dynamic quality reporting through cost functions is desirable and left as future work.

Thirdly, in Chapter 5 we have implemented the proposed method on an artificial event log introduced in Section 4.2. They demonstrated the applicability of our method, but additional experimental evaluation could provide additional insights. An important point to note here is that we have only evaluated the quality of the results in this thesis by looking athow well they are generated from proposed plug-in. However, theseplug-ins are flawed, and the resulting process models may not correspond to what a process expert would consider a high-quality analysis. Therefore, to properly evaluate the practical usability of the proposed methods, it would also be necessary to perform case studies on real-life event logs where the resulting cost and quality analysis are evaluated by process experts.

Besides the aforementioned limitations, this research has not look into the field of cost simulation, hence, simulating how changes in business decisions will affect the operational cost is not withinthe scope of this research.The visualization proposed in this thesis is based on the simple, traditional 2D visualization. Undoubtedly, more advanced visualization techniques can be found, with the advantage of being more representative for analysis and more user-friendly.

REFERENCE

[1] T. Altiok, *Performance analysis of manufacturing systems*, 1st ed.: Springer-Verlag New York, 1997.

[2] R. R. Fullerton, F. A. Kennedy, and S. K. Widener, "Management accounting and control practices in a lean manufacturing environment," *Accounting, Organizations and Society,* vol. 38, pp. 50-71, 2013.

[3] M. Hammer and J. Champy, *Reengineering the Corporation: Manifesto for Business Revolution, A*: Zondervan, 2009.

[4] J. Mendling, M. zur Muehlen, and A. Price, "Standards for workflow definition and execution," *Process-Aware Information Systems,* p. 281, 2005.

[5] D. Sinclair and M. Zairi, "Benchmarking best-practice performance measurement within companies: using total quality management," *Benchmarking for Quality Management & Technology,* vol. 2, pp. 53-71, 1995.

[6] D. Sinclair and M. Zairi, "Performance measurement as an obstacle to TQM," *The TQM Magazine,* vol. 7, pp. 42-45, 1995.

[7] G. Chryssolouris, *Manufacturing systems: theory and practice*, 2nd ed.: Springer New York, 2006.

[8] A. D'Angelo, M. Gastaldi, and N. Levialdi, "Performance analysis of a flexible manufacturing system: A statistical approach," *International journal of production economics,* vol. 56, pp. 47-59, 1998.

[9] K. Kerdprasop and N. Kerdprasop, "Performance Analysis of Complex Manufacturing Process with Sequence Data Mining Technique," *International Journal of Control and Automation,* vol. 6, pp. 301-312, 2013.

[10] J. Vetter, "Performance analysis of distributed applications using automatic classification of communication inefficiencies," in *Proceedings of the 14th international conference on Supercomputing*, 2000, pp. 245-254.

[11] M. Zhou, "Modeling, analysis, simulation, scheduling, and control of semiconductor manufacturing systems: A Petri net approach," *Semiconductor Manufacturing, IEEE Transactions on,* vol. 11, pp. 333-357, 1998.

[12] B. Klefsjö, H. Wiklund, and R. L. Edgeman, "Six Sigma seen as a methodology for total quality management," *Measuring Business Excellence,* vol. 5, pp. 31-35, 2001.

[13] T. C. Powell, "Total quality management as competitive advantage: a review and empirical study," *Strategic management journal,* vol. 16, pp. 15-37, 1995.

[14] T. H. Davenport and J. E. Short, "The new industrial engineering: information technology and business process redesign," 1990.

[15] M. Hammer, "Reengineering work: don't automate, obliterate," *Harvard business review,* vol. 68, pp. 104-112, 1990.

[16] H. Harrington, E. Esseling, and H. Van Nimwegen, *Business Process Improvement Workbook: Documentation, Analysis, Design and Management of Business Improvement*: New York: McGraw-Hill, 1997.

[17] S. Y. Son, "A Process Mining Based Approach to Complex Manufacturing Process Flow Analysis: A Case Study," Master, Management Engineering, Ulsan National Institute of Science and Technology (UNIST), 2014.

[18] W. M. P. van der Aalst, T. Weijters, and L. Maruster, "Workflow mining: Discovering process models from event logs," *Knowledge and Data Engineering, IEEE Transactions on,* vol. 16, pp. 1128-1142, 2004.

[19] W. M. P. van der Aalst, "Discovery, Conformance and Enhancement of Business Processes," 1st ed: Springer-Verlag Berlin Heidelberg, 2011.

[20] W. M. P. van der Aalst, M. H. Schonenberg, and M. Song, "Time prediction based on process mining," *Information Systems,* vol. 36, pp. 450-475, 2011.

[21] W. M. P. van der Aalst and A. J. M. M. Weijters, "Process mining: a research agenda, Computers in Industry," *vol,* vol. 53, pp. 231-244, 2004.

[22] W. M. P. van der Aalst, H. A. Reijers, A. J. M. M. Weijters, B. F. van Dongen, A. K. A. De Medeiros, M. Song, and H. M. W. Verbeek, "Business process mining: An industrial application," *Information Systems,* vol. 32, pp. 713-732, 2007.

[23] S. Son, B. Yahya, M. Song, S. Choi, J. Hyeon, B. Lee, Y. Jang, and N. Sung, "Process mining for manufacturing process analysis: a case study," in *Proceeding of 2nd Asia Pacific Conference on Business Process Management, Brisbane, Australia*, 2014.

[24] A. Rozinat, I. S. M. de Jong, C. W. Gunther, and W. M. P. van der Aalst, "Process mining applied to the test process of wafer scanners in ASML," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on,* vol. 39, pp. 474-479, 2009.

[25] J. Park, D. Lee, and J. Zhu, "An integrated approach for ship block manufacturing process performance evaluation: Case from a Korean shipbuilding company," *International journal of production economics,* vol. 156, pp. 214-222, 2014.

[26] R. Anupindi, S. D. Deshmukh, J. A. Van Mieghem, and E. Zemel, *Managing Business Process Flows*, 3rd ed.: Pearson, 2011.

[27] P. J. Lederer and S. K. Rhee, "Economics of total quality management," *Journal of Operations Management,* vol. 12, pp. 353-367, 1995.

[28] A. Livingston, G. Jackson, and K. Priestley, "Root causes analysis: Literature review," *HSE Contract Research Report,* 2001.

[29] W. Nauta, "Towards Cost-Awareness in Process Mining," Citeseer, 2011.

[30] M. T. Wynn, J. De Weerdt, A. H. ter Hofstede, W. M. P. van der Aalst, H. A. Reijers, M. J. Adams, C. Ouyang, M. Rosemann, and W. Z. Low, "Cost-aware business process management: a research agenda," 2013.

[31] M. T. Wynn, W. Z. Low, A. H. ter Hofstede, and W. Nauta, "A Framework for Cost-Aware Process Management: Cost Reporting and Cost Prediction," *J. UCS*, vol. 20, pp. 406-430, 2014.

[32] M. T. Wynn, H. A. Reijers, M. Adams, C. Ouyang, A. H. ter Hofstede, W. M. van der Aalst, M. Rosemann, and Z. Hoque, "Cost-informed operational process support," in *Conceptual Modeling*, ed: Springer, 2013, pp. 174-181.

[33] W. M. P. van der Aalst, B. F. van Dongen, J. Herbst, L. Maruster, G. Schimm, and A. J. M. M. Weijters, "Workflow mining: a survey of issues and approaches," *Data & knowledge engineering*, vol. 47, pp. 237-267, 2003.

[34] M. Song and W. M. P. van der Aalst, "Towards comprehensive support for organizational mining," *Decision Support Systems*, vol. 46, pp. 300-317, 2008.

[35] W. M. P. van der Aalst, A. Adriansyah, and B. F. van Dongen, "Replaying history on process models for conformance checking and performance analysis," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 2, pp. 182-192, 2012.

[36] W. M. P. van der Aalst, *Process mining: discovery, conformance and enhancement of business processes*: Springer Science & Business Media, 2011.

[37] J. Scott, *Social network analysis*: Sage, 2012.

[38] S. Wasserman and K. Faust, *Social network analysis: Methods and applications* vol. 8: Cambridge university press, 1994.

[39] M. Park, M. Song, T. H. Baek, S. Son, S. J. Ha, and S. W. Cho, "Workload and Delay Analysis in Manufacturing Process Using Process Mining," in *Asia Pacific Business Process Management*, ed: Springer, 2015, pp. 138-151.

[40] M. Adams, M. T. Wynn, C. Ouyang, and A. H. M. ter Hofstede, "Realisation of cost-informed process support within the YAWL workflow environment," in *Asia Pacific Business Process Management*, ed: Springer, 2015, pp. 3-18.

[41] S. Douwe P. Flapper, L. Fortuin, and P. P. Stoop, "Towards consistent performance management systems," *International journal of operations & production management*, vol. 16, pp. 27-37, 1996.

[42] M. J. Lebas, "Performance measurement and performance management," *International journal of production economics*, vol. 41, pp. 23-35, 1995.

[43] M. Dumas, M. La Rosa, J. Mendling, and H. A. Reijers, *Fundamentals of business process*

management, 1st ed.: Springer Berlin Heidelberg, 2013.

[44] A. De Toni and S. Tonchia, "Performance measurement systems-models, characteristics and measures," *International journal of operations & production management,* vol. 21, pp. 46-71, 2001.

[45] M. Hudson, A. Smart, and M. Bourne, "Theory and practice in SME performance measurement systems," *International journal of operations & production management,* vol. 21, pp. 1096-1115, 2001.

[46] D. Medori and D. Steeple, "A framework for auditing and enhancing performance measurement systems," *International journal of operations & production management,* vol. 20, pp. 520-533, 2000.

[47] K. G. Mola and H. Parsaei, "Dimensions and measures of manufacturing performance measurement," in *Computers and Industrial Engineering (CIE), 2010 40th International Conference on*, 2010, pp. 1-6.

[48] C. L. Yang, S. P. Chuang, and R. H. Huang, "Manufacturing evaluation system based on AHP/ANP approach for wafer fabricating industry," *Expert Systems with Applications,* vol. 36, pp. 11369-11377, 2009.

[49] M. Yurdakul, "Measuring a manufacturing system's performance using Saaty's system with feedback approach," *Integrated Manufacturing Systems,* vol. 13, pp. 25-34, 2002.

[50] N. Dhafr, M. Ahmad, B. Burgess, and S. Canagassababady, "Improvement of quality performance in manufacturing organizations by minimization of production defects," *Robotics and Computer-Integrated Manufacturing,* vol. 22, pp. 536-542, 2006.

[51] A. M. Ghalayini, J. S. Noble, and T. J. Crowe, "An integrated dynamic performance measurement system for improving manufacturing competitiveness," *International journal of production economics,* vol. 48, pp. 207-225, 1997.

[52] E. Amrina and S. M. Yusof, "Key performance indicators for sustainable manufacturing evaluation in automotive companies," in *Industrial Engineering and Engineering Management (IEEM), 2011 IEEE International Conference on*, 2011, pp. 1093-1097.

[53] R. R. Mohan, K. Thiruppathi, R. Venkatraman, and S. Raghuraman, "Quality Improvement through First Pass Yield using Statistical Process Control Approach," *Journal of Applied Sciences,* vol. 12, p. 985, 2012.

[54] T. L. Albright, R. W. Ingram, and J. W. Hill, *Managerial accounting: Information for decisions*, 4th ed.: Cengage Learning, 2005.

[55] T. Altiok, *Performance analysis of manufacturing systems*, 1st ed.: Springer-Verlag New York, 1997.

[56] R. S. Kaplan and A. A. Atkinson, *Advanced management accounting*, 3rd ed.: Prentice Hall

New Jersey, 1998.

[57] P. Brewer, R. Garrison, and E. Noreen, *Introduction to Managerial Accounting*, 7th ed.: McGraw-Hill Education, 2015.

[58] R. S. Kaplan, and Steven R. Anderson. (November 2004) Time-Driven Activity-Based Costing. *Harvard Business Review 82, no. 11*. 131-138.

[59] E. Noreen, "Conditions under which activity-based cost systems provide relevant costs," *Journal of Management Accounting Research*, vol. 3, pp. 159-168, 1991.

[60] M. Tse and M. Gong, "Recognition of idle resources in time-driven activity-based costing and resource consumption accounting models," *Journal of applied management accounting research*, vol. 7, pp. 41-54, 2009.

[61] C. T. Horngren, J. W. T. Harrison, and M. S. Oliver, *Accounting*, 11th ed.: Pearson Education Limited, 2012.

[62] J. W. T. H. C.T. Horngren, M.S. Oliver *Accounting*, 11th ed.: Pearson Education Limited, 2012.

[63] U. S. Karmarkar and R. C. Pitbladdo, "Quality, class, and competition," *Management Science*, vol. 43, pp. 27-39, 1997.

[64] U. S. Kannarkar, "Quality Management," *Working Paper, Center for Manufacturing and Operations Management, William E. Simon Graduate School of Business Administration, University of Rochester, Rochester, New York 14627*, 1991.

[65] E. R. Ziegel, "Quality Management: Tools and Methods for Improvement," *Technometrics*, vol. 37, pp. 468-469, 1995.

[66] N. Gupta, "Analysis on the Defects in Yarn Manufacturing Process & its Prevention in Textile Industry," *Analysis*, vol. 2, pp. 45-67, 2013.

[67] Jeffrey A.Hoffer, V.Ramesh, and H. Topi, *Modern Database Management*, 11th ed.: Prentice Hall, 2012.

[68] R. J. C. Bose and W. M. P. van der Aalst, "Abstractions in process mining: A taxonomy of patterns," in *Business Process Management*, ed: Springer Heidelberg, 2009, pp. 159-175.

[69] G. T. Lakshmanan, S. Rozsnyai, and F. Wang, "Investigating clinical care pathways correlated with outcomes," in *Business process management*, ed: Springer Heidelberg, 2013, pp. 323-338.

ACKNOWLEDGEMENTS

Time flies so fast when I think of the first day at UNIST in August of 2014 like just a few days ago, but still long enough. It has been two years since I started my study at UNIST. It is definitely a meaningful two years of my life. Looking back whole time I have spent at UNIST, I would like to express my sincere thanks to whom warming my heart during cold winters at UNIST, cheering me up in my harsh time in research and supporting me pursuing Master degree.

First of all, I would like deeply to express my gratitude to my beloved advisor, Professor Minseok Song (송민석교수님께감사합니다). Thank you for your advice since the day I knew nothing about process mining to this moment when I have just completed my thesis about process mining-driven performance analysis in the manufacturing process. You are not only my advisor but also my mentor who introduced me to the interesting research topics of performance analysis, cost and quality mining in manufacturing process and instructed me how to struggle and overcome difficulties occurring during the research process. I greatly appreciate your kind support and careness about my research and my life. Thanks for your understating and tolerance of mistakes I have ever made. I have learned and benefited greatly from your insightful knowledge, working ethic and scientific methods in conducting research. I would not have been able to achieve this far without your constant guidance, endless patience, and encouragement.

I would like to show my sincere appreciation to my first advisor in Korea-Professor Wooje Cho for his kindness and encouragement. I would like to thank all UNIST instructors giving lectures to enrich my knowledge. I am especially grateful to Professor Changyong Lee, Professor Marco Comuzzi fortheir valuable advice, contructive suggestions during the preparation of this thesis.

Hereby I heartily thanks my labmates. The first labmate is Minsu Cho. He is a senior helping me a lot at the beginning days in the laboratory at UNIST. I would also like to thank other members of AIM Lab–Minjeong Park, Hojeong Yi, DohyeonKim ,Yonghyeok Lee - for all their help in research and for together time we had during lunches, dinner, coffee break. Thanks to closed friends in my homeland, Kim Loan and Mai Lan, for always being there for me. Thanks to VN-UNIST friends, especially to Dr. Cao Truong Hai Nguyen, Dr. ThienBinh Nguyen, Dr. Duc Tam Ho, Van Ty Nguyen, Mau Dong Nguyen, Minh Quan Tran for their friendship and encouragement during my Master program.

Thanks to my future husband because you have not shown up yet, so I smoothly gained bachelor degree, successfully defended my master’s thesis, and will perhaps pursue Ph.D. program.

Special thanks to my aunt, as well as my idol, Dr. Thi Phuong Nga Nguyen for her love and kindness. She helped me a lot during my research, also always beside to encourage me when I was in harsh time. She is the one I can always ask for help without explanation.

Last but not least, I express my heartfelt gratefuless to my parents for their endless love, support and encouragement. Thanks to my beloved brothers, my grandparents, my aunts, as well as my uncles for their constant love, care, and advice give me strength to keep on moving forward. You all make me never feel lonely and strong always so that I completed my research abroad. This thesis is dedicated to my parents.

--☺☺--

LỜI CẢM ƠN

(Acknowledgements in Vietnamese)

Không khí oi bức, nắng nóng rát của những ngày đầu hạ ở miền đồi núi Ulsan làm cho con người ta lưu luyến làn gió xuân tươi mát, nhung nhớ những ngày thu lá vàng mơ mộng và thèm khát cái lạnh thấu xương của những ngày đông băng giá. Tae (Hồng) đang viết những lời này trong tâm trạng đầy xúc động khi bồi hồi nhớ lại ngày đầu tiên đặt chân đến học tập và nghiên cứu cho chương trình thạc sĩ tại viện khoa học và công nghệ quốc gia Ulsan (UNIST). Đó là những ngày thu cuối tháng tám, năm 2014. Thấm thoắt gần hai năm đã trôi qua! Vẫn câu nói cũ Tae hay dùng, thời gian trôi nhanh như ngựa điên chạy, như chó dí ngoài đường. Chặng đường đó không quá dài, nhưng cũng không quá ngắn đủ để con người ta làm nên một vài điều nho nhỏ, góp nhặt lại để làm nền tảng lớn cho quá trình nghiên cứu trong tương lai. Lời cảm ơn này, Tae xin được viết riêng cho những người thân và những người bạn thân thiết của Tae.

Cuốn luận văn thạc sĩ đang được hoàn thành ở bước cuối cùng. Từ nơi xa xôi đất khách quê người, bằng cả trái tim, con xin cảm ơn Ba Mẹ người đã luôn yêu thương, ủng hộ con vô điều kiện. Con yêu Ba Mẹ nhất trên đời! Cảm ơn hai em trai “soái ca” và “soái nhi” đã luôn bên cạnh và cổ vũ Tae. Cảm ơn ông bà, các dì, dượng, cô, chú và thím đã luôn theo dõi bước đường con đi, luôn động viên con.

Lời cảm ơn tiếp theo xin gửi đến những người bạn của Tae nơi quê nhà cho những lần thăm hỏi. Đặc biệt gửi lời cảm ơn tới cô bạn, Kim Loan cho những đêm thức trắng cùng nhau, cho những ngày tháng chường mặt chường tai ra chịu đựng những cung bậc cảm xúc thất thường của Tae. Cảm ơn chị bạn Mai Lan cho những ngày tay trong tay du hí, cho những lần chia sẻ nỗi niềm, và cho những lời khuyên bổ ích.

Cảm ơn gia đình mờ nờ (matnet family), nơi những yêu thương và chia sẻ đã gắn kết những con người không chung dòng máu thành một gia đình. Nơi đó Tae có Bố Hải, Má Nam, Chị 3 Ánh Đông, bà Thím Ty, Mợ Quân và ông Chú Bình. Cảm ơn Bố Má cho những lần quan tâm, thăm hỏi. Cảm ơn Chị 3 và bà Thím Ty cho những lời sỉ nhục cố ý để Tae biết sống là phải mạnh mẽ đạp lên dư luận. Cảm ơn Mợ cho những câu chuyện cười nửa vời ngây ngất. Cảm ơn mama tổng quản của con, ông Chú Bình, chú đã luôn lắng nghe, giúp đỡ và trả lời cho những câu hỏi ngu ngơ, khó đỡ của con. Cảm ơn những chuyến đi lầy lội cùng nhau để ta thấy đời rất đẹp và ta đang còn trẻ lắm. Cảm ơn những người bạn, người chị, người anh, người em tại UNIST vì những sẻ chia cùng nhau trong những ngày tháng xa nhà.

Lời cảm ơn đặc biệt dành cho bà dì già Nguyễn Thị Phương Ngà, vừa là dì, vừa là bạn thân vì sự yêu thương, chăm sóc, lắng nghe, và chia sẻ. Cảm ơn ông dượng già háp Hồ Đức Tam vì những quan tâm cho hai dì cháu, những đêm ca hát tưng bừng để ta thấy đời còn nồng nàn lắm. Cảm ơn dượng vì đã yêu và luôn ở bên cạnh người con yêu. Cảm ơn dì dượng Tám!

Lời cuối cùng, Tae gửi đến cho chồng tương lai. Cảm ơn anh, chồng tương lai của em vì anh không đến sớm nên em đã tốt nghiệp cấp ba, đại học suôn sẻ và giờ đây em đã tốt nghiệp thạc sĩ. Tương lai nếu anh không đến có khi em sẽ học lên tiến sĩ, rồi sẽ thành giáo sư. Tae xin lỗi nếu đã phớt lờ anh trong suốt hai mươi lăm năm qua cuộc đời của Tae.

--☺☺--