

# A hybrid cost estimation framework based on feature-oriented data mining approach <sup>☆</sup>



Narges Sajadfar <sup>a</sup>, Yongsheng Ma <sup>b,\*</sup>

<sup>a</sup> University of Alberta, Edmonton, Canada

<sup>b</sup> 4-9 Department of Mechanical Engineering, University of Alberta, Edmonton, Alberta T6G 2G8, Canada

## ARTICLE INFO

### Article history:

Received 9 May 2014

Received in revised form 2 April 2015

Accepted 4 June 2015

Available online 2 July 2015

### Keywords:

Cost estimation

Feature modeling

Data mining

ERP

Welding feature

## ABSTRACT

This paper presents an informatics framework to apply feature-based engineering concept for cost estimation supported with data mining algorithms. The purpose of this research work is to provide a practical procedure for more accurate cost estimation by using the commonly available manufacturing process data associated with ERP systems. The proposed method combines linear regression and data-mining techniques, leverages the unique strengths of the both, and creates a mechanism to discover cost features. The final estimation function takes the user's confidence level over each member technique into consideration such that the application of the method can phase in gradually in reality by building up the data mining capability. A case study demonstrates the proposed framework and compares the results from empirical cost prediction and data mining. The case study results indicate that the combined method is flexible and promising for determining the costs of the example welding features. With the result comparison between the empirical prediction and five different data mining algorithms, the ANN algorithm shows to be the most accurate for welding operations.

© 2015 Elsevier Ltd. All rights reserved.

## 1. Introduction

Due to increasing manufacturing product complexity, variations, as well as dynamic supply chain management, manufacturing companies need to coordinate their order acceptance process dynamically; hence, they must estimate costs based on timely information resources among their business partners. The customer inquiry management has to be done accurately and quickly to provide effective feedback.

Currently, different computer-aided systems are used for collaborative manufacture and aimed to increase efficiency and flexibility of information flows while also keeping sensitive collaboration loops with the different business partners. Enterprise Resource Planning (ERP) is the most important information management technology that supports order-fulfillment and is capable of creating the links between product documents, schedules, and other forms of communication. Different ERP software tools are used to share information among different partners for daily manufacturing activities, through which many of the process functions are available to support collaborative manufacturing, such as

customer order product configuration, process planning, financial decision making, accounting, and supply chain coordination; ideally, all of which are unified and managed from a single system [1]. However, ERP packages do not have the sufficient functions for market-oriented cost estimation processing with dynamic production scenarios, even though such functions play a main role in business success [2]. Cost estimation (CE) is one of the key factors that create integration between internal production management system and market-oriented dynamic informatics functions.

Manufacturing costs, in many companies, are not accounted for systematically and accurately due to the complexity and the constant changes of the processes involved in production and the lack of data collection schemes. The process model in an ERP package contains certain details of manufacturing processes, such as the types and setups of machining features, cutting conditions, expected productivity, past costs, and status of job completion. Most ERP systems can estimate process costs based on the required time for the machining process (i.e., labor cost), materials, and power consumption. However, preparing the complete information for a specific manufacturing process is difficult. Developing a generic and comprehensive data model for manufacturing processes in ERP has been a major issue that is being studied to achieve the expected production performance [2]. One limitation of most ERP systems is that the actual variations of processes, such

<sup>☆</sup> Handled by C.-H. Chen.

\* Corresponding author.

E-mail addresses: [sajadfar@ualberta.ca](mailto:sajadfar@ualberta.ca) (N. Sajadfar), [yongsheng.ma@ualberta.ca](mailto:yongsheng.ma@ualberta.ca) (Y. Ma).

as the different welding jobs, are not well defined and documented. Such technical challenges are particularly true in companies with small batches and high variations of product models, which cause overwhelming difficulties in accurate cost estimation. Even with a complete set of process definitions, using the historical data from similar operations requires an expert judgment to determine the factors that correlate to the respective costs.

Traditionally, the cost of a product is calculated based on the cost of product development, time spent on manufacturing processes and equipment used, and the energy and materials consumed. In turn, these factors are influenced by the parts' geometrical relations, accuracy of dimensions, as well as material and surface characteristics. On the other hand, the modern cost paradigm focus on customer inputs, product similarities among configurations, and flexible markups [3]. There has been a considerable amount of research available on cost estimation for manufactured goods, specifically about the manufacturing processes, such as turning and milling cost estimation. However, many process costs have not been consistently defined yet, such as the cost of welding operations in variation workshops. Empirical methods are commonly used for manufacturing cost estimation. Such methods are difficult to adapt to different jobs and estimation can only be given by experienced personnel [2].

## 2. ERP solutions

ERP systems have certain financial functions for cost management. A typical ERP system records and tracks product and process related data, such as raw material, set-up time, run time, number of machines and labor time for cost calculation. A disadvantage of this data tracking method is that many detailed features and parameters are not fully considered in cost calculations; therefore, the accuracy of the cost results is questionable. In addition, many cost items can only be calculated when all of the parameters are in place. This constraint requires a lot of time to obtain the necessary data and even makes timely cost estimation too expensive to be realistically implemented.

To solve such problems, ERP vendors are developing more accurate costing functions. For example, EPICOR™ [4] is working on the development of cost forecasting and planning in addition to the calculation of manufacturing costs; the budget process is defined based on the manufacturing process plan in multiple cost-calculation scenarios. Then, the financial planner uses different methods, such as examining history and trends and making a depreciation analysis, to do the forecasting [4].

In an ERP system like EPICOR™, the cost estimation module still uses the traditional method for quotation generation and final cost determination for a company's manufacturing products and processes. An ERP system conducts cost calculation based on four categories: labor, material, overhead, and service, which can provide the general information necessary to determine cost. To have a more accurate cost estimation and calculations, customization of the software tool is required because more attributes need to be extracted. As an example, Figs. 1 and 2 show the cost attributes of a specific part in EPICOR™ and Visual Info™ [34] from a company which has gone through a transition of ERP package change from one to another.

## 3. Literature review on cost estimation methods

Although product cost estimation has important roles in company decision-making, it usually remains as a key issue in the manufacturing industry. The time required to collect data and the available information limitations impose essential constraints in choosing the most suitable cost estimation method.

Many cost estimation techniques can be initially categorized into qualitative and quantitative methods [5]. Qualitative cost estimation uses a heuristic method in the case of difficulty in finding the accurate cost based on parametric product data [6]. Usually, this technique requires historical cost data and cost engineering experts to identify the similarity of products. The most common qualitative techniques for product cost estimation are implementations of a case-based methodology, a decision support system, and analogical cost estimation.

The quantitative cost estimation approach is typically focused on analytical cost calculation functions and the exact parameters summarized from the records of business transactions for costing purpose [5]. Analytical cost estimation can be very detailed and accurate for complex products. This method, basically, decomposes product manufacturing processes into cost factors based on their operation methods, features, surface quality requirement, key process dimension and geometrical tolerances, and supporting activities. In order to manage the tedious cost elements more efficiently and yet to keep the accuracy of cost estimation, the correlated cost patterns can be identified and categorized with product configuration variations and predefined procedures. In other words, analytical analysis can be extended to a feature-based approach.

In this paper, a feature-based cost estimation approach is proposed as the cost estimation mechanism to identify cost-related patterns that are associated with certain functions and manufacturing processes [5]. The following review sections are focused on the research works related to this approach.

The first research about feature recognition goes back to Lyc Kyprianou's PhD thesis (Shape Classification in Computer-Aided-Design) at University of Cambridge in 1980 [35]. By the end of the 1980s, Shah and Rogers [7] introduced the feature-based modeling system and its requirements conceptually because feature concept mapped very well with engineering semantics with software engineering object-oriented approach. Features have been widely used by the process planning community since that time. One year later, Emmerik and Jansen [8] introduced a set of graphical user interfaces for feature-based modeling. The feature-based engineering informatics approach has gained more and more penetration into computer-aided software tools, such as those in CAD and CAM domains. However, the feature concept had not been used for determining cost purpose until Wierda [9,10], who was one of the first researchers, used the concept of the feature in qualitative and quantitative cost estimation. In 1990, he introduced a new tool that supported users to effectively control costs. He defined the "life cycle cost" by three phases in the model; including conceptual design, materialization, and detailed design. He also defined two types of cost information: design rules and manufacturability information. Design rules refer to the historical design information used to describe a framework for a current design. Manufacturability information refers to the current design information and geometry information of a product; this information is used to extract the product cost data. In his cost model, the target cost was considered in each phase. As a result, the design was based on cost, and the cost was based on information generated from the feature reorganization program [9]. In 1991, Wierda expanded "life cycle cost" to include feature-based cost modeling to fill the gap between the design and the process plan. In addition, he described the advantages of calculating the costs per feature [10].

By the end of the 1990s, more techniques were introduced for using the feature-based cost estimation. Zhang et al. [11] launched the feature-based cost estimation by using back-propagation neural networks, especially for packaging products. They introduced the definition and quantification of the cost-related features, and then used the back-propagation neural networks to figure out the relationship between the cost-related features of product

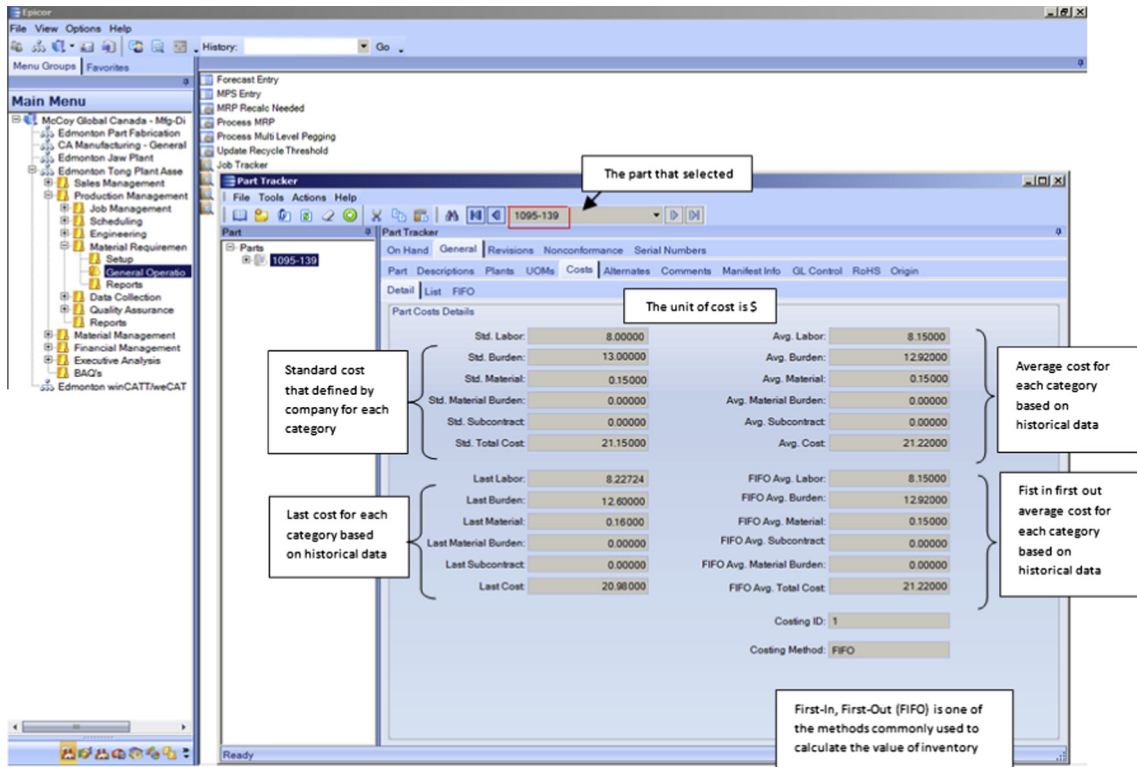


Fig. 1. Typically-defined cost attributes in EPICOR™.

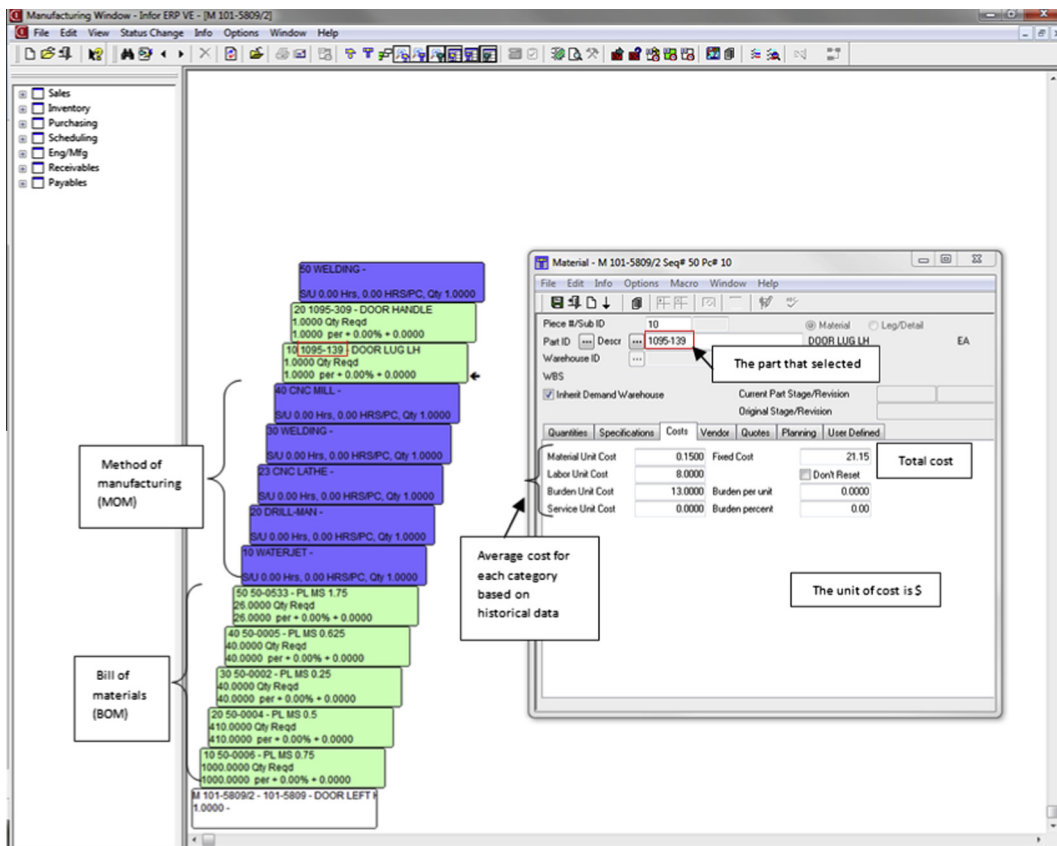


Fig. 2. An example set of cost attributes defined in Visual Info™.

packaging cost. Ou-Yang and Lin [12] suggested new tools for feature-based cost estimation in the early stages of the product lifecycle. They extracted factors that might affect production costs in an attempt to reduce the costs in the design phase. At the end, their method calculated the cost based on products' shapes and the accuracy of the extracted factors. Leibl [13] designed a program-based system for feature-based cost estimation which was integrated with the CAD system to extract the geometrical and non-geometrical information from the design models. Then the product's cost was calculated based on the details that were extracted. This system has to use other design documents to make cost calculation, as it cannot do a cost estimation just based on the CAD model alone. Brimson [14] used the feature-based cost estimation as an alternative activity-based costing (ABC). In his approach, the product's features have to be identified first, and then the working steps or the required activities of each feature can be recognized afterward. Finally, the cost of each activity has to be calculated.

In the last two decades, many researchers tried to improve the efficiency and accuracy of feature-based cost estimations based on specific manufacturing operations. Masmoudi et al. [15,16] used a feature-based concept to do analytical and parametric welding cost estimations with computer-aided tools. They described a cost estimation model that uses preparation features and welding features to approximate an effective welding time and, thus, estimate the cost. They used forward reasoning calculation with multi-parameter inputs, and they captured more details than they would if they had taken the traditional calculative approach. So, the limitation of this approach was that in order to make a parametric cost estimation, all of the features data was required, but in practice, not available.

Alternatively, data mining based cost estimation has emerged as a new approach to improve the accuracy and consistency of cost estimation. Currently, a large portion of the software for cost estimation (SCE) is using data mining algorithms or artificial intelligence (AI) techniques [17]. Lum et al. [18] reported about the effects of data mining techniques on cost estimation. They described that data mining techniques are not always implemented better than traditional cost estimation. It has been reported that a combination of data mining techniques could provide accurate SCE for specific products or processes [18].

#### 4. The proposed hybrid cost estimation methodology

In this paper, a hybrid cost estimation framework is proposed. A parametric approach is used for forward-reasoning cost estimation while a data mining approach is used as a backward cost estimation. These two routes are merged by a weighted combination formula for the ease of application and better accuracy of prediction. The advantage of this framework is that historical cost data can be used and matched with data mining results which are obtained from the backward approach. This data mining route could predict cost estimations, even if none of the data has been provided.

It is important to know that the accurate costs to be incurred should be related to the current plant capacity status and the forecast, which can only be made available from the real shop floor processes. Until recently, only few cost estimation systems consider this aspect [19,20]. Wei and Ma [21] introduced the capacity feature concept to fill in the gap between customer orders and resource management, which is considered an important input source in cost estimation. The goal of this research is to introduce a combination method of parametric feature-based cost calculation and the data mining algorithms, which can be used as a flexible and practical tool for ERP system implementation. Its application is demonstrated by a case study involving feature-based welding activities.

In more detail, as shown in Fig. 3, the method proposed under this framework includes two main modules: empirical analysis and DM analysis. The initial input of the cost estimation system is the historical data, which is usually available in many companies but in different table forms. To unify the data structures and create a characteristic representation of cost components and calculation algorithms associated, a new feature-based cost estimation model has been introduced by authors [22,23].

The authors define a *cost feature* concept as a unique class in the unified feature modeling system to address the characteristics of cost-related data, restrictions, and dependency relations, which are closely related to the product configurations and the manufacturing processes involved. By using the product and manufacturing process information with the support of historical data information, different cost features are extracted according to a set of consistent data structures, i.e., object classes, or templates, to assist complete data gathering. The type definitions of cost features are organized into a pre-defined library, making them available for typical common applications with feature-based cost estimation. The cost feature class diagram is to be introduced in Section 5.

#### 5. Definition of cost feature

As described in the previous publications [22,23], a *cost feature* is a set of an object's characteristic attributes and their related functions and servicing methods, which can be associated and constrained to represent the pattern of semantics for the purpose of cost engineering. The cost feature has to consider the variety of cost-related information aspects including quality, operation sequence and time, design, equipment, materials, functions of application, and geometry and non-geometry details, such as tolerance, surface finish, and working allowance. In addition, each of these aspects has its properties and behaviors that must be taken into consideration for cost feature implementation and applications. Fig. 4 illustrates the partial relations among different features for cost engineering, in which the process cost feature and product batch cost features are presented.

The concept of "*cost feature*" is intended to manage diverse data elements and application variation using "abstraction" and "classification" techniques, which are well known in software engineering domain. By introducing "*cost feature*" as a generic "class" type and further details into sub-classes of different application costs, the common shop-floor costing data can be systematically modeled, represented and processed according to the object-oriented methodology. Hence, the "*cost feature*" becomes a common template. Ideally, the *cost feature* class definition has generic applicability such that one class can be applied for many different instances; therefore, the designers do not need to model the feature structure again. The reusable feature template can be created via one time effort by knowledge or software engineers. The end-user of the software tools like the designers or engineers on the shop floor do not need to model the abstracted class template themselves.

For extracting the cost feature, reference data is required that includes specific design and manufacturing information, such as the number of available machines in the shop floor, the capacity of each machine, the operation that can be performed by each machine, the available human resources that can work with each machine, and the sequence of the machining process. Then, some functions are required to extract the manufacturing features and the related information from the design feature model. Then cost features are mapped to manufacturing process features within the ERP system. The advantage for designer/engineer to use the cost feature scheme is its reusability of the data and the consistent application software interfaces.

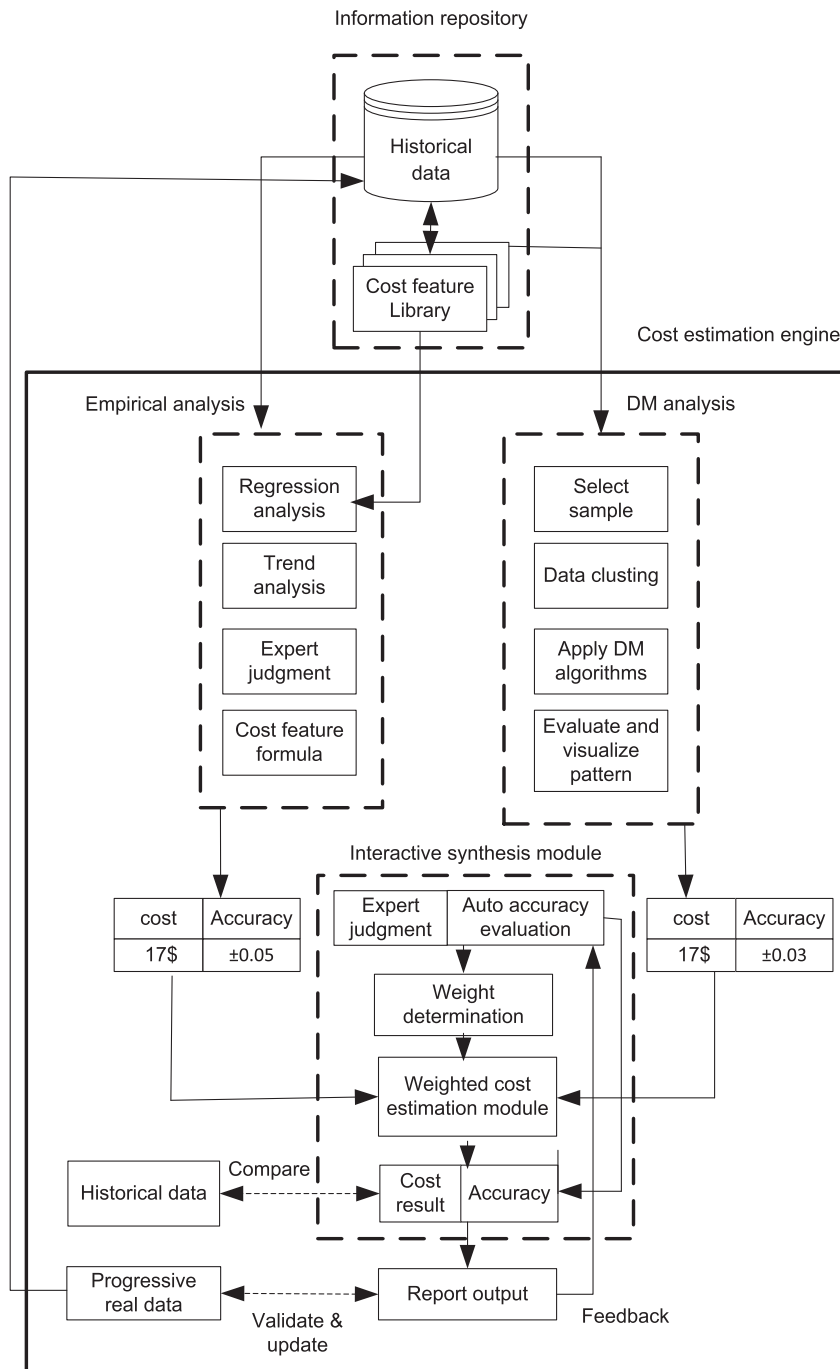


Fig. 3. A hybrid cost estimation method.

The quality of generic *cost feature* definitions have to be proven and matured by cycles of software development, and the application functions are supposed to cover the whole life span of its intended applications. The more “generic” the cost feature definition is, then more “adoptability” is for the use of its property elements and methods. Such object-oriented software engineering approach has been well accepted in the field. To effectively apply *cost feature* concept in specific applications, sub-classes of cost-feature need to be further defined to cater for variations of different product configurations and manufacturing processes. These sub classes’ generic adoptability can be also gradually achieved via semantic modeling.

With the sub-classes of the cost features, all components can be divided into different categories according to their principal

functions, geometry information, similarities, and process plans. As a result, a feature library can be used to classify cost-feature families into a tree under a common parent class. It can also use a feature-mapping technology [24,25] to create a classification for each manufactured product. The proposed method in this work uses pre-defined feature objects to create links between each feature template (class) and its instances (details of costs for the data items).

In the following part of this subsection, the authors use a typical torch welding process as an example to show the cost feature definition procedure. According to [15,12] for a welding process cost feature, which can be a child class of the process cost feature, the follow properties and the related functional methods are suggested:

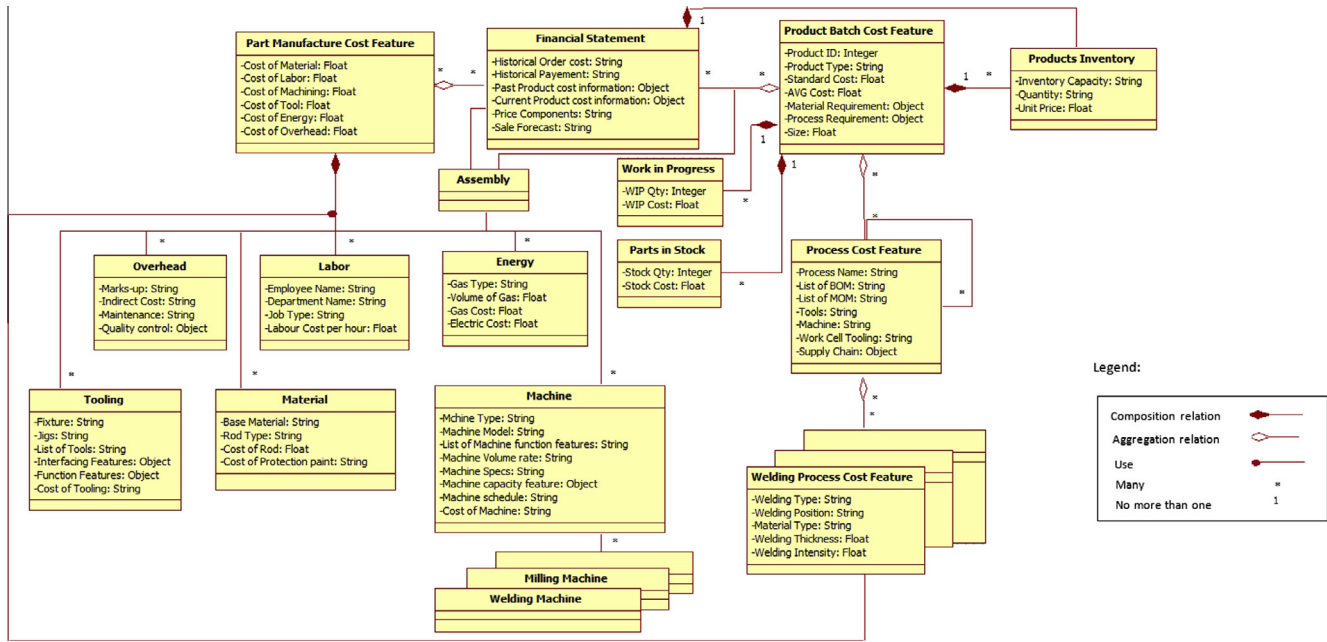


Fig. 4. Cost feature definition in UML format.

- $L_{ij}$ : welding length
- $SS$ : welding section area
- $\rho$ : rod output rate
- $d$ : filling metal density
- $\eta$ : process efficiency
- $D_a$ : welding wire consuming rate with protection gas
- $\phi$ : operator efficiency
- $p$ : degree of weld position complexity
- $M_h$ : unit time cost (\$/h) for each machine hour
- $S_h$ : setup cost (\$/h) for each machine hour
- $V_s$ (): evaluate weld volume:  $V_s = SS \times L_{ij}$
- $m_a$ (): calculate amount of welding wire:  $m_a = \frac{\rho \times V_s \times d}{\eta}$
- $t_{arc}$ (): calculate net welding time:  $t_{arc} = \frac{m_a}{D_a}$
- $T_{ij}$ (): work out welding operating time:  $T_{ij} = p \times \frac{t_{arc}}{\phi}$
- $C_{ij}$ (): work out machining cost for a welding operation:  $C_{ij} = M_h \times T_{ij} + S_h$

Table 1 illustrates the cost component variables identified in a cost feature template for a type of welding operation. Regarding extracting information from parts' geometry and process features, some of the weld information has to be extracted from CAD/CAM model so that the welding time, welding length, volume of welding wire (rod), and the welding tool as well as its control can be worked out. Such feature extraction can be done automatically when the relations are defined. Basically the extraction of information can be achieved by programed algorithms with APIs of the CAD/CAM tool.

**6. Empirical cost estimation process**

To support the hybrid approach, in this proposed framework, an alternative cost estimation function has also been constructed with

the inputs from the historical data and the cost feature data sets collected as illustrated in Fig. 3. In other words, the values of the cost feature variables can also be collected from historic data as the initial input for both the empirical estimation as well as the data mining (DM) analysis. Then, automatic cost variable recognition can be carried out. Realistically human involvement in justification and acceptance of data tolerances is expected. Engineers can also interactive input through friendly and consistent EXCEL spreadsheet user interfaces. Hence, the application barrier for this proposed cost-estimation method does not exist.

Usually, companies use the parametric cost calculation formula to determine the cost of each product. However, they are not considering all of the products' features when they use this formula. As a result, their cost calculations are not accurate. To solve this problem, this work uses empirical analysis to become confident that all of the products' features are considered to achieve an accurate cost estimation.

The empirical cost analysis method has been developed based on the approach suggested by Masmoudi et al. [15,16]. Regression analysis with expert judgment is carried out to develop a feature-based cost calculation formula. An empirical analysis is one of the two complimentary cost estimation methods to estimate product costs and their accuracy from existing data.

Fig. 5 shows the proposed empirical cost estimation process built into the cost estimation engine. The empirical cost estimation module uses the existing ERP data and interpolates data to analyze the weights and coefficients of cost component initial variables from the cost feature input. Regression analysis is a useful tool for evaluating the weights of variables and the coefficients among them. Regression curves can show the data strength and their relations. By using the above analysis, the relative statistic cost weight predictions, coefficients can be extracted out of the initial variable

Table 1  
An example cost feature template for a welding operation.

Feature ID	Feature type	Process step	Process type	Process description	Welding sections	Welding operating time	Machining cost	Material cost
xx-01	T-joint	1	Shielded metal arc welding	End to end with vertical borders	(e, α, z, g, t, h, form)	$T_{ij}$	$C_{ij}$	$M_{ij}$

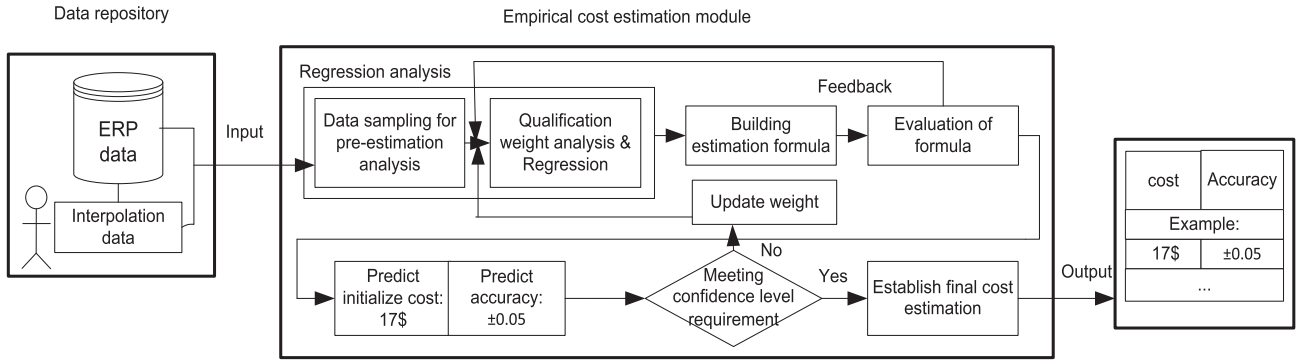


Fig. 5. Empirical cost estimation process.

set; afterward, those contributing and independent variables are quantitatively used to form a linear regression cost estimation formula.

A cost data set is defined as  $\{C_{total}, x_1, x_2, \dots, x_n\}$  where the total cost of the product is denoted by  $C_{total}$ ; it is a differentiable function of  $n$  input component variables:  $x_1, x_2, \dots, x_n$ .

The goal of regression is to draw a function through the data set that can pass data as much as possible. To find the best-fit function of the data, the regression has to analyze the relationship between the variables that are denoted by  $w_i$  to illustrate the coefficient among  $x_i$  and  $C_{total}$ , which means how changes in  $x_i$  can affect  $C_{total}$ .

$$C_{total} = W_1x_1 + W_2x_2, \dots, + W_nx_n + \varepsilon$$

We need to find the best set of  $\{w_i\}$  to minimize the error  $\varepsilon$  and build a more accurate formula for cost estimation. To achieve this goal, the data related to each variable has to qualify the product data with high precision. For example, observing the amount of welding wire consumed in 40 products with their costs, and the relation between the two parameters is shown in Fig. 6. Similarly, after analyzing the cost data correlation with each variable, we can use its coefficient to build the cost estimation formula.

After evaluation of coefficients, the cost estimation formula needs to be updated with the pool of the existing data, and then, it can predict the initial cost value for a new product or order and its accuracy. Along that way, the estimated confidence level is constantly checked. If the predicted costs cannot meet the confidence level required, the weights of the regression analysis have to be updated with regression refinement. This process can be iterated several times to ensure that the cost estimation formula and the level of accuracy are acceptable.

With the data set studied in this work, five hundred products were selected as sample data elements for testing regression analysis for the purpose of cost estimation. As shown in Table 2, the output of regression analysis between the total cost  $C_{total}$  and  $\{X_i\}$  shows that Multiple  $R$  (correlation coefficient) is 0.99. The standard range of Multiple  $R$  is between  $-1$  and  $1$ , in which  $1$  is showing a

strong relation among the variables. In addition, R Square (coefficient of determination) is 0.97. The range of R Square has to be between 0 and 1, and 1 means that the regression line passes exactly through the data. Fig. 7 shows the scatter plot of the multiple regressions that have been generated by the analysis. Note that the products have been reorganized according to the individual total costs.

The formula extracted from regression was:

$$\begin{aligned} \text{Welding cost} = & 0 + (\text{Labor cost per hour} \times (-0.059562)) \\ & + (\text{Amount of welding wire} \times 1.44) \\ & + (\text{paint cost} \times 1) + \text{Gas volume consumed} \times 10.5) \\ & + (\text{Machine cost} \times 4.2) \\ & + (\text{o verhead cost} \times (-0.621)) \end{aligned}$$

### 7. DM cost estimation process

In contrast, as another branch of the hybrid system, the data mining (DM) algorithms are applied to the sample historical data to analyze and extract the patterns for each data set. After evaluating the cost patterns, the DM module will estimate the cost components and their accuracies for new product cost inquiries.

Working out the most accurate cost estimation can be seen as an optimization problem in which the most accurate estimate is sought after. Data mining is a powerful technology that can be used in cost estimation. Indeed, there are many data mining algorithms, and each is suitable for one kind of data analysis. Some widely used techniques such as linear regression can describe data coefficients and behavior. Data mining can extract knowledge from certain amount of data as reference, transfer the data into useful processing rules after data training and testing, and discover useful information and hidden patterns by analyzing new sample data. In the recent decade, data mining has been widely used in order to extract data patterns.

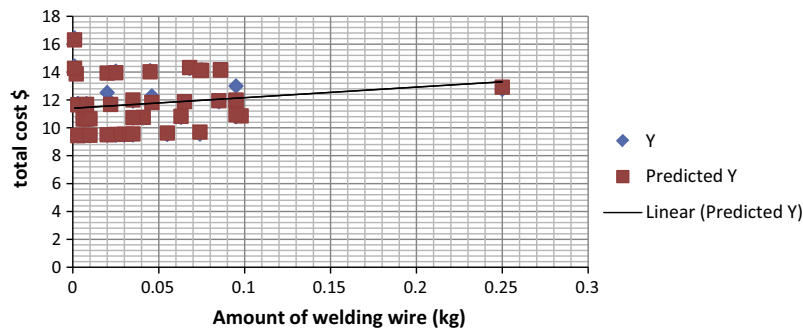


Fig. 6. Relation between total cost and amount of welding wire.

**Table 2**  
Regression statistics result.

Multiple R	0.99
R square	0.97
Adjusted R square	0.85
Standard error	0.33
Observations	500

The proposed DM cost estimation method has several interfaces with the product domain and the process domain. Fig. 8 illustrates the DM cost estimation configuration. The product domain contains information about design configurations, modules, parts, design features, and materials. The process domain defines manufacturing processes, such as machining operations and their related machines, tools, and labors involved.

The DM cost estimation module is shown on the right side of Fig. 8. It can calculate the cost of each specific manufacturing feature based on the information of the product and process domains, as well as the ERP historical data. To achieve this goal, the proposed DM module uses ERP historical data to categorize cost factors by product families based on product configurations, materials, and manufacturing processes. The dominant cost factors are generically abstracted into a cost feature model. In the next step, the DM module analyzes the cost data to associate the cost values with cost feature variables. Then an appropriate data mining algorithm is selected and used to extract the full cost feature patterns (i.e., object instances). As a result, the selected data mining algorithm is associated with the sample data product family. In implementation, the user interface will show the confidence level of each algorithm to the user, and based on that, the user can select the best algorithm for each product family (i.e., group).

## 8. Data mining implementation via the cost engineering engine

For implementing the proposed DM cost estimation method, cost-related data needs to be constantly saved into the ERP database. Several data mining algorithms have been implemented and tested via a free software tool, WEKA, in order to extract and process the significant amount of data in an efficient manner. WEKA has been selected in this paper because it is a machine learning software tool that provides a comprehensive collection of well-tested algorithms for automatic classification, regression, clustering, and feature selection [26].

Usually, software tools for cost estimation (SCEs) use different Artificial Intelligent (AI) algorithms and techniques; such as linear regressions, multiple regressions [27], and artificial neural networks (ANNs) [28,29] which are commonly applied in many types of cost estimation software. Gunduz et al. reported that regression algorithms are more suitable for numerical data, and ANN presents more accurate data with a lower error percentage [27]. In this research, beside using the afore mentioned techniques, KNN and MP5 data mining algorithms have also been added, which are

suitable for both numerical and non-numerical data and multiple relations [30].

In the next section, as a thorough case study, DM algorithms are used to generate new rules and patterns for welding cost estimation. Then, the results of the best DM algorithms are saved into a model to apply to cost estimation. Further, the user has to select the most effective cost estimation algorithm based on confidence level. Confidence level can be defined by comparing actual and estimated cost values based on the mean magnitude of relative error (MMRE). MMRE is a standard function used in SCE that is applied to evaluate the performance of cost estimation on a training data set [31]. Where  $n$  denotes the number of manufacturing parts used for cost estimation and MRE is equal to the following:

$$MRE = \frac{|\text{Actual Cost} - \text{Estimated Cost}|}{\text{Actual Cost}}$$

$$MMRE = \frac{1}{n} * \sum_{i=1}^n MRE$$

## 9. A case study – welding process cost estimation

So far, this paper has presented a preliminary cost estimation framework. A case study that works with welding features has been carried out, welding feature templates are defined as a cluster of related cost feature classes. Note that other than supporting the development of functional modules, as indicated in Fig. 8, a feature library has been constructed to organize the templates used for collecting and analyzing cost data related to welding features. Such templates are useful to classify, sort, and analyze cost factors in future data mining processes.

### 9.1. Definition of welding feature

Welding feature is new type of manufacturing features defined by the authors, which have not been formally defined in the literature thus far. The welding feature is a child class of the *manufacture feature* (see Fig. 9), and it contains a set of extracted welding-related process data, which is associated with welding geometric entities and their characteristic parameters, technical process setup parameters, and related resources. In addition, welding-related geometric entities are further related to product design features and machining derivatives, such as welding paths, which are angles that have to be defined within any welding feature.

Referring to the methods section of the welding feature class in Fig. 9, the welding process is executed with three main sub-processes: *preparing-welding()*, *run-welding()*, and *after-welding()*. They interact internally with the feature properties as shown in the Fig. 9. For example, the preparing process is defined with the setup information; it extracts technical

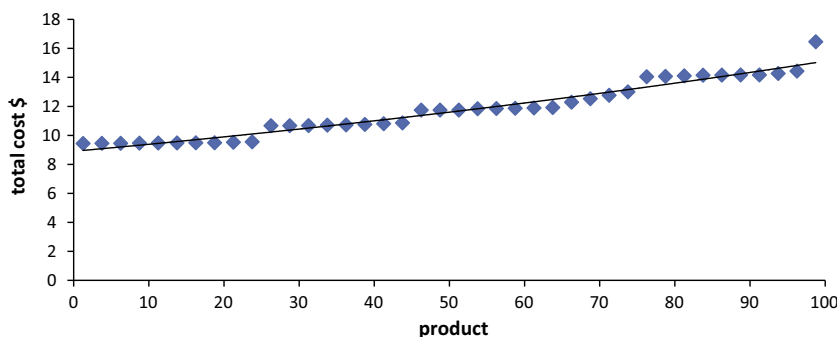


Fig. 7. Scatter plot of the multiple regression analysis for 100 sample products.

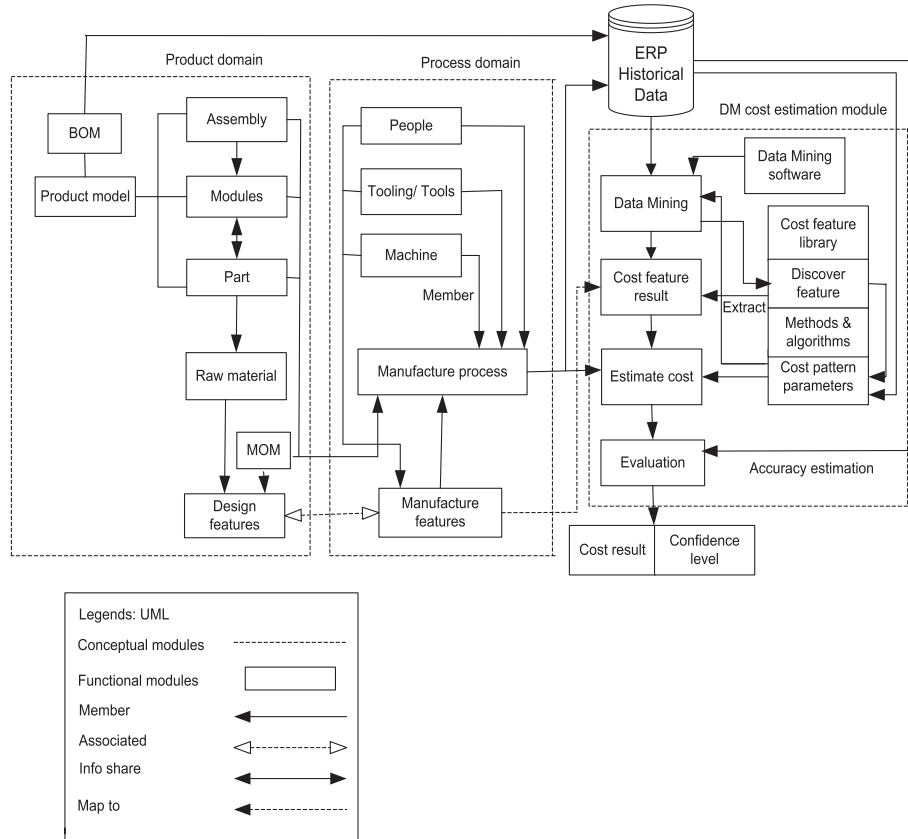


Fig. 8. DM cost estimation configuration.

parameters from handbooks and generates the necessary drive curves and path segments as well as characteristic geometric parameters from product design and structure entities. Now look into the product unified feature model block in Fig. 9. To instantiate welding features, their properties need to be extracted from manufacture and design feature data. The *run-welding()* methods uses the information of welding equipment resources from workshop process planning and operation model via their pointers, such as the welding machine and welding tools. Finally, *after-welding()* method is supposed to define and guides operational application to do cleaning, inspection, and test tasks.

9.2. The data structure of welding feature

In this research, the following data structure of welding feature class is suggested as shown in Fig. 10. The indicative parameters are explained below:

- $L_1, W_1, H_1$ : length, width, and height of first piece
- $L_2, W_2, H_2$ : length, width, and height of second piece
- $T_1$ : time of weld
- $T_2$ : time of setup
- $T_3$ : time of paint
- $T_4$ : time of clean
- BM: Base Material
- WThick: weld thickness
- WType: weld type
- Tech: technology used in welding
- WPosition: welding position
- WProcess: welding process
- WW: amount of welding wire
- PP: amount of protection paint

- GT: gas type
- GV: gas volume rate
- OF: operating factors
- StartP: the location of start point
- EndP: the location of end point
- form: form has three positions: End to End in V; End to End with vertical borders; and End to End in X.

According to [15] Geometric Parameters:  $GP = f(e, \alpha, \beta, g, t, \text{form})$ ; where form has three positions: End to End in V; End to End with vertical borders; and End to End in X. Fig. 11 illustrates these three positions. According to the current definition, P1 represents the geometric information of the first piece such as length, width, and height. P2 represents the geometric information of the second piece that needs to be welded to the first piece. TP includes the technical parameters of welding, which includes base material, weld thickness, weld type, technology used in welding, welding position, welding process, amount of welding wire, amount of protection paint, gas type, gas volume rate, operating factors, the location of start point and end points of welding. GP is the geometric parameters of welding, which includes shape characteristics. Table 3 illustrates the data structure of the welding feature properties.

9.3. The basic functions of welding cost feature

By using the data structure of welding feature, we can define welding cost feature (see Fig. 12), and its related functions as follows:

- Cost of material ( $C_{mt}$ ) = (amount of welding wire  $\times C_{Rod}$ ) + (amount of protection paint  $\times C_{Protection}$ ) + (gas volume rate  $\times C_{Gas}$ )
- Cost of labor ( $C_{lab}$ ) =  $(t) \times R_{labor} \times \text{Operating Factor}$

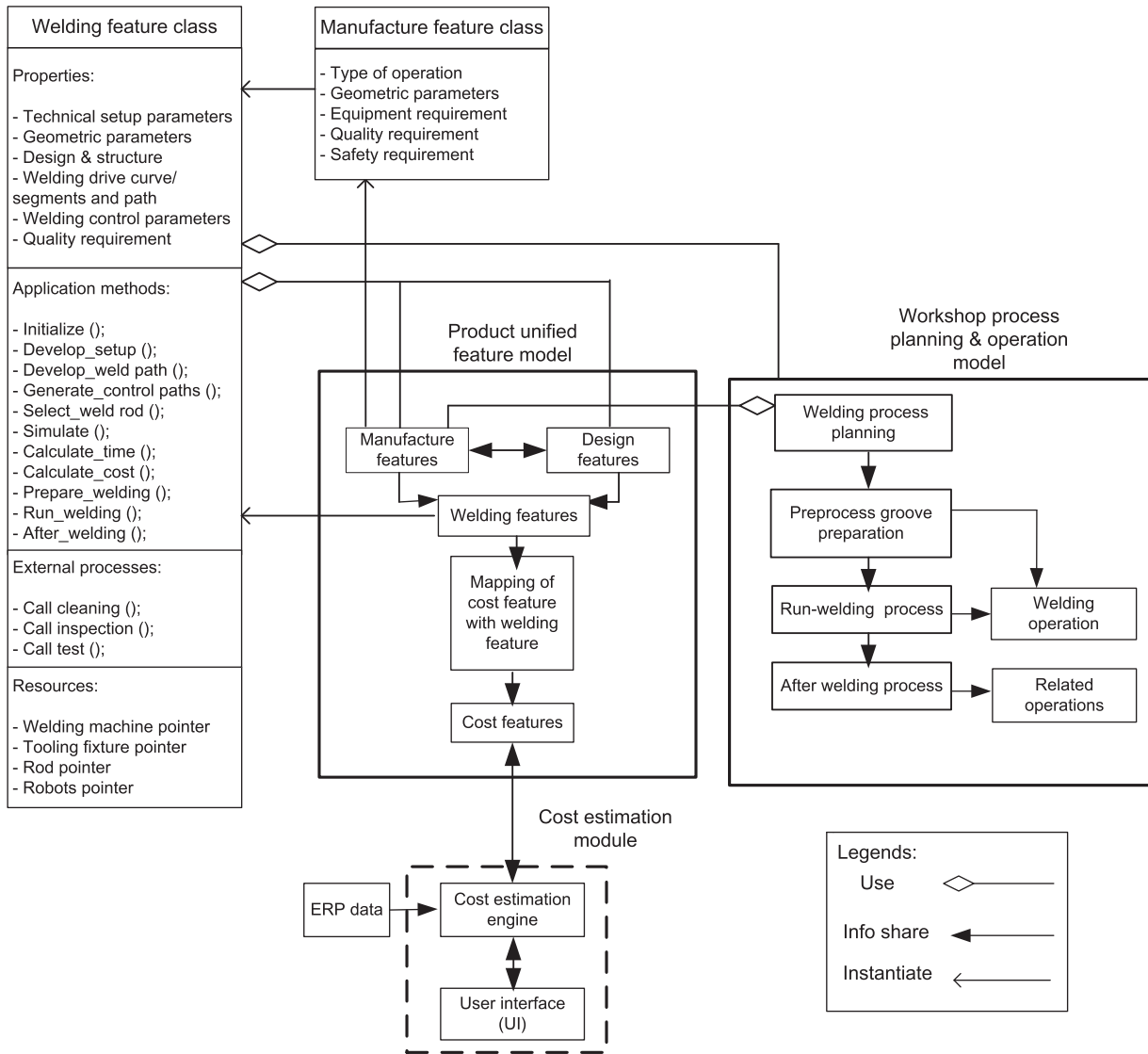


Fig. 9. Welding feature and cost feature structure relations.

- Cost of machine ( $C_{mch}$ ) =  $(C_{occp} \times T1) + (C_{setup} \times T2) + (C_{clean} \times T3) + (C_{paint} \times T4)$
- Cost of welding ( $C_{weld}$ ) =  $f(\text{Type of welding, welding condition})$
- Cost of overhead ( $C_{OH}$ ) =  $f(C_{administration}, C_{depreciation}, C_{infrastructure})$
- CF (:): cost feature:  $CF = [C_{mtl}, C_{lab}, C_{mch}, C_{weld}, C_{overhead}]$

feature-mapping schematic between the welding and cost features is presented.

$C_{Rod}$	$C_{Pro}$	$C_{Gas}$	0	0	0	0	0	0	0	0
0	0	0	$C_{Lab}$	0	0	0	0	0	0	0
0	0	0	0	$C_{occp}$	$C_{Set}$	$C_{Clean}$	$C_{Paint}$	0	0	0
0	0	0	0	0	0	0	0	$C_{Weld}$	0	0
0	0	0	0	0	0	0	0	0	0	$C_{OH}$

Mapping matrix

9.4. Mapping of welding manufacture feature and the associative cost feature

To illustrate the cost feature association method, a feature-mapping system has been used. Feature-mapping system can map the cost features with other types of features in cost estimation via a set of data structures, and this data has been derived from a CAD module or an ERP system. The mapping function is defined as:

$$MF = [WF, CF, f]$$

where

$$f(WF) = CF$$

WF means welding feature, CF is the cost feature, and  $f$  represents the function-mapping algorithm. In the matrix below, a

$$\begin{matrix}
 WW \\
 PP \\
 GV \\
 t^*OF \\
 T_{weld} \\
 T_{set} \\
 T_{clean} \\
 T_{paint} \\
 f_{weld} \\
 f_{OH}
 \end{matrix}
 *
 \begin{matrix}
 C_{mtl} \\
 C_{lab} \\
 C_{mch} \\
 C_{weld} \\
 C_{overhead}
 \end{matrix}
 =
 \begin{matrix}
 C_{mtl} \\
 C_{lab} \\
 C_{mch} \\
 C_{weld} \\
 C_{overhead}
 \end{matrix}$$

Feature array      Cost feature

Welding feature class
<b>Properties:</b> - Geometry information of first work piece: $L_1, W_1, H_1$ - Geometry information of second work piece: $L_2, W_2, H_2$ - Welding times: $T_1, T_2, T_3, T_4$ - Technical parameters - Welding geometric parameters .....
<b>Constraints:</b> - $L_1 > 5\text{cm}$ - $90^\circ < \alpha < 180^\circ$ - $W_{\text{thick}} < 10\text{cm}$ .....
<b>Functions:</b> - Get-parameters (); - Save-parameters (); - Calculate-welding-time (); - Initiate-feature (); - Initiate-UI (); .....

Fig. 10. Data structure of welding feature class.

10. Prototype results

To test the advantages of the proposed framework feature-based cost estimation, a typical problem was formulated as discussed in Section 9 and the computer program was

Table 3

Data structure of welding feature properties.

Welding feature properties	Abbreviation	Options
Base Material	BM	Inch, SQIN
Weld Thickness	WThick	mm
Weld Type	WType	Grove, Seam, Joint
Technique of Grove	Tech Grove	U, square, flare, bevel, V
Technique of Seam	Tech Seam	Single side, double side, all round
Technique of Joint	Tech joint	Square butt joint, V butt joint, Lap joint, T-joint
Weld Position	WPosition	Axis (x, y, z)
Weld Process	WProcess	SMAW, SAW, GMAW, GTAW
Gas type	Gas type	CO <sub>2</sub> , argon, argon/CO <sub>2</sub> mix
Operation factor	OF	(SMAW, 30%), (GMAW, 45%), (MCAW, 55%), (GMAW, FCAW, and SAW processes, 100%)
Start point	StartP	(x, y, z)
End point	EndP	(x, y, z)

Welding cost feature class
<b>Properties:</b> - Geometry information of first work piece: $L_1, W_1, H_1$ - Geometry information of second work piece: $L_2, W_2, H_2$ - Welding times: $T_1, T_2, T_3, T_4$ - Technical parameters - Welding geometric parameters .....
<b>Constraints:</b> - $C_{\text{lab}} = \$24 \text{ per hr}$ - $C_{\text{mch}} < T_1 * 3\$$ .....
<b>Functions:</b> - Get-parameters (); - Calculate-cost of material (); - Calculate-cost of labor (); - Calculate-cost of machine (); .....

Fig. 12. Data structure of welding cost feature class.

Weld feature objects	Diagram
End to End in V - e: height of thickness - t: heel or height of the flat - $\alpha$ : chamfer opening top angle	
End to End with vertical borders - e: height of thickness - g: clearance apace	
End to End in X - e: height of thickness - $\alpha$ : chamfer opening top angle - $\beta$ : chamfer opening bottom angle - g: clearance apace - t: heel or height of the flat	

Fig. 11. Geometric parameters of welding in different welding form [15].

implemented. Visual Studio 2012 was used to implement the cost estimation engine as a functional module acting as a driver in the prototype program; and an SQL Server 2012 was used to store and manage our features and ERP data.

To obtain preliminary results, a sample set of welding-related operation data with 500 welding parts has been selected, currently manufactured by McCoy FARR, ranging from simplistic to complex. This test case is large enough to serve as a good benchmark for our proposed formulation, and yet small enough to be tractable, thus allowing us to test our hypothesis. The test case has been selected to mimic real welded parts and a real welding feature-based cost

**Table 4**  
Sample cost estimation based on regression analysis.

Part ID	Joint type	Labor cost per hour (\$)	Welding time (h)	Operating factor (%)	Amount of welding wire (kg)	Weld rod cost (\$)	After-weld paint cost (\$)	Shield gas volume rate (m <sup>3</sup> /h)	Shield gas cost per cubic meter (\$)	Welding machine cost (\$)	Overhead (\$)	Primary weld cost (\$)	Regression weld cost (\$)	Predicted cost (\$)
A101	butt-joint	24	0.16	0.4	0.03	1.44	0.25	0.62	10.5	0.48	0.65	8.43	9.438	9.427
A102	butt-joint	24	0.25	0.4	0.075	1.44	0.25	0.62	10.5	0.75	0.8	9.65	10.859	11.17
A103	butt-joint	24	0.33	0.4	0.062	1.44	0.25	0.62	10.5	0.99	0.9	10.5	11.85	11.849
A104	butt-joint	24	0.5	0.4	0.75	1.44	0.25	0.62	10.5	1.5	0.98	12.62	14.156	14.155
A105	T-joint	24	0.16	0.4	0.06	1.44	0.25	0.62	10.5	0.48	0.66	8.44	9.449	9.449
A106	butt-joint	24	0.33	0.4	0.03	1.44	0.25	0.62	10.5	0.99	0.75	10.47	11.742	11.742
A107	butt-joint	24	0.16	0.4	0.25	1.44	0.25	0.62	10.5	0.48	0.82	11.1	12.763	12.52
A108	Lap-joint	24	0.33	0.4	0.046	1.44	0.5	0.84	10.5	0.99	0.76	10.54	12.288	11.801

estimation problem as close as possible to reality, ensuring the validity of the predicted results.

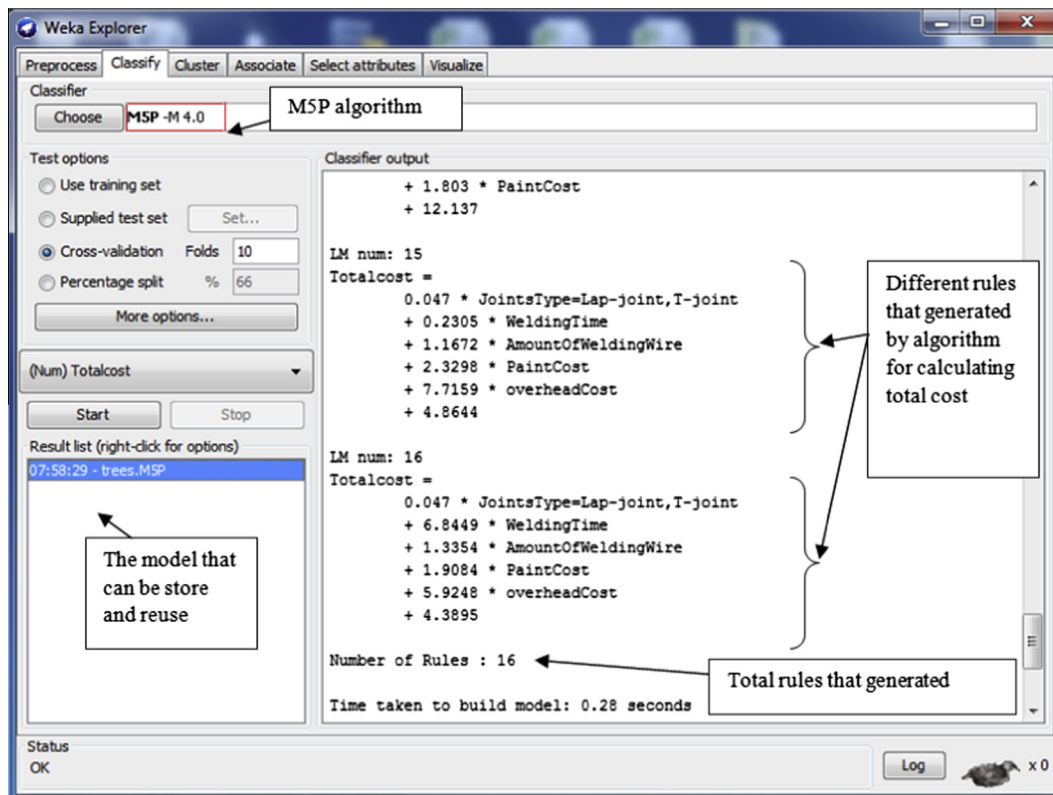
In the first step, data from the company’s ERP system for data analysis has been extracted. Recognizing the trend data, reasonably established by common numerical analysis, is the next step. Data trends can illustrate the overall pattern of data changes, which is suitable for comparing different groups of data. To determine data trends, more details need to be added, such as machining and overhead costs. Besides adding more details, the new welding cost is calculated based on a regression cost formula, which has been introduced in Section 6. The data analysis shows that the new cost increases 7–9% compared to the primary company’s data.

The regression welding cost estimation is presented below in Table 4, beside the primary welding cost. As a result, accurate cost estimation based on a forward approach is used to cross validate the DM cost estimation process.

Based on existing information for welding costs and features, a design of a set of rules (tailored to a specific implementation) is

needed that allows to define the guidelines used to implement the DM cost estimation technique consistently, for the case study. The set of rules we defined is the following:

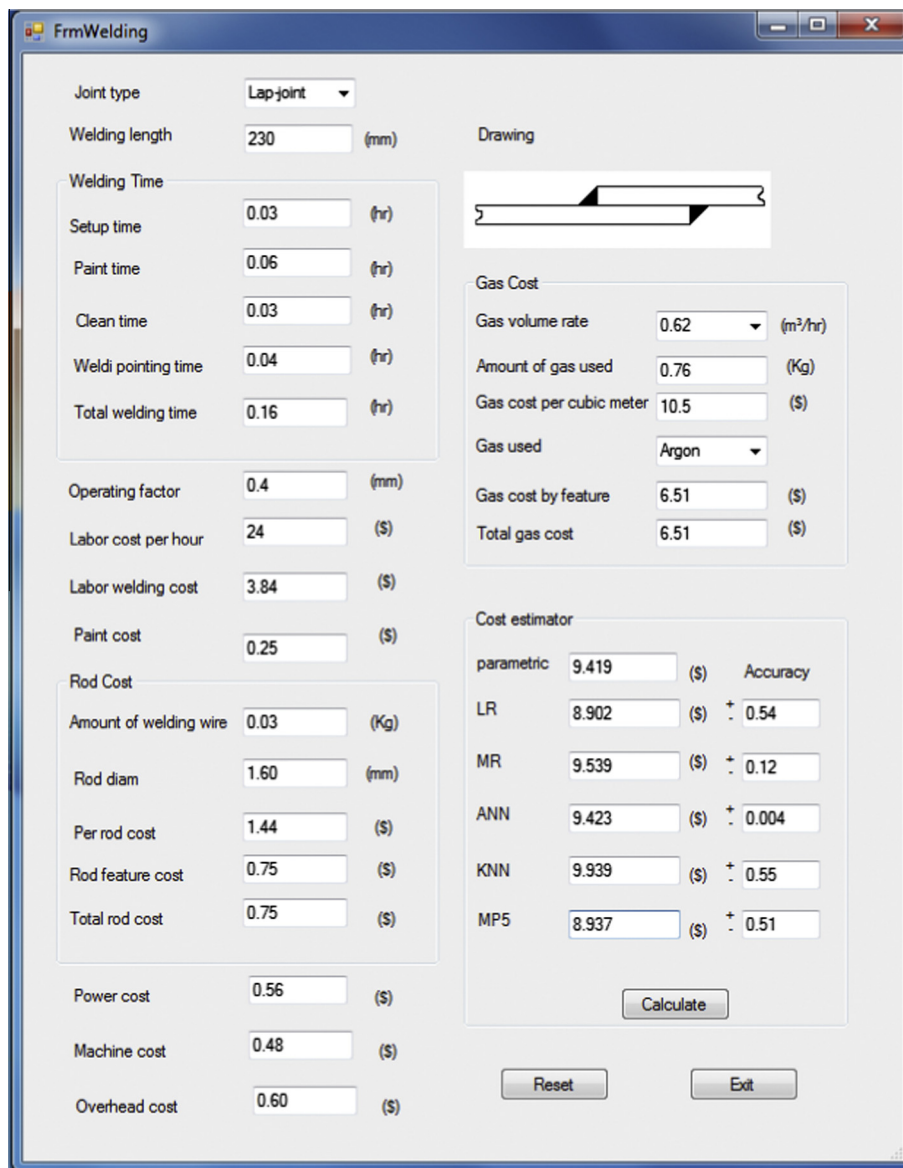
- Rule 1: Joint type: {butt-joint, Lap-joint, T-joint}
- Rule 2: Labor cost: {hourly = \$24}
- Rule 3: Welding time: {setup time, length of weld, welding sections, complexity, machining volume rate}
- Rule 4: Operating factors: {FCAW = 40%}
- Rule 5: Amount of welding wire: {welding volume, filling metal density}
- Rule 6: Rod cost: {per rod 7018 = 1.44\$}
- Rule 7: Paint cost: {welding sections, joint type (butt-joint = 0.25\$, Lap-joint = 0. 5\$, T-joint = 0.25\$)}
- Rule 8: Gas volume rate: {nozzle size 1/2 in. = 0.62 (m<sup>3</sup>/h), nozzle size 5/8 in. = 0.84(m<sup>3</sup>/h)}
- Rule 9: Gas cost: {per cubic meter = 10\$}
- Rule 10: Machine cost: {welding time \* 3\$}



**Fig. 13.** Implementing DM algorithm (MSP) in WEKA.

**Table 5**  
Accuracy comparison among the five different algorithms (%).

Models	N	LR			MR			ANN			KNN			MP5		
		Training data	Testing data	MMRE	Training data	Testing data	MMRE	Training data	Testing data	MMRE	Training data	Testing data	MMRE	Training data	Testing data	MMRE
Cluster 0	52	78	53	53.4	74	68	33	84	92	16.8	72	86	51.6	85	95	33
Cluster 1	130	68	43	52.1	69.3	29.6	52	82.3	96.2	15.6	69.3	35	65.6	79.6	92.3	25.3
Cluster 2	78	82	85	45	85	89.6	16.2	79.6	93.6	16.5	86	70.3	39.6	74.6	89.6	35.6
Cluster 3	43	76	69	33	63.3	46.2	39.3	85.6	85	12.3	64.3	35.7	52.3	89.7	92.6	55.6
Cluster 4	19	68	73	52.3	69.6	36.1	23.6	80.3	86.3	16.3	82.6	73.5	33.5	86	92.3	38.9
Cluster 5	55	64	35	59.6	67.4	42.5	17.6	78.6	92	19.3	68.9	75	59	84	85.6	18.9
Cluster 6	123	74	46	43.8	72.6	39.6	56.4	83.6	93.5	25.2	74.8	48.6	45	84.6	91.6	47.6
AVG		73	58	48	72	50	34	82	91	17	74	61	50	83	91	36



**Fig. 14.** User Interface (UI) for welding cost estimation.

- Rule 11: Overhead cost: {0.05 \* direct cost}

The proposed feature-based cost engine uses the following steps to prepare a DM cost estimation:

1. Reorganizing and setting cost feature data extracted from the ERP system.
2. Implementing data mining clustering algorithm via WEKA: Expectation Maximization (EM) selected as a data mining clustering algorithm, which models a data set based on linear

combination and normal distribution [32]. EM divided a data set into seven clusters, and then each cluster was used as an input for cost estimation. Eighty percent of each cluster has been used as training data set and 20% for testing the data set.

3. Implementing data mining algorithms via WEKA: Linear Regression (LR); Multiple Regression (MR); Artificial Neural Network (ANN), k-Nearest Neighbors (KNN), and the Meteorological parameter section (MP5). For instance, Fig. 13 illustrates the implementation of the MP5 algorithm in WEKA. This model, generated for each cluster, was saved and reused for further cost estimation.

As mentioned previously, five hundred welding parts have been used as an input for five generated data mining models. The comparison between the results of five different algorithms in the case study is shown in Table 5. Clearly, ANN and MP5 have better results for correct prediction. In addition, ANN has a better confidence level compared to the other algorithms. To optimize the result of models, the trends of MMRE were analyzed. LR can be optimized by adding MMRE to the cost estimation. However, other models do not have a same trend, so the result is sometimes less than the actual cost and sometimes more. In this paper, the suggested algorithm for welding cost estimation is the artificial neural networks (ANN) algorithm.

Fig. 14 illustrates more extended welding cost information, which includes welding time, rod, gas, power and machine costs. This kind of information should be provide by user, however if some data is not available, the UI will be fill it based on historical data. The UI allows cost engineers to review and capture all welding cost features; then the cost can be calculated based on parametric cost estimation and five data mining algorithms. Finally, the company can select the best algorithm based on the company's strategy for cost estimation.

## 11. Discussion

The proposed system is based on a semi-automatic hybrid method, where categorizing the product features based on their shapes and characteristics has to be done by engineers. Potentially, this task could be assisted by a feature recognition program [33], which is not the focus of this study. Note that the proposed *cost features* are not similar to CAD design or geometry features. They represent a group of classes from a different engineering application domain. However, they could be associated by feature mapping mechanism as introduced in [36]. Current CAD feature definitions can be applied but they are not versatile enough. More customer-oriented feature development is necessary via CAD APIs.

When feature-based empirical models and data mining algorithms are used, the estimated cost results can be different. The user can use both methods in parallel at the beginning to build up confidence over each method and make judgment on the selection of the results. These two cost estimation methods can be validated with the historical data, and they could be also combined with a weighted sum mechanism, as shown in Fig. 3. The weight given to each method is based on expert judgment from the validation results or is suggested by a system's accuracy evaluation algorithm. Then, ideally, a combined method can be customized based on the accuracy of each method. The empirical approach has the most accurate cost estimation if all variable are available, but this approach requires more data and time (for gathering data) for the accurate cost estimation. On the other hand, data mining approach is preferred when some variable are missing or new product needs cost estimation.

Therefore, the proposed framework offers a mechanism to support the gradual adoption of automated cost estimation method

with the moderation of human experience; and over a period of adoption, the approach can be stabilized and, potentially, has accurate cost estimation with adaptive adjustment capability in real implementation.

The example company's ERP data description is not the same as this work's welding and cost feature-definitions. Data was preprocessed based on the definitions proposed, to keep the consistency of data input to the prototype algorithm. This issue exists because in some instances, the actual manufacturing operations applied are not the same as the process plans, causing some discrepancy in the testing model. Finally, some feature data values do not exist in ERP system and had to be obtained from other sources. For example, to find the thickness of welding specifications of the seam, information has to be extracted from a CAD drawing weld symbols.

## 12. Conclusion

This paper presents a cost estimation framework with feature-based empirical data regression and data mining algorithms. The theoretical significance of proposed method is the new feature-based hybrid approach with linear regression and data mining. The method leverages feature-based engineering unique strengths and creates a mechanism to discover cost features and validate the cost estimation by crosschecking the results of the two approaches. The final estimation function can tap the user's confidence levels for each of the approaches such that the application of the method can be phased in gradually in reality by building up the data mining capability. The case study shows an application to welding products, and it demonstrates that the approach is capable of producing satisfactory results for different welding-feature-based parts. The proposed model has proven to be repeatable, robust, and accurate for the range of features applied. In addition, as shown in the case study, a set of rules or guidelines used to fill the gaps in missing data. Such rules are complimentary with specific applications while the built-in features consider general patterns of welding parameter estimation. This cost estimation methodology could be extended beyond just cost engineering, e.g. the total manufacturing time evaluation.

There are significant advantages in implementing the proposed framework. The first is the capability to work out the estimated costs with known confidence measure for different components. Engineers will spend less time and can deal with more complexity in cost engineering application compared to the traditional analytical cost-estimation formula. The proposed reusability of cost feature patterns and the recursive accuracy-enhancement process will prove the applicability of the proposed approach over time. On the other hand, the cost estimation from DM based on accumulated historical data can serve a calibration tool for the empirical model, which can support new part cost estimation without detailed exhaustive data. This proposed method makes it feasible to increase cost estimation accuracy and to shorten the cost feedback time within an ERP system via customized UIs so that the decision-makers and customers can be timely informed. The experience of the report case study has supported this argument and the method provides practical guidelines to front-line cost engineers.

However, there are some limitations of which the reader should be aware. Predefinition of manufacturing features is the first limitation so that the manufacture process data can be consistently managed and analyzed. Such consistent feature definitions need one-time semantic modeling effort in any adopting enterprise. However, reusing the definitions with repetitive applications is highly recommended and hence the implementation effort barrier would be reduced significantly. On the other hand, data mining needs to be supported with available and comprehensive historical data. In the current implementation case, ERP data was translated

into a different data structure in order to support data mining and feature data extraction. Ideally, with the popular adoption of the proposed scheme and well-defined feature, ERP vendors via the use a converter or following a compliance standard can interface the ERP data structures automatically. Clearly, a lot of future work is necessary.

### Acknowledgements

The authors would like to acknowledge that the technical case study original data was provided by McCoy Drilling and Completions (Farr) Company. This research work was jointly supported by two continued NSERC discovery grants, a MITACS accelerate cluster internship project, and a NSERC CRD grant; they were co-sponsored by McCoy. This research effort was fully carried out at the University of Alberta, Edmonton, Canada.

### References

- [1] L. Shaul, D. Tauber, CSFs along ERP life-cycle in SMEs: a field study, *Ind. Manage. Data Syst.* 112 (3) (2012) 360–384.
- [2] C. Li, ERP packages: what's next?, *Inf Syst. Manage.* 16 (1999) 31–35.
- [3] D. Veeramani, P. Joshi, Methodologies for rapid and effective response to requests for quotation (RFQs), *IIE Trans.* 29 (10) (1997) 825–838.
- [4] Epicor, "Epicor Enterprise Resource Planning Overview". <<http://www.atc.gr/files/Brochures/Epicor-Enterprise-Resource-Planning-Overview-ENS.PDF>>.
- [5] A. Niazi, J.S. Dai, S. Balabani, L. Seneviratne, Product cost estimation: technique classification and methodology review, *J. Manuf. Sci. Eng.* 128 (2) (2006) 563–575.
- [6] R. Roy, S. Kelvesjo, S. Forsberg, C. Rush, Quantitative and qualitative cost estimating for engineering design, *J. Eng. Des.* 12 (2) (2001) 147–162.
- [7] J.J. Shah, Mary T. Rogers, Functional requirements and conceptual design of the feature-based modelling system, *Comput.-Aided Eng. J.* 5 (1) (1988) 9–15.
- [8] Van M.J.G.M. Emmerik, F.W. Jansen, User interface for feature modelling, in: *Computer Applications in Production and Engineering, Proceedings of the Third International IFIP Conference, CAPE '89, 1989*, pp. 625–632.
- [9] L.S. Wierda, Design-oriented cost information: the need and the possibilities, *J. Eng. Des.* 1 (2) (1990) 147–167.
- [10] L.S. Wierda, Linking design, process planning and cost information by feature-based modelling, *J. Eng. Des.* 2 (1) (1991) 3–19.
- [11] Y.F. Zhang, J.Y.H. Fuh, W.T. Chan, Feature-based cost estimation for packaging products using neural networks, *Comput. Ind.* 32 (1) (1996) 95–113.
- [12] C. Ou-Yang, T.S. Lin, Developing an integrated framework for feature-based early manufacturing cost estimation, *Int. J. Adv. Manuf. Technol.* 13 (9) (1997) 618–629.
- [13] P. Leibl, Cost calculation with a feature-based CAD system using modules for calculation, comparison and forecast, *J. Eng. Des.* 10 (1) (1999) 93–102.
- [14] J.A. Brimson, Feature costing: beyond ABC, *J. Cost Manage.* 12 (1998) 6–13.
- [15] F. Masmoudi, W. Hachicha, Z. Bouaziz, A new feature-concept applied in cost estimation model for a weld assemblage: additional Information, *Munich Personal RePEc Archive (MPRA)* (2007) 149–162.
- [16] F. Masmoudi, Z. Bouaziz, W. Hachicha, Computer-aided cost estimation of weld operations, *Int. J. Adv. Manuf. Technol.* 33 (3–4) (2007) 298–307.
- [17] Z. Abbasi Khalifelu, F. Soleimani Gharehchopogh, Comparison and evaluation of data mining techniques with algorithmic models in software cost estimation, *Procedia Technol.* 1 (2012) 65–71.
- [18] K.T. Lum, D.R. Baker, J.M. Hihn, The effects of data mining techniques on software cost estimation, in: *Engineering Management Conference, 2008, IEMC Europe 2008, IEEE International, IEEE, 2008*.
- [19] M. Eklin, Y. Arzi, A. Shtub, Model for cost estimation in a finite-capacity stochastic environment based on shop floor optimization combined with simulation, *Eur. J. Oper. Res.* 194 (1) (2009) 294–306.
- [20] P. Feldman, A. Shtub, Model for cost estimation in a finite-capacity environment, *Int. J. Prod. Res.* 44 (2) (2006) 305–327.
- [21] J. Wei, Y.-S. Ma, Design of a feature-based order acceptance and scheduling module in an ERP system, *Comput. Ind.* 65 (1) (2014) 64–78.
- [22] Y.-S. Ma, N. Sajadfar, L. Campos Triana, A feature-based semantic model for automatic product cost estimation, *IACSIT Int. J. Eng. Technol.* (2014) 109–113.
- [23] N. Sajadfar, L. Campos Triana, Y.-S. Ma, Interdisciplinary semantic interactions within a unified feature model for product cost estimation, *Int. J. Mech. Eng. Mech.* (2014) 10–19.
- [24] Z. Liu, W. Bu, J. Tan, Motion navigation for arc welding robots based on feature mapping in a simulation environment, *Robotics Comput.-Integr. Manuf.* 26 (2) (2010) 137–144.
- [25] L.M. Cao, X.W. Wang, Q.L. Zeng, Study on cost modeling based on feature mapping technology, *Adv. Mater. Res.* 295 (2011) 830–833.
- [26] Mark Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, *The WEKA data mining software: an update*, *ACM SIGKDD Explor. Newsl.* 11.1 (2009) 10–18.
- [27] M. Gunduz, L. Onur, E. Ozturk, Parametric cost estimation system for light rail transit and metro trackworks, *Expert Syst. Appl.* 38 (3) (2011) 2873–2877.
- [28] K.-K. Seo, A methodology for estimating the product life cycle cost using a hybrid GA and ANN model, *Artificial Neural Networks-ICANN 2006*, Springer, Berlin, Heidelberg, 2006, pp. 386–395.
- [29] F. Gharehchopogh, Neural networks application in software cost estimation: a case study, in: *International Symposium on Innovations in Intelligent Systems and Applications (INISTA 2011)*, pp. 69–73.
- [30] X. Wu, V. Kumar (Eds.), *The Top Ten Algorithms in Data Mining*, CRC Press, 2009.
- [31] M. Jørgensen, A critique of how we measure and interpret the accuracy of software development effort estimation, in: *First International Workshop on Software Productivity Analysis and Cost Estimation*, Information Processing Society of Japan, Nagoya, 2007.
- [32] O.A. Abbas, Comparisons between data clustering algorithms, *Int. Arab J. Inf. Technol. (IAJIT)* 5 (3) (2008).
- [33] S. Prabhakar, M.R. Henderson, Automatic form-feature recognition using neural-network-based techniques on boundary representations of solid models, *Comput. Aided Des.* 24 (7) (1992) 381–393.
- [34] [http://www.infor.com/product\\_summary/erp/visual/](http://www.infor.com/product_summary/erp/visual/).
- [35] L.K. Kyprianou, *Shape Classification in Computer-Aided Design*, PhD Thesis, Christ's College, University of Cambridge, UK, 1980.
- [36] Y.-S. Ma, *Semantic Modeling and Interoperability in Product and Process Engineering*, Springer, London, 2013.