Available online at www.sciencedirect.com
ScienceDirect
Procedia CIRP 50 (2016) 665 – 671
www.elsevier.com/locate/procedia

26th CIRP Design Conference

Innovative Design and Analysis of Production Systems by Multi-objective Optimization and Data Mining

Amos H.C. Ng[1,2*], Sunith Bandaru[1] and Marcus Frantzén[1,3]

[1]*Production & Automation Engineering, School of Engineering Science, University of Skövde, Sweden*
[2]*School of Engineering, Jönköping University, Sweden*
[3]*Volvo Car Corporation, Sweden*

* Corresponding author. Tel.: +46-500-448541. *E-mail address:* amos.ng@his.se

Abstract

This paper presents an innovative approach for the design and analysis of production systems using multi-objective optimization and data mining. The innovation lies on how these two methods using different computational intelligence algorithms can be synergistically integrated and used interactively by production systems designers to support their design decisions. Unlike ordinary optimization approaches for production systems design which several design objectives are linearly combined into a single mathematical function, multi-objective optimization that can generate multiple design alternatives and sort their performances into an efficient frontier can enable the designer to have a more complete picture about how the design decision variables, like number of machines and buffers, can affect the overall performances of the system. Such kind of knowledge that can be gained by plotting the efficient frontier cannot be sought by single-objective based optimizations. Additionally, because of the multiple optimal design alternatives generated, they constitute a dataset that can be fed into some data mining algorithms for extracting the knowledge about the relationships among the design variables and the objectives. This paper addresses the specific challenges posed by the design of discrete production systems for this integrated optimization and data mining approach and then outline a new interactive data mining algorithm developed to meet these challenges, illustrated with a real-world production line design example.

Peer-review under responsibility of the organizing committee of the 26th CIRP Design Conference

Keywords: Production Systems; Multi-Objective Optimization; Data Mining.

1. Introduction

Designing a production system involves many complex decisions over various phases of the project period in order to satisfy the strategic objectives of the manufacturing company. These design decisions include equipment sizing, layout, level of automation, workload allocations, internal and external material logistics, for a new facility or for the re-configuration of an existing one can pose big challenges to the designer/manager (or decision maker, DM) because of the complex combinations and interactions among the system entities. While ideally all designers want their systems to be "optimally" designed, but in most of the practical design situations, to select the optimal values of the design variables of the system entities so as to achieve the desired overall performance of the system is a very difficult, if not an impossible task. This is particularly relevant to production systems design as the number of design variables is very often numerous, which making the determination of which variables are influencing and should be included in the analysis activity too difficult to be answered. On the other hand, a fact that is easily overlooked by many designers is that production system design is actually a multi-criteria decision-making (MCDM) problem by nature. Firstly, practically all production development projects are subject to a cost constraint so that the system requirements (e.g. production capacity) can only be fulfilled within a fixed budget. A trade-off situation can very easily be encountered, for instance, when certain design solutions suggest higher performance machinery with higher investment cost can replace some lower performance, but

Peer-review under responsibility of the organizing committee of the 26th CIRP Design Conference
doi:10.1016/j.procir.2016.04.159

cheaper machinery. In other words, a production system design problem always involves at least two objectives: production capacity and cost. Secondly, if a designer considers more than one performance measures of the system, for example, system throughput (TH), Work-In-Process (WIP) and cycle time (CT), according to Factory Physics [1], then how the decision variables affect these performances will become a complicated issue. Take a very simple design decision for discussion: how to allocate the inter-station buffers, i.e., how many and where inter-station buffers are allocated, can be the crucial decision variables that affect TH and WIP as two conflicting objectives in designing a production line.

Computer simulation has been described as the most effective tool for designing and analysing complex production systems in industry, especially when the analysts are interested in the dynamics of the system and the system performance of a given design [2]. Simulation should be utilized to predict the impact on system performance in advance, early in the conceptual phase of production system design, before any important design decisions have been made [3, 4]. When designing a production system in the conceptual phase, there are often many different design alternatives for the DM to select. To succeed in building an effective production system in terms of maximized TH and minimized required investment, a decision support tool can be utilized in aiding the responsible DM to select the best possible configuration of the system. Simulation by itself, however, does not suggest solutions of how the system should be configured to achieve the desired performance – it merely provides the DMs with valuable information so that their decisions to select the final solutions for implementation can be based on [5]. Since the number of different possible alternatives can be huge when considering all the combinations of the related decision variables, it can be impractical or even impossible to find an appropriate system configuration with an acceptable performance level by using a trial-and-error approach with experiments on a simulation model alone [4]. As explained in [6], simulation for performance evaluations alone is often insufficient – a more powerful approach is to combine simulation with optimization when analyzing complex manufacturing systems. By using multi-objective optimization (MOO) techniques, trade-off solutions for configurations of the system can be found and then analyzed when designing a production system [7].

The aim of this paper is to introduce a production systems design approach that explicitly integrates interactive decision-making and rule visualization into a general simulation-based optimization system for supporting decision making. Particularly, the approach will put emphasis on how interactive decision making and rule visualization can fit perfectly into a MOO framework. As will be explained in Section 2, unlike ordinary optimization approaches for production systems design, which several design objectives are linearly combined into a single mathematical function, MOO that can generate any design alternatives and sort them into an efficient frontier can enable the designer to have a more complete picture about how the design decision variables, like number of machines and buffers, can affect the performance of the system. Because of the multiple optimal design alternatives generated, they constitute a dataset that can be fed into some data mining algorithms for extracting the knowledge about the relationships among the design variables and the objectives (see Fig. 1). Such kind of knowledge obtained from the optimization and the post-optimality processing is important for confident decision making, as will be revealed in a real-world conceptual design study in Section 4. Section 3 addresses the specific challenges posed by the design of discrete production systems and then outlines a new interactive data mining algorithm called Flexible Pattern Mining developed to meet these challenges.

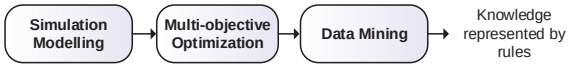

Fig. 1. The key components of the proposed approach.

2. Multi-Objective Optimization

In this section, we consider how a production system design problem is transformed into a multi-objective optimization problem (MOP). Consider a general MOP consisted of M objective functions, $f_i(x)$, $i=1,\ldots,m$, which can be minimized and maximized with **x** as a decision (design) vector, consisted of n decision variables, x_i, within their respective lower bounds (x_i^L) and upper bounds (x_i^U). Mathematically,

$$\text{Maximize/Minimize } \mathbf{F}(\mathbf{x})=\{f_1(\mathbf{x}),\ldots, f_m(\mathbf{x})\} \quad (1)$$

Subject to $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$,
where $x_i^l \le x_i \le x_i^u$ and $i = 1,2, \ldots n$.

For a continuous MOP, we call $\Phi = \prod_{i=1}^{n}[x_i^l, x_i^u] \subset \mathbb{R}^n$ as the design space. $\mathbb{Z}^m$ is called the objective space with the mapping $\mathbf{F}: \Phi \to \mathbb{Z}^m$ that evaluates $f_1(\mathbf{x}),\ldots, f_m(\mathbf{x})$, for $\mathbf{x} \in \Phi$. For a production system design problem, x_i can be either continuous or discrete. While processing times, availability are continuous by nature, they can be restricted to be discrete in most practical situations. Similarly, layout geometry and conveyor lengths can be formulated as either continuous or discrete. However, buffer capacities, number of workers/pallets/AGVs and other nominal variables are always discrete.

In many MOPs, where the objectives $f_i(x)$ are in conflict with each other, finding a single best optimal design is impossible because improving one objective would deteriorate the other. This gives rise to the concept of Efficient Frontier (*EF*) that denotes the best trade-off in $\mathbb{Z}^m$ with respect to $f_i(x)$. The definition of dominance is essential for an optimization to find and compute **x** that constituent *EF*: considering only Min$\{f_1(\mathbf{x}),\ldots, f_m(\mathbf{x})\}$, a design vector $\mathbf{x_1}$ is said to dominate another solution vector $\mathbf{x_2}$, denoted as $\mathbf{x_1} \preccurlyeq \mathbf{x_2}$, iff $f_1(\mathbf{x_1}) \le f_i(\mathbf{x_2})$ $\forall i \in \{1, \ldots, m\}$, but $\exists j \in \{1, \ldots, m\}$ s.t. $f_1(\mathbf{x_1}) < f_i(\mathbf{x_2})$. A design vector $\mathbf{x^*} \in \Phi$ is Pareto-optimal if $\nexists \mathbf{x}$ that dominates $\mathbf{x^*}$. In other words, the Pareto-optimal set, *PS*, is consisted of $\mathbf{x^*}$ that are non-dominated to each other in $\mathbb{Z}^m$. Equivalently, an objective vector $\mathbf{z^*} \in \mathbb{Z}^m$ is Pareto-optimal if the design vector corresponding to it is Pareto-optimal [8].

With the above formulation, it is not difficult to understand an optimization-based design approach will involve a process for finding the *PS* that forms *EF* so that a DM can make a final selection on which $\mathbf{z^*}$ to choose. The MOO literature refers to this post-optimality decision-making process to be based on some higher-level information, which is also very common to involve some preferences of the DM that were not formulated into the objectives in the MOP.

With a simulation-based optimization approach, all function evaluations are done through the simulation model, giving only estimation of $f_i(\mathbf{x})$. In other words, $\mathbf{F}(\mathbf{x})$ can only be estimated by $\hat{\mathbf{F}}(\mathbf{x})$ through the performance values obtained from simulation replication runs. Stochastic simulation will give different $\hat{\mathbf{F}}(\mathbf{x})$ in each replication so the estimated output objective function values used for the optimization are actually $\hat{\mathbf{F}}(\mathbf{x}) = \{(\sum_{j=1}^{r} f_{1,j}(\mathbf{x}))/n, \dots, (\sum_{j=1}^{r} f_{m,j}(\mathbf{x}))/n\}$. Therefore, MOO is in general more computational demanding than single-objective optimization, requiring many more function evaluations [9]. The fact that stochastic simulation requires multiple replications to reduce the uncertainty of the objective functions evaluations has further increased the computational burden for finding **x***. In [9], a summary of several important topics for future MOO research is given, including: (1) automatic, on-line adaptation of tuning parameters; (2) hybrid MOO and local search strategies; (3) algorithms that can provide good performance with few function evaluations; (4) efficient algorithms for many (>3) objectives; (5) the incorporation of user preferences into the algorithms. The DM is not always interested in the whole *EF* has led to the idea that efficiency of optimization can be improved by incorporating the preference(s) so that algorithm search will be focused in the interested area(s) of the *EF*. This concept of acquiring the preferences from the DM to guide the search of trade-off solution is a central theme in MCDM research. There are a variety of techniques and numerous examples in the MCDM literature for helping DM to select among the trade-off solutions [8, 10]. Nevertheless, the fact that many MCDM research efforts only put focus on assisting the DM to choose among the solutions by performing analysis on the objectives (i.e. selecting **z***) have ignored two important facts: (1) before a decision is made, very often a DM wants to know if there are any patterns/rules that relate the decision variables (**x***) to the **z***; (2) the decision space is equally important in decision making as there may happen that several **x*** can lead to the same **z***. Fig. 2 illustrates the concept of finding pattern of solutions in PS in the decision space that lead to the *EF* in the objective space.

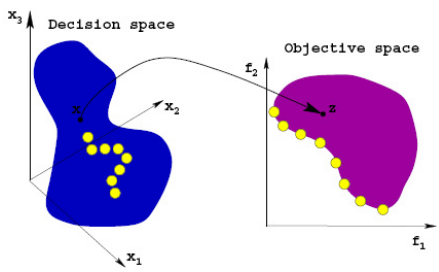

Fig. 2. Mapping between **x** in the decision space and **z** in objective space; also showing a pattern of **x*** that leads to the solutions on the efficient frontier (adapted from [11]).

The idea of deciphering knowledge, or knowledge discovery, by the post-optimality analysis of Pareto-optimal solutions from MOO was first proposed by Deb [11]. He coined the term innovization (innovation via optimization) to describe the task of discovering the salient common principles present in the Pareto-optimal solutions so that deeper knowledge/insights on the behaviour/nature of the problem can be gained. The innovization task employed in earlier publications involved the manual identification of the important relationships among decision variables and objectives that are common to the obtained trade-off solutions. Recent studies using data mining techniques so that innovization procedures can be performed automatically have been shown to be promising in various engineering design problems [12, 13].

The uniqueness of innovization, as has been shown in several engineering applications, is in using advanced data analysis to decipher salient properties from the optimization data generated, and not data that are already exist in a data source. Such concept of using advanced visualization and data mining techniques to products, engineering design and operation research applications has emerged to be an important research topic [14, 15]. As a matter of fact, by integrating the concept of innovization with simulation and data mining techniques, the innovization task can be used effectively for the analysis and decision-making support in the system design/development of industrial-scale production or supply-chain systems. Related studies that use MOO for production systems design can be found in [16] but more recent advance in combining advanced visualization and data mining into the optimization-based design process can be found in [17-20].

3. Knowledge Discovery in MOO

Data mining of MOO datasets can reveal interesting properties about the solutions and, in general, can help discover domain-specific knowledge [11, 13]. Based on the representation of the knowledge gained, data mining methods can be categorized as: (i) descriptive statistics, (ii) visual data mining, and (iii) machine learning methods. Descriptive statistics are simple univariate and bivariate measures that summarize the location, spread, distribution and correlation between the variables in the form of numbers, but the knowledge extracted is explicit, meaning that it is compact and can be easily stored, transferred or parsed programmatically. On the other hand, visual data mining methods convey knowledge visually through graphs, heatmaps, coloured clusters, etc., which are all implicit knowledge representation, meaning that they lack a formal notation and hence prone to subjectivity. Machine learning methods are capable of learning or discovering knowledge from data in both implicit and explicit forms. For example, support vector machines and neural networks generate implicit knowledge models, whereas decision trees give explicit rules.

3.1 Interactive Data Mining

While several data mining methods already exist for numerical data, most of them are not tailored to handle MOO datasets that come with inherent properties that distinguish them from ordinary datasets. Firstly, the presence of two different spaces, objective space and decision space, adds a certain degree of difficulty in discovering knowledge of relevance. Since DMs are often interested in the objective space, it would be beneficial to develop data mining methods that operate in the decision space, but at the same time, take the structure of solutions in the objective space into consideration. Furthermore, DMs usually express preferences to certain regions of the objective space. Data mining methods for MOO datasets should be able to take these preferences into account.

A knowledge-driven optimization-based (KDO) design process is therefore more than finding the *PS* that forms the *EF* for a MOP. It involves a knowledge extraction process that tries to explain, in some explicit knowledge representation format, the relationships between **x*** and **z*,** as illustrated in Fig. 3.

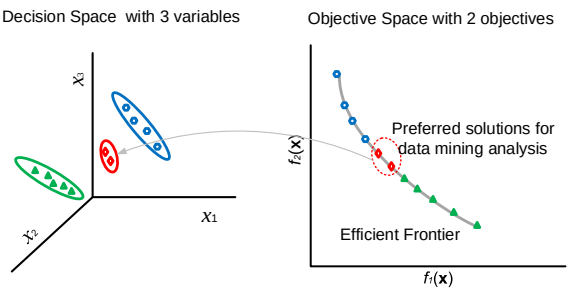

Fig. 3. Unlike innovization, interactive data mining in MOO allows the DM to select and analyze preferred solutions in **z** in the objective space to find the pattern of **x***. Also important to note is that FPM allows the analysis of any **z**, including solutions that are not on the EF.

Although most data mining techniques can be directly applied to process the optimization datasets, they do not respect the fact that optimization data involve two distinct spaces, objective space and decision space. For example, visual data mining methods either ignore one of the spaces or deal with them separately. This makes the involvement of a DM difficult if the mappings between the decision space and the objective space are interested, like Fig. 2. The distance-based regression tree learning approaches proposed in [19, 20] are the only methods that come close to achieving this. The shortage of such interactive data mining methods is the biggest hurdle in the analysis of MOO datasets. On the other hand, discreteness in optimization data also presents new challenges to most of the existing data mining methods. Unlike engineering design problems that usually work with continuous variables, production system design problems are most often consisted of discrete (integer) variables. There are many decision variables that are either inherently discrete, like buffer capacities or practically discrete, e.g. processing times, albeit theoretically are continuous variables, are forced to change in steps of seconds. The following deficiencies are observed when dealing with optimization data containing discrete variables:

- With visual data mining methods, discreteness may lead to apparent but non-existent correlations between the variables.
- Most distance measures used in both visual and non-visual methods are not applicable to ordinal and nominal variables. For instance, “Machine Option 1” and “Machine Option 2” is usually not quantifiable.
- Data mining methods that use clustering may also result in superficial structures.
- Association rules are a suitable form of representation for discrete, nominal variables. However, since association rule mining is unsupervised, it is difficult for a DM to get involved in the knowledge discovery process.

With the above discussions, it is argued that there is a need to develop some new, customized interactive data mining methods to tackle the specific challenges posed by the design of discrete production systems. In the next section, we introduce a new algorithm called Flexible Pattern Mining (FPM) which is an extended version of Sequential Pattern Mining (SPM).

3.2 Flexible Pattern Mining

The term sequential pattern was first used in the context of market basket data analysis [5]. An instance in such a dataset typically consists of the transaction IDs and the items bought together by a customer in each transaction. An item-set is a non-empty set of items and a sequence is an ordered list of item-sets. Each row in a transaction table is therefore a customer's complete sequence. A customer is said to support a sequence if it is contained within that customer's complete sequence. The goal of SPM is to generate an exhaustive list of all sequences that are supported by a predefined number of customers.

Table 1. An example of discrete optimization dataset. The wildcard * can take any integer between 1 and n, assuming some values for * in solution k.

Solution ID	Variable 1 (x_1)	Variable 2 (x_2)	Variable 3 (x_3)	Variable 4 (x_4)	...	Variable n (x_n)
Solution 1	d_{1*}	d_{2*}	d_{3*}	d_{4*}	...	d_{n*}
Solution 2	d_{1*}	d_{2*}	d_{3*}	d_{4*}	...	d_{n*}
Solution 3	d_{1*}	d_{2*}	d_{3*}	d_{4*}	...	d_{n*}
...	...	...	...	...	...	...
Solution k	d_{12}	d_{21}	d_{33}	d_{42}	...	d_{n2}
...	...	...	...	...	...	...
Solution $N-1$	d_{1*}	d_{2*}	d_{3*}	d_{4*}	...	d_{n*}
Solution N	d_{1*}	d_{2*}	d_{3*}	d_{4*}	...	d_{n*}

A typical discrete optimization dataset is shown in Table 1. The entry d_{ij} represents the *j*-th discrete option of the *i*-th variable. A SPM algorithm can be applied to such a dataset by considering each solution vector as a customer sequence, each variable as a separate transaction, and each discrete variable option as a different item.

The Apriori algorithm [21] is one of the earliest and most popular of sequential pattern mining techniques in literature. Given a dataset and a minimum support value, the algorithm first generates a list of all one-item-set sequences. Those sequences that meet the minimum support condition form the set L_1 which stands for frequent one-item-set sequences. Next, the algorithm makes use of the downward closure property or Apriori principle. This also means that frequent sequences with higher number of item-sets can be formed by combining smaller frequent sequences. The candidate generation step combines various one-item-set sequences to form candidate two-item-set sequences, of which those satisfying minimum support form a new set L_2. The process repeats till L_k is empty, i.e. no frequent *k*-item-set sequences can be found. A final step, called the maximal phase, prunes all non-maximal frequent sequences from $\bigcup_k L_k$. A sequence is maximal if it is not contained in any other sequence and the maximal sequences that remain after pruning are called sequential patterns.

Sequential pattern mining is extremely useful for finding exact patterns in the dataset. For example, given the simple one-dimensional dataset $x_1 = \{1,1,2,2,2,2,3,3,10,10,10\}^T$, it can detect that the patterns ($x_1 = 1$), ($x_1 = 2$) and ($x_1 = 10$) have supports 3, 4 and 3 respectively. However, it cannot detect the more important rule ($x_1 < 10$) which has a support of 9. To address this inflexibility of SPM, we propose FPM as an extension. Like SPM, FPM extends the concept of market basket data to MOO datasets by considering each solution as a customer and each variable as a transaction containing

individual items (variable values). However, to enable the generation of rules through the same Apriori algorithm, the optimization data in Table 1 is transformed into a truth table. The columns of this table correspond to each of the n variables being less than (<), equal to (=) and greater than (>) each of its discrete options. If n_1, n_1, … and n_n are the number of discrete options for x_1, x_2, : : : and x_n respectively, then the total number of columns in the truth matrix are $3(n_1 + n_2 + \ldots + n_n) - 2n$. The second term comes from the fact that no variable x_i can take a value less than d_{i1} or greater than d_{in_i}. The transformed truth vector for one of the solutions from Table 1 is shown in Table 2.

Table 2. The truth vector of solution k from Table 3, as required for FPM.

Solution ID	$x_1 = d_{11}$	$x_1 > d_{11}$	$x_1 < d_{12}$	$x_1 = d_{12}$	$x_1 > d_{12}$	$x_1 < d_{13}$	···	$x_1 < d_{1n_1}$	$x_1 = d_{1n_1}$	→
Solution k	0	1	0	1	0	1	···	1	0	
→	$x_2 = d_{21}$	$x_2 > d_{21}$	$x_2 < d_{22}$	$x_2 = d_{22}$	$x_2 > d_{22}$	$x_2 < d_{23}$	···	$x_2 < d_{2n_2}$	$x_2 = d_{2n_2}$	→
Solution k	1	0	1	0	0	1	···	1	0	
→	$x_3 = d_{31}$	$x_3 > d_{31}$	$x_3 < d_{32}$	$x_3 = d_{32}$	$x_3 > d_{32}$	$x_3 < d_{33}$	···	$x_3 < d_{3n_3}$	$x_3 = d_{3n_3}$	→
Solution k	0	0	0	0	1	1	···	1	0	
···	···	···	···	···	···	···	···	···	···	
→	$x_n = d_{n1}$	$x_n > d_{n1}$	$x_n < d_{n2}$	$x_n = d_{n2}$	$x_n > d_{n2}$	$x_n < d_{n3}$	···	$x_n < d_{nn_n}$	$x_n = d_{nn_n}$	
Solution k	0	1	0	1	0	1	···	1	0	

FPM combines the merits of SPM, classification trees and hybrid learning. Firstly, like SPM it is unsupervised and can be applied to the whole dataset in the absence of user preferences. Secondly, like classification trees, it generates rules which are much more flexible than exact patterns. Thirdly, like hybrid learning, it does not require the user to provide class labels. Sequential pattern mining is usually applied to the whole dataset. However, often the decision-maker is interested in a preferred region of the objective space. DMs sometimes have vague preferences which makes it difficult to provide a single reference point. A more natural and interactive approach is to allow the user to "brush" over solutions using a mouse input to select a region of interest in the objective space. The modified Apriori algorithm is applied only to the selected set of solutions, but the patterns discovered in the process can be visualized over all solutions. In case the user chooses to express preference by brushing over solutions, an additional processing can be performed on the rules. By sorting the rules according to the ratio of support in the selected solutions to that in the rest of the solutions, a unique ordering of the rules may be obtained. The topmost rule in such an ordering then represents the most distinguishing property of the selected region. We illustrate this through a real-world application example in Section 4.

4. Real-world Production Line Design

This section reveals how the proposed approach has been used to help an automotive manufacturer in deciding a future production line design, instead of their current practice of using static Excel spreadsheet to estimate the number of machines and buffers that are needed for every production stage, which does not consider the dynamics of system as a whole. In this design scenario, there are many different parameters that affect the decisions in earlier phases, such as quality (Q), cost, space, flexibility, competence/complexity (C) and automation level. Regarding quality, some finer machining may need a certain type of specialized machines which are inflexible, but able to keep higher process quality. When considering only flexibility, multi-purpose machines are in general better option as they can perform several types of operations and sometimes more space-efficient. The automation level required at each machine will determine whether there is an option between a fully automated material handling such as gantry robots or other types of lower-cost manual handling. Automation level is also associated with the labour cost (L), since L in some countries is lower and a manual solution may be the appropriate choice. Furthermore, the competence required for some workstations (or stations hereafter) may not be possible if there is a shortage of well-educated personnel. Fig. 4 shows the complete simulation model in FACTS Analyzer [4], which is composed of about 22 operation stages. Fig. 5 shows one stage in more details. The operation stages have one to three different options regarding the types of stations to be chosen. However, there are also various choices among automation level, buffer sizes and the number of inspection stations at different operation stages. An example of an operation stage is shown in Fig.5, showing the buffers "B1" and "B2" have changeable buffer sizes, mainly for the transportation and to absorb the variability between the operation stages.

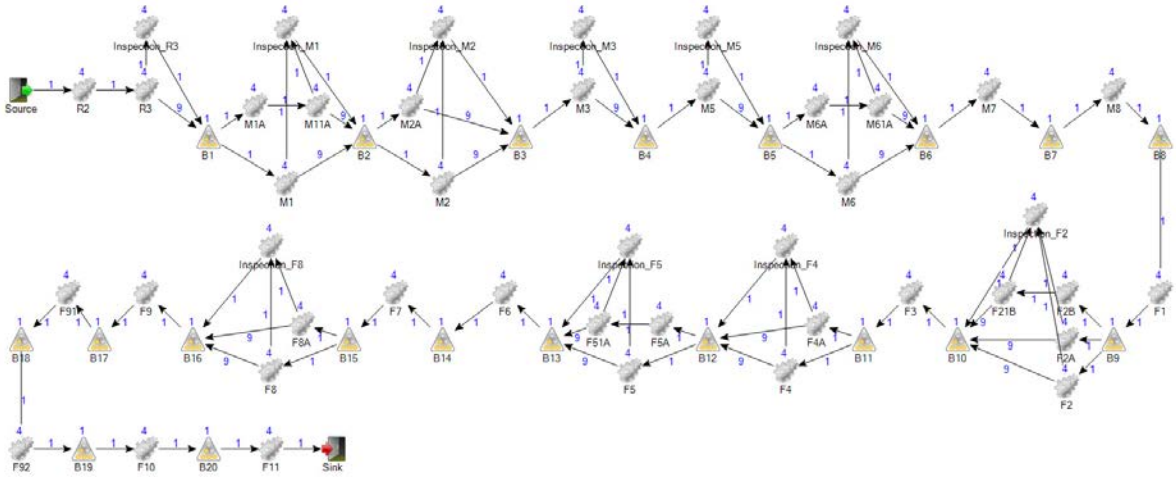

Fig. 4. The complete model in FACTS Analyzer for conceptual design of an automotive machining line.

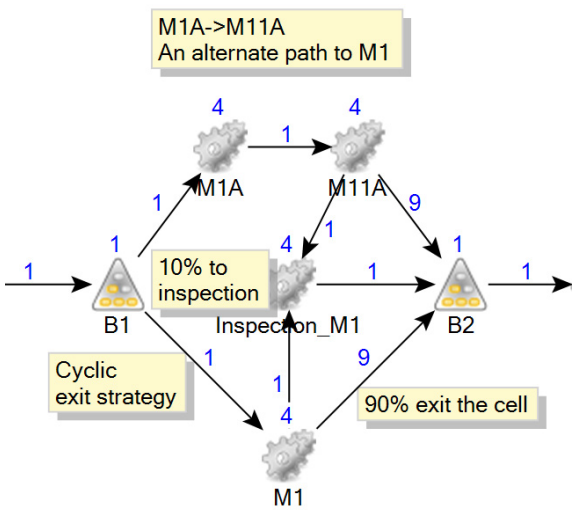

Fig. 5. An operation stage in the conceptual design showing alternative options: one specialized station M1 or two standardized stations (M1A and M11A). The numbers on the arrows can have different purposes: with a cyclic exit strategy the outputs from the buffers are randomly pass to the first non-blocked station. The arrows with label 9 and 1 represent 90% output to the output buffer and 10% to inspection.

In Fig.5, there are two main options: (1) two specialized stations named M1A and M11A and (2) one standard/flexible station, M1. The standard station is cheaper to buy, but the specialized stations are faster and therefore there is a trade-off between the options. Furthermore, all the stations can be consisted of parallelized operations. The company requested that a combination of these two options can exist so that the decision variables are the number of parallel machines in each station with M1A and M11A controlled by an inequality constraint in the form of: M1+M1A×M11A ≥ 1. Every inspection (quality control) station in a process stage is also possible to be parallelized if needed making some additional decision variables to the optimization problem. The optimization has been formulated with *TH*, *INV* (total investment) and *WIP* as a 3-objective optimization problem:

$$\mathbf{F}(\mathbf{x}) = \begin{cases} Max.\, f_{TH}(\mathbf{x}) \\ Min.\, f_{INV}(\mathbf{x}) \\ Min.\, f_{WIP}(\mathbf{x}) \end{cases}$$

Fig. 6 and 7 illustrate two different ways to "brush over" the Pareto-optimal objective vectors in the *INV-TH* plot in order to generate the rules presented in Table 3 by highlighting the selected solutions with various colour. The optimization was run using the well-known NSGA-II algorithm as in [11,12], but any other efficient MOO algorithms can also be used. Each colour represents one set of **z*** that is selected by the DM as inputs to FPM. While the DM in Fig. 6 is interested on what form the 6 separate clusters in the objective space, Fig. 7 represents a more realistic decision-making activities to analyse the solutions that can be used for the final design, with respect to what TH level is desired.

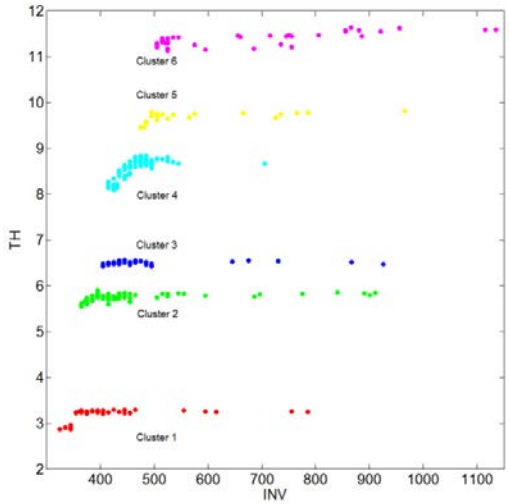

Fig. 6. EF with respect to the 3 objectives; highlighted with the solutions "brushed" for different FPM runs.

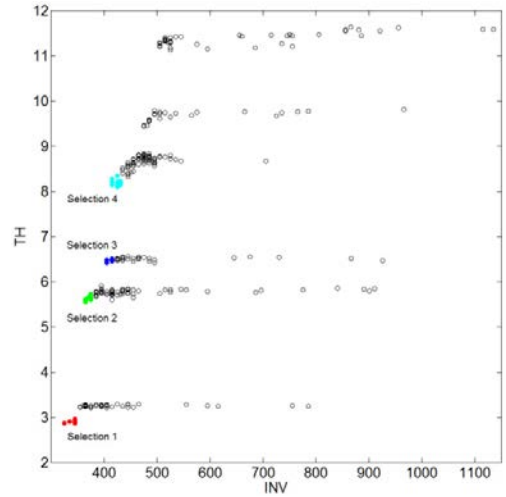

Fig. 7. Preferred solutions "brushed" for different FPM runs for deeper analysis of their properties in the decision space.

The FPM results, showing composite, topmost rules (significance > 80% and ratio of selected and unselected solutions that match the rules > 60%) of the selected solutions from Fig. 6 and 7 are shown in Table 3. When checking the rules in the table, the DM has learnt which are the most important stations that primarily determine the performance of the whole line (note that R3, M3, M6 and M5 repeatedly appear in the topmost significant rules). When looking at the rules that separate Selection 1 to 4, it can be easily see the combination of R3 and M3 mostly determine the TH level of the line. In terms of the combination of specialized station and flexible ones. The rules have shown some very important knowledge – specialized machine (M1A & M11A) are more optimal for low volume production (TH) whereas standardized/flexible machine (M1) is favourable when higher TH is desired. This finding is somewhat counter-intuitive to the DM and impossible to be found by any current industrial design practice.

Table 3. Rules generated by using FPM on the highlighted MOO data.

Selected **z**	Rules on decision variables (x_i)	Other Indexes
Cluster 1	R3=1 ∩ M6<3 ∩ M3<3 ∩ M5<3 ∩ B7=1 ∩ M1=0 ∩ B13>1 ∩ F7=1 ∩ M7=1 ∩ M1A=1 ∩ M11A=1	Q<113 L <37 C<111
Cluster 2	R3=2 ∩ M6=3 ∩ M1=0 ∩ F7=1 ∩ M1A=1 ∩ M11A=1 ∩ M7=1	Q <125 L <41 C <123
Cluster 3	M5=3 ∩ M3=3 ∩ R3=2 ∩ M6=3 ∩ M1>0 ∩ B13=1 ∩ B7>1 ∩ M1A>0 ∩ M11A>0	Q>122 L>40 C>120
Cluster 4	M1=3 ∩ R3=3 ∩ F7=2 ∩ M1A=0 ∩ M11A=0 ∩ M7=2 ∩ M5>2 ∩ B13=1 ∩ M3>2 ∩ B7>1 ∩ M6>2	Q>128 L>42 C>126
Cluster 5	M1>3 ∩ R3=3 ∩ M3>3 ∩ M5>3 ∩ B19>1 ∩ M6>3 ∩ F9>1 ∩ M1A=0 ∩ M11A=0 ∩ M7=2 ∩ B9=1 ∩ F7>1	Q>143 L>47 C>141
Cluster 6	R3>3 ∩ F5=2 ∩ M1>3 ∩ F3>1 ∩ F9>1 ∩ M3>3 ∩ M5>3 ∩ M6>3 ∩ M1A=0 ∩ M11A=0 ∩ B19>1 ∩ M7>1 ∩ B9=1 ∩ F7>1	Q>155 L>51 C>153
Selection 1	M3=1 ∩ M5=1 ∩ R3=1 ∩ M6 <3 ∩ B7=1 ∩ B13>1 ∩ M1=0 ∩ F7=1 ∩ M7=1 ∩ M1A=1 ∩ M11A=1	Q <110 L <36 C <108
Selection 2	M5=2 ∩ M3=2 ∩ R3=2 ∩ M1=0 ∩ F7=1 ∩ M7=1 ∩ M1A=1 ∩ M11A=1	Q=116 L=38 C=114
Selection 3	M1=1 ∩ M5=3 ∩ M3=3 ∩ R3=2 ∩ M6=3 ∩ B13=1 ∩ F7=1 ∩ M7=1	Q=125 L=41 C=123
Selection 4	M1=1 ∩ R3=3 ∩ M5=3 ∩ M3=3 ∩ M6=3 ∩ F7=2 ∩ B13=1 ∩ M7=1	Q=131 L=43 C=129

5. Conclusions & Future Work

This paper introduced a "knowledge-driven" production systems design approach that synergistically combines simulation, MOO and data mining so that a designer can understand better the relationships among the decision variables and design objectives. Although apparently this approach could be implemented with different MOO and data mining algorithms, this paper addressed the specific challenges posed by a large amount of discrete design variables so that a new data mining called FPM, based on SPM commonly used in market basket analysis, has been developed. A real-world production line sizing and selection problem found in an automotive manufacturer has been used to illustrate how useful, explicit knowledge has been obtained from the optimization and the post-optimality processing to support more confident decision making. It is believed that the same approach and algorithm, FPM, can also be applied effectively to other problem domains that involve a large set of discrete or mixed variables, like complex engineering systems design. Our current work is focused on developing a FPM-embedded, "knowledge-driven" optimization algorithm that can use the knowledge gained during the optimization to automatically enhance the performance of the algorithm in an on-line basis.

Acknowledgements

This work is partially financed by Knowledge Foundation (KKS), Sweden, through the IDSS project and AB Volvo through their Academic Preferred Partnership in Virtual Manufacturing. The authors gratefully acknowledge their provision of the research funding.

References

[1] Hopp WJ and Spearman ML. Factory physics: foundations of manufacturing management, 2nd ed., Irwin McGraw-Hill Higher Education, Burr Ridge, IL; 2000.

[2] De Vin LJ, Ng AHC, Oscarsson J. Simulation based decision support for manufacturing system life cycle management. Journal of Advanced Manufacturing Systems, Vol.3, No.2; 2004; 115–128.

[3] Urenda M, Ng AHC, Svensson J. Simplification and aggregation strategies applied for factory analysis in conceptual phase using simulation. In Proceedings of the 40th Winter Simulation Conference; 2008; 1913-1921.

[4] Ng AHC, Bernedixen J, Urenda M, Jägstam M. Factory flow design and analysis using simulation-based multi-objective optimization and automatic model generation. In Proceedings of the Winter Simulation Conference 2011, 11-14 Dec. Phoenix, Arizona; 2181-2193.

[5] Hollocks BW. The impact of simulation in manufacturing decision making. Control Engineering Practice, 3(1); 1995; 105-112.

[6] Ólafsson S, Kim J. Simulation optimization: Exploring New Frontiers. In Proceedings of the 34th Winter Simulation Conference; San Diego, California; 2002; 79-84.

[7] Ng AHC, Svensson J, Urenda M. Multi-Objective Simulation Optimization for Production Systems Design using FACTS Analyser. In Proceedings of the 2'nd Swedish Production Symposium (SPS'08), Stockholm, Sweden, 2-3 November 2008.

[8] Miettinen K. Nonlinear multiobjective optimization. Kluwer Academic Publishers 1999; p. 11.

[9] Simon D. Evolutionary Optimization Algorithms: Biologically Inpsired and Population-Based Approaches to Computer Intelligence. New Jersey: John Wiley & Sons, Inc; 2013; 557-558.

[10] Malakooti B. Operations and Production Systems with Multiple Objectives. New Jersey: Wiley & Sons, Inc. 2014.

[11] Deb K, Srinivasan A. Innovization: Innovating design principles through optimization. In Proceedings of the 8th Annual Conference on Genetic and Evolutionary Computation, GECCO 2006, ACM; 2006. pp. 1629-1636.

[12] Bandaru S, Deb K. Towards automating the discovery of certain innovative design principles through a clustering-based optimization technique, Engineering Optimization 43 (9); 2011; 911-941.

[13] Deb K, Bandaru S, Greiner D, Gaspar-Cunha A, Tutum CC. An integrated approach to automated innovization for discovering useful design principles: Case studies from engineering, Applied Soft Computing 15; 2014; 42-56.

[14] Graening L, Sendhoff B. Shape mining: A holistic data mining approach for engineering design. Advanced Engineering Informatics 2014; 28(2): 166-185.

[15] Meisel S, Mattfeld D. Synergies of Operations Research and Data Mining. European Journal of Operational Research 2010; 206(1): 1-10.

[16] Feyzioğlu O, Pierreval H, Deflandre D. A simulation-based optimization approach to size manufacturing systems. International Journal of Production Research 2005, 43(2); 247–266.

[17] Dudas C, Frantzén M, Ng AHC. A synergy of multi-objective optimization and data mining for the analysis of a flexible flow shop. Robotics and Computer-Integrated Manufacturing 2011; 27 (4); 687-695.

[18] Ng AHC, Dudas C, Deb, K. (2011). Simulation-Based Innovization using Data Mining for Production Systems Analysis. In Evolutionary Multi-objective Optimization in Product Design and Manufacturing, L. Wang, A. Ng, K. Deb (eds.), Springer, 401-430.

[19] Dudas C, Ng AHC, Pehrsson L, Boström H. Integration of data mining and multi-objective optimisation for decision support in production systems development, International Journal of Computer Integrated Manufacturing 2014; 27 (9); 824-839.

[20] Dudas C, Ng AHC, Boström H. (2015) Post-optimality analysis of multi-objective optimization data using decision trees. Journal of Intelligent Data Analysis 2015; 19(2); 259-278.

[21] Agrawal R, Srikant R. Fast algorithms for mining association rules, in: Proceedings of the 20th International Conference on Very Large DataBases, VLDB, 1994; 487-499.