INTERNATIONAL DESIGN CONFERENCE - DESIGN 2006
Dubrovnik - Croatia, May 15 - 18, 2006.

DESIGN 2006

PRODUCT AND PROCESS IMPROVEMENTS BASED ON DATA MINING OF SHOP-FLOOR INFORMATION

T. Wuenscher, D. G. Feldmann and D. Krause

Keywords: customised and complex products, product development, assembly, disturbance, online analytical processing, data mining, continuous improvement

1. Introduction

Companies are permanently forced to reduce costs as well as the time to customer and simultaneously assure the expected quality. To do so, improvement potential has to be exploited and disturbances in the production process have to be avoided.

However, as we know from our research projects in the aircraft and ship-building industry, in some areas the improvement potential is unexploited and often similar disturbances occur. Pfeifer describes that 60% of the failures in manufacturing are known or at least similar [Pfeifer 2001]. We discovered that this is partly due to the shop-floor information, including information about disturbances, customer claims and suggestions of improvement, not being used systematically.

In this paper, the reasons for not using shop-floor information effectively and the drawbacks of existing approaches are explained. Following this, we give an overview about our basic concept, which is to systematically collect shop-floor information, consolidate it and make it useable for corrective actions as well as during product customisation. In the third section, a selection of the underlying methods is presented in more detail.

2. State of the Art of Using Shop-floor Information

During our research projects, we analyzed the reasons for the suboptimal usage of shop-floor information by structured interviews and workshops with employees from different departments. What has always been primarily mentioned is that the following boundary conditions of the named industries impede or at least aggravate a systematic use of shop-floor information:

- High product- and process complexity due to the variety of systems, huge product dimensions and the number of parts and components
- Highly customised, single- and small batch products with many technically complex requirements in combination with late changes by the customer, even after start of production

Based on these conditions, three main problems regarding the learning from shop-floor information have been derived. The description focuses on disturbances, but because of their similarities it is valid for customer claims and suggestions of improvements as well.

The first reason is the missing possibility to consolidate information about similar disturbances. Changing a part or component will eliminate a specific cause for an individual disturbance. But, in the case of low-volume assembly and a highly customised product, only a few parts or components are identical. Therefore an analysis regarding typical disturbances, which is based on part or component numbers, is not effective. Similar disturbances can only be avoided if the systematic causes of the disturbances are known.

The second reason is that only a very rough analysis of the disturbances can be done automatically, because the disturbances are described in text form without any restrictions regarding the usable terms. In addition, disturbances are being registered by workers who are not being specifically qualified for an accurate description of disturbances and workers get a time credit for registering disturbances; this leads to a trend to even register minor disturbances. It is aggravating in this context that the effects of the disturbances are not being quantified. Furthermore it is difficult to get a comprehensive insight into the production because the information about the disturbances, customer claims or suggestions of improvements are handled in separate tools.
The third reason is the misleading calculation of potential savings. Generally, it is worth eliminating the root causes of a disturbance if the costs for recurring disturbances offset the costs for changing the product definition. Due to extended requirements regarding certification (especially in the aircraft industry) and documentation, the time and effort for a product is high. In contrast, the potential savings are quite low in the low-volume assembly if only specific components are being considered and not a class of components.
Computer Aided Quality Assurance Systems (CAQ-Systems) should contribute to the solution of these problems, because they specialise on the processing of quality- and disturbance information. An evaluation of commercial CAQ-Systems has revealed that the analysis of the information is basically limited to control charts (statistical process control), failure frequency or Pareto charts. The generated indicators sometimes have less informative value and cannot drill-down for detailed analysis of problems [Nyendick 2005]. A more sophisticated approach in terms of analysing methods has been developed by Schrems, but it is specialised on the high-volume manufacturing and considers quantifiable measures like the deviation of the nominal diameter of machined parts [Schrems 2001]. All in all, these approaches focus on quantitative inspection features of the manufacturing domain. Hence the analyzed tools and methods do not address the needs of companies with a low-volume assembly line. Data which can be gathered in the assembly is mainly categorical and cannot be collected automatically - especially in manual assembly.
Other approaches are based on the FMEA method. Edler, for example, presents a concept where a Design-FMEA is being augmented with data from the use of the product, in order to get help on the development of the causal interrelations and by the objective risk assessment. Although this is a promising approach, it seems non-applicable for the use in the customer-specific variant design, because of the big effort to do a FMEA. Kmenta et. al. developed a so-called mixed-model FMEA, whereby defects in the assembly should be reduced by asking questions “based on historical assembly, service or warranty problems” [Kmenta, Cheldelin et al. 2003]. This is a promising approach even though the questions that are being asked are less detailed and no immediate actions are being taken if serious problems have been detected.

3. Solution for a better use of shop-floor information

What can be concluded from the named situation in the industrial companies regarding the use of shop-floor information? Basically, two aspects have to be improved by developing and implementing new methods and tools. First of all, taking the complexity of the product and process into account, a better knowledge about the situation in the assembly is needed. Revealing hidden relationships, getting more detailed and quantified information about typical problems of past orders and enabling a company wide and easy access to this information are corresponding challenges. The second challenge is to make sure that, based on this knowledge, the correct consequences and actions are being taken.

3.1 Basic Concept

Our basic concept is to systematically collect the data about disturbances and customer claims as well as suggestions of improvements (Figure 1). This data, which is stored in separated operative databases, has to be consolidated and rated, which is being done in a data warehouse. To rate the collected information, it is necessary for the effects of a disturbance or customer claim to be quantified. With the aid of so-called business intelligence methods, serious problems are detectable and order-specific information can be found. Based on this information, it is far easier for the employees to analyse the root causes of these special problems and initiate either order independent, immediate actions or rather preventive actions during handling of new customer orders.

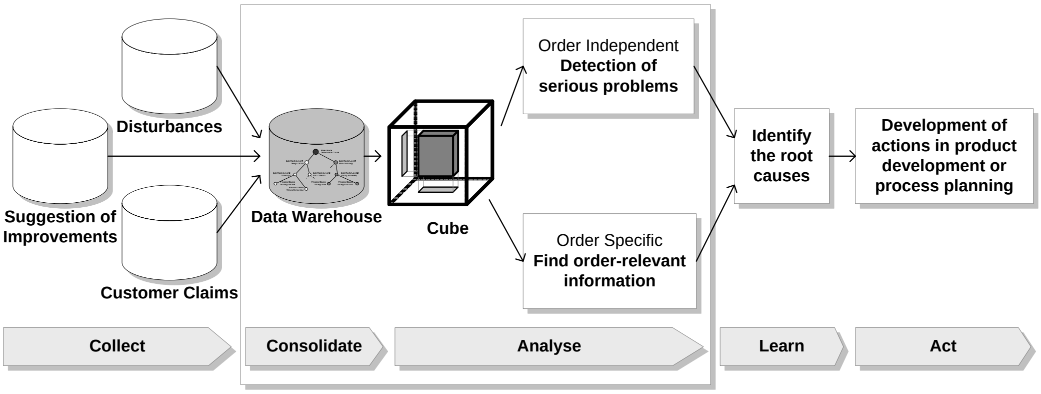

Figure 1. Basic Concep

Table 1 shows a morphology of actions that gives an overview about the broad range of these actions. For instance, based on the information, not only product changes can be helpful to prevent disturbances but also adapting capacity planning in the design office can be a solution.

Table 1. Morphology of Actions

Attribute	Characteristic				
Application area	Actual product		Product variant	Product generation	
Department	Engineering	Process-planning	Manufacturing	Assembly	Service
Changed item	Product	Process	Combination of product and process	Capacity	Time frame
Response time	Preventive			Reactive	
Effectiveness	Immediately			Delayed	
Duration	Temporal			Permanent	

In this paper we focus on how to consolidate and analyse the shop-floor information to derive more effective actions. Refer to our previous work for a more detailed description of the whole concept and the underlying processes [Wünscher and Feldmann 2005].

3.2 Multidimensional, Hierarchical Data Model and Data Warehouse

After the basic concept for the aircraft and ship-building industry has been generated, we have collected the requirements on the data model, developed and verified it.
But before going into detail, it will be explained what we subsume into the term shop-floor information. Generally it is information about disturbances, customer claims and suggestions of improvement. For us disturbances are every kind of deviation from the planned manufacturing- and assembly activities. A typical example for a disturbance in the aircraft industry is that a worker wants to fix a wiring, but he is not able to do so because a bracket obstructs the access. The causes of disturbances are wide ranging, e.g. a suboptimal assembly process, a wrong design or part distribution. Further causes for disturbances are omission, incorrect installation or installing the wrong part [Cheldelin and Ishii 2004]. The root cause for the mentioned example could be that during development of the wiring its surrounding area has not being taken into account carefully.
A speciality in the aircraft and ship-building industry is that the customer inspects major subcomponents of the final product before delivery. The resulting customer claims have similar or equal attributes than disturbances, e.g. *location*, *product* or *customer order*.

Even suggestions of improvements from the shop floor worker have similar attributes; whereas disturbances cause actual costs, not implemented suggestions of improvement have imputed costs.
To derive more effective actions, this information has to be detailed, precise and quantified. In addition, the information should be available throughout the company and on every hierarchy level. Because of the particular circumstances of low-volume assembly, the main challenge is to consolidate the information about similar disturbances and so on. However, not only consolidated information but also information of an individual disturbance is needed during the learning process.
In order to meet these requirements, the data model has to be multidimensional and hierarchical. In this model, the data can be viewed as a n-dimensional data cube. A data cube consists of facts, dimensions and hierarchies. Facts, like *disturbance costs*, are numerical values, which are in the focus when consolidating the information. Unlike the facts, the dimensions are descriptive and allow different views of the facts. If the dimensions have a vertical relationship of the single elements, they are called concept hierarchies.
In our research projects we have initially generated this data model within several workshops with industrial experts from the product development, process planning and manufacturing domain. Following that, we filled the concept hierarchies based on an analysis of a huge amount of historical disturbance descriptions. The outcome is, for example, that the number of disturbance categories, which are needed to categorise all disturbances are now less than one quarter of the number of the analysed disturbances.

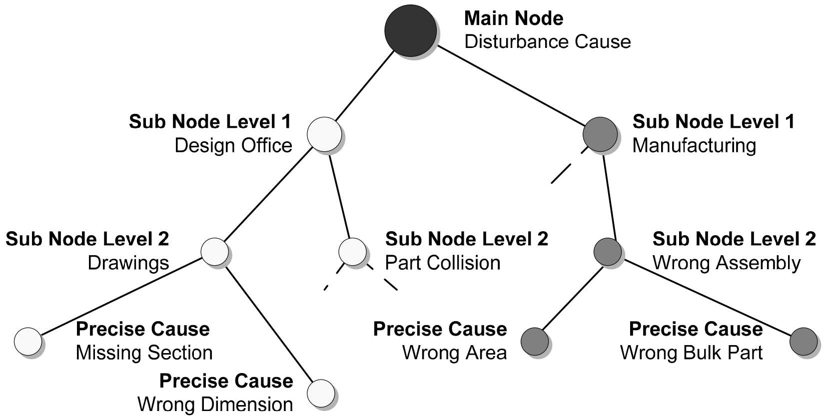

Figure 2. Concept Hierarchy *disturbance cause*

Finally we verified the data model and the concept hierarchies based on twice as much actual disturbance descriptions as of the first phase. After some minor revisions, we are convinced that our approach is appropriate to categorize all disturbances. Our concept hierarchies are sorted according to their process, but many others are also possible, e.g. outcome-based classification of Hinckley [Hinckley 2001]. However, we have learned during this phase that the most effort is needed to define an unambiguous, easy to understand concept hierarchy with the right level of detail. But, nevertheless, we are convinced that this categorisation is the best possibility to consolidate similar disturbances, customer claims or suggestions of improvements. In addition, following our positive experiences in the aircraft industry, we want to widen our scope and start a research project in the special machine industry.
In our approach, a data warehouse is being deployed to store the data. Characteristic for a data warehouse is the separation from the operative data sources, so that information can be consolidated and in addition starting an analysis on the data does not influence the performance of the operative data sources. Furthermore, historical data is being saved in the data warehouse, which allows analysing trends. Based on the stored information, there are numerous methods available to analyse the data. Basically it can be distinguished between methods that are hypotheses based and hypotheses free, whereas a hypotheses is a user-query on the data. Hence in the first case, the user has an

impression what he is looking for in the database and can define an appropriate query. In the second case, the user is not defining queries and instead a tool is almost autonomous uncovering unknown pattern and relationships, which the user has to assess regarding correctness and plausibility.

3.3 Online Analytical Processing (OLAP) Methods

The described multidimensional data model usually contains a lot of data. But if using this information company wide, it is necessary for many users to be able to intuitively analyse the data with short response times. This is the function which can be done by Online Analytical Processing methods.
There are many operations available: For filtering the data in a data cube, the operations of slice and dice have been developed. As can be seen in Figure 3, the slice operator generates a two dimensional table by keeping one value fixed. The more general case is to define an upper and lower threshold for each dimension; the result is a sub-cube.

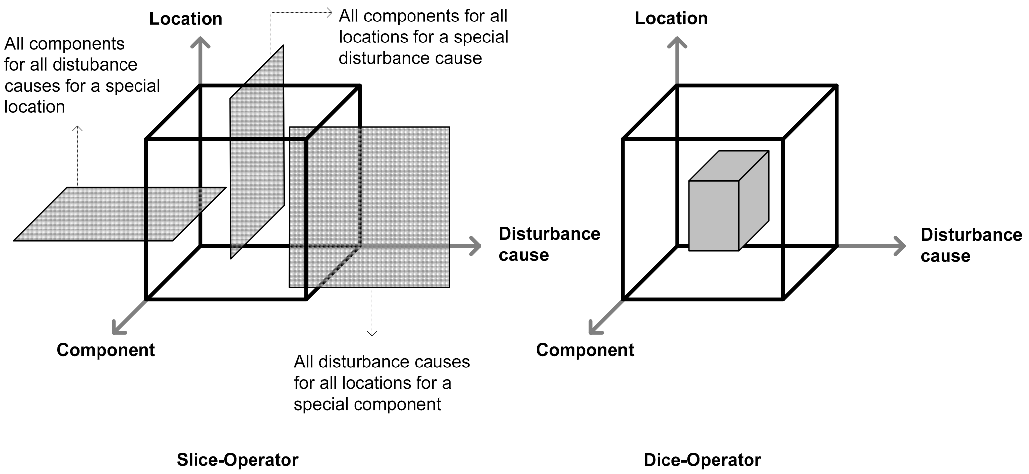

Figure 3. Slice and Dice Operations; based on [Kemper, Mehanna et al. 2004]

Roll-up and drill-down are operations to navigate within concept hierarchies (Table 2). The role of roll-up is to reduce the level of detail of a dimension by consolidating the facts of a hierarchy; drill down does the opposite. Table 2 exemplifies the roll-up and drill-down operator.

Table 2. Roll-up and Drill-down Operations on a *disturbance costs* by *production process* Table

Process Dimension	Product A	Product B	Product C	Product D
Production Process	1400 €	1000 €	2000 €	1200 €
	Roll-up ⇧		⇩ Drill-down	
Process Dimension	**Product A**	**Product B**	**Product C**	**Product D**
Part Manufacturing	400 €	350 €	700 €	400 €
Structural Assembly	450 €	350 €	600 €	350 €
Component Assembly	550 €	300 €	700 €	450 €

In our concept, these OLAP operations will be rather used for order specific, preventive actions within the product development and production department. How this will be done is being shown by the following sample queries: *Which are the most critical customer claims of a specific customer during his last project?* This is of interest when handling a new customer project to create preventive actions. *Which customer requirements caused the main disturbances in the assembly?* Knowing this will help to prevent disturbances when a new customer has the same requirement like a known customer. *Which*

disturbances have been caused by a specific group of components at a specific location in the product? Typically a designer is responsible for a specific group of components and a specific location. Giving the possibility to analyse what kind of disturbances are being caused by this area will help a designer to prevent failures when adjusting a related component.
In our research projects in the aircraft industry, we have validated this method by integrating sample data into a software prototype. The discussions with the domain experts have shown that an OLAP based analysis is an important cornerstone to get a better insight into the typical assembly-problems.

3.4 Data Mining Methods

The major drawback of Online Analytical Processing is that the aggregation of data is only possible for numerical data and that the user has to decide which dimensions and measures should be analysed. Data Mining in contrast is a rather automatic approach of analyzing the data, which is being used to uncover hidden patterns in large databases. Basic data mining functions are classification, prediction, clustering and association analysis. One challenge within the analysis with the industrial background is to characterize classes of disturbances, customer claims and suggestions of improvements. In this section it will be described which data mining methods can support this analysis efficiently.
Deviation-Based Outlier Detection, especially OLAP Data Cube Technique
An Outlier is a data object, which is inconsistent or dramatically different from the rest of the data space. There are many methods available to detect outliers. One method is an extension of the OLAP method whereby, with the aid of pre-computed measures, serious discrepancies are identified and highlighted on every level of detail of the data cube [Han and Kamber 2001].
The main advantage is that the user will be guided through the analysis and will be able to detect relevant data with less time and effort. Adjusted to the computation of shop-floor information a typical analysis will be executed as follows. First of all the user has to decide which dimensions, such as *disturbance cause* or *location*, shall be taken into account and which measures, such as *disturbance costs*, he is interested in. The system will display the result as a table where the outliers are highlighted, so that he can interactively detail the information by a drill-down operation on this dimension.
Data Generalization
The stored information in a database is typically very detailed and there are numerous data sets. To get a better insight into the stored data, it is necessary to consolidate the data. One method to do so is the Attribute-Oriented Induction (AOI). The first step of the AOI process is to compile the data that is relevant for an analysis (task-relevant data). After that, a consolidation of the data is being realized by removing or generalising attributes of the table. To take an example, imagine that the analyst wants to know which typical problems are being caused by the orders of a special customer.

Table 3. Source Table

ID_number	Type	Component	Costs	Source_of_incident
5444566	Part has been damaged	Air Condition	1200	Customer claim
3214565	Scratch on the surface	Primary structure	950	Disturbance
3545568	Scratch on the surface	Primary structure	1530	Customer claim
…	…	…	…	…

Table 4. Consolidated Table after Removing and Generalisation of Attributes

Type	Component	Costs
Scratch on the surface	Primary structure	2480
Part has been damaged	Air Condition	1200

Then the first step is to generate a table with disturbances and claims of that special customer, which is may be filtered for a special order or for a fixed period of time (Table 3). In this example, data sets can be consolidated by removing the attribute *source of incident* because it does not matter who has detected a problem. In addition, the *id_number* is not appropriate to distinguish the different data sets

(Table 4). After generating the consolidated table, the information can be represented by table bar charts, pie charts, 3D cube views or even rules.
The main problem of the AOI is that the user has to guess which attributes have the main impact on the consolidation of data sets. Whereas this will be possible for small tables, it will be almost impossible for tables with many data sets, many attributes or many levels of abstraction within concept hierarchies. To automate the manual task, the analytical characterization methods have been developed. In this approach, the relevance of an attribute is being valued based on its probability to be useful to distinguish the class from others [Han and Kamber 2001].

Association Rule Mining

Association rules show which attribute-value conditions in a data set are frequently together. The best known example is the market basket analysis, where it is of interest which items are frequently bought together by the customer. This method can be used to find an answer to the question: *Which components are frequently involved together in disturbances?* This information can be used to figure out which departments have to improve their cooperation. An example for an association rule is which component classes are involved in a disturbance:

component (D, “brackets”,”primary structure”) ⇒ component (D, “air condition”)

[support = 33 %, confidence = 67 %]

The two criteria support and confidence express the usefulness and certainty of the discovered rules. The support of a rule shows how often the considered components appear in the database in relation to total number of data sets. Confidence is a criteria that relates the support of left sided items within the rule to the support of all components. In the example (Table 5) that means: if brackets and primary structure are involved in a disturbance, in 67% of the incidents air-condition is also involved.

Table 5. Example of an Association Rule with Calculated Support and Confidence

Disturbance	Involved components
1	electrical systems, brackets, air condition, primary structure
2	brackets, air condition, primary structure
3	electrical systems, hydraulic, brackets, air condition
4	air condition, primary structure
5	electrical systems, primary structure
6	electrical systems

Support of X = {brackets, air condition}: 3 of 6 = 50 %
Support of R = {brackets, primary structure, air condition}: 2 of 6 = 33%

Confidence of “brackets, primary structure ⇒ air condition” = 2 of 3 = 67% [=Support (R)/Support (X)]

Typically, many rules are discovered during an analysis. So that for the support and the confidence a general threshold can be defined to reduce the effort for analysing the plausibility of mined rules.

4. Summary and Outlook

Because of the high customisation and product- and process complexity in the aircraft, ship-building or special machine industry, disturbances in the assembly are common. However, to win through on a competitive market these companies need robust and lean processes.
In this paper, we have presented a data mining based approach to discover improvement areas especially of the assembly. Basis for the analysis are the shop-floor information which are being created anyway. Based on this information, OLAP Operations, Deviation-Based Outlier Detection, Attribute-Oriented Induction and Analytical Characterization as well as Association Analysis are being deployed in our approach.
Although the examples that have been shown are quite simple, sophisticated tools are available to implement these methods in an industrial application area. Based on these tools, a software-prototype

is being developed that will be used to process real information and to demonstrate the usefulness of these methods.
It has to be emphasized that the systematic collection, analysis and provision of consolidated information cannot automatically achieve an improvement. Spear mentions that one of the success factors of the Toyota Production System is the extensive, direct observation of the assembly [Spear 2004]. Our approach does not contradict this; with our approach we even support this idea by giving hints where to look in more detail.
In addition, to derive effective corrective actions, the related processes and organisational circumstances have to be considered as well as the methodical support and motivation of the people involved.
Finally, it is important to point out that the presented approach tends to set the focus onto the origin of the disturbance. Unfortunately, this cannot be avoided if an individual disturbance should be reduced significantly. This is a dilemma because to expose the originator of problems will reduce his willingness to find a creative and common solution. To overcome this dilemma is a basic management task, where motivating people and offering the right incentives is essential. To analyse these soft aspects in more detail and to identify possible solutions will be one of our upcoming research activities.

References

Cheldelin, B. and K. Ishii, "Mixed Model Assembly Quality: An Approach to prevent human errors", 2004 ASME International Mechanical Engineering Congress & Exposition, Anaheim, California USA, November 13 - 19, 2004.
Han, J. and M. Kamber, "Data mining: concepts and techniques", Academic Press, 2001.
Hinckley, C. M., "Make No Mistake", Productivity Press, Portland, OR, 2001.
Kemper, H.-G., W. Mehanna, et al., "Business Intelligence - Grundlagen und praktische Anwendungen: eine Einführung in die IT-basierte Managementunterstützung", Vieweg, Wiesbaden, 2004.
Kmenta, S., B. Cheldelin, et al. "Assembly FMEA: A Simplified Approach for Evaluating Assembly Errors", 2003 ASME International Mechanical Engineering Congress & Exposition, Washington, D.C., November 16 - 21, 2003.
Nyendick, M., "Vom Leidsystem zum Leitsystem. Qualitäts- und Produktionskennzahlen", Quality Engineering (10/2005), 2005.
Pfeifer, T., "Qualitätsmanagement: Strategien, Methoden, Techniken", Hanser, München [u.a.], 2001.
Schrems, O., "Die Fertigung als Versuchsfeld für die qualitätsgerechte Produktoptimierung", Ima, Stuttgart, 2001.
Spear, S., "Management à la Toyota", Harvard Business manager, 08/04, 2004.
Wünscher, T. and Feldmann, D. G., "Nutzung von Erfahrungswissen aus der Produktherstellung zur Optimierung der Produktentwicklung", 16. Symposium „Design for X", Neukirchen, 13. and 14. October 2005.

Dipl.-Ing. oec. Thomas Wuenscher
Research Assistant
Hamburg University of Technology
Institute for Product Development and Engineering Design
Denickestr. 17
21071 Hamburg, Germany
Tel.: + 49 40 42878 2149
Fax.: + 49 40 42878 2296
Email: wuenscher@tuhh.de