Your Knowledge Partner™

Improving Quality for Automated Manufacturing

Megaputer Industry Solutions

Tan-Shei David Wu

Megaputer Intelligence
120 West Seventh Street, Suite 310
Bloomington, IN 47404 USA
+1 812-330-01

www.megaputer.com

Contents

Industry Background

With the introduction of the WTO and continuously increasing globalization, manufacturing competition is becoming more fierce than ever. Providing quality is just as important as lowering cost and delivery on time. As a manufacturer, what if you knew which piece(s) of process equipment is the source of quality problems delaying your entire shipment? Could you use this information to reduce down time, reduce cycle time, improve throughput, and improve capital asset utilization? Could you use this information to differentiate your company with quicker turnaround, reduced waste, and better quality? Would any of these improvements make your customers happier and move your company ahead of the competition?

Industry Situation

"High Technology Manufacturing" refers to the present day manufacturing that is highly automated with complex production processes. Examples of high tech manufacturing can be found across various industries including semiconductor, electronics components, aerospace, and pharmaceutical industry. These industries are characterized by immense capital allocation, high product value and rapid changes in product mixes. **Due to the tremendous complexity of high technology manufacturing processes, understanding what went wrong is no longer a simple and straitforward question.**

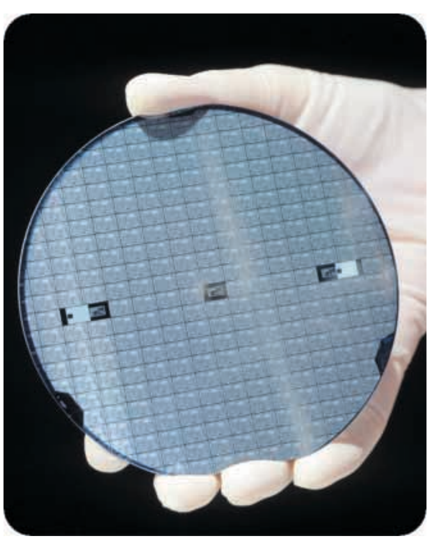
To maintain ultra-precisions in manufacturing, most controls and processes are computerized. Manufacturers implemented IT systems such as SPC (statistical process control) to track and provide real-time information through all phases of operation from conceptualization and design, to the total manufacturing cycle. As a result, enormous amount of raw data is generated. With limited assistance from statistical tools, manufacturing engineers often sift through hundreds or even thousands of charts and graphs in hope for spotting a trace of irregularity that leads to the source of manufacturing problems.

What manufacturers need are solutions that automatically and intelligently assist production engineers to narrow down and pinpoint the sources of manufacturing problems. Yet, each manufacturer utilizes different data sources and is seeking diverse equipment processing patterns. This requires custom analytical solutions that can be used to discover specific patterns that suit each manufacturer's unique need. Custom solutions can be integrated into the existing IT system to automatically provide answers to questions raised by complex high tech manufacturing processes.

Case: Background

Megaputer Intelligence carried out an equipment analysis project for a world leading semiconductor foundry. The company operates highly automated and complex manufacturing plants in numerous countries, selling silicon chips to electronics companies and chip design firms worldwide.

Semiconductor foundries are contract manufacturers of silicon chips for electronics companies and fabless chip design companies. The foundries encounter hundreds and even thousands of chip designs from customers at any one time. Given the myriad of distinctly different chip designs, foundries compete with core competencies in superior production quality and process efficiency. Both factors critically rely upon a foundry's ability to continuously improve manufacturing processes – quicker than anyone in the industry.



Case: Situation

The semiconductor manufacturer wants to uncover the root causes attributed to high defect rates in certain production lots. Typically when a new chip design or a new process technology is introduced, defect rate is initially high. The challenge is to quickly bring up the yield rate into acceptable range for mass production and drive continuous enhancement in order to improve quality and lower cost.

To assist the process improvement, the manufacturer has implemented sophisticated IT systems that track and record a huge amount of semiconductor engineering data for versatile analytical charting and reporting in production and development. The more sophisticated the manufacturing process, the more data is required to be analyzed, and the more difficult it becomes to conduct troubleshooting, particularly concerning yield enhancement. Everyday, production engineers manually screen through dozens of charts and statistics seeking irregularities in the production process. The main challenges encountered in the current process are that:

1) Engineers cannot easily assimilate the tremendous amount of information produced and summarized by the IT system.
2) The current system employs simple statistics to prioritize charts into a list for review. The list is neither sufficient nor effective in prioritizing charts and statistics. Often engineers would catch the real root-cause of poor yield only after reviewing dozens of charts and statistics appeared in the front end of the list.
3) Basic statistics does not reveal the complex intersectional effect caused by nonlinear relationships or multi-factor interactions in a reliable and timely fashion.
4) Data dimensions involved in the yield analysis could be up to thousands. The data attribute varies from quantitative (ordinal), and qualitative (categorical) to date-time stamps.
5) False alarms raised by the IT systems waste precious engineering resources.

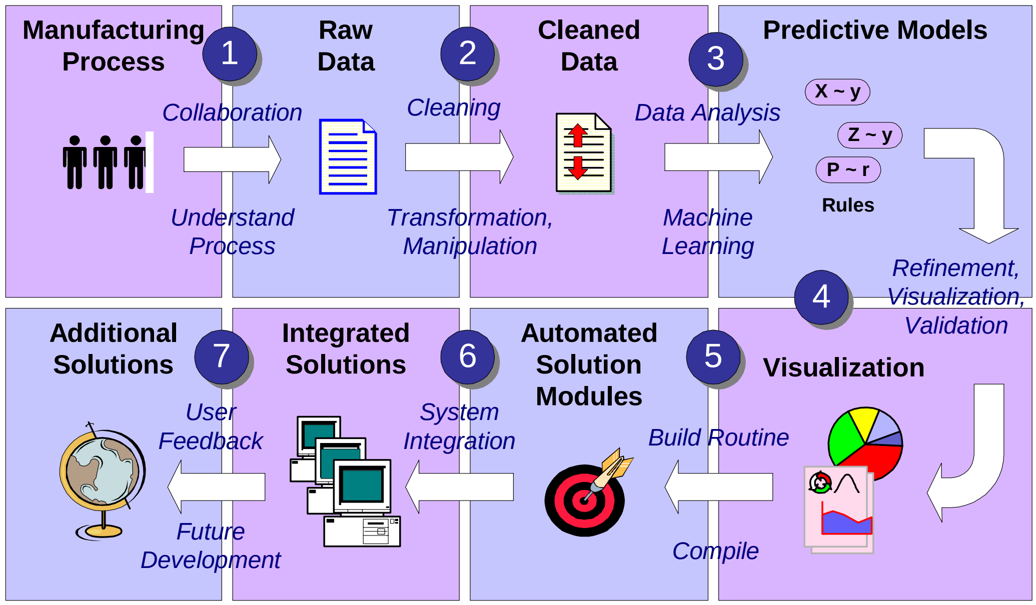
In short, for multi-dimensional data analysis, the manufacturer needs powerful analytical solutions that can be integrated into their existing IT system to pinpoint the most susceptible causal relationships in a heap of engineering data in a reliable and timely way.

Case: Business Value

The identification of root causes of poor production quality can be valuable to the semiconductor company in the following ways:

- Speeds up the trial production enhancement and reduces ramp-up time for new chip designs or production processes.
- Reduces equipment and human resources wasted during the longer trial production runs.
- Improves the capital asset utilization, captures financial float and improves ROI from the reduced ramp up time.
- Creates an unfair strategic advantage over the competitors by being able to offer the production of new chip designs in much shorter time frame.
- Commands value-pricing structures for customers who need quick access to market
- Identifies sources of quality issues for long-term improvement.

Implementation Approach



The overall objective is to create automated analytical solutions that can be easily accessed by production engineers to routinely enhance manufacturing processes. The end result is an array of analytical solutions integrated within the existing IT system. The end-user is only required to activate the solutions and adjust analysis configuration settings. The solution system automatically performs the analysis and produces useful reports. The implementation process is a self-looping cycle. After the end-users experience the integrated solutions, user feedbacks are important for fine-tuning the existing solutions and creating additional solutions that enhance business processes. The approach consists of series of steps.

1) Fully understand the manufacturing process to create data analysis solutions that are truly useful and applicable to the manufacturer.
2) Understand the raw data, and transforming the raw data into a suitable format.
3) Apply machine learning algorithms to generate predictive models. The predictive models are hypothesized, created, tested, and validated.
4) Create visualizations to illustrate and validate findings from the predictive models. Sometimes visualizations themselves are useful in establishing patterns and association.
5) Summarize the analysis methods in step three and four. Using the Megaputer analytical solutions development platform, the analysis methods are compiled into an editable and automated solution module that performs the exact analysis.
6) Integrate the solution into the existing IT system for quick access by the end-users. Integration reduces learning curve and offers routine access to the analytical solutions by the end-users.
7) Fine-tune the solution modules and create additional useful solution modules based on end-user experience and feedback.

Understanding the Process and Analytical Need

The very first step in creating useful solutions for manufacturing is to fully understand a company's unique manufacturing process. Collaboration among experts in the manufacturing process and experts in data analysis is necessary to fully understand the situation and analytical needs. Experts from both sides then work together to break down and understand the manufacturing process and to map out feasible and applicable analyses.

For the semiconductor manufacturer, the company is most concerned with which equipment operated at what time periods attributed to higher product defect rate. This information enables manufacturing engineers to quickly pinpoint the source of trouble and reduce production defect. The problem is complicated due to several factors. First, hundreds of different pieces of production equipment are used for one single product. Second, combinations of different production lines and equipment could be used for different production lots. Third, the same equipment operated during different time periods could contain dissimilar configurations to suit different production processes, thus influencing the final product. Further mounting the complexity, the defect rate of affected products may not be uniformly high. Even if specific equipment is actually faulty, defective products are produced intermittently. Hence on average, the resulting defect rate associated with the faulty equipment may seem to be within the normal range.

Together with close cooperation between industry experts and data analysis experts from Megaputer, several solutions are developed based on combinations of modular analytical components. These solutions become the most useful when integrated into existing IT systems to streamline production diagnosis and enhancement.

Understanding the Data and Data Transformation

Often, the original data is collected and managed for purposes other than data mining analysis. More, manufacturing data is often cryptic and requires domain expertise to relate to realistic situations. Data manipulation, cleaning, and transformation are almost always required preceding actual analysis in a real world situation.

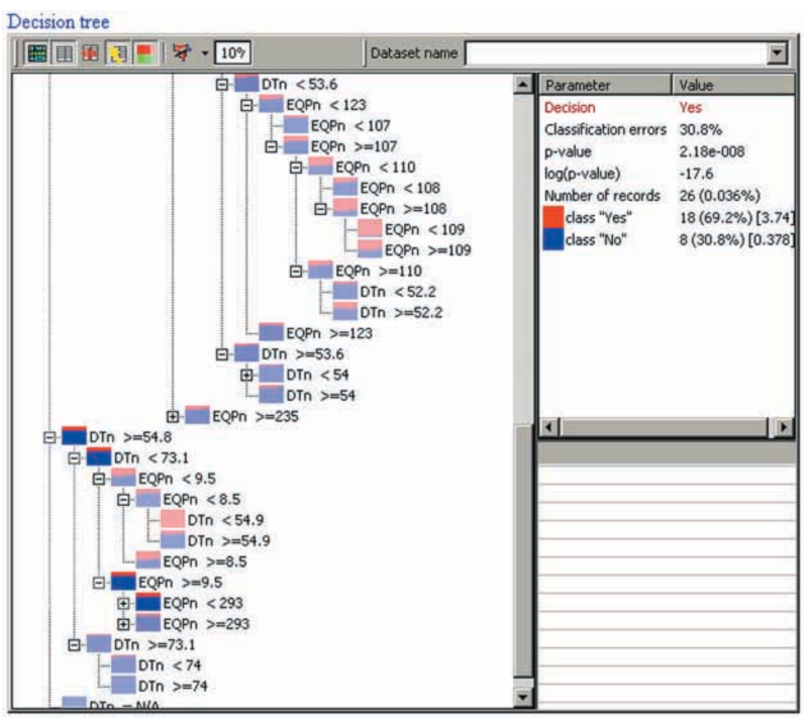
The original manufacturing data for the semiconductor company was collected mainly for statistical reporting. The data was stored in a format not suitable for data analysis. Data cells were represented in thousands of fields and relatively small number of record rows. With collaboration manufacturing experts, the data was broken down, and transformed into a format that was suitable for knowledge discovery while retaining all the original information. One interesting note was that some useful information is actually embedded in the data field headers instead of data cells. Derivatives of the original data were created to better represent useful features of the data. For example, a derivative representing equipment idle time in between two process steps may be more important than the absolute time when the equipment operated.

Combined Analysis

Typically in the analysis of complex operations, multiple analytical techniques are employed to examine the same data. The reasons are two fold. Firstly, different analytical engines offer different perspective on the same sets of data. Secondly, analytical techniques could be used sequentially. Analysis results from one analytical engine could be further analyzed as the input for another engine.

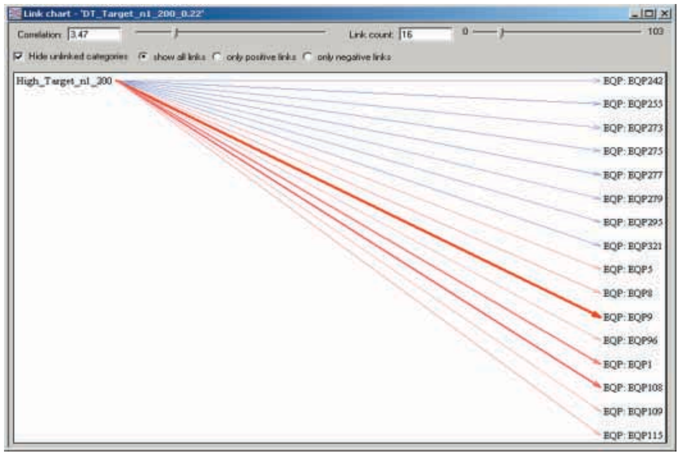
Decision Tree Analysis

A combination of data mining techniques is sequentially applied to the semiconductor manufacturing data to uncover valuable patterns. First, decision trees analysis is applied to discover broader sets of suspicious equipments and operating time periods that may attribute to high defect rate. Based on the Shannon's information gain theory, the decision trees analytical engine intelligently split the dataset into smaller subsets using the equipment and operating time as criteria in order to segregate the most suspicious equipments and operating time periods. The result is a set of splitting rules that visually looks like a tree with stems and leaves. The splitting stems represent a decision made to segregate data; the leaves represent a subset of data that fits the splitting criteria. A final decision is made on the leaves to place the subset into greater categories: in this case, high or low defect. The chart below shows the visual representation of the decision tree rule discovered in the analysis.



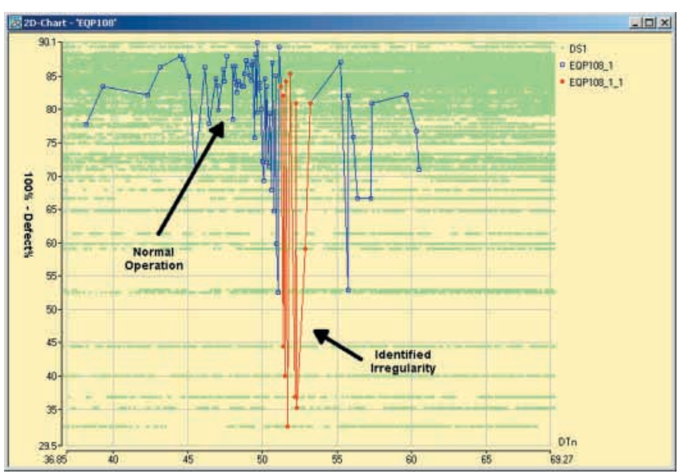
Link Chart Analysis

Next, output from the decision trees analysis is used as input for the link chart analysis to further narrow down the findings to identify the faulty equipments and operating time periods. Link chart analysis is also useful in ranking the identification based on strength of association to high defect and statistical significance. The link chart below identifies certain equipment that best fits the suspected causal pattern. With the combination of decision trees and link chart analysis, the most suspected equipments and operating time periods are identified and ranked in the order of relevance.



Validation with Visualization

Visualizations and graphs are not data mining techniques, but they are useful in intuitively spotting irregularities and visually verifying results discovered by data mining techniques. In the semiconductor company case, two-dimensional graphs with a twist turned out to be sufficient to validate the findings visually. Based on findings from decision tree analysis and link chart analysis, subsets of data representing these identified equipment and time are isolated into separate datasets for further examination or reporting. These isolated data records are then plotted on the background of all processes to allow engineers to visually confirm the findings. The chart below shows an identified equipment resulted in high defect rate in an intermittent fashion. Identified time periods are also shown in the graph in a different color – red. Green dots plotted in the background are other equipments in different processes.

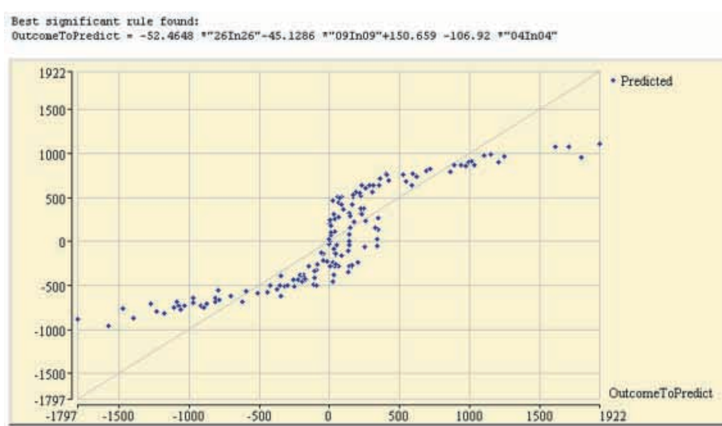


Find Laws Analysis

Engineering development in manufacturing frequently entails performing experiments, collecting data, and performing empirical analysis. Life would have been simple if everything fits nicely on a strait line in a two dimensional space. However, empirical analysis in the real world often involves approximating non-linear equations in a multi-dimensional vector space. Empirical analysis of this nature is time consuming by trial-and-error and regression methods.

The PolyAnalyst Find Laws analytical engine solves this problem. Find Laws is based on Symbolic Knowledge Acquisition Technology (SKAT), an algorithm developed by Megaputer Intelligence. Using methods of evolutionary programming, it develops high-degree rational expressions and other mathematical constructions, which can represent a wide variety of phenomena and are much better at representing nonlinear dependencies than are linear regression and similar techniques. In addition, unlike neural networks, SKAT outputs a readable symbolic expression, so that a human can read and understand how the model works. The evolutionary algorithm uses an idea of so called Generalizing Transformations to base new expressions upon previously developed ones, keeping and evolving whichever expression best fits the data. Thus, Find Laws works without knowing the form of the dependence.

In the semiconductor manufacturing case, the company hopes to discover empirical formulas that describe a product attribute by electromagnetic measurements of its raw material – silicon wafers. The relationship is non-linear, and probable independent variables consist of more than two hundred measurements. The problem is solved by first applying the PolyAnalyst Find Dependencies engine to identify the most influential independent variables from the more than two hundred measurements. Next, the Find Laws analytical engine is applied to generate a non-linear formula that describes the relationship between the dependent and independent variables. The resulting formula would be valuable in assisting manufacturing engineers to identify types of raw materials that perform better for specific types of products.



Solution Development and Integration

Data analyses are most useful to manufacturers when these analyses are automated and fully embedded within the manufacturers' existing IT systems. Megaputer Intelligence offers analytical component modules and a solutions development platform that adequately suits this purpose.

The solutions development platform facilitates rapid custom development of automated and reusable solution modules that are easily editable and expandable. Custom solution modules could be tailor designed for discovering unique patterns that are the most useful for individual companies. Because the solution modules could be embedded as part of the existing IT system, the learning curve for the end users is minimal. Production engineers and even business managers can access these automated solution modules to enhance the production process on a routine basis.

For the semiconductor manufacturing case, Megaputer worked with manufacturing experts to design several analytical solution modules. Several analytical processes are described in previous sections. The solution modules are binary objects developed on the Megaputer's analytical solution development platform. Each solution modules accesses several standard PolyAnalyst analytical COM modules to perform the specific analysis. These solution modules were then to be integrated with the manufacturer's IT system to provide easy access and automated analysis for the production engineers to perform process enhancement.

Conclusion

The case illustrates the overall implementation cycle of applying data mining technologies in high technology manufacturing. The analysis described above can be combined with existing engineering analysis to obtain a more holistic picture of different engineering and manufacturing issues.

In the semiconductor manufacturing case, results from the data analysis are benchmarked with findings from independent engineering investigation. The actual root cause of problem often appears within the top three most probably causes identified by the automated data analysis. These findings are isolated from hundreds of unique equipments and thousands of process steps operated during a period of two months. Instead of reviewing thousands of charts to instinctively identify irregularities, the manufacturing engineers can now just evaluate graphs for the most probable causes identified by the data analysis solution.

It is important to note that this type of solution modules integration provides an entire new source of knowledge. The solutions replace previously time consuming and unscientific process of manufacturing diagnostics. Not only do the solution modules save tremendous time and resource, they also enable manufacturers to gain significant strategic advantage over their competitors by being able to enhance production more quickly and efficiently.

Corporate and Americas Headquarters
Megaputer Intelligence Inc.
120 West Seventh Street, Suite 310
Bloomington, IN 47404
TEL **+1.812.330.0110**; FAX **+1.812.330.0150**
EMAIL info@megaputer.com

Europe Headquarters
Megaputer Intelligence Ltd.
B. Tatarskaja 38
Moscow 113184 Russia
TEL **+7.095.951.8079**; FAX **+7.095.953.5731**
EMAIL info@megaputer.com