



A novel design of OA-SPQC system for Steel Manufacturing Process

Veena N. Jokhakar*
M.Sc.(I.T) Programme,
Veer Narmad South Gujarat University,
Surat, India
veena.jokhakar@gmail.com

Dr S.V. Patel
Computer Science Department,
Veer Narmad South Gujarat University,
Surat, India
patelsv@gmail.com

Abstract--Almost all industries need to perform effective data analysis of their business process data. However, methodologies for such analysis will differ depending on the type of data to be analyzed. A steel manufacturing process consists of sequence of processing units having many complex sub processes. In each processing unit of steel manufacturing, various sensors are used to sense and collect data for the control of the processes as well as to maintain and improve the quality of the steel product. Processing as well as analysis of this data carries lot of significance. As conventional methods of analysis of these data are neither suitable nor efficient due to huge and diverse data to be analyzed, we must apply methods that make best use of modern technologies to obtain desired results. This paper presents design and implementation of an effective analysis system 'OA-SPQC' (Online Analysis- Statistical Process and Quality Control) System that makes use of statistical process control, quality control and Online analytical processing (OLAP) with the goal of analyzing process and quality control parameters effectively and further to provide flexibility to the user for necessary analysis in a typical mild steel manufacturing process.

Keywords – Data warehouse, Online analytical processing, Statistical process control, Statistical quality control, Cube.

I. INTRODUCTION

In today's business environment, all organizations need to collect, organize and interpret the wide variety of information available in their business. There is always a need to measure, understand and control the variables that affect the business process. Recent advances in computers and manufacturing techniques have made it easy to collect and store all kinds of data in manufacturing enterprises [1]. Organizations around the world need access to timely and accurate data to discover and distribute a useful business insight that in turn offers competitive advantage of better, faster and reliable business decisions to organizations.

Business Intelligence (BI) is gathering a significant momentum across organizations and industries in recent times to replace conventional methods of data analysis and decision-making. Using BI applications now it is possible to access and analyze information in almost every dimension of business efficiently. BI is basically a blend that consists of various technologies like data warehouse with all its components wherein the data is extracted, integrated, cleansed, transformed and then loaded in to a relational multidimensional database. Information and knowledge representation technologies like data mining, OLAP (Online Analytical Processing), Statistical analysis, Executive information systems, decision support system etc.

Organizations may use different BI tools and technology to develop different types of business intelligence applications. Some of the features that BI tools support are typically data analysis (both real time and historical data), web based interactive reporting, OLAP and alert messaging or notification services. However it is required that the tool should be easy to implement and manage and at the same time flexible enough to accommodate future changes and complexities of business.

In the paper, we present a design and implementation of such system that performs the data extraction, cleansing and then perform multidimensional analysis using Online analytical processing (OLAP) for a steel manufacturing company.

The paper has been organized as follows: Section 2 describes the steel coil manufacturing process and highlights

the importance of analysis of various data during each sub process. Section 3 describes different ways of data analysis. Section 4 Shows the design of the system for analysis with OLAP and SPC/SQC, Section 5 shows the implementation and results and Section 6 concludes the paper.

II. STEEL MANUFACTURING PROCESS

Steel manufacturing companies produce steel products like coils, plates, pipe etc. The process involves various steps to convert an iron ore to steel or customized steel. Typically the process starts from making of bricks/balls of iron ore in HBI/DRI (Hot-briquetted iron / Direct reduced iron). The ore produced from HBI/DRI are then melted in electric furnaces (Electric Arc Furnace), which is then called a heat. By adding different combinations of elements the molten iron converts it to molten steel in LF (Ladle furnace). This liquid steel takes the form of slabs in caster by putting high-pressure cold water on it. These slabs then take the form of coils or pipes or sheets in further mills. Further in Service center, there are many customized services and process that are performed according to the need of the customer like smoothening, galvanizing etc.

Taking an example a heat is made by melting iron ore(HBI/DRI), which is converted into a slabs weighing around 20 -25 T of around 10 meters each in length, which again is converted into a coil, pipes or sheets of approximately same weight but 100 to 1800 meters in length. Now this coil can further be processed into small packs or be still elongated to almost 2500 meters to be converted into pipes for transporting gas, water, etc. Figure 1 shows the whole process discussed. During these process huge number of control and process parameter data are collected. In some situation, due to the dynamically changing parameters, sometimes they may not adhere to the band and hence difficult to identify, which may cause problems or defects in further processing of the product.

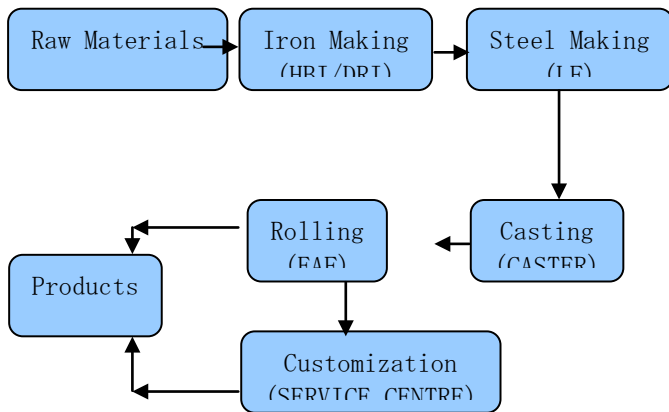


Figure 1. Steel coil-making process.

These data are generated and stored on continuous basis. Quality control department using different techniques then analyzes these data.

III. SOME METHODS OF DATA ANALYSIS

Different techniques may be used for the purpose of analysis of such data. Some of them are as follows:

A. Using Spreadsheets/Excel:

A simple method of statistical analysis can be done using excel. However it has many limitations. Human Error is also expected to happen leading to a big serious mistake in decision-making. People can mislead others by intentionally showing compromising results. Further it is not automated system; hence it is not possible to see real time results.

B. Using OLTP:

Many On-line transaction-processing systems exist currently that allow analysis of data by report generation. However they are not user friendly because the user has to see the fixed format of report every time. Moreover data as would be stored in RDBMS and future enhancements for decision support systems and creating intelligent systems in future is difficult. As the huge amount of data exists in diversified data stores, this kind of static analysis is not effective way for analysis.

C. Using Data mining:

This technique has been found promising and it is gaining popularity. Data mining provides a way for data exploration, mining & Analysis by Automatic or Semi-Automatic means of large quantities of data in order to discover meaningful patterns, rules and knowledge. Kesheng Wang [1] describes one of the implementations of DM techniques in manufacturing and their implementations on product design and manufacturing. M li et al. [2] presents the preliminary results of a data mining study of a production line involving hundreds of variables related to mechanical, chemical, electrical and magnetic processes involved in manufacturing coated glass. The study was performed using two nonlinear, nonparametric approaches, namely neural network and CART, to model the relationship between the qualities of the coating and machine readings. Much work has been done involving neural networks to predict state variables of a blast furnace. Bulsari et al. [3] observers good results when using multi-layered feed-forward artificial neural network (ANN) models to predict the silicon content of hot metal from a blast furnace. As time-dependencies (time lags) between each of

the various inputs used and the output (silicon content) were found to be significant, each input was lagged from the output by a set amount. The amount was determined by the correlation between the input variable and the output.

D. Using OLAP:

On-line analytical processing systems are way of multidimensional analysis wherein the user is provided with various ways of data presentation like drill-down and roll up, multidimensional analysis etc. OLAP analysis on a multidimensional model involves selecting various slices of the data cube (i.e. fixing the value of a dimension by selecting a record from the corresponding dimension table) and aggregation of facts with respect to a dimension [4]. There are different types of OLAP. Multi-dimensional OLAP (MOLAP) is where data is stored in a multidimensional cube. The storage is not in the relational database, but in proprietary formats. Relational OLAP (ROLAP) is a relational back-end wherein operations of the data are translated to relational queries. Hybrid OLAP (HOLAP) is the Integration of MOLAP and ROLAP. Desktop OLAP (DOLAP) is a simplified version of MOLAP or ROLAP. K F Tsangl et al. [5] shows a data mining system (DMS), which is incorporated into emerging technologies including on-line analytical processing (OLAP). Wang Cai et al. [6] also shows a solution that urged the transform from Traditional Information Management System to the Information Analysis Decision System. Some data mining tools are bundled with online analytical processing (OLAP) packages, while others are designed to run against relational databases or data warehouses. Still others are extensions of standalone analysis packages and run against local data tables such as that found in a spreadsheet environment [7]. One may extend the output of OLAP with special effects to make it more efficient like, zooming, tool tips etc. Still others use OLAP as extensions of standalone analysis packages and run against local data tables such as that found in a spreadsheet environment [5].

OLAP is one of the most effective ways of analysis as the user may look at the data the way he wants and on the any dimension he wants. One can look analyze the data in different dimension at any granularity level. This helps the decision makers of manufacturing firm personnel to closely study the production performance of the industry.

E. Using Statistical Analysis:

Statistical analysis is also used to analyze huge amount of data where the nature of data is numeric. Gary Fisher et al. [8] Shows a method and a system for controlling a manufacturing process statistical process control data or other statistical indicators of performance from the production process and specification data are displayed in tables. SPC (Statistical Process Control) is the primary analysis tool of quality improvement. It is the applied science that helps you collect, organize and interpret the wide variety of information available. Control charts demonstrate how consistently the processes are performing, and whether one should, or should not, attempt to adjust it.

Statistical analysis is an efficient way of analyzing data with various process control functions but is not enough for producing known or hidden information from the data nor does this gives a way for user to perform analysis in any dimension.

As we see that though the above approaches are for analyzing data, their applicability differs from application to application. If we wish to carry out analysis of continuous real time data, we need to it would be necessary to combine some of the above approaches. Further we would like to provide a solution that is user friendly. Hence the approach stated in the paper tries to combine the strength of OLAP and Statistical analysis. OLAP as stated could provide a powerful way to perform multidimensional analysis and statistical process and quality control charts can be used for measurements and analysis of real time processes. Here the data cube is created based on the users request. According to the data, the best suitable statistical chart is selected and according to it the cube is created for multidimensional analysis. It is a very effective way that satisfies all the needs of user like detail analysis, performance, memory management, and user friendliness.

IV. DESIGN OF OA-SPQCS

The OA-SPQCS has been designed to work with various types of data, i.e. control, process and other non-technical parameters. A non-technical parameter like delay may be analyzed for detecting the reason of delay, like electrical or mechanical issues etc.

Hence, we make use of various aspects of the SQL 2005 for the required purpose such as SQL server Analysis services and particularly used AMO analysis Object model for building cubes on the fly, MDX multidimensional query for extracting the data from the cube, SSRS SQL server reporting services, RDL report definition language for data representation and XMLA and ADOMD for presentation of analysis. It would create a dynamic and temporary cube on the fly for statistical analysis using different types of control charts like Histograms, Pareto Charts etc with the analysis display of the statistical calculation according to the chart. The design consists of 5 modules as under:

A. Data Grabbing and view selection:

The analysis that we need to perform needs the data from all the plants. But as the data generated from these plants are stored in diversified manner, we need to extract and integrate the relevant data in some common platform. Hence, the system has a view selection and complex selection module that helps the data selection procedure. Complex selection will be having contents like CRM Complex, Service center etc. New complex can be added dynamically by adding it to the combo fill configuration file.

Line selection will be having list of all the lines of selected complex e.g. CRM complex having lines like PKL1, PKL2, CGL1 etc. As per the data storage architecture of the company, all data of a particular LINE, such as Galvanizing, Picking etc are stored in their respective database like ORACLE or SQL server and in respective servers. The system makes use of various views from the database end. View is the entity, which works as a data source for this system. It will capture data from production table. There are basically three types of views for each line.

The data are stored in views as per the data availability for the line. For eg, some line may not have data related to delay. The data stored in view are of 3 types.

Process Details – these contain data of every coil of every 5 sec for all the controllable parameters of the coil as per Location.

Delay Data – These contain the delay data related to the coil Summary Data – Contains the summarized data as per time.

Meta data of this views will be stored in one table which will have fields like id, complex, data source name, line, userid, password, catalog. So whenever any new view or line comes end user has to enter information about it into table. User has to select one view for further analysis in the system.

B. Parameter Selection and Filtration tool:

Each view has different parameters. The entire parameter list is listed in a tree view. The user may hence choose more than one attributes from the same. Later one will have to select the filtration criteria. Fig 2 shows the partial flow for the same.

- Filtration tool:** User can select particular data by selecting particular time, or specific values, ranges to the attributes. Filtration operators will be equal to, less than, greater than, in between etc. User can use any operator with any parameter. Then query builder will build query for the criteria selected as shown in figure. These filtration operators are provided according to the type of the attribute selected for analysis.
- Query Builder:** will build query based on the type of SPC/SQC charts suitable for the attributes and views selected for the analysis, which shall be acknowledged to the user.

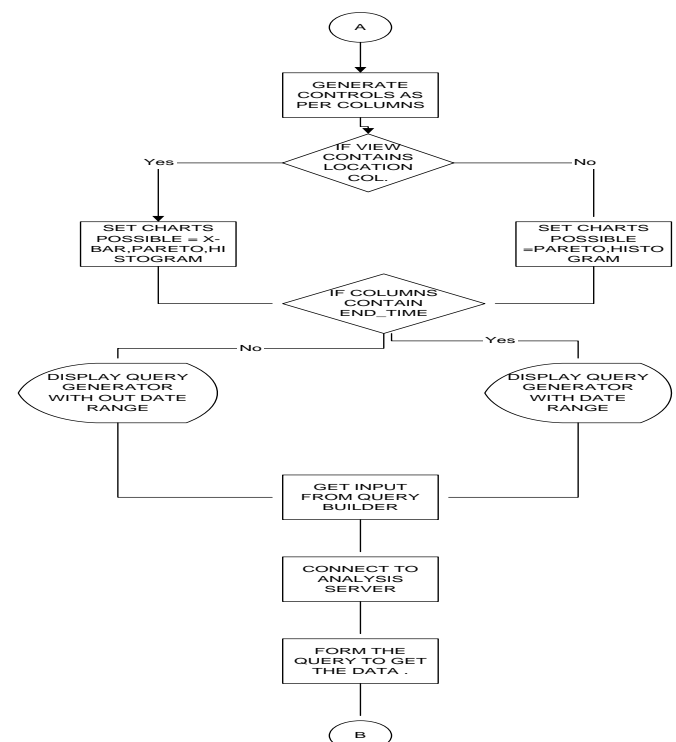


Figure 2. Partial flowchart for filtration tool.

C. Cube Creation:

After data selection & configuration, a cube is generated in background. Cube is an entity, which stores pre-calculated data. In this case cube will store all the data related to different charts such as histogram, X-bar, Pareto as shown in fig 2.

D. User Interface:

User has to select one dimension from the list. This list will contain all the dimensions in the cube depending on the

chart. If Histogram is selected, then all the dimensions having ranges will be displayed.-In case of x-bar chart, all dimensions having the subgroup will be displayed -And in case of Pareto chart, all dimensions will be displayed.

E. Statistical Analysis:

Statistical analysis component has different calculations for each chart Calculation for Histogram and Pareto chart: Histogram charts has following calculation: Mean, median, min, max, skewness, sigma, UCL, LCL, Internal width.

Calculation for X-bar chart: UCL, LCL, Sigma, min, max, range of all data, Average of Range, Range per data, Avg. per data. Various types of analysis can be done like process is unstable, all the data are correct, some data are wrong, some data are incorrect, probably data has been taken are over less period of time and Problem with some data.

V. METHODOLOGY AND RESULTS

OA-SPQC would help the company to improve the quality of products and performance of the plant It does a lot of number crunching and analysis on the data that is gathered from various lines of the plant and do the further analyses of the data on the various measures. The data is available in a huge amount. Every 5 seconds 1 record with more than 100 parameters is stored in one or more tables from all the lines in the plant. This leads to approx of 20000 records per day per line.

The data that has been stored needs to be analyzed by various dimensions with various measures with any amount of granularity that would allow the user to visualize the whole situation in a readable and user friendly manner, therefore the system uses graphs to display the huge analyzed data in a readable format by the use of charts or graphs.

For the cube generation process, a page contains all the controls as per the user-selected chart. The Controls to be displayed and the format in with they are to be displayed is controlled by configuration files. An example of the structure of the file is as follows: Element = DWCombo, Range = No, Modes= Yes, etc. Element; Describes the Control Type, Range; tells whether to display 2 textboxes for FROM and TO, Modes; tells whether to display a check box as AUTO for Automatic System criteria to be taken, and heading; displays as heading of the page.

Cube creation, as discussed starts with creation of data source. If cube name already exists, it will be dropped and new data source will be created with the same name. Data source view is created depending upon the attributes selected by user. Different data table are created and data type is checked for, it is numeric then user is asked for chart selection. If chart type selected is xbar then subgroup data table is created. Minimum, Maximum, Internalwidth for each attribute will be calculated and according to those subgroups will be created. Keys for data table will be defined then fact data table and relationships will be created or if chart type is histogram then range data table is created. Minimum, Maximum, Internalwidth for each attribute will be calculated and according to those subgroups will be created.

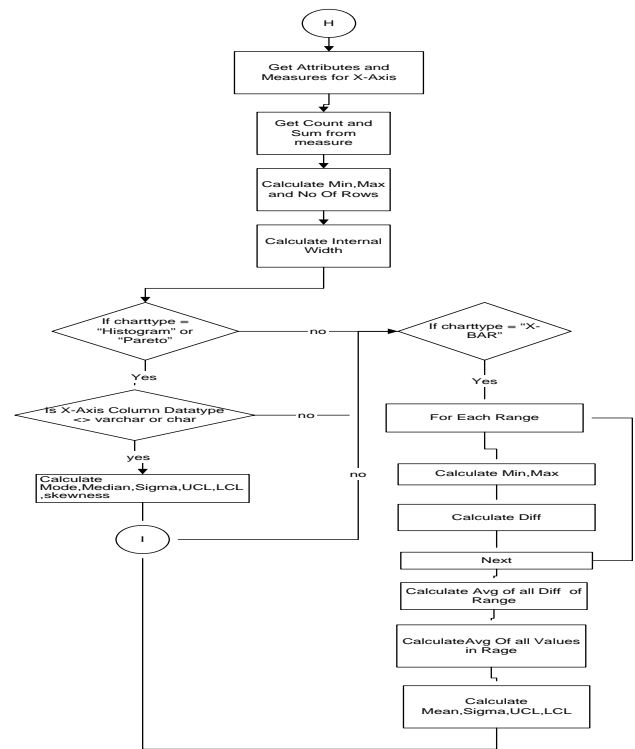


Figure 3: Automatic chart type selection algorithm flowchart

Each range will be attribute of subgroup data table. Keys for data table will be defined then fact data table and relationships will be created and if chart type is Pareto or data type is varchar then only simple data table without range and sub group will be created. The figure 3 shoes the flowchart of the selection steps of chart types.

Following steps are followed to create the user interface

- User selection for x-axis and chart type will be taken
- Measures and attribute of given x-axis dimension will be collected.
- Gets minimum and maximum value for series axis by using statistics.
- Generation of dataset: If chart type is histogram then get ranges/attributes. Generate MDX query with all measures and then generate dataset and fields. If chart type is pareto then simply generate MDX query by count of given attribute and dimension attribute and then generate dataset and fields. If chart type is x-bar then get subgroups for each subgroup get measure for average and generate MDX query by measure and dimension attribute.
- Report Definition Language will be generated for display of the datasets. In RDL file there are many parts, which will be combined to generate chart. Creation of RDL starts with selection of data source, Creating Data Set, Creating body of the chart, Creating Chart Data, Creating Category axis and Value axis, Creating Category groupings, Creating Series groupings, Creating Plot area, Formatting report, Setting Chart type as shown in fig 4. Fig. 4 plots the X-Bar graph for furnace_pressure for a particular coil. Also calculates the UCL, LCL, Mean etc.

Statistical Analysis on the data shows that some data are wrong, some data are incorrect, probably datas taken are over less period of time, Problem with some data, Most of the values are on the left side of the center values, Most of the values are on the right side of the center values are learnt. Similarly for X-bar chart conclusions like Process Is Stable,

Check for next Constraint, Process is totally stable, There was sampling from multi-stream process, Process Mean has been shifted Or There is a problem in capturing the data etc may be learnt as shown in fig 5.

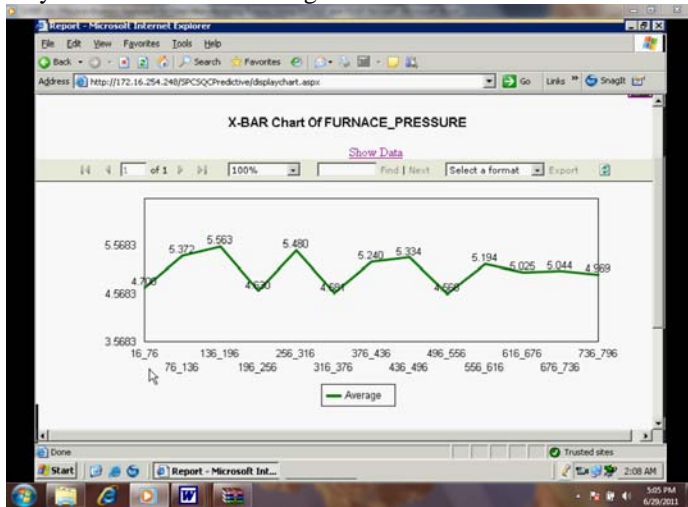


Figure 4: A chart showing user flexible approach

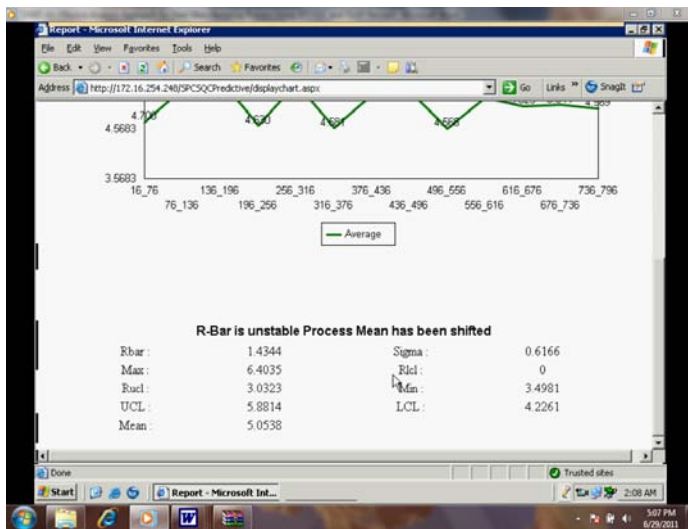


Figure 5: Statistical reading for the data

By following this approach one may be able to perform multidimensional analysis on manufacturing company's process data and view the various control charts. This will also not occupy lots of memory as only the needed data is grabbed and analyzed. These graphical data may also be converted into .csv or in table form as shown in figure6. Moreover the cube structure may be deleted automatically after few days of non-use or on user's request.

VI. CONCLUSION

A design of OA-SPQC system, a new methodology for on-line multidimensional and statistical analysis was presented in the paper for a typical steel manufacturing unit. The relevant data for analysis is collected from diversified

databases (which might be continuously updated) and cleansed. Then according to the requirement of user, he may select the data relevant to him. Depending upon the selection of data and the users need; OA-SPQC selects the best suitable process and quality control charts. The data cube for multidimensional analysis is created on the fly dynamically. Dimensions, fact table and measures are based upon the selection of the statistical chart and the data. Hence, the approach combines the concepts of OLAP and Statistical process and quality control. Further, it is a very flexible and user friendly as it is totally user driven and data driven. The future enhancement of this is to extend the same to do predictive analysis using data mining, to take decisions on real time and make the system more intelligent.

VII. REFERENCES

- [1]. Kesheng Wang, "Applying data mining to manufacturing: the nature and implications", Journal of Intelligent Manufacturing, Vol. 18, Number 4, 487-495, 2007
- [2]. M. Li. S. Sethi, J. Luciw, K. Wangner, "Mining Production Data With Neural Network & CART", Proc. IEEE Int. Conf. Data Mining, 2003.
- [3]. Bulsari, Abhay and Saxen, Henrik. "Classification of blast furnace probe temperatures using neural networks.", Journal Steel Research. Vol. 66., ISSN 0177-4832, Vol. 66, 1995, pp. 231-236.
- [4]. Gergely Pintér, Henrique Madeira and Marco Vieira, István Majzik and András Pataricza, "Integration of OLAP and data mining for analysis of results from dependability evaluation experiments", Int. J. Knowledge Management Studies, Vol. 2, No. 4, 2008.
- [5]. K F Tsang, H C W Lau, S K Kwok, "Development of a Data Mining System for Continual Process Quality Improvement", Proc. IMechE Vol. 22 Part B: J. Engineering Manufacture, 2007
- [6]. WANG Cai, TANG Hui-jia, "Application research of manufacturing fault analysis base on warehouse and OLAP", Journal of Railway Computer Application, ISSN 1005-8451, Issue 6, pp 35-37, 2007
- [7]. Richard Kittler and Weidong Wang, Data Mining For Yield Improvements Proc. Int'l. Conf. on Modeling and Analysis of Semiconductor Manufacturing, Temp, Arizona, 2000, pp. 270-277
- [8]. Gary Fisher, Mineral Ridge, Ohio, Mark Clark, "Process Control Method for improving manufacturing operations", Patent no. 54440478, Mercer Force Company, issue date 8/8/1995.