

DEVELOPMENT OF AN OLAP- FUZZY BASED PROCESS MINING SYSTEM FOR QUALITY IMPROVEMENT

G.T.S Ho, H.C.W Lau

Department of Industrial & Systems Engineering, mshenry@inet.polyu.edu.hk

Abstract: Currently, companies active in the development of high-tech products has become more and more complex in the age of mass customization. Not only do they have to focus on improving product quality, but rather on gaining experience to modify the current processes in order to streamline the integrated workflow. A real-time process mining system (R-PMS) is developed to analyze the proposed XML based process data for discovering the hid-den relationship between processes. The new feature of this system is the in-corporation of the process mining engine, which is characterized by the combined capabilities of the Online Analytical Processing (OLAP) and fuzzy logic (FL), to form a robust framework for highlighting the undesirable process set-ting and parameters for further improvement in a real-time manner. The simulation results indicate that the OLAP based fuzzy approach is generally superior to those of conventional methods which offer higher flexibility on production process management with decision support ability. In this paper, the de-tailed architecture and a case study are included to demonstrate the feasibility of the proposed system.

Key words: Online Analytical Processing, Fuzzy Logic, Extensible Markup Language

1. INTRODUCTION

As organizations become more conscious, management of processes and process data with temporal context is gaining increased attention [5]. Process mining can be seen to contribute to this. It aims at extracting information

Please use the following format when citing this chapter:

Ho, G.T.S., Lau, H.C.W., 2006, in IFIP International Federation for Information Processing, Volume 228, Intelligent Information Processing III, eds. Z. Shi, Shimohara K., Feng D., (Boston: Springer), pp. 243–258.

from event logs to capture the business process as it is being executed [15]. Process management is primarily concerned with the integration of task and context knowledge in application. Processes vary from place to place and from organization to organization. Future management systems are expected to incorporate process management that enable the operations staff to shift its focus from managing equipment to managing processes. Management will become a distributed, co-operative problem-solving activity [9]. Many enterprise work very hard to produce goods and services to a high standard. In order to do so, every process must be effective and efficient, i.e., do the right things and do the things right. As a result, knowledge and information sharing within the enterprise become a must in a fast changing market environment. The utilization of information technology (IT) is taking up momentum to meet this challenge. In particular, data mining (DM), artificial intelligence (AI) and distributed object technology have achieved significant attention for achieving agility of manufacturing system, which has played an important role in transforming quality to new generation.

The objective of this research is to develop the real time process mining system for continual quality enhancement. This paper is divided into three main sections. Section one describes the framework of proposed process mining engine- OLAP based Fuzzy approach. Section two examines the process mining engine on how to improve the finished quality in any process continuously. Section three describes the procedures of system development and its findings of implemented system. Section four concludes the paper and discusses further improvement of R-PMS.

2. RELATED STUDIES

In order to attract and retain customers as well as business partners, organizations need to provide their services (i.e., execute their processes) with a high consistent, and predictable quality [4]. Many companies work very hard to produce goods and services to a high standard. In order to do so, every process must be effective and efficient, i.e., do the right things and do the things right. Poor process decisions from any individual may lead to poor customer satisfaction and the ultimate goal is to achieve better collaboration for making right decision anytime in any enterprise member. A process is a series of steps or sequence of business activities the outcome of which to achieve customer satisfaction by providing the customer with what they need, when they require it and in the manner which they expect [12].

In fact, the recent trend of global manufacturing is to implement system infrastructure that allows analysis being performed on vastly distributed data according to the corporate objectives in order to make decisions on elements

of business strategies. In the area of information-based management, the key of success is to recognize the company's competitive advantages and weaknesses with the support of information technologies for decision support. For the business intelligence of an enterprise, there are only about 20 % information can be extracted from formatted data stored in relational database. The remaining 80 % information is hidden in unstructured or semi-structured documents. Recently, data mining technology, which aims at the conversion of clusters of complex data into useful information, has been under active research [1] [7] [8] [10]. In this project, the On-Line Analytical Processing (OLAP), which is based on data mining technologies, would be deployed as the tool for knowledge discovery to ensure efficient process interactions in the production workflow.

Recent years have seen a number of publications related process improvement and process mining. An intelligent system, which is able to improve an organization's current performance by mining and understanding the historical process data, is still an area that requires more in depth study and investigation. The issue is addressed in this paper with the incorporation of various computational intelligence techniques including Extensible Markup Language, Online Analytical Processing and Fuzzy reasoning. This paper describes the development of a Real time Process Mining System (R-PMS) architecture that enable process engineers to drill down and monitor the quality of finished products in different levels. The advantages of this system over conventional production expert system can be characterized as follows: (a) In-depth and fast analysis can be performed on distributed process data; and (b) Real-time decision support is provided for eliminating the number of failure products.

3 INFRASTRUCTURE OF A REAL-TIME PROCESS MINING SYSTEM

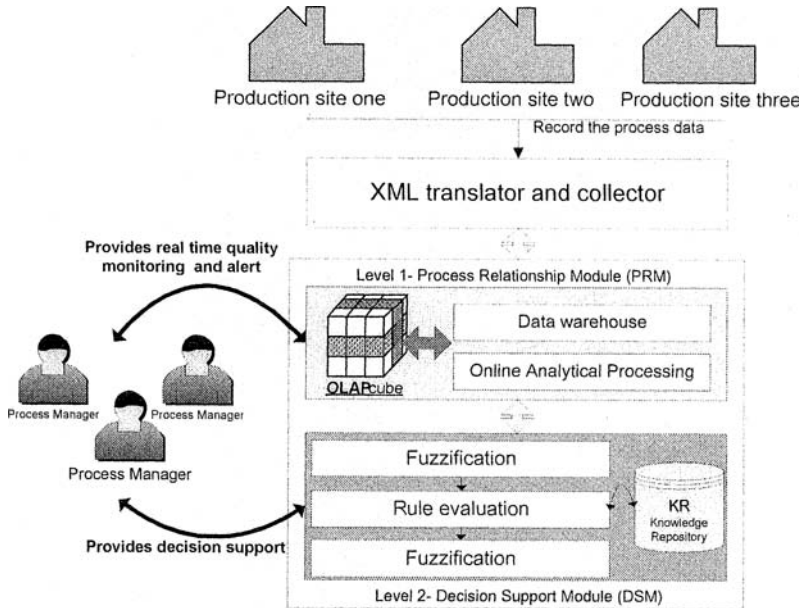


Fig 1 Infrastructure of Real- time Process Mining System (R-PMS)

The infrastructure of the proposed R-PMS is shown in Fig.1. R-PMS consists of two levels: level 1- Process Relationship Module (PRM) and level 2- Decision Support Module (DSM). It is designed for capturing the process quality data from different departments and converting into knowledge to support real- time quality control and continual improvement. In fact, it allows process and quality engineers within the production line to access the process mining engine to retrieve the updated current inspection status and suggested corrective action in terms of optimized control parameters. The system also allows collaboration and data sharing and provides engineers with the ability to generate the defect reports through the common interface. The proposed system can be divided into server and client sides with the server side as the centralized process mining engine and client side referring to different process engineers along the production sites. To support and realize the total quality management for the production of high- quality goods, every process engineers must keep monitoring the process quality. The system will also provide some corrective actions provided that the existing process is out of control in order to ensure “Do it right at the first time” in each process. It is developed for mastering the continual quality enhancement for high- quality products based on the proposed OLAP based Fuzzy approach.

3.1 LEVEL ONE- PROCESS RELATIONSHIP MODULE (PRM)

The distributed process data within the enterprise is interchanged through the Inter-net and the purpose of the XML translator is to integrate the enterprises' existing data models and stored in a centralized data warehouse.

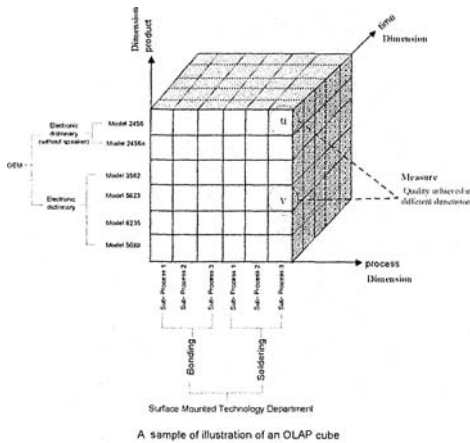


Fig 2. A sample of illustration of an OLAP cube

Whilst the process data is stored in the data warehouse, the further step is to conduct data analysis and reporting in order to alert or assist quality or process engineers in solving quality problems. The technique of Online Analytical Processing (OLAP) is suggested in this module. It facilitates timely access and manipulation of the process data and the application to drill down into data to obtain further information. The OLAP acts as a multidimensional data model which organizes process data into a hierarchy that represents levels of details on the data. It allows engineers to analyze, create and performs process quality reports online. This powerful query engine assists engineers to find and retrieve the defected statistics or performance measurement with the dimensions of time, process, operator and project which can keep monitoring the performance of production line in a real time manner.

The essential characteristic of OLAP is that it performs a numerical and statistical analysis of data and the data is organized in multi-dimension. In OLAP data model, it consists of descriptive data (dimensions) and quantitative value (measures) which builds up the OLAP data cube (Fig 2).

In the OLAP data cube, it builds up with two elements: fact table and dimension. In the fact table, the measures are defined for data analysis and used the defined measures to create user-defined measures (calculated members) which are used on data analysis, e.g. a control chart for keeping a continuing record of a particular quality characteristic in a 'real-time' manner. In the dimension table, different dimension levels are defined to use on different views of OLAP data cube which allows engineers to browse the process quality or control charts based on time-to-time variation, process-to-process variation and within-product model variation

3.2 LEVEL TWO- DECISION SUPPORT MODULE (DSM)

After the quality problem is highlighted in PRM, Some modifications of operating process parameters will be suggested to minimize the defect during production. Most of the process parameters settings are decided by experienced process engineers and controllers and it is a hard task for them to select the optimal settings when different specifications and requirements are required by customers. Moreover, the parameters setting must be adjusted or fine-tuned due to the variation of quality achieved in previous process. The purpose of this module is to provide decision support in each sub-process by using fuzzy technique. In creating decision support functionality, a mechanism, which is able to combine and coordinate many sets of diversified data into a unified and consistent body of useful information, is required. In larger organizations, many different types of users with varied needs must utilize the same massive data warehouse to retrieve the right information for the right purpose. Whilst data warehouse is referred as a very large repository of historical data pertaining to an organization, data mining is more concerned with the collection, management and distribution of organized data in an effective way. The nature of a data warehouse includes integrated data, detailed and summarized data, historical data and metadata. Integrated data enable enterprise members to easily and quickly look across vistas of data. Detailed data is important when they wishes to examine data in its most detailed form while historical data is essential because important information are hidden in this type of data. The data flow between PRM and DSM has been depicted in Figure 3. It should be noted that when the abnormal trends of achieved quality are obtained, the OLAP will extract the crisp values and convert the data into recognized format for generating suggestions in DSM. OLAP can also act as a bridge between PRM and DSM by passing the right data for

further analysis. After the process of fuzzy inference is conducted, the crisp output of suggested operating parameters will be packaged as an XML based document and distribute to related enterprise members for carrying adjustment of operating parameters. The query result can also be published on the web page in order to provide decision functionality required by organizations anywhere. Thus, the R-PMS assists the engineers or related staffs to make efficient data analysis on the proposed infrastructure and to assimilate the analysis results to different locations.



Fig 3. The process data flow between PRM and DSM

A number of factors or rules must be firstly identified by engineering experts in the organization that may affect the quality of finished products and stored in knowledge base. The knowledge base contains the knowledge related to the remedial actions when the quality problems and defects are found. The fuzzy rules are stored and defined as a conditional statement in IF-THEN form, e.g. IF rate of deposition is short THEN time of sputtering is adjusted to slightly long. This kind of linguistic variables include fuzzy sets, such as extremely low, low, medium and high which can greatly reduce a number of production rules.

The core of this module is fuzzy inference; it is the process of formulating the mapping from a given input to an output using fuzzy logic. The mapping then provides a basis from which decisions can be made, or patterns discerned (MathWorks 2002). Mamdani's fuzzy inference method is commonly used in engineering application and is performed four steps: fuzzification, rule evaluation, aggregation of the rule outputs and defuzzification. For the Decision Support Module (DSM) of R-PMS, the first step is to extract the crisp values obtained in PRM (the achieved quality is within the upper or lower limits), and uses pre-defined fuzzy sets to determine the relevant degree of these inputs. After all the inputs have been fuzzified, they will be applied to the antecedents of the fuzzy rules stored in the knowledge base. As the rules defined have multiple antecedents, the fuzzy operator is used to obtain the single truth value. The classical fuzzy operation union and intersection are used to evaluate the conjunction and disjunction of the rule antecedents and calculated as below.

OR fuzzy operation

$$\mu_{A \cup B}(x) = \max[\mu_A(x), \mu_B(x)]$$

AND fuzzy operation

$$\mu_{A \cap B}(x) = \min[\mu_A(x), \mu_B(x)]$$

where μ_A and μ_B represent two fuzzy sets A and B on universe X

The outputs of all rules will be unified in order to generate a single fuzzy set for defuzzification. Among several defuzzification methods, Centre of Gravity (COG) is adopted and its equation can be expressed as below.

$$\text{COG} = \frac{\sum_{x=a}^b \mu_A(x)x}{\sum_{x=a}^b \mu_A(x)}$$

This equation represents the centre of gravity of fuzzy set, A on the interval, ab can be calculated over a sample of points along x axis.

4 CASE STUDY

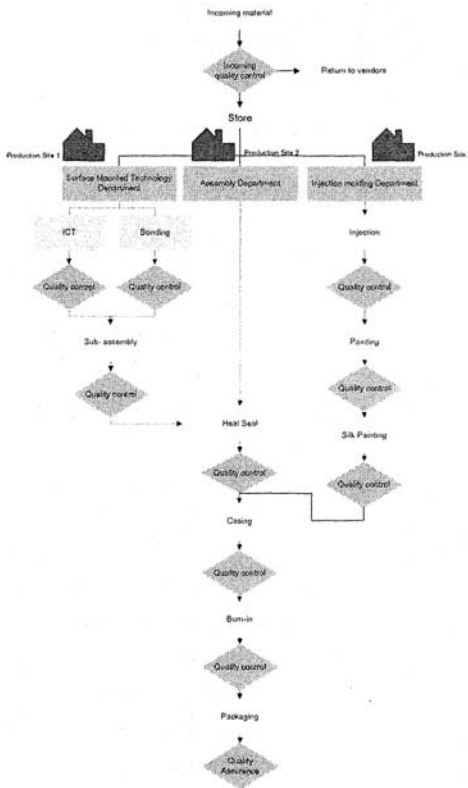


Fig4. The workflow of electronic dictionary manufacturer

In order to validate the proposed concept, structure and techniques, a rapid conceptual prototype of R-PMS was developed. The prototype was validated in electronic product manufacturer- GSL and shown to have more expressive power that renders its applicability in more realistic applications involving quality enhancement. GSL is proficient in designing and manufacturing a wide range of hand-held electronic products for consumers to acquire and to utilize information in a convenient and fast manner for education, entertainment, data storage and communication purposes. It designs and manufactures a wide range of products including electronic dictionaries, personal digital assistant (PDA), translators and electronic organizers. GSL currently employs over 3001-6000 people in China and Hong Kong. GSL founded in 1988 and launched of the first Instant-Dict electronic dictionary EC1000 in 1989. GSL established the brand as

“Instant-Dict” and start OBM early rather than transforming from OEM or ODM to OBM.

4.2 PROBLEM DESCRIPTION

The case study focuses on improving three production sites' current performance by reviewing the way its processes are organized and taking proactive actions in order to avoid the cost of rework and ensure the finished product that performs at or near perfection. The current process flow chart of electronic dictionary manufacturing is illustrated as Figure 4. There are three production sites which are responsible for manufacturing some parts of electronic dictionary. One failure or quality problems found in any production site should cause the finished product- electronic dictionary rework or wasted. In the traditional approach, all of the process quality data is captured in different production sites and stored in proprietary and fragmented discrete database systems. Check sheets, problem and work distribution logs are the tools for quantifying the number of incorrect area of activities which occur in the areas of blockage. However, this approach cannot provide quality engineers to identify the relevant process data in a real time manner and no decision support for workers when quality problems are addressed. In response to the demands of the marketplace, GSL has recognized the need to change the way they operate. Improving and reorganizing the processes by which a business achieves its customer satisfaction is a powerful mechanism for change. In order to conduct data analysis from vast amount of process data in a real time manner, it needs to organize the data into different levels with different views in order to discover the relationship between processes. In addition, documenting issues surrounding the process, GSL considers such problems as duplication of effort, unnecessary controls and poor communication. GSL supports the view that XML can be used as a universal data standard for replacing the flat files exchange between production sites. It helps to sort the process data into logical order and present them in a readable format, outing next steps to be taken as the review progresses. A Real time Process Mining System (R-PMS) is then proposed to obtain the quality work piece in each production site and achieve the continual quality enhancement of the integrated workflow. The major tasks of R-PMS are (i) to assist the setup of a corporate database; (ii) to help the extraction of useful information from the database; and (iii) discovering the trends and relationships in the data in order to take proactive actions using that information.

5 PROPOSED ROADMAP FOR IMPLEMENTATION OF R-PMS

In order to develop a useful generic process mining approach, the following phases were taken into consideration in designing the R-PMS.

Phase 1: Infrastructural design of R-PMS

Based on the findings of this study, an approach is then formulated for the step-by-step development of the R-PMS. The hybrid Module (OLAP based Fuzzy approach) is an integral component of the R-PMS and can be considered as the most critical research element of this research. In particular, the inclusion of Fuzzy and OLAP will achieve a high performance of decision support functionality. The main development tools for the hybrid module includes the OLAP and Fuzzy packages plus other supporting programming and computational intelligence tools, which will be adopted by the research personnel to work out the design of the hybrid model. Computing tools related to object technology (Visual Basic) and Fuzzy (using a tool called MATLAB Fuzzy Toolbox) are already available as they have been in use over the past years by different investigators.

It is important that the hybrid model will integrate with the object technology to ensure the smooth running of the whole system. With the strengthening of the capability of the system realized by the augmentation of the OLAP based Fuzzy technology, the R-PMS can perform tasks which are normally seen to be difficult to achieve, such as fine tuning of process parameters based on varied finished quality. This can ensure that a more effective monitoring of the production workflow where early alertness of potential quality problems and the quick adaptation to improve are crucial.

Phase 2: Prototyping of R-PMS

With the availability of the framework of R-PMS, a prototype can be developed based on the infrastructural details and the design methodologies resulting from the Phase 1 research. Basically, there are primarily two modules within the R-PMS, namely the PRM and DSM. Building up a centralized relational data warehouse is the first step for the storage of distributed process data. It maintains the necessary information on PRM and building up a linkage between them.

Preparation for Process Relationship Module (PRM)

Before implementing OLAP approach on real-time process quality monitoring and reporting, the relationship of the fact tables must be defined in the OLAP server. In the data cube, the dimension, measures and calculated member is defined as follows.

Dimension

In “Project” dimension, the “Customer ID”, “Project No” and “Product barcode” fields are used to trace the reasons of failed products based on the complains from specified customers.

In “Process” dimension, the “ProcessID” field is used to find the quality information of work piece based on the specified manufacturing process.

In “Time” dimension, the “RecordDate” and “Time” field are used to find the defect statistics or quality performance of specified process based on different period of time.

In “Operation” dimension, the “MachineID” and “OperatorID” fields are used to find the quality characteristics of work piece machined by different operators and machines.

Measures

The “Product Acceptance_level”, “Frequency_of_rework” and “Frequency_of_scrap” fields are defined as measures that are used to provide the overview of quality records.

Calculated Member

The calculated member of OLAP is used to calculate the mean, standard deviation, accumulative defect statistics required for plotting the real-time two dimensional control charts.

Preparation for Decision Support Module (DSM)

The initial rule repository for this fuzzy module is given by the experience of expert. The behavior of the skilled operators is captured from the operation log book or interviews. Different membership function is determined by a single expert in the organization and must vary between 0 and 1. In general, there are several types of membership functions available, including Gaussian distribution function, the sigmoid curve, and quadratic and cubic polynomial curves. In this case study, trapezoid and triangular membership functions (Figure 5) are chosen to represent all process parameters as they provide an adequate representation of the expert knowledge. The expert knowledge related to the bonding, plating, painting

and casing production line is then defined as “if-then” rules, which are easily implemented by fuzzy conditional in fuzzy logic.

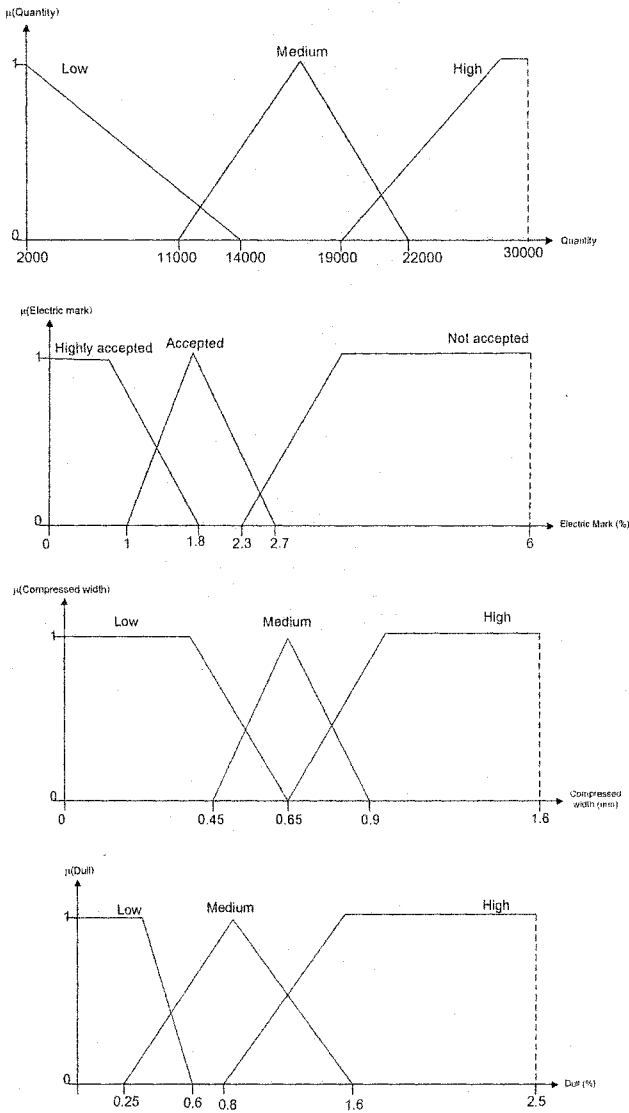


Fig 5. The fuzzy sets of the case example

Phase 3: Site-testing and monitoring the change process

The third phase is concerned with the overall site evaluation of the system. It is important that the R-PMS is able to be linked with other information systems and the integrated system is to be field-tested by the actual end-users in order to determine the possible problems when operating in a practical industrial environment. Basically, R-PMS is to be linked with other systems of the dispersed network such as user interface, system database and information update. Before this evaluation process, the application software programs have to be modified to suit the actual situation. The project team members will work closely with the potential end-users to ensure that the developed software can actually meet the practical requirements of the enterprise as well as other enterprises. A closely-monitored progress and continuous feedback and comments from the end-users will be proactively checked and followed. It is expected that a number of “bugs” will be found and subsequently, significant software refinement, updating and modifications of the original prototype programs and substantial site tests are all needed to cope with the possible flaws of the system. This is a continuous and arduous process until satisfactory results are obtained. It is also important to identify what further modification of the package is necessary before finalizing the design of R-PMS. As this phase covers substantial analytical effort and huge amount of programming work related to modifications and additions of source code as well as the successful application to production systems, duration of one year is needed.

6 EVALUATION

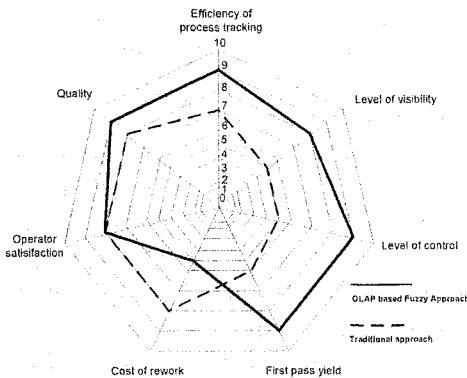


Fig 6. The radar chart for prototype for evaluation

There is a radar chart which illustrates the important categories of performance and makes visible concentrations of strengths and weakness compared with the traditional approach of electronic dictionary manufacturer implemented. Each spoke is subdivided into number of increments established in the rating scale. The scoring range is defined as 0 to 10 with 10 being full performance. The operators, process engineers and quality engineers were invited to participate in the prototype evaluation. The response rate was about one- fifth of the population. 136 responses are included in the summary statistics for the evaluation of R-PMS

For the radar chart (Figure 6) shown as above, those subjective attributes are re-served to the measurable dimensions which are (1) First pass yield (2) Level of control (3) Cost of rework (4) Level of visibility (5) Quality (6) Operator satisfaction (7) Efficiency process tracking. Referring to the radar chart, it is found that the first pass yield is enhanced. Operators found that R-PMS helps taking proactive actions to eliminate the quality problems in various areas which lead to poor customer satisfaction. Furthermore, it provides higher visibility of process because engineers can browse the quality trend in a real time manner.

In summary, the benefit of the system is concerned with improving quality of finished products in order to achieve cost-effectiveness and competitive advantages. However, there are other implications arising from the proposed model. The implementation of this system paves the way for a fundamental shift of enterprise strategy and sharpens the competitive edge of the company in the volatile and ever-changing industry.

7 CONCLUSION

This paper proposes an integrated system, R-PMS, which incorporates the concept of data mining and artificial intelligence, to form a robust approach for quality enhancement. It was pointed out that the OLAP based Fuzzy approach is feasible for discovering the hidden relationship and providing suggestions from vast amount of captured process data. The major contribution of the proposed system is to improve the way a discrete set of business activities is organized and managed. Further re-research on the structural configuration of the system is needed in order to further enhance its benefits. In general, this model paves the way for a novice approach to deal with the quality management by using process mining with proposed innovative information technologies. It is expected that this proposed system

will support manufacturing of quality finished products as to fulfill the customer satisfaction.

ACKNOWLEDGEMENTS

The authors wish to thank the Research Office of the Hong Kong Polytechnic University for the support of this research project.

REFERENCES

- [1] Berson, A. and Smith, S.J., (1997) Data Warehousing, Data Mining, & OLAP, McGraw-Hill, New York
- [2] Dick, Kevin (2000). XML A Manager's Guide: Addison- Wesley.
- [3] Erasala, Naveen., Yen, David C., and Rajkumar T.M. (2003), "Enterprise Application Inte-gration in the electronic commerce", Computers Standards & Interfaces, Vol 25, Issue 2, May 2003, Pages 69-82.
- [4]Grigori, Daniela., Casati, Fabio., Castellanos, Malu., Dayal, Umeshwar., Sayal, Mehmet. and Shan, Ming- Chien., (2004), "Business Process Intelligence", Computers in Industry 53 (2004) 321-343.
- [5] Hwang, San- Yih., Wei, Chih- Ping. and Yang, Wan- Shiou., (2004), "Discovery of temporal patterns from process instances", Computers in industry 53 (2004) 345-364.
- [6] MathWorks (2002). Fuzzy Logic Toolbox User's Guide. The MathWorks. Inc
- [7] Michael, L.G. and Bel, G.R., (1999), "Data mining - a powerful information creating tool", OCLC Systems & Services, Vol.15, No. 2
- [8] Peterson, T. (2000) Microsoft OLAP unleashed, 2nd edition, Sams Publishing, Indianapolis.
- [9] Ray, Pradeep. (2000). Cooperative management of enterprise networks: Kluwer Academic/ Plenum Publishers, New York.
- [10] Robert, S.C., Joseph A.V. and David B., (1999) Microsoft Data Warehousing, John Wiley & Sons.
- [11] Salvato, G., Leontaritis, P., Zelm, M., Rivers- Moore and Salvato, D., "Presentation and exchange of business models with CIMOSA-XML". Computers in Industry, Vol 40, Issues 2-3, November 1999,125-139.
- [12] Sarah, Cook. (1996). Process improvement: A Handbook for managers: Gower Publishing Limited, USA.
- [14] Tseng, Frank S.C. (2004).. " Design of a multi- dimensional query expression for document warehouses". Information Sciences (available online).
- [15] Van der Aalst W.M.P., Van Dongen B.F., Herbst J., Maruster L., Schimm G. & Weijters A.J.M.M. (2003). Workflow mining: A survey of issues and approaches. Data & Knowledge Engineering 47 (2003) 237-267.