Proceedings of the 5th World Congress on Intelligent Control and Automation, June 15-19, 2004, Hangzhou, P.R. China

第五届全球智能控制与自动化大会 2004年6月15 - 19日, 中国杭州

Research On the Application of Data-mining For Quality Analysis In Petroleum Refining Industry

Jiang Chen
Department of Automation
Tsinghua University
Beijing, 100084, People's Republic of China
chenjiang02@mails.tsinghua.edu.cn

Quanyi Fan
Department of Automation
Tsinghua University
Beijing, 100084, People's Republic of China
fqy-dau@mail.tsinghua.edu.cn

Bowen Xu
Department of Automation
Tsinghua University
Beijing, 100084, People's Republic of China
Yjb-dau@mail.tsinghua.edu.cn

ABSTRACT-**It is a main target of the petroleum refining industry to achieve a high product quality. It is widely admitted that there are some limitations of traditional product-quality -monitoring methods. Data-mining (DM) is a method to get useful information, which other regular methods cannot find, from enormous data. Data warehouse (DW) is the best way to store and manage massive enterprise data and provide a strong support to data analysis methods. This paper presents a new framework to deal with quality analysis, which combines the soft sensor, DM and DW. It is promising to overcome the limitations of soft sensor and apply soft sensor in quality analysis in the petroleum refining industry.**

Index Terms- data mining (DM) , data warehouse (DW) , soft sensor, quality analysis

I. INTRODUCTION

The refining industry is currently facing a difficult situation, characterized by increasing profit margins due to surplus refining capacity and increasing oil price. Simultaneously, market competition and stringent environmental regulations are forcing the industry to perform extensive modifications in its operation. Nowadays, advanced process engineering tools are extensively used by refineries to boost output effectively. Such tools range from advanced process control to corporate long-term planning, passing through process optimization, scheduling, and short-term planning. Advanced control often has its operating range arbitrarily constrained due to the existence of unmeasurable operating constraints. In general, operators deal with such constraints using feeling and experience, which implies that a reasonably wide safety margin must be maintained, resulting in nonoptimal operation [1].

The quality of production is the main one of the constraints. There are three main methods to measure quality: offline analysis, online instruments and soft sensor. Offline analysis cannot be used in process control. Online instruments are precise, but the high cost, large lag and difficulties in maintenance limit the application of such instruments. Soft sensor is a promising technique to solve the problem. It can calculate the quality from other measurable parameters, but it still cannot meet the requirement of accuracy in control under certain conditions. There are some successful applications of the soft sensor technique in industry but it's far from being widely used in petroleum refining industry, because the process of refining is very complex and the soft sensor models cannot fit under all conditions.

This paper tries to study this problem of data mining (DM) based on data warehouse (DW) and offer a framework of application. The main procedures, advantages and difficulties of its application are analyzed.

II. DATA MINING AND DATA WAREHOUSE

In order to implement this new idea, a review of date mining and data warehouse should be firstly conducted.

1. DATA MINING (DM)

Data Mining (DM) refers to extracting or "mining" interesting knowledge from large mounts of data. The knowledge is connotative, undiscovered useful information[2].

The purpose of Data Mining is to discover knowledge from large amounts of data. In the past decision-making support systems, the knowledge and rules of knowledge repository have been set up by specialists and programmers. The tasks of data mining are to discover the undiscovered knowledge from large mounts of data, and it is a process of taking knowledge automatically. The knowledge discovered by Data Mining can be shown as concepts, rules, regularities, patterns, constraints, or visualizations.

The process of Data Mining fundamentally consists of three stages: data preparation, mining operation, results presentation and explanation. The task of discovering knowledge can be viewed as the rotation of these three stages[3].

- Data Preparation: This stage can be further divided into three stages: Data Integration, Data Selection and Data Transformation. Here, Data Integration is to remove noise or irrelevant data and combine multiple

data sources. Data Selection refers to retrieve data relevant to the analysis task from the database. Data Transformation is to transform or consolidate data into forms appropriate for mining by performing summary or aggregation operations.

- Data Mining: This stage performs the actual mining operations. It consists of the following stages: (1) Set up assumptions: Discovery-Driven Data Mining and Verification-Driven Data Mining. The former lets Data Mining system set up assumptions for users, the latter let users themselves make assumptions about the knowledge hidden in the database. (2) Choose appropriate tools (3) Operation of mining knowledge (4) Verify the discovered knowledge.
- Knowledge Presentation: This stage analyzes the extracted information according to the end user's decision purpose, distinguishes the most valuable information and sends them to the decision-maker by decision support systems. In this stage, visualization and knowledge representation techniques are used to present the mined knowledge to users.

2. *DATA WAREHOUSE*

The term data warehouse (DW) was firstly introduced in [4] and was used to describe a subject oriented, integrated, nonvolatile, and time variant collection of data, in support of management's decision-making. Subject oriented, integrated, nonvolatile and time variant are the four main features of data warehouse.

The technique of DW includes: warehouse design, data extraction, data storage, data integration, warehousing specification & maintenance, optimizations, etc.

Fig. 2 shows the architecture of a DW application.

III. FRAMEWORK OF APPLICATION

DW and DM still have less application in petroleum refining industry of China, especially DM. A new framework of DM application in quality analysis is presented and its steps and problems are analyzed.

Fig. 3 presents the framework of a DM application based on DW:

Now, each step of this framework will be described following:

1. STEPS OF APPLICATION:

A. BUILDING OF DW AND PRODUCTION DATA MART

Building of DW and production data mart is the prerequisite to this framework. Building DW is a complex project, which is time-consuming and expensive. If quality analysis is the only intention of DW, apparently building DW is uneconomical. But DW is powerful and can bring much more benefits to enterprise than this. So building DW is the trend of enterprises including the petroleum refining industry. DW is usually considers as an existing source and it is ready for access.

The worst thing is that there is no DW available and also no planning to build it when this framework is to be implemented. At that time, the alternative is building production data mart, which is the direct object of this framework and is much easier to build.

Methods to building DW have been discussed intensively in the literature and a number of powerful software have been devised. All that one should do is to plan the DW carefully and choose the right software to implement it. So in this paper, this issue will not be discussed in detail.

B. RELATIVITY ANALYSIS OF VARIABLES

This is a very important step in this framework and it has not been performed in quality analysis before.

In the process of petroleum refining, there are many factors influence the quality, such as materials, product equipments, technical process, manipulation method, environment, etc. It's difficult to determine which factors are more important. On the other hand, it is unreasonable to choose all factors.

The common way to choose variables in building soft sensor model is to analyze the mechanism or by experience. At most time, the mechanism is too complex for a reliable analysis, and then the most useful is experience. As we

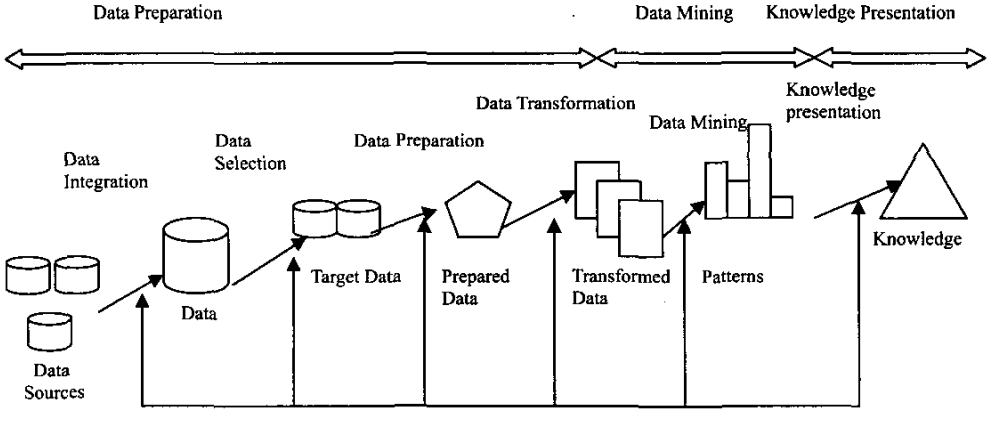

Fig. 1: Data Mining Process

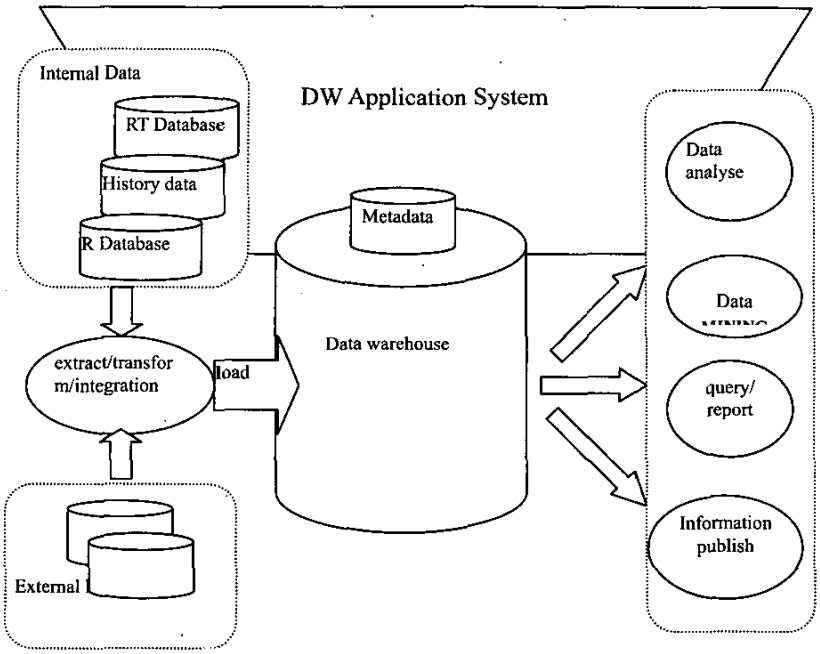

Fig. 2: DW Application Architecture

know experience is never right all the time and often misleads us and disables us to find the important factors which are thought to be trivial by experience.

In this framework, data-mining is used to analyze the relativity of variables by association rule analysis. Usually, association rule analysis is used in mining relativity of discrete variables, but almost all the variables discussed here are continuous. As mentioned in [5], they can be discretized by several methods. By this approach, the most relative variables can be chosen and then the next step can be proceeded to.

C. *MINING SOFT SENSOR MODEL*

After the variables are chosen in the last step, the next step then is to mine the soft sensor model. In traditional soft sensor modeling process, neural network is widely used to train the model. As for data mining, this process is often call as *prediction*. Neural network is one of the important algorithms of prediction in data mining.

In the traditional soft sensor technique, the method of data set selection is to sample data in specified time span, process them and leave a part of data which are useful. Obviously, this data set can't cover all kind of variance such as season, product scheme, petroleum quality, etc. As a result, the model trained by this data set tends to bring precise outputs in some conditions but it may be not appropriate in other situations.

In this framework, the DM technique based on DW is utilized, and then with the aid of massive data which cover

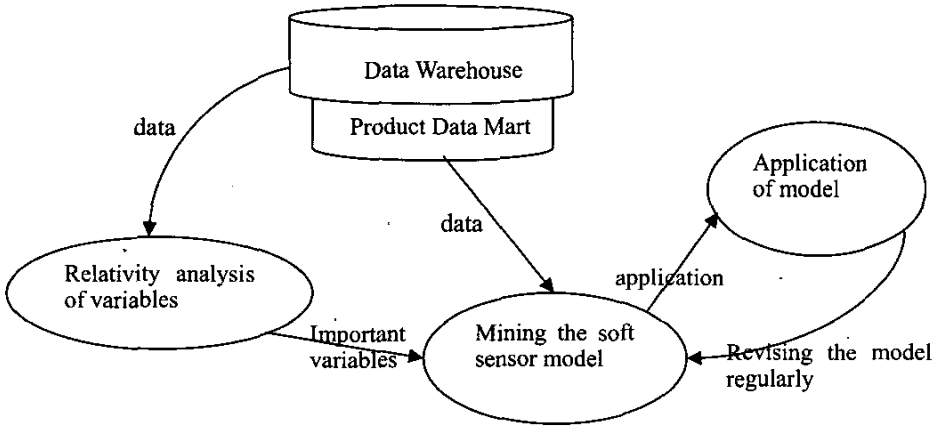

Fig. 3: Application Framework

almost all kinds of conditions in the DW, a more stable model or a set of models can be trained.

D. REGULAR REVISING OF MODEL

Unfortunately, there is no model which can cover all conditions. Consequently, we must revise the model according to the variation of conditions. The traditional soft sensor technique also has online revising function. However, it's difficult to determine the revising method. The revising method itself is fixed after being determined. Furthermore, the structure of the model cannot be revised by this approach. Therefore, such revising method often cannot adapt themselves to the variation of conditions.

In this framework, the DW-based DM is used to revise the model. Certainly, the common revising method has to be used to achieve online revising. The whole model including the online revising method can be revised by the DM technique. This process of revising is the same as modeling. In the following analysis, we can see why this method is useful and practical.

2. ADVANTAGES OF THIS FRAMEWORK

In the last section, all steps are described, and subsequently, it has been shown that why this framework is useful. With the collaboration of data warehouse and data mining, this framework has following advantages:

A. AUTOMATION OF DATA PROCESSING.

In the common DM and soft sensor technique, data processing is usually difficult and time-consuming. With the aid of DW technology and corresponding software, almost all part of this process can be performed automatically.

B. VARIABLE SELECTION AND DATA CAPACITY.

The most important factors which influence the precise of training model are variable selection and data capacity used. The selection of variables in traditional soft sensor technique is often limited within a narrow range, because this selection is typically based on experience and some useful variables are often omitted. The capacity of data used in traditional soft sensor technology is also often limited, because the data is often sampled only when we tend to train the model, and certainly it cannot cover all range of conditions. DW gives us a solution to these problems. In DW, the snowflake model can be used to fetch specific data set conveniently. Almost all kind of variables which possibly influence the quality can be taken into consideration. Considering massive history data have been stored in DW, we have abundant data to train the model.

C. AUTOMATION OF MODEL REVISING.

In traditional soft sensor technique, once the model is trained, the structure of it will never be changed, because the model adjustment can only be accomplished by experts and usually it's impossible to invite the experts to do it regularly. In this framework, the automation of model revising can be attained.

The main reason why experts are needed to revise the model in traditional method is the difficulty of data selection. But with the aid of DW software, this process can be much easier. Furthermore, because the data are nicely arranged in DW, the DM process can easily get required data from it. The DM process can be automatically done once it is designed. As a result, this framework achieves the automation of model revising.

Furthermore, this framework can obtain another kind of model revising, that is, the model can be revised before the precision of the model exceeds the limit. The way to do it is that a set of models can be mined from the DW according to different conditions which cannot be brought into the models, and then we can use different models in specific conditions. Theoretically, all important factors should be used in the model. However, to satisfy the online-measurable requirement of the model, some important factors cannot be used to train the model, but they can be used to mine a set of soft sensor models and according to the variation of these factors, corresponding models are adopted.

3. DIFFICULTIES OF THIS FRAMEWORK

The advantages of this framework have been discussed in the previous section, but some difficulties of this framework should also be aware of.

A. BUILDING OF DW

Building of DW itself is a complex project, and it's the foremost difficulty of this framework. As mentioned in the description of the first step, DW is very important to enterprise's informatization and then building DW is the necessary choice of petroleum refining industry. As a result, the DW is assumed to be available when developing this framework, and what one should do is just to use it. Otherwise, building production data mart is much easier than building data warehouse.

B. CHOICE OF MINING ALGORITHM

With regard to the mining of soft sensor model, neural network is a reasonable choice. Neural network still has many variations and so it's not very straightforward to choose the right one. After all, there are many successful modeling cases to refer to. The relativity analysis of variables is not the same thing. There are rare cases about relativity analysis of continuous variables, so it's difficult to choose the right algorithm. As mentioned in [5], the continuous variables can be discritized, but the choice of parameters is still a problem.

C. PERIOD OF MODEL REVISING

One of the advantages of this framework is that the model can be revised periodically. But problem is then brought up, that is, how to determine the period of model revising? The process of petroleum refining is a slow and big-lag process, along with the online-revising function of model, the life of one model will not be too short and then the period of model revising can be long. Due to the same reason, if the model is revised only after the precision of model exceeds the limit, the correct result will also be delayed for some time. As mentioned in the third step, the model can be revise in advance according to the other influent factors. It is still difficult to choose the actual time required.

IV. CONCLUSIONS

In order to overcome the limitations of traditional soft sensor technique in quality analysis of petroleum refining

industry and implement optimal control at last, a new framework at the view of DW-based DM is presented and its advantages and difficulties are analyzed. Although this framework is based on the quality analysis of petroleum refining industry, it can also be applied as a common framework in process industry or other industry.

REFERENCES

[1] Lincoln F. Lautenschlager Moro, Process technology in the petroleum refining industry-current situation and future trends, computer & chemical engineering, 27 (2003) 1303-1305

[2] Apte, C. Data Mining: an industrial research perspective. IEEE Computational Science & Engineering.1997, 4(2) : 6～9

[3] Bhavani M.Thuraisingham, Marion G.Geruti. Understanding data mining and applying it to command, control, communications and intelligence environments. The 24th Annual International Computer Software and Applications Conference (COMPSAC) .2000:171-175

[4] W.H.Inmon. Building the Data Warehouse, Second Edition. John Wiley&Sons,Inc. 1996

[5] Jiawei Han, Micheline Kamber. Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers.Inc. 2001