Industrial Diagnoses by Hyperspace Data Mining

Nianyi Chen and Dongping Daniel Zhu
Zaptron Systems, Inc. 415 Clyde Avenue, Suite 198,
Mountain View, CA 94043 USA
Email: dan@zaptron.com

Wenhua Wang
Salomon Smith Barney, 250 West Street 8th Floor, New York, NY 10013 USA

Abstract

Data mining offers an effective solution to product and process diagnosis in that it requires no additional equipment or capital investment and it does not cause interruption to production and operations. The most important role of data mining is the ability to separate data into different classes so that an accurate model can be built for fault detection, recognition and prediction. The proposed data mining methodology consists of data separability, hidden projection, back mapping, feature selection and reduction, and model building. The solution can be quantitative or qualitative depending on the pattern of the original data set. When combined with other soft computing techniques, data mining method can provide Application examples are briefly described to demonstrate the efficacy of the new method in diagnosis.

I. Introduction

Basic Diagnostics

A fault is an abnormal state of a machine or a system, including dysfunction or malfunction of a part, an assembly, or the whole system. The occurrence of a *fault* is associated with a number of *factors*, which in turn are related to a number of *symptoms*. Fault diagnostics is the study of the relationship of fault, factors and symptoms, and it is to predict and control the performance of a system, be it a communication system, a semiconductor manufacturing equipment, or even a human body.

For equipment fault diagnosis, sensor data are usually used to compute a number of device parameters in hope to detect, isolate, and recognize (classify) fault in the equipment. Mathematically, this involves two transformations: one from measurement space to feature space, another from measurement space to decision space.

Objectives of Fault Diagnosis

Efforts in fault diagnostic research and development seek to provide advanced warning and prediction on consumable devices, to avoid parts replacement at the last minute before a scheduled batch production job. In some equipment setups, it is necessary to classify the failure patterns (modes) in the various hardware and software configurations. The ultimate objective of diagnosis is to offer the capability to automatically schedule preventive maintenance and corrective maintenance.

In the current Internet age, devices and equipment are or can be connected to and accessed over the Internet. This new scenario requires that diagnosis products must be web-based to provide remote diagnosis and fixing. Agent-based techniques are being developed that enable access, communication, diagnosis and control.

Diagnostic Techniques

Fault diagnostics is a complex issue that calls for many different techniques to offer effective and efficient solutions. The most native method, called first-principle method, is based on fault physics and chemistry of the underlying process and materials. The fault tree method is used in root case analysis and reliability analysis to define relationships among symptom, factors and events (failure). Probabilistic, estimation and information theoretical approaches are used in fault detection and isolation, while self-learning, fuzzy expert system methodology is utilized in automated fault explanation and rule discovery (why it happens) where numeric and linguistic information can be fused together. In more recent years, to overcome the difficulty in obtaining exact physical, chemical and statistical models of a system under study, neural networks and real-time data mining techniques are used for fault recognition by using real-time measurement data.

Data mining techniques [1] [2] [5] [6] [7] offers an effective solution to product and process diagnosis, by using a number of computational methods in pattern recognition, artificial neural networks and regression analysis. Data mining method is especially suitable for processing complex data sets in *non-linear*, *highly noisy*, and *multivariate* environment, and to build effective models for fault isolation, recognition, and prediction.

Complex data are very common in many practical systems, such as in industrial diagnosis and optimal controls, new product and new materials designs, medical diagnostics and treatment, geological exploration, environmental monitoring, economical and financial analysis, and so on. Therefore, development of advanced data mining techniques is a challenging and important task.

Traditional linear regression method is not appropriate for processing such complex data sets. Professor Lofti Zadeh of UC-Berkeley has stated that different techniques of soft computing are synergistic rather than competitive. This point of view is in agreement with our experience in applying pattern recognition and other computation techniques to industrial optimization and diagnosis. We have developed an innovative methodology of data mining for processing industrial data that has been proved to be very effective in various applications. The applications include, but not limited to, process design optimization, diagnosis, and pattern recognition. In this paper we present its principle and application to product and process diagnosis. The proprietary techniques have been built into a software suite, MasterMiner™, that provides a set of effective and user-friendly tools to solve real-world, complex data mining problems where various intrinsic data structures (models) manifest themselves.

Section 2 review various computational methods used for diagnosis with discussions on their advantages and limitations. Section 3 describes the proposed hyperspace data mining techniques including data separability, envelope method, feature selection and reduction, auto-boxing method. Section 4 shows a few real-world examples of using MasterMiner in diagnostic applications.

II. Problem Background

By nature, a diagnostic problem is an optimization problem, and methods in pattern recognition and data mining can be used to offer effective solutions. Most pattern recognition methods are based on the computerized recognition of the multidimensional graphs (or their two-dimensional projections) of the distribution of samples from different classes in a multidimensional space. Independent variables (often called system input, features or factors) influencing the target (dependent variable or system output) are used to span a multidimensional space.

We can describe samples of different classes as points with different symbols in these spaces. Various pattern recognition methods can be used to "recognize" the patterns shown in the graph of *distribution zones* of different samples. In this way, a mathematical model can be obtained that describes the relationship (or regularity) among targets and factors. If we adjust criterion of classification, semi-quantitative models describing the regularities can be found, if noise is not too strong.

Unlike various regression methods (linear regression, nonlinear regression, logistic regression, etc.) or the artificial neural networks (ANN) [4] that provide *quantitative* solutions, pattern recognition methods often provide *semi-quantitative* or *qualitative* solutions to classification. This is of course a limitation of pattern recognition methods. However, this is not always a disadvantage, because many data sets exhibit strong noise, and a quantitative calculation would be *too precise to* present them. Besides, practical problems in many cases are of the "yes or no" type, and pattern recognition is especially suited to offering adequate solutions to them. For example, a problem may be "whether the fault will occur or not", or "whether an intermetallic compound will form or not."

A number of common pattern recognition methods have been built into MasterMiner software. They include Principal Component Analysis method (PCA), Fisher method (Fisher), Partial Least Square method (PLS) [3], Sphere-Linear Mapping method (LMAP), Envelope method (Envelope), Map Recognition method (MREC), and Box-Enclosing method (BOX). PCA, PLS and Fisher methods are traditional pattern recognition methods, and their principles are described in standard textbooks. In general, limited separation is achieved by traditional PCA or Fisher method when the data exhibit strong non-linearity. The other four methods listed above are developed specially for processing complicated data sets.

In the sphere-linear mapping (LMAP) method, computation starts by moving the origin of the initial data coordinate system to the center of sample points of class "1", followed by finding a hyper-ellipsoid to enclose all sample points of class "1". By a *whiten transformation*, this hyper-ellipsoid is changed to a hyper-sphere. The multidimensional graph of the data points is then projected onto a series of two-dimensional planes to form a number of 2-dimensional maps on the computer screen.

III. A Hyperspace Data Mining Method

Data mining [1] [5] [6] is the process of discovering meaningful new correlations, patterns and trends by sifting through large amounts of data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques. Data mining in fact is an optimization technique, and it has found practical applications in diagnosing products and process in various industries, including steel making, power generators, petro-chemical, materials design and manufacturing, drug screening and production, and operations management [1] [2].

Data Separability by MREC

The data separability test of MasterMiner is designed to investigate the possibility of separating data points from different populations or clusters in a hyper-space. If the data are separable, it may be possible to build a mathematical model to describe the system under study. Otherwise, a good model can not be built from the give data and more data or data processing is needed.

MREC (map recognition by hidden projection) is a method used to choose the "best" projection map from a series of hidden projection maps. Here "best" implies that the separation of sample points of different classes is better than those obtained from other maps. In MasterMiner, MREC is used together with the "auto-square" method to provide the so-called MREC-auto-square solution. In each projection map, sample points of class "1" are automatically enclosed by a square frame (as shown in Fig. 1), and a new data set is formed that contains only sample points within this "auto-square." This new data set will be used to in model building for diagnosis or failure recognition.

After this auto-box operation, a second MREC is performed on the new data set to obtain a new "best" projection map where sample points remaining in the auto-square are separated better into 2 classes. After series of such hidden projections, a complete (close to 100%) separation of sample points into different classes could be realized. It has been shown that MREC-auto-square method is much more powerful than the traditional patter recognition method.

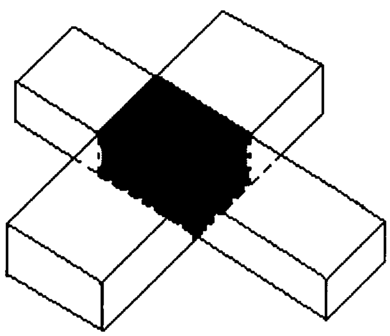

Fig.1 Polyhedron formed by square tunnels.

The physical meaning of MREC-auto-square method is explained as follows: each "auto-square" represents actually a square "tunnel" in the original multidimensional space, and several such tunnels would form a *hyper-polyhedron* in this space by intersection of all tunnels, as shown in Fig. 1. This hyper-polyhedron, enclosing all "1" sample points, defines an optimal zone in the multidimensional space, if all or most of the samples of class "2" are separated from this zone because they are located outside of this hyper-polyhedron.

Source of inseparability: If the result of separability test is not very good, one has to consider the source of the inseparability of a data set. Generally speaking, there are two kinds of source of inseparability in data: (1) the data are too noisy, and (2) the form of the optimal zone (the distribution region of "1" samples in a multidimensional space) is so complicated that it cannot be described by a single *convex* hyper-polyhedron. Many practical examples have confirmed these two types of inseparability. An effective approach to the second type of inseparability is by "local view" treatment whereby one can cut a multidimensional space into several subspaces to achieve better data separability in each of the subspaces.

Back Mapping

Since the hidden projection by MREC transforms data from the original measurement (or feature) space into a number of other orthogonal spaces, we need to back map the transformed data into the original feature space to derive mathematical models for practical use. A method called PCB (principal component backing) [2] has been developed whereby a point representing an unknown sample from a low-dimensional principal component subspace is back-projected to the high-dimensional space of original features. In PCB, non-linear inverse mapping and linear inverse mapping are used to obtain an accurate solution to predicting sample points in the optimal region.

Let **X** be a standard training set with n samples and m features, and let **Y** be the sample set in the PC (principal component) space corresponding to **X** in the original feature space. We have

$$\mathbf{Y} = \mathbf{XC}$$

Here the columns of **C** are the eigenvectors of the covariance matrix **D** ($\mathbf{D} = \mathbf{X}^T\mathbf{X}$) for the training set **X**. The 2-d subspace of PCs consisting of **Cu** and **Cv** is defined as the main map where samples are assumed to be completely classified. A point P in the main map represents the unknown sample, and it is described by two variables, y_{pu} and y_{pv}, respectively. In general, p is expected to be an optimal sample if its neighbors are optimal points. In order to back transform an unknown sample point to the original space, i.e., to find $\mathbf{X}^*_p$, one has to determine its boundary conditions; otherwise, an uncertain solution will occur. Two types of conditions are proposed for PCB: non-linear inverse mapping and linear inverse mapping.

In non-linear inverse mapping (NLIM), let the error function **E** be defined as

$$E = \left(\left(\sum_{j=1}^{n} d_{pj}\right) \right)^{-1} \sum_{j=1}^{n} \frac{(d_{pj} - d_{pj}^*)^2}{d_{pj}}$$

where

$$d_{pj}^* = \left(\sum_{k=1}^{m} (x_{pk} - x_{jk})^2\right)^{1/2}$$

$$d_{pj} = [(y_{pu} - y_{ju})^2 - (y_{pv} - y_{jv})^2]^{1/2}$$

Here d_{pj} is the distance from the unknown sample represented by p to all known samples in the subspace defined by PCs two coordinates, u and v, and is the same distance in the original feature space. Non-linear optimization method is utilized to compute the values of z_{pk} that minimizes the error function **E**. The solution using NLIM boundary condition is only an approximation, because the parameters obtained in this way depend to some extent on the trial coordinates in the original space.

Linear Inverse Mapping (LIM): Besides the 2-d subspace of PCs consisting of $\mathbf{C}_u$ and $\mathbf{C}_v$, there exists an $(m\text{-}2)$-dimensional subspace of PCs consisting of $\mathbf{C}_i$ (i = 1, 2, ..., m and $i \neq u, v$), since **C** is derived from the covariance matrix **D** (m x m). When the projection of point p, which is described by y_{pv} and y_{pv} in the main map, are determined with y_{pi} (i = 1, 2, ..., m and $i \neq u, v$) in the $(m\text{-}2)$-dimensional subspace, a set of simultaneous *linear* equations cab be obtained as

$$y_{pk} = \sum_{j=1}^{n} c_{jk} x_{pj}$$

Where $k = 1, 2, ..., m$, and the set of linear equations can be solved for the parameters of the unknown sample point corresponding to point p. The linear inverse mapping will always produce an exact solution. As to the projection of point p, in general one can let point p be at the center of the region containing the largest number of known optimal samples so that the unknown sample has properties similar to its neighbors, the known optimal samples.

Envelope Method

Envelope method is another method powerful for separating sample points into different classes. In this method, no projection is made. In computation, a series of *hyper-planes* automatically create a smallest *hyper-polyhedron* that encloses all sample points of class "1", and form a new subset of data that include only the sample points within the hyper-polyhedron. Since data separation takes place in the multidimensional space directly, the results of separation is usually very good, provided the original classification of different class in multidimensional space is well define.

The rate of separability R is defined as R = (1-N2/N1). Here N_1 is the number of the sample points of class "1", and N2 the number of "2" points remaining within the hyper-polyhedron and not separated from '1' points. If R > 90%, the separability is "excellent." If 90% > R > 80%, the separability is "good." If 80% > R > 70%, the separability is "acceptable." If R < 70%, the separability is "unsatisfactory." If the largest R is less than 80%, one needs throw away samples from the given data set to improve separability before feature reduction and modeling. Separability test by "Envelope" method is simpler than that by MREC method. If the separability by Envelope method is satisfactory, the MREC operation will not be necessary. But if the results of separability test by envelope method are not good, one should perform a MREC operation since the projection maps by MREC can offer useful information about the inseparability of the data set under study.

Interchange of classes: To determine if a data set is completely separable, it is needed to interchange sample points of type "1" and type "2", and perform "separability test" again by using the same procedure. This will produce a second rate of separability R. The final separability of a data set is determined by the largest R value obtained in the two tests.

Auto-box for Concave Polyhedron

Since both the MREC and envelope method can only form a *convex hyper-polyhedron*, they can not separate the sample points of different classes if the distribution zone is not a *convex* but a *concave* polyhedron. In these cases, another technique, the BOX method shown in Fig. 2, can be used as an effective solution. When there are some sample points of class "2" still remaining within the hyper-polyhedron after applying the envelope method, a smallest multidimensional "box" will be built by MasterMiner to enclose all sample points of class "2" (or class "1").

We can build a box within the hyper-polyhedron in an effort to enclose all "2" sample points. If the box so built contains no sample point of class "1" (or only very few points of class "1"), rather good separation will be achieved by throwing away those sample points (mostly of type "2") in the box. In fact, BOX of MasterMiner is a tool for virtual data mining. It can also be used to separate sample points of class "2" from that of class "1". BOX method itself has limited use, but it becomes a very useful tool in virtually mining the data space when combined with the MREC or Envelope geometrical method.

Fusion of Data Mining with Other Methods

The effectiveness of MasterMiner in processing complicated data sets depends not only on data mining techniques, but also on other computational methods. It derives effectiveness from using data mining and pattern recognition techniques, and achieves reliability by using other techniques, such as artificial neural networks and regression methods.

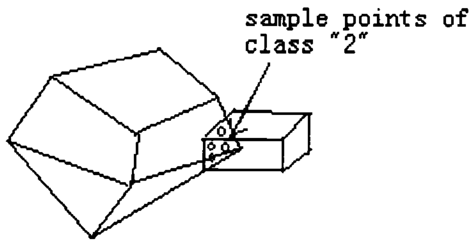

Fig.2 Principle of the Envelope-Box Method

For instance, we can use the set of inequalities obtained by MREC method, which describe the boundary of optimal zone, as the boundary condition for the mathematical model obtained by ANN. This will compensate for the effect of inaccurate and unreliable prediction by ANN.

We can also use the analysis result by MREC to help select a nonlinear function for regression computation. For example, in one process optimization project we have used quadratic terms of the original features as additional terms for nonlinear regression. In this case, a nonlinear function with 9 feature terms was first created that gave a low PRESS (predication residual error squared sum) value of 0.15. However, the result by MREC indicates that the principal components PCA(1) and PCA(2) give very good separability. Therefore it is reasonable and feasible to build a nonlinear function based on these two features, instead of the original four features. In this way, a final nonlinear function with fewer feature terms was obtained that had similar prediction performance (PRESS was 0.128) as the first non-linear model with 9 terms.

Solutions to Multi-Target Problems

Many practical problems are multi-target problems in nature. For example, industrial diagnosis and optimization tasks require high quality, low cost and high yield at the same time. In MasterMiner, the multi-target problems are addressed by the following two methods:

(1) Treat samples satisfying all targets as members of class "1", and the rest of samples as members of class "2". Then design a *desirability* function **F**(t), that is suitable for target values t*1*, t2, ..., t*n*, to represent the optimal condition or desirability of the problem: **F**(t) = **F**(t*1*, t2, ..., t*n*), and use **F**(t) as an alternative target for the original problem.

(2) Treat every target independently and build a mathematical model for each separately. Take the intersection of all optimal zones as the overall optimal zone of the original problem.

IV. Diagnostic Application Examples

The objective of industrial diagnosis is to find the fault or bottleneck that reduces production yield or product quality by analyzing available data. Data mining is a powerful solution to this task, since is requires no additional investment in equipment, and causes no interruption to operations.

Case 1. ST14 steel plate making

An iron and steel company uses an advanced steel rolling system made in Germany to produce more than 100,000 tons of ST14 steel plate per year. The product, ST14 steel plate, is used for automobile body. The problem is that the deep pressing property of the ST14 plates is off-spec and the company wanted to improve the product quality so that the fraction of the off-spec product is reduced.

The steel plate manufacturing process can be divided in 5 stages: blast furnace, steel-making, casting, hot rolling, and cool rolling. There are 20 factors in each of these 5 stages, and the total number of factors is 5x20 = 100. With the given field data, we used MasterMiner to have identified 2 major factors in the last stage (cool rolling) that have most influence on the product quality: (1) Nitrogen (N_2) content, and (2) the distance between the two cool rollers. Data mining has found better diameters for the two rollers that could improve the quality without changing the system setup. After adjusting the roller diameters acccordingly, the fraction of off-spec product was reduced by 5 times.

Case2. Bottleneck in Gasoline Productivity

In oil refineries, distillation towers are used where crude oil is fed through an inlet to the tower inside which heat is applied to produce gasoline as well as other petrochemical products. An oil refinery factory requested us to find a solution to increasing the yield of gasoline in a reduced-pressure distillation tower. Data mining techniques are used to process the data record from the distillation tower so as to find and fix the bottleneck in fractionation. We found that the length of the cooling coil on top of the tower was too short. In this process, an unexpected result was first obtained: the pressure at the tower top increases if the action of vacuum pump intensifies. This conclusion seemed to be in direct contradiction with common knowledge in practice. But this is indeed true in this special case. After study, we understand it as follows: given the tower an as open system, when the action of vacuum pump intensifies, a large quantity of oil vapor will flow up. However, if the water cooling coil on the tower top is not adequately designed, the pressure may be increased. So the cooling coil is the "bottle neck" and its length should be adjusted accordingly. After increasing the cooling coil length, the gasoline yield was indeed increased by 10,000 tons per year.

Case 3 Ethylebenzene Production

In petroleum refineries, ethylbenzene is a product in platinum-reforming plans where a catalyst cracking process takes place. Naphta is fed through an inlet to a reactor which is linked to a fractionnation tower where heat and platinum catalyst are applied to produce ethylbenzene. A chemical analysis on the down-flow of the ethylbenzene tower at an oil refinery indicated that the yield of ethylbenzene was rather low. The diagnostic task was to isolate and fix the problem (fault) in the workship to increase ethylbenzene yield. Data mining techniques were applied to diagnose the equipment setup of the the platinum reforming workshop, and we detected that the position of the inlet was not optimized. After changing the position of inlet from 99-th plate to 111-th plate, the yield of ethylbenzene was increased by 35%.

IV. Conclusions

This paper presents basic background on diagnostics, reviews technologies that are used for diagnostic purposes, and proposes a hyper-space data mining methodology. Data mining offers an effective solution to product and process diagnosis when combined with other pattern recognition and statistical methods. The most important role of data mining is the ability to separate data into different classes so that an accurate model can be built for fault detection, recognition and prediction. The proposed data mining methodology consists of data separability, hidden projection, back mapping, feature selection and reduction, and model building. When data exhibit linearity, a quantitative model can be built. When they exhibit non-linearity, a semi-quantitative or qualitative model can be obtained. Application examples are briefly described to show the efficacy of the proposed data mining method for product and process diagnosis.

References

[1] N. Y. Chen and D. Zhu, "Data mining for industrial design and diagnosis," Proceedings of the First International Conference on Multisensor-Multisource Information Fusion, Las Vegas, NV, July 6-9, 1998.
[2] H. Liu Y. Chen and N. Y. Chen, "PCB Method applied to material design - computer aided synthesis of a super-conductor," Journal of Chemometrics, Vol. 8, pp. 429-443, 1994.
[3] P. Geladi, B. R. : Analytica Chemica Acta, vol. 185, No. 1, 1986
[4] J. Holland, "Adaptation in nature and artificial intelligence," The University of Michigan Press, Michigan, 1975.
[5] Xiaohua Hu and Nick Cercone, "Data Mining via Discretization, Generalization and Rough Set Feature Selection," Knowledge and Information Systems: An International Journal, Vol. 1, No. 1.
[6] U. Fayyad, "Knowledge discovery and Data mining: towards a unifying framework," Proc of KDD-96, Menlo Park, CA, AAAI Press, p. 82-88, 1996.
[7] S. Anand, EDM, "a general framework for data mining based on evidence theory," Data and Knowledge Engineering, Vol. 18, pp. 189-223, 1996.