



Binary ant colony optimization applied to variable screening in the Mahalanobis–Taguchi System

Edgar Reséndiz*, Luis A. Moncayo-Martínez, Georgina Solís

The Technological Institute of Saltillo, Saltillo, CP 25280, Mexico

ARTICLE INFO

Keywords:

Binary ant colony optimization
Dimensionality reduction
Mahalanobis–Taguchi System
Variable screening

ABSTRACT

This work presents the application of the Mahalanobis–Taguchi System (MTS) to a dimensional problem in the automotive industry. The combinatorial optimization problem of variable selection is solved by the application of a recent version of binary ant colony optimization algorithm. Moreover, a comparison with respect to binary particle swarm optimization algorithm is also presented and a discussion regarding the numerical results is given.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Mahalanobis–Taguchi System (MTS) is a relatively new collection of methods proposed for behavior patterns diagnosis and prognosis corresponding to multivariate data sets. An initial methodology was proposed by Genichi Taguchi, who is well known for his controversial ideas and methods used in the design of experiments. Taguchi's philosophy defines “quality” as the loss that a product causes to society and the effect of non-quality products is measured through the so called “loss function” which involves financial loss to the industry specifications, that is to say the language of money related to language engineering specifications. Furthermore, “off-line” quality control consists of experimentation in order to improve product design and process. In order to achieve rapid improvement over costs and quality by optimizing product design and manufacturing processes, several statistical techniques and engineering procedures can be combined. A very important concept which should be taken into consideration is “robustness” that makes a product/process insensitive to factors effects or variables that can not be eliminated affecting in this manner the final product quality.

Nowadays with the rapid growth of technology, organizations have the ability to collect a large amount of data required for processes analysis but more importantly for modern business it means to deal effectively with the vast amount of data, i.e. the extraction of important information from large databases, (Rai, 2009; Taguchi & Jugulum, 2002). The MTS is a prediction and diagnosis method which is used to analyze data patterns in multivariate systems. If one is interested on the system improvement or it is looking forward to optimize a system it is necessary the deep understanding

of those variables and noise conditions which are really affecting the system performance since not all variables have the same level of impact (Cudney, Hong, Jugulum, Paryani, & Ragsdell, et al., 2007; Cudney, Jugulum, & Paryani, 2009, 2006; Cudney, Paryani, & Ragsdell, 2008; Kim, Tsui, Sukchotrat, & Chen, 2009).

The Mahalanobis–Taguchi System has gained great interest because it has proved its efficiency in the performance of systems diagnosis and prognosis using data sets and correlated features (Ragsdell, Cudney, & Paryani, 2007) even though the simplicity of MTS is considered ad hoc in the sense that it is a method that was developed without the use of any underlying statistical theory (Kim et al., 2009). Mahalanobis–Taguchi System is mainly developed within four steps: (1) the building of the Mahalanobis space (MS), (2) MS validation, (3) identification of useful variables and (4) prediction and system diagnosis. (Cudney et al., 2007, Cudney, Jugulum, & Paryani, 2006; Rai, 2009; Rai, Chinnam, & Singh, 2008; Taguchi & Jugulum, 2002).

Regarding the optimization strategy we focus on the ants colony optimization (ACO) which is a novel, nature-inspired meta-heuristic that mimics the real foraging behaviour of ant colonies and the concentration/evaporation of chemical substances called pheromones (Dorigo & Stützle, 2010).

When ants begin looking for food, they explore the forage area randomly depositing a quantity of pheromones all along their trail. As soon as an ant finds food, it returns to the nest using the path built on its forward trip, reinforcing the pheromones deposited over it. While other ants are looking for food or when ants leave the nest, they tend to choose (in probability) trails that have a strong concentration of pheromones. As a general rule: the higher the quantity of pheromones over a trail, the higher the probability that ants follow it. The quantity of pheromones concentrates over a path because they are a function of time, i.e. if pheromones are not reinforced in short periods of time, they evaporate. It is clear that the shorter the distance from the nest to the food, the faster the

* Corresponding author.

E-mail addresses: eresendiz@its.mx (E. Reséndiz), lmoncayo@its.mx (L.A. Moncayo-Martínez), gsolis@its.mx (G. Solís).

ants complete the tour, thus the concentration of pheromones is higher in short trails than in long ones. As pheromones over long trails are not reinforced at the same rate, they evaporate.

As in real ants, the artificial ones (*A*), also called agents, cooperate to find solutions to hard combinatorial problems by reacting to changes in their environment using an indirect communication process based on artificial pheromones (τ), thus near – and optimum solutions emerge from agent’s cooperative interaction. The problem to be solved is represented by a graph in which the pheromones are deposited over the set of either vertices or edges. An initial and a final condition represent the nest and the food, respectively. Ants build a solution by stepping from vertex to vertex based on: (a) a probabilistic decision rule that is a function of the quantity of τ over either vertices or edges and the heuristic information (η) which is a source of information that is not related to ants’ parameters and gives *A* the opportunity of exploiting the specific knowledge of the problem; and (b) the set of problem constraints.

The ACO meta-heuristic has three basic procedures that can take place simultaneously or one at a time. The first procedure is the *Construct Ant Solution* which allows *A* to build a solution, i.e. *A* travels around a graph which represents the problem. They know they have to stop travelling when the final condition is reached. The *Apply Local Search* is the second sub-process. This is optional and aims to improve the ant’s solution by applying a local search. Finally, in the third procedure called *Update Pheromones*, the pheromones trails are reinforced according to the solution’s “quality” and are evaporated by an evaporation factor (ρ) to avoid stagnation due to an indiscriminate concentration of pheromones.

2. Methodology

This section presents a short overview of the hybrid methodology using particle swarm optimization developed in Pal and Maiti (2010) followed by the binary ant colony algorithm proposed in Jang (2011) used as an alternative optimizer to solve the combinatorial optimization problem of variable screening.

2.1. Review of hybrid Mahalanobis–Taguchi system

The hybrid MTS methodology consists on the following main steps:

1. Population data pre-classification.
2. Mahalanobis distance (MD) calculation.
3. Screening of variables.
4. Integer programming formulation problem.
5. Application of binary particle swarm optimization for combinatorial problem solving and identification of useful variables.

In the first stage of this methodology data are collected and classified in two sets which are commonly called the *healthy* group, which stores observations conforming to specifications, and the *unhealthy* group containing observations not conforming to specifications. Once the two data sets are built a scaled MD is used as metric to build a baseline multivariate space. This scaled MD is defined as follows:

$$MD = \frac{1}{k}(\mathbf{z}^t R^{-1} \mathbf{z}) = \frac{1}{k} D$$

where

- \mathbf{z} = column vector of standardized variables
- R = Correlation matrix
- k = Number of variables
- D = Standard Mahalanobis–Distance : $\mathbf{z}^t R^{-1} \mathbf{z}$

The next step is screening of variables. In the standard MTS this is done by application of Taguchi orthogonal arrays experimentation together with Signal-to-Noise Ratio as a metric for differentiating between the results of experiments. An alternative approach is to define and optimize a mixed integer programming model. Therefore the concept of misclassification is introduced. Misclassification is caused due to wrong identification of observations which motivates the definition of the following two types of misclassification:

1. Classifying *healthy* observations as *unhealthy*.
2. Classifying *unhealthy* observations as *healthy*.

Thus, a function to measure the total misclassification reads:

$$TWFM = w_1 \frac{n_1^e}{n_1} + w_2 \frac{n_2^e}{n_2}$$

where

- n_1^e = number of *healthy* observations classified as *unhealthy*
- n_2^e = number of *unhealthy* observations classified as *healthy*
- n_1 = number of healthy observations
- n_2 = number of unhealthy observations
- $w_1 = c_1 / (c_1 + c_2)$
- $w_2 = c_2 / (c_1 + c_2)$

c_1 and c_2 are costs associated with misclassifications 1 and 2, respectively. The values for n_1^e and n_2^e are computed as follows:

$$n_1^e = \sum_{j=1}^{n_1} MD_j^1 \quad \text{s.t } MD_1^2 \leq MD_j^1$$

$$n_2^e = \sum_{j=1}^{n_2} MD_j^2 \quad \text{s.t } MD_{n_1}^2 \leq MD_j^2$$

where the scaled Mahalanobis distances are sorted in ascending order, i.e.,

$$MD_1^1 \leq MD_1^1 \leq \dots \leq MD_{n_1}^1$$

$$MD_1^2 \leq MD_2^2 \leq \dots \leq MD_{n_2}^2$$

With the misclassification criteria just described the following mathematical model for variable screening must be solved in order to detect those variables influencing the process.

Let $\mathbf{x} = (x_1, x_2, \dots, x_p)^t$ be a p dimensional vector where

$$x_i = \begin{cases} 0 & \text{if variable } i \text{ is not selected} \\ 1 & \text{if variable } i \text{ is selected} \end{cases}$$

The optimization problem reads as follows:

$$\text{Min } f(\mathbf{x}) = \alpha \underbrace{\left(w_1 \frac{n_1^e}{n_1} + w_2 \frac{n_2^e}{n_2} \right)}_{f_1(\mathbf{x})} + \beta \frac{p_{\text{selected}}}{p} \tag{1}$$

subject to

$$\sum_{i=1}^p x_i \leq p, \quad \sum_{i=1}^p x_i = p_{\text{selected}}, \quad f_1(\mathbf{x}) \leq f_1^{\text{max}}$$

We refer to Pal and Maiti (2010) for more details regarding the mixed binary integer programming model.

The solution to this problem can be obtained by the application of a binary algorithm. In Pal and Maiti (2010) binary particle swarm optimization, proposed by Khanesar, Teshnehlab, and Shorahdeli (2007), is used. In this paper we implemented the same BPSO algorithm as well as a recent version of binary ant colony optimization (NBACO) proposed by Jang (2011) which is shortly described next.

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات