Chapter 4

Hybrid Algorithm Applied to the Identification of Risk Factors on the Health of Newly Born in Mexico

María Dolores Torres
Autonomous University of Aguascalientes, Mexico

Aurora Torres Soto
Autonomous University of Aguascalientes, Mexico

Carlos Alberto Ochoa Ortiz Zezzatti
Juarez City University, México

Eunice E. Ponce de León Sentí
Autonomous University of Aguascalientes, Mexico

Elva Díaz Díaz
Autonomous University of Aguascalientes, Mexico

Cristina Juárez Landín
Autonomous University of Mexico State, Mexico

César Eduardo Velázquez Amador
Autonomous University of Aguascalientes, Mexico

ABSTRACT

This chapter presents the implementation of a Genetic Algorithm into a framework for machine learning that deals with the problem of identifying the factors that impact the health state of newborns in Mexico. Experimental results show a percentage of correct clustering for unsupervised learning of 89%, a real life training matrix of 46 variables, was reduced to only 25 that represent 54% of its original size. Moreover execution time is about one and a half minutes. Each risk factor (of neonatal health) found by the algorithm was validated by medical experts. The contribution to the medical field is invaluable, since the cost of monitoring these features is minimal and it can reduce neonatal mortality in our country.

DOI: 10.4018/978-1-4666-0297-7.ch004
INTRODUCTION

There are several factors that can be used as health indicators of societies, but infantile mortality is the one that better describes life conditions (Blum & Langley, 1997) since it depends on nutrition, general care, educational level, quality of medical services, etc.

This work is centered on the application of artificial intelligence to the neonatal risk factor determination, because this problem is not only of social relevance, but also corresponds to a field called feature subset selection that has been widely studied since the last sixty years.

Some researches show that this problem has been addressed by different approaches. For example, Torres, Torres, Ponce, and Diaz (2004) used a combination of logistic regression and correlation and then in 2005 the same researchers approached the problem by graphical modeling (Torres, Torres, Ponce, & Torres, 2005).

The determination of risk factors during pregnancy is very important to prevent neonatal mortality because it allows medical staff to act on time. “Approximately two thirds of neonatal deaths correspond to precocious ones, reflecting mainly problems of quality on attention during childbirth, asphyxia and malformations; the remaining of deaths is caused mainly by infectious problems, premature birth, and low weight on born” (Solis, Mardones, Castillo, & Romer, 1993). However, most of these deaths can be prevented by means of simple and cheap interventions according to Dawudo in his study “Neonatal mortality: Effects of selective pediatric interventions” (Dawudo & Effiong, 1985).

Because of the importance of such researches, results are usually validated by experts of the medical area. In this study results were evaluated by pediatricians from “Instituto Mexicano del Seguro Social” (IMSS).

The feature subset selection problem consists on choosing a subset from all variables who describes a phenomenon, by the elimination of features with low discriminative and predictable information (Pelikan, Sastry, & Cantu-Paz, 2006). An appropriate feature selection is important on removing irrelevant data, increasing learning accuracy, and improving comprehension of results (Blum & Langley, 1997).

On the case of the risk factors during pregnancy and childbirth, to know the variables that rebound on the health of children could not only to reduce risk of childbirth but also to prevent baby from death.

The approach proposed in this work, combines two different techniques; one taken from the world of evolutionary computation and the other taken from the wonderful world of mathematics. The techniques are the genetic algorithm and the logic-combinatorial approach called testor analysis. The objective is to combine both tools to find the subset of variables that impact the most on the health state of the newly born. To prove the effectiveness of this approach, results were validated by medical experts. Physicians verified each resulted variable and agreed with the importance degree assigned by the system.

IDENTIFICATION OF RISK FACTORS DURING PREGNANCY

Nowadays in medicine like in any other science, the use of computational tools to make a better use of data gathered about a specific phenomenon results essential. Pediatrician from all over the world have always had the need to be well informed about maternal antecedents and previous and actual obstetric record, when acquire the responsibility of a new patient (Hobel, Hyvarinen, Okada, & Oh, 1973); therefore, many researchers have focused on the identification of risk factors of neonatal mortality, stillbirth and morbidity, in order to improve prognosis and to prevent complications.

This work has the goal of identify the features that impact the most on the health state of the