IDEA GROUP PUBLISHING
701 E Chocolate Avenue, Hershey PA17033, USA
Tel: 717/533-8845; Fax: 717/533-8661; URL: www.idea-group.com
ITB7264

Chapter XV

# Heuristic Knowledge Discovery for Archaeological Data Using Genetic Algorithms and Rough Sets[1]

Alina Lazar
Wayne State University, USA

*The goal of this research is to investigate and develop heuristic tools in order to extract meaningful knowledge from archeological large-scale data sets. Database queries help us to answer only simple questions. Intelligent search tools integrate heuristics with knowledge discovery tools and they use data to build models of the real world. We would like to investigate these tools and combine them within the genetic algorithm framework. Some methods, taken from the area of soft computing techniques, use rough sets for data reduction and the synthesis of decision algorithms. However, because the problems are NP-hard, using a heuristic approach by combining Boolean reasoning with genetic algorithms seems to be one of the best approaches in terms of efficiency and flexibility. We will test our tools on several large-scale archeological data sets generated from an intensive archaeological survey of the Valley of Oaxaca in Highland Mesoamerica.*



This chapter appears in the book, *Heuristic and Optimization for Knowledge Discovery*, edited by Ruhul Sarker, Hussein Abbass and Charles Newton. 

# INTRODUCTION

## Archaeological Knowledge Discovery Problem

Anthropologists interested in ancient societies of Highland Mesoamerica, Valley of Oaxaca, have used intensive archaeological survey in order to study the state formation. Since these archaeological surveys were begun in the 1960s, the computer was an essential tool because of the large quantity of data resulting from the surveys. After the data was collected, it was placed on punch cards and the additional results were published in several books (Blanton, 1989; Blanton, Kowalewski, Feinman, & Appel, 1982; Kowalewski, Feinman, Finsten, Blanton, & Nicholas, 1989) along with extensive site maps. The reason behind this archaeological survey was to find answers to the following questions: What were the characteristics of Mesoamerican agricultural systems? What role did hydraulic agriculture play in prompting or facilitating the growth of large population centers? When was irrigation first introduced? What was the nature of these population centers? When and where did urbanism first arise? What decision making structures and adaptations were necessary to facilitate these changes? (Blanton et al., 1982).

Our goal for the proposed research is to integrate evolutionary learning tools into the knowledge discovery process and to apply them to the large-scale, archaeological spatial-temporal data produced by the surveys. This heuristic based approach used here will employ rough set concepts in order to represent the domain knowledge and the hypotheses.

While answers to the questions above can possibly be found by investigating the large-scale database resulting from the archaeological survey, this database contains over 6,000 regional sites and over 2,000 residential sites at the Monte Albán urban center. Each site is comprised of one or more components and can be occupied in one or more archaeological periods, spanning a period from approximately 9000 B.C. to 1500 A.D. Thus, the total spatial and temporal scope is so vast as to make manual interpretation a difficult if not impossible task. In addition, each temporal and spatial instance of a site component can be described in terms of several hundred variables of different types. We can clearly see a gap between data generation and data understanding here. Tools and techniques from artificial intelligence can be used to fill this gap and to aid in the extraction of emergent patterns hidden in the data, as is shown by Reynolds (1994, 1999).

## Heuristics

Uninformed or blind search, which processes and evaluates all nodes of a search space in the worst case, is not realistic here because of time constraints that are closely related to the dimension of the data. Generally, the search space increases exponentially with problem size, thereby limiting the size of problems which can realistically be solved using exact techniques such as exhaustive search. An alternative solution is represented by heuristic techniques, which can provide much