Predictive Modeling of the Indus Civilization Port Sites in the Gujarat: Site Location Through Rules-Based Predictive Modeling

Roy Mathur

Department of Archaeology
University of York
United Kingdom
rm515@york.ac.uk

Abstract

The Indus Valley Civilization is the most geographically extensive yet least researched culture of all of the Old World Civilizations, with many new sites yet to be explored. The Gujarat of northwest India is of particular interest as its coast projects deeply into the Arabian Sea offering access to other Old World Civilizations to the west such as Mesopotamia. This research, in common with landscape archaeology, takes a holistic approach to the environment and requires a polymath approach drawing from many different fields including computer science, geography, pre-history, archaeology, remote sensing, and GIS. This work is designed for an investigation of the role of rivers in the Harappan maritime trade network. By utilizing artificial intelligence driven predictive modeling, it should be possible to conclusively identify the roles of existing sites and to determine the possibility of finding new ones, thereby furthering the understanding of intra-cultural trading links.

1 Introduction

The subject of this research is the role of rivers in the Harappan maritime trade network of the Gujarat area of northwest India. Part of the research consisted of confirming that existing sites were indeed used as beaching, dock or port areas and in locating new sites. Due to the limited availability of high resolution imagery from satellite remote sensing and aerial photography, it sis hoped that predictive modeling will be of some use in extrapolating further information from a fairly poor data-set.

It should be noted that this short paper is a small sub-set of an on-going project. As such, it is presented here only as a brief example of how artificial intelligence can assist in solving archaeological problems—in this case site location—using simple techniques that any researcher can use.

2 Predictive Models

Before discussing the building of the predictive model, a brief definition of the term "predictive model" in terms of this research and what is hoped to be accomplished by building one must be established. Simply speaking, a predictive model is a comparison tool. The underlying premise for all predictive models is that particular kinds of archaeological sites tend to occur in the same kinds of place" (Renfrew and Bahn 2004:93).

Predictive modeling has been seen most frequently in archaeological use in the USA for tasks such as, for example, the location of prehistoric sites in the Shawnee National Forest in southern Illinois (Renfrew and Bahn 2004:93). Its importance is becoming more apparent in Europe as can be seen from a multiple-period site location exercise in the Regge Valley Project in the Netherlands (Brandt et al. 1992) and the recent review of archaeological predictive modeling from the Netherlands (Van Leusen and Kamermans 2005).

3 Data Sources for the Predictive Model

A by-product of the GIS component of this research is a table of possible coastal sites and riverine ports in the Gujarat. The data has been obtained from the Harappan ports project of the Indian Space Agency (Thakker 2000) and Greg Possehl's gazetteer of Harappan sites (Possehl 1999:27-845) as well as many other documentary sources.

Physically the table can be described as a comma separated plain text file. In the GIS model this was formerly used to supply site location points on the various ArcView map layers of the region. By constantly adding data to this table through further research, it is hoped that it will also eventually be used as the basis of a predictive model for identifying features in the landscape common to Harappan port sites. Subsequently, the model may offer an insight into the role of rivers in the Harappan maritime trade network and perhaps even assist in the discovery of new sites.

In order to create the predictive model, the first process is to examine what attributes uniquely identify a Harappan port or harbor. These need not only to be physical landscape features, but caninclude geopolitical features, such as nearness to other Harappan port sites and archaeological features, such as presence of anchor stones, as at Lothal. The attributes chosen will eventually become fields in a database table. As is common with predictive modeling in archaeology, humanistic factors such as viewsheds are also to be considered.

In the modern parlance of database terminology, "data mining" techniques are used to produce a "data warehouse" of "non-live" data. Then an expert system takes over further complex query processing.

4 Why Use an Expert System?

With archaeology in general, some data types lend themselves well to artificial intelligence (AI) analysis techniques, more specifically, the use of expert systems. In this research, the type of expert system eventually chosen can be described in AI terms as a knowledge-based, forward-chaining expert system. Forward-chaining simply means data driven rather than hypothesis driven (backward-chaining). This is a top-down approach where data is filtered through a set of rules. Although expert systems have a reputation for brittleness when faced with input outside their narrow parameters and have been described as "idiot savants" (Fritz 2002:24), the data here, is already confined to narrow and specific limits, therefore this supposed disadvantage will actually be a benefit. The mechanism of this expert system will be more fully described below.

It is also thought beneficial to greatly automate the expert systems technology in this research because of the need to process many different factors pertaining to site location. To manually process this information is a waste of project resources and the ability to apply the many archaeological rules to the large data set through automation is an attractive proposition (Crevier 1993:158).

5 Some Examples of Expert Systems

Expert systems in science are not new. One example of very well-known and successfully tested so-called brittle expert system is called MYCIN, developed by the Stanford Medical Center as a tool for diagnosing infectious disease (Horn 1986:5). Unfortunately, for legal reasons MYCIN was never officially implemented.

While expert systems in archaeology exist outside CRM, they seem used on somewhat micro rather than macro scales. For instance, a brief web search located only one expert system. This is an expert system used to identify Coriosolite coinage of the Coriosolites, one of several tribes in Brittany at the time of Caesar's campaigns in Gaul (Hooker 1996).

The expert system eventually designed for this research most closely resembled, in intent at least, an older system used by a geological prospecting company called Prospector (Duda et al. 1974-1983). Although not built for archaeology, this system is a site location program for mining that has successfully found ore deposits where land-based surveying and remote sensing have failed (Horn 1986:10). Used by knowledgeable users, in this case geologists, it bases its search assumptions on past characteristics favorable to mineral exploitation (University of Surrey 2006).

In the case of this research, the expert system will consist of a data model of the ideal site, then similar environmental and cultural indicators will be searched for elsewhere in the landscape.

6 Lack of a Standard Predictive Tool in Archaeology

While looking at archaeological predictive modeling in general, it is apparent that there are no off-the-shelf software available for general, multi-environment site location. Most predictive models used today are proprietary, non-portable systems consisting of non-automated manual procedures combined with the use of statistical software, spreadsheets, and GIS. In other words, each research team creates its own unique model that can be used for that particular research only.

Predictive modeling is more of a conceptual set of technique rather than a simple tool that any archaeologist can download or purchase and simply use. The main reason, and herein lies the main criticism of predictive modeling, is that in order to be accurate, the models need to be specific and narrow in scope.

The model being designed for this project, however, is general in its design, but specific in its actual use. By this it is meant that it can allow a knowledgeable archaeologist, for example, an expert in British Roman archaeology, to provide data and expertise relating to their own field of research in order for the model to work accurately.

Thus, one of the goals of this project is to attempt to produce a simple, off-the-shelf, predictive model program for site location. This will be accomplished by creating a predictive model using expert systems technology. The expert system component of the research will be discussed shortly after a brief overview of the model is presented.

7 Overview of the Model

The following flowchart is a very high level look at the model within the scope of the project in general. The area within the border represents the predictive model.

By combining various data sources discussed earlier, it is possible to form the beginnings of a powerful site location predictive model. The form of the actual model basically consists of raw data passed through various summing and averaging algorithms, then a module which re-filters the data using artificial intelligence expert system technology, followed by a module that factors in expert user expectations before final results are produced.

The program can be used in one of two ways. First, as can be seen, there are two lists. The first list (A) is used to create the perfect representative site, and the second list (B) contains all the candidate sites. After processing, this second list is compared to the perfect representative site. The output is an ordered list with the sites most like the perfect representative site nearest the top. The second way that the program can be used is simply as a tool to create an average representative site if no candidate sites are available.

The core of the final implemented predictive model consists of programs written using the Perl programming language. This has been chosen for a number of reasons, the chief ones being that it includes almost every component necessary for text processing, it is Open Source and

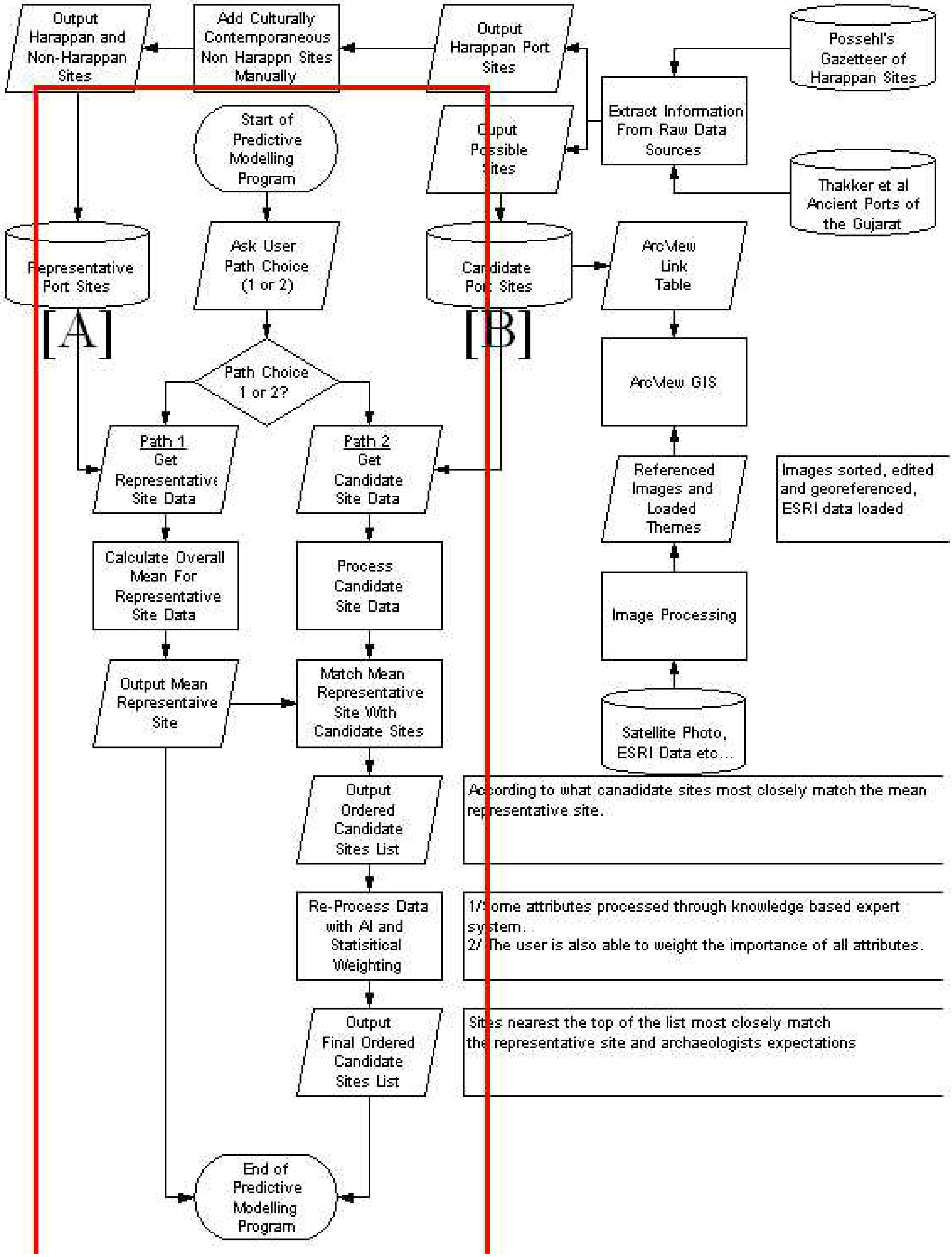

Figure 1. Flow chart with the predictive model represented within thick border.

free, and, finally, Perl runs on most operating systems. This means that the final product will be deployable on any operating system without the need for change. It will run, for example, in exactly the same way on Windows, Unix, and Mac computers. Pseudo-code based on BASIC, however, is necessary in the planning stages to experiment with the logic of the methodology used.

A prototype of the predictive model designated FindArc 0.1a has been created. This will later be developed into a fully implemented production system. This prototype system is a test and thought experiment to investigate the usefulness of expert system technology to this research and to archaeology in general, particularly in the field of reconnaissance archaeology. A description, therefore, of the first incarnation of FindArc follows.

8 A Brief Example: How Does FindArc 0.1a Work?

For the sake of an extremely simple test, soil micro-morphology provides an apt example of a test case for expert system techniques.

There are basically two states for each deposition type in soil micro-morphology when relating it to the occupation level of a settlement in the model. These states are "Yes" or "No".

To further elaborate this point, an example database contains data describing the only two possible states for each type of deposition:

- Is there primary cultural deposition? "Yes" or "No"
- Is there secondary cultural deposition? "Yes" or "No"
- Is there tertiary cultural deposition? "Yes" or "No"

Inference rules, also known as the rule base, form the basic archaeological rules that tell the program what to make of the available data. In this example, a rule is written that infers that if all three forms of deposition are present, the site has been used or occupied more than if there is only primary deposition:

- First only = Low
- First two = Medium
- All three = High

At this point, part of the program called an inference engine loops through the database applying the above rules to its search and performing an action when it finds one of the rules to be true:

- search through above database records
- apply inference rules to each record
- add result to record

By using the pseudo code, it is possible to see the whole process in action. In the database there are rows that look like this:

- Site_ID Prim_Dep Second_Dep Tert_Dep
- SiteA Y N N
- SiteB etc...

The rules base act on these rows. The rules base could be implemented in the following coded sub-routine:

```
RULES:
IF pd = "Y" AND sd = Y and td = "Y" THEN
occupation = "high"
IF pd = "Y" AND sd = Y and td = "N" THEN
occupation = "medium"
IF pd = "Y" AND sd = N and td = "N" THEN
occupation = "low"
```

This skeleton pseudo-code can easily be translated into any computer programming language with relatively few changes. There can, of course, be dozens of rules examining each data attribute and an inference engine to control the flow of all the database records through all the rules.

9 Weighting Attributes

As well as supplying the base data to the model, the last section of the model also allows the expert user to weight each attribute. The user can give each data attribute or database field an importance value of between, for example, one and ten. This rating defines the importance of each attribute in relation to all other attributes within the model. These subjective statistical weightings are then used to alter the values of each data attribute in each record before the final sorting occurs. In other words, the expertise of the end user in his or her particular field of archaeology increases accuracy of the model.

10 Summary and Conclusions

To sum up, the prototype model takes in raw data, uses averaging, summing, expert system technology, and statistical weighting to process bulk data and find probable sites based on both common archaeological concepts and end-user-specific archaeological expertise. Based on this research, it is possible to determine that an off-the-shelf AI driven software product would be useful in archaeological predictive modeling for site location. Evidence for this conclusion includes the following points.

1. This model is written specifically to look for new sites and, as has been seen, some attributes lend themselves well to AI. This can help automate the analysis of basic, easily understood archaeological concepts.
2. The model uses the local expertise of the end-user to make the results meaningful, specific, and narrow in scope.
3. The implemented model is portable; that is, because it is written in Perl, it can be used on Windows, Unix, and Mac computers.

References Cited

Brandt, Roel, Groenewoudt, Bert J., Kvamme, Kenneth L. 1992. An experiment in archaeological site location: Modeling in the Netherlands using GIS techniques. Theme issue, "Analytical Field Survey," *World Archaeology* 24(2):268-282.

Crevier, Daniel. 1993. *AI: The Tumultuous History of the search for Artificial Intelligence*. New York: Basic Books.

Duda, R., Hart, P. E., Nilsson, N. J., Reboh, R., Slocum, J., Sutherland, G., and Gasching, John. 1974-1983. *Prospector*. Artificial Intelligence Center, Stanford Research Institute (SRI) International Menlo Park, California, USA.

Fritz, S., ed. 2004. *Understanding Artificial Intelligence*. New York: Scientific American and Byron Preiss Visual Publications, Inc.

Hooker, John. 1996. *The Coriosolite Expert System*. http://www.writer2001.com/exp0003.htm (Last accessed 29 May 2006).

Horn, Michael van. 1986. *Understanding Expert Systems*. USA: Bantam Books Inc.

Possehl, Gregory L. 1999. *Indus Age: The Beginnings*. Philadelphia: University of Pennsylvania Press.

Renfrew, Colin and Bahn, Paul. 2004. *Theories Methods and Practice*. Fourth Edition. London: Thames and Hudson.

Thakker, P. S., Raval, M. H., and Dasgupta, A. R. 2000. Ancient ports of Gujarat. *GIS Development.net,* Geospatial Application Papers: Archaeology: Site Prediction, February. (Last accessed 01 June 2005]) http://www.gisdevelopment.net/application/archaeology/site/archs0004.htm

University of Surrey. 2006. *Prospector*. School of ECM, University of Surrey, Guildford, Surrey GU2 5XH, UK. (Last accessed 16 June 2006). http://www.computing.surrey.ac.uk/AI/PROFILE/prospector.html

Van Leusen, Martijn and Kamermans, Hans, ed. 2005. *Predictive Modelling for Archaeological Heritage Management: A Research Agenda*, Chapter 11. Amersfoort, Netherlands: National Service for Archaeological Heritage.