Automated Classification of Stone Projectile Points in a Neural Network

E.S. Lohse
Department of Anthropology
Idaho State University, USA

C. Schou
Technology Innovation Center
Idaho State University, USA

R. Schlader
Technology Innovation Center
Idaho State University, USA

D. Sammons
College of Education
Idaho State University, USA

See the CD for the extended version

Abstract

We have built a working prototype for online classification of stone projectile points in a neural network. The initial application involves specimens drawn from the North American Pacific Northwest cultural area. The computing system environment for hardware is designed for I386 architecture. Software is coded in VB.net and C# for the DLLs. The current database design is not software specific; however, it requires a robust relational database server. The auto-classification system consists of three stages. Stage 1 is the classification system, with software that allows users to submit images of artifacts or actual specimens that are digitized by lab staff. Stage 1 generates projectile point classifications with specimens assigned to recognized types and is a .NET standalone application. Stage 2 consists of release of a typological descriptive report to system users, including a full image inventory of submitted and classified specimens with attached statistical probabilities of type assignment. Stage 3 is a web-based application hosted on the Technology Innovation Center system that serves as the educational system for public access and study. This paper presents the practical difficulties and successes encountered in automating stone projectile point classification in a neural network, which offers potential for development of a creative, thinking classification system and a rich, accessible, secure reference database.

Introduction

In 1985, Lohse produced a standard classification of stone projectile points from the Columbia Plateau region of North America. He used measurements taken from a collection of 600 points broken into eighteen identified types to classify over 1600 stone projectile points spanning a period of 8000 years (Lohse 1985). This classification was performed as a discriminant analysis. All specimens were photocopied, traced by hand, outline landmarks plotted, and then recorded on a digitizer. Type classifications were made with high confidence levels. All assignments were then checked against classifications made by experts, and were found to be quite accurate.

The process by which the 1985 typology was produced was tedious and long. With improvements in computer technology, especially imaging and data analysis, it is now possible to create a system that can automatically accomplish what took several researchers weeks to accomplish twenty years ago. In addition, there is now a demand from go vernment agencies, researchers, and archaeological contractors to complete typological assignments of points quickly, accurately, and inexpensively, with the data maintained and available over a long period of time and for a large geographical region. Accordingly, the Archaeological AutoClassification System (AACS) was developed (Lohse et al. 2003).

The initial research problem before the AACS is the typing of stone projectile points, but it can eventually be applied to other archaeological and non-archaeological object sets for which recognition and typing is important.

Archaeological Auto-Classification System Working as Virtual Analyst

Our current activity for the AACS Project has focused on obtaining authentic data to fill our database. By "authentic data," we mean information about projectile points from well-dated stratigraphic contexts. This allows us to produce a "clean" set of data that exactly reproduces the classification produced by Lohse (1985). The AACS first incorporated information from the same projectile points as those used by Lohse in the 1985 study.

Neural Networks and Archaeological Classification

The "brains" behind the AACS is a neural network, nicknamed SIGGI. Through training, SIGGI can replicate the actions of archaeological experts in identifying types. SIGGI has implemented a projectile point classification system, which replicates the results developed by Lohse (1985). There are several steps in training SIGGI and in making sure that SIGGI produces valid results. First, SIGGI must be given the necessary information and then be allowed to train itself on that information. Next, SIGGI must move beyond the initial training set of data and be able to classify new and original data sets correctly. Finally, SIGGI's thinking processes must be evaluated for quality, clarity and accuracy, not only to assess the validity of SIGGI's typology but also to gain insights into how human typologists think and work.

Feeding SIGGI

SIGGI's brain is an artificial neural network (ANN). An ANN is composed of a series of artificial neurons, or nodes, which are connected by a series of edges that possess an attribute known as a weight. The artificial neuron produced reflects the structure of a biological neuron (Buckland 2002). Like a biological brain, the ANN is composed of the intelligent combination of artificial neurons and weights. A typical ANN consists of three layers: an input layer, a hidden layers, and an output layer. The input layer consists of a set of data items that the ANN acts upon. The output layer is the set of valid responses that the ANN produces. Our input layer consists of information pulled from an image of a projectile point, while the output layers consists of the information set and classification of that point. The hidden layer is simply an intermediate stage in the actual analysis where the data are manipulated (Buckland 2002).

Weights are set values used by the ANN to modify the inputs. The key to having an intelligent ANN is identifying the proper combination of weights. Combination requires a two-step process. The first step has each artificial neuron performing a summation of all the inputs modified by the weight of the path that the specific input took to reach the neuron. This summed value is then fed through the second step or activation function. The activation function is a mathematical equation that evaluates the summed value of the inputs to the neuron and decides whether or not the neuron will fire. The activation function calculates the value passed to the next stage of the process. Typically, there are two types of activation function: a step function and a sigmoid function. SIGGI uses the sigmoid function (hence the name, "SIGGI"). The sigmoid function uses an equation to produce a value within a given range of values, rather than the all-or-nothing approach of the step function (see Lohse et al. 2003).

Once the activation of one layer is complete, the next layer performs the same calculations until the data reaches the output layer. In the output layer, the neurons pass the final values back to the program, and the process ends. These raw values themselves may then be further modified or may be used as they are, depending on what the ANN is directed to perform. In our case, we take the values, pass them through a normalization functions, which produces a normalized distribution across the domain of the outputs. This identifies a set of variables and establishes a set of statistical probabilities. The output of the ANN is the probability of inclusion in each projectile point type, rather than simply a set of activations of neurons (Buckland 2002).

Training SIGGI

There are two ways to train an ANN: supervised and unsupervised training. SIGGI utilized supervised training, since supervised training required implementing only one AI algorithm. A typical supervised training runs feedback from a set of predefined input that modifies the weights by a set value that is fed back into the network in reverse, or back propagation (Buckland 2002). Back propagation feeds the data set through the network and calculates the error between the expected values and the actual values returned from the ANN. The error value is then fed through a set of equations at each layer in the network. The equation for that layer then modifies the weights that feed into each layer so that the next that set of inputs is fed into the ANN, the ANN is more likely to produce the desired values. This process fits a line to the data set so that a predictive model is generated. In essence, the ANN is conducting a discriminant analysis.

The variables on which the ANN conducts its discriminant analysis are created through a process of image input and manipulation. Images of projectile points are converted to grayscale and then to black-and-white bitmaps. The backgrounds are removed and the edge distortions are smoothed. Finally, an outline of the point is created and submitted to a "tokenizer," a function of the ANN in which the outline is converted into a series of tokens (line segments with connecting end points). The distance and direction of line segment are packages into a vector, and these vectors are the tokens for input into the ANN. The accuracy of the system is dependent upon the number of tokens generated. The original discriminant analysis (Lohse 1985) generated eighteen line segments, but this was an insufficient number for SIGGI to successfully discriminant among point types. The final version of SIGGI establishes 100 tokens on each image, providing sufficient resolution to separate out the different types.

The Archaeological AutoClassification System and SIGGI On-line

SIGGI initially trained on the same set of projectile points used by Lohse (1985) in his original discriminant analysis. SIGGI has been able to successfully replicate that study. It is a working prototype for online classification of stone projectile points by a neural network. Our initial application uses projectile point specimens drawn from the Pacific Northwest cultural area, but the system is extensible to other cultural areas and other artifact classes. The ultimate goal of the AACS, and SIGGI's typology, is to create a web site on which researchers can obtain valid, reliable classifications of projectile points they have recovered or hold in collections, conveniently, on-line, and without the expense of either shipping points to the experts or bringing experts to the points.

The autoclassification system consists of three interrelated stages. Stage 1 is the classification system, with software that will allow users to submit images of artifacts or actual specimens that are digitized by lab staff. This stage generates projectile points classifications with specimens assigned to recognized types. Stage 2 consists of release of a typological descriptive report to system users, including a full image inventory of submitted and classified specimens with attached statistical probabilities of type assignment. Stage 3 is a web-based application hat will be an educational venue for public access and study.

Through these stages, we envision four different classes of on-line users of the AACS: government agencies, researchers, Native Americans, and the general public (Lohse et al. 2003). The look, options, and access of the AACS will be different for each of these four user groups. While some of the functions will be common to all four groups, some groups have very specifics interests and agenda for the information contained in the AACS. Government agencies will be most interested in having the ACCS analyze and classify projectile points which the agencies submits, and in having the AACS store data about the points securely and confidentially. Agencies will also request the AACS to produce summative temporal and spatial information about the projectile points. Academic researchers will have similar interests, with the added function of requesting images and descriptive information about specific projectile points from the database. Native American tribes could also interact with the AACS for similar reasons, but will be especially concerned that the information in the AACS databases remain secure and confidential, and most importantly, that locational information not be released. The fourth type of user, the general public, will be able to access the AACS for educational purposes, using the site like a key (what kind of point did I find?) or for general information (where are these points found? How old are they?). The AACS must be able to provide complete and valid information even to the lay-users, so that the complexity of their questions is limited only by their own knowledge and not by shortcomings or constraints of the AACS.

Conclusions

Our neural agent, SIGGI, has been trained to accurately reflect the thinking of archaeological typologists. SIGGI produces valid types in an automatic online environment, a confidential validated database that preserves information as images and as text fields, and insi ghts into how analysts think. Because SIGGI effectively learns, analysis of how SIGGI makes decisions and manipulates data can lead to new insights regarding redefinition of types and definition of new types, and may well revised how archaeologists consider doing typology. Input from archaeologists interacting with SIGGI will be included in the data being considered and will be reflected in types assigned and their evaluation. SIGGI will correspond with archaeological consultants in real time, responding to their inquiries and helping to enhance the operation of their own classification systems.

References

BUCKLAND, M., 2002. AI Techniques for Game Programming. Premier Press, Cincinnati, Ohio.

LOHSE, E.S., 1985. Rufus Woods Lake Projectile Point Chronology. In Campbell, S. (ed.), Summary of Results: Chief Joseph Dam Cultural Resources Project, Washington:317-364. Report to the U.S. Army Corps of Engineers. Office of Public Archaeology, University of Washington, Seattle.

LOHSE, E.S., SCHOU, C., STRICKLAND, A.W., SAMMONS, D. and SCHLADER, R., 2003. Information Data Archives: Management, Research and Information Distribution in a Secure, Controlled Environment. Paper presented at the World Archaeological Conference, Washington, D.C. (http://imnh.isu.edu/wac5/, June 20, 2003).