Outline of Thesis Proposal

David Ezra Sidran

Thesis: that using machine learning methods it is possible to construct an artificial intelligence system capable of making human-level tactical military decisions. Specifically, the method involves using clustering classification to identify similar tactical positions from a database of historical battles analyzed by subject matter experts (SMEs) and implementing the corresponding solution as recommended by the collective wisdom of the SMEs.

I. Implementing the System.

TIGER (our test-bed program) is a robust and fully featured application that runs in the Windows environment. It employs a graphical user interface including dialog boxes, drop-down menus, a floating/docking toolbar and point and click drawing layers. It supports the importation of graphical bitmap files (BMPs) from which elevation and terrain can be extrapolated and saved. It supports positioning and editing 12 standard military unit types. It supports 3D Line of Sight (LOS) calculations. It supports least-weighted path calculations optimized for terrain, elevation, unit type and LOS. It supports the calculation of tactical lines and frontages and the implementation of the five canonical offensive maneuvers. It supports the calculation of optimal Line of Departure (LD) positions. It produces descriptions of tactical positions in 'English predicate statements' and outputs them in HTML format.

II. Survey of Conceptual Clustering Algorithms.

Knowledge Acquisition via Conceptual Clustering (unsupervised learning by observation) has been researched for over twenty years and has produced a number of well known techniques:

a. Partitioning Methods which use an iterative relocation technique that assign objects to groups (usually employing either the *k-means* or the *k-medoids* algorithms) This method assumes comparable feature space axes in Euclidean space. This technique is best employed for finding spherical-shaped clusters in small databases.

b. Hierarchal Methods which can be separated into either *agglomerative* (bottom-up) and *divisive* (top-down) approaches. The agglomerative method successively merges groups while the divisive method successively divides groups.

c. Density-based Methods 'grow' clusters of objects until a distance threshold is met.

d. Grid-based Methods employ a multi-layered hierarchy that quantizes the object space into a grid structure. Distance thresholds are commonly employed for clustering and "binning" is used to avoid dimension differences.

e. Model-based Methods form clusters based on the 'best fit' of a model.

III. Choice of Appropriate Learning Algorithm.

Our specifications and concerns for an appropriate clustering and learning algorithm include:

a. We have no *a priori* knowledge of the number of *attributes* that will be used to describe a tactical situation or 'scenario' (i.e., an 'object' or 'instance'). Indeed, we anticipate that the number of attributes will continuously increase during our research. Furthermore, the attributes may not comparable – indeed, this will almost certainly be the case - and therefore, it would be inappropriate to employ a method that is dependent upon comparing feature space axes.

b. We have no *a priori* knowledge of the number of classes that will divide our historical database.

c. We are not optimistic about any approach that uses Euclidean or Manhattan distances to separate scenarios into clusters because of (a) above.

d. Our database of tactical situations will continuously grow during our research, yet we wish to quickly arrive at a 'closest match' to the current scenario under consideration from our database. And, even if this changes in the future, we need to make our "best guess" based on what we know now.

IV. Experiments and Validation of Research by SMEs

We propose conducting a series of experiments that will be validated by SMEs:

a. Sequentially entering 40 unique scenarios into TIGER, observing how each new scenario is classified and validating the returned 'closest matching' scenario.
b. Repeating experiment (a) above but entering the scenarios in a different order thereby observing the stability of the classification system.
c. Dividing input scenarios into two categories (pre-1900 and post-1900) and observing the results to determine if the system has any historical bias[1].
d. Repeating experiments (a) and (b) above but each time removing a random scenario and observing the results.

Because of the above considerations we have opted to first implement the ClassIT algorithm,[2] which is an extension of COBWEB, a "conceptual learning algorithm that performs probability analysis and takes *concepts* as a model for clusters."[3]

---

[1] Historically, military theory can be divided into 'classical' and 'modern' eras.

[2] Genarri, J. H., & Langley, D. F. (1989). Models of Incremental Concept Formation. *Artificial Intelligence*, 11 - 62.

[3] Han, Jiawei, and Micheline Kamber. *Data Mining Concepts and Techniques*. New York: Morgan Kaufmann Publishers, 2006.