Association for Information Systems
AIS Electronic Library (AISeL)

AMCIS 2000 Proceedings

Americas Conference on Information Systems (AMCIS)

2000

Designing Optimal Knowledge Base for Integrated Neural Expert Systems

Dat-Dao Nguyen
California State University - Northridge, datdao.nguyen@csun.edu

Dennis S, Kira
Concordia University, kira@vaz2.concordia.ca

Designing Optimal Knowledge Base For Integrated Neural Expert Systems

Dat-Dao Nguyen, Department of Accounting & MIS, California State University, Northridge
<datdao.nguyen@csun.edu>
Dennis S. Kira, Department of Decision Sciences & MIS, Concordia University, Montreal Canada
<kira@vax2.concordia.ca>

Abstract

This paper reports on the design of an optimal knowledge base for integrated Artificial Neural Network (ANN) and Expert Systems (ES). In this system, an orthogonal plan is used to define an optimal set of examples to be taken from a problem domain. Then holistic judgments of experts on these examples will provide a training set for an ANN to serve as an initial knowledge base for the integrated system. Any counter-examples in generalization over the new cases will be added to the training set to retrain the network in order to enlarge its initial knowledge base.

Introduction

Literature of cognitive psychology points out that most experts cannot express explicitly how the decisions were reached (Nisbett and Wilson, 1977; Ericsson and Simon, 1980). As such, the knowledge acquisition for conventional Expert Systems (ES) has to deal with seemingly incomplete and/or inconsistent production rules. The machine induction approach has been used in ES technology to deduce new knowledge from sample solutions of a problem domain. However, most machine induction techniques are hindered by the inadequacy of traditional statistical methods and/or the requirement of a large sample set to represent a problem domain.

This paper reports on the design of an optimal knowledge base for an integrated system of Artificial Neural Network (ANN) and ES. In this system, an orthogonal plan is used to define an optimal set of examples to be taken from a problem domain. Then holistic judgments of experts on these examples will provide a training set for an ANN. With its ability in pattern recognition and function approximation, the ANN will learn decision patterns and production rules in the sample set to build the system's initial knowledge base and then generalize this rule base to new cases. Any counter-examples in generalization will be added to the training set. Then the network is retrained to enlarge its knowledge base. Functioning with this initial knowledge base, the integrated system can overcome difficulties encountered in the knowledge engineering of conventional ES when dealing with incomplete and/or too large a production rule set

Limitations of Conventional ES

The goal of knowledge acquisition for ES is to obtain the necessary information to represent knowledge by production rules in the form of cause/effect, situation/action, if-then-else (Chorafas, 1990). But discovering the experts' mental models could be very difficult. There is evidence that people generally have little or no introspective access to higher order cognitive processes. People are rarely able to give retrospective reports on what they were thinking about in solving a problem (Ericsson and Simon, 1980). In addition, literature has reported on the limitations of human cognition. The rationality of human thinking in formal logical sense can be questioned and some aspects of deductive reasoning prove to be problematic for human beings. People have difficulty in handling combinations of uncertain evidence in multi-dimensional decision problems (Kahneman, Slovic, and Tversky, 1982).

Within these cognitive limitations, one faces with difficulties in building conventional ES since these systems require structured rule sets but experts cannot always express explicitly the rules and procedures used in arriving at a solution. However, experts can supply suitable examples of problems and solutions reflecting their conceptualization of a domain. It has been found that human beings in general do have ability to make holistic judgments (Slovic and Lichtenstein, 1971), and expertise is based mostly on the recognition of patterns in the problem space (Chase and Ericsson, 1982). Hidden behind the holistic assessment of problem alternatives, there always exits implicitly a logic on the association between input-output patterns. These characteristics make the machine induction with ANN become an appropriate technique to discover decision patterns from a set of holistic assessments.

Machine Induction with Integrated ANN/ES

Machine induction offers the possibility of deducing new knowledge. It can list all factors that influence the decision, without understanding their impacts, and induce a rule that works successfully. The method needs only pre-classified examples and the consideration of all samples in the domain. Apparently, one of its major disadvantages is in the requirement of a database containing sufficiently documented cases structured around human knowledge on a problem domain. In addition, the induced rules can be too large or too complex leading to unintelligibility. Most machine induction techniques rely on traditional linear statistical methods with strong assumptions on the behavior of data.

To overcome the inadequacy of traditional techniques in learning nonlinear relationship and to avoid imposing *a priori* restrictions on the data, the use of ANN could be an appropriate approach in machine learning and ES building. This approach is particularly useful in recognizing the input-output patterns in heuristics, which cannot be expressed explicitly by the experts. An ANN does not require *a priori* elaborate models or pre-specified probability distribution function of the data in order to learn the underlying discriminant patterns and association.

In theory, as a universal function approximator, an ANN can map the relationship between input-output patterns, and learn probabilities and statistical distributions from data (Cybenko, 1989; Hornik et al., 1989. Consequently, ANN can discover the associations between groups, elements of the domain space and those of the problem space from past data and generalize over new cases. In practice, an ANN can offer the advantage of computer execution speed. The ability to learn from cases and train the system with data rather than to write programs may be more cost effective and even more convenient when frequent updates are needed (Medsker, 1995). In applications where rules are unknown, an ANN may be able to represent those rules implicitly as stored connection weights.

However, previous attempts at the integration of ANN and ES (Gallant, 1988; Medsker, 1995) are hindered by the availability of examples and the cognitive effort of experts in providing judgment on these examples. The cognitive limitation of human experts makes the assessment of a large sample impractical if not erroneous. The representation of tradeoffs in an appropriate rule set for conventional ES is extremely difficult. Taking into account the tradeoffs in their assessment, experts might provide judgments that seem to be contradictory and/or inconsistent with formal logic. These limitations could be overcome with an optimal training set for the ANN containing holistic assessments of experts on a sample defined by an orthogonal plan.

Orthogonal Plan as an Initial Knowledge Base

In the Analysis of Variance, an orthogonal main-effect plan (Addelman, 1962) permits the study of several factors without going into every combination of factor levels. It has been estimated that, by using an orthogonal design, a set of 49 holistic assessments of production rules can cover the dimensionality of decision problems having up to eight criteria, with seven levels per criterion (Barron and Person, 1979). For example, in a four-criterion problem with five levels per criterion, instead of making 4^5 or 1024 assessments to define a set of production rules for a conventional ES, with an orthogonal plan one needs only 25 assessments to capture the main effects of problem factors. The orthogonal plans not only help to reduce the burden of information processing on experts, but also serve to define initial knowledge bases for integrated systems of ANN and ES.

This paper investigates the implementation of an optimal knowledge base designed by an orthogonal plan in an integrated system of ANN and ES. Such integration will help to overcome the difficulties often encountered by conventional ES technology. Within the proposed framework, one starts with an appropriate orthogonal plan to define a set of basic examples of the problem domain. Holistic judgments of an expert on these basic examples constitute a training set for the ANN. Holistic assessments intend to overcome the difficulty of the expert in explaining explicitly the production rules of his/her heuristics. Once the decision patterns of this set are learned, the network acquires an initial knowledge base on the problem domain. After being trained, the network will be used to generalize over new cases. Any wrong prediction will constitute a counter-example to be added to the training set. Consequently, the network will be retrained to learn new decision patterns emerging from the addition of counter-examples to the training set. One notes that the integrated system can be put into production even it has learned only partial information on the problem domain. Since the system is trained with basic patterns of the problem domain, it can generalize well over new cases. The process of "learning by doing" will continue as needed to acquire a robust knowledge base for the integrated system. Over time, the system will enrich its knowledge base with additional decision patterns as it learns from its own production and the experts' opinion.

Illustration

Table 1. The Problem Domain

NPV of Cash Flow	$[1.0 2.0 3.0 4.0 5.0] millions
Initial Investment	$[2.5 2.0 1.5 1.0 0.5] millions
Market Growth Rate	[fair good very-good]
Capability to Market	[fair good very-good]
Prospect of Technical Success	[fair good very-good]

The following illustrates the implementation of an optimal knowledge base in an integrated system of ANN and ES. The problem relates to project evaluation and economic appraisal of proposals for new products in a manufacturing company. In the appraisal process, the quantitative data are converted to categories to avoid the difficulty in dealing with data on a continuous scale. There exist tradeoffs among criteria and no single criterion absolutely dominates any others. In this study the appraisal is taken by three senior managers, indicated as Expert A, Expert B, and Expert C, respectively. These experts rate each proposal on a scale ranging from 0 to 100.

For this problem domain, a complete set of 864 production rules would be needed to build a knowledge base for a conventional ES and to make such ES functioning. With the integrated ANN/ES proposed in this study, an initial knowledge base is build by using an orthogonal plan to define a sample of only 24 examples of production rules taken from the problem domain. To validate the performance of the integrated system, two test sets, each containing five out-of-sample examples, are set up.

In this study, an ANN is configured with five input nodes, one for each criterion, five hidden nodes and one output node. This network uses a backpropagation algorithm with sigmoid transfer function, a learning rate of 1 and a momentum of .9. The training and testing tolerance are set at .1 and .3, respectively.

For Expert A, the network converges after 119 training epochs with a Root Mean Square of Errors (RMSE) of .0304. In the generalization over the two test sets, its RMSE are .0170 and .02275 respectively. For Expert B, the network converges after 104 training epochs with an RMSE of .0333. In the generalization over the two test sets, its RMSE are .0381 and .0473 respectively. In these cases, the experts judge that the system has provided satisfactory generalizations on both test sets.

For Expert C, the network converges after 2425 training epochs, with an RMSE of .0334. On the first test set, it makes two wrong predictions with an estimate of 81.70 for the example having a preference score of 70 and an estimate of 79.42 for the example having a preference score of 65. These two counter-examples are added to the training set and the network is retrained to learn new patterns in these examples. In the second run, the network converges at 1769 epochs with an RMSE of .0369. On the second test set, it makes only one wrong prediction with an estimate of 93.34 for the example having a preference score of 60. The network training for this case takes longer time to learn the peculiarity in decision patterns and production rules of Expert C. However, the knowledge base of the system and its predictability for preference of Expert C increase over time as it learns more about decision patterns of this expert.

Conclusion

This study has shown that an integrated ANN/ES can perform well with an initial knowledge base designed with the orthogonal plan and holistic assessments. This optimal sample contains only a subset of all possible production rules of a problem domain. Consequently, less cognitive effort is required from experts since they make judgments on a set of fewer examples. In addition, the holistic assessment does help experts to overcome the difficulty in expressing explicitly the production rules of their heuristics.

To acquire a robust knowledge base for the integrated system, one may consider aggregating judgments of multiple experts in a composite training set. In a future work, we will show that such aggregation it provides a consensus solution for the group.

References

Addelman S. "Orthogonal Main-Effect Plans for Asymmetrical Factorial Experiments," *Technometrics*, (4:1), 1962, pp. 21-46.

Barron F.H. and Person H.B. "Assessment of Multiplicative Utility Functions via Holistic Judgments," *Organizational Behavior and Human Preference*, 24, pp. 147-166.

Chase W.G. and Ericsson K.A. "Skill and Working Memory," in G.H. Bower (Ed.), *The Psychology of Learning and Motivation*, *Vol.16*. New York: Academic Press, 1982.

Chorafas D. *Knowledge Engineering*. New York: Van Nostrand Reinhold, 1990.

Cybenko G. "Approximation by Superpositions of A Sigmoid Function," *Mathematics of Control, Signals, and Systems*, 2, 1989, pp. 303-314.

Ericsson K. and Simon H.A. "Verbal Reports As Data," *Psychological Review*, 87, 1980, pp. 215-251.

Gallant S.I. "Connectionist Expert Systems," *Communication of the ACM*, (31:2), 1988, pp. 152-169.

Hornik K., Stinchcombe M,. and White H. "Multi-Layer Feedforward Networks Are Universal Approximators," *Neural Networks*, 2, 1989, pp. 359-366.

Kahneman D., Slovic P., and Tversky A. (Eds.) *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press, 1982.

Medsker L.R. (1995). *Hybrid Intelligent Systems*. Boston: Kluwer Academic Publishers.

Slovic P. and Lichtenstein S. (1971). Comparison of Bayesian and Regression Approaches to the Study on Information Processing in Judgment. *Organizational Behavior and Human Performance*, 6, pp. 649-774