Using Tactic-Based Learning (formerly Mentoring) to Accelerate Recovery of an Adaptive Learning System in a Changing Environment

Alice Armstrong
The George Washington University
Department of Computer Science
Washington, DC
piffle@gwu.edu

Peter Bock
The George Washington University
Department of Computer Science
Washington, DC
pbock@gwu.edu

Abstract

Tactic-Based Learning (TBL), formerly referred to as mentoring, is a selection policy for statistical learning systems that has been initially tested with a Collective Learning Automaton that solves a simple, but representative, problem. To respond to an immature stimulus that does not yet have a high-confidence response associated with it, TBL hypothesizes that selecting a response that has been designated as useful by a different, but nonetheless well-trained stimulus, is a better strategy than selecting a random response. TBL does not use any feature analysis in search of an appropriate response. Previous results [1] show that TBL significantly accelerates learning of a static problem, especially when several stimuli share the same response, i.e., when broad domain generalization is possible. This paper shows that TBL also increases the speed of recovery when the problem changes abruptly after the learning agent has mastered the initial state of the problem.

1. Introduction

In Collective Learning Systems (CLS), a Collective Learning Automaton (CLA) learns the appropriate response for each stimulus by selecting responses until one of them emerges as statistically optimal, guided by feedback from an evaluating Environment (Bock 1976). Generally, CLS theory ignores what has already been learned by other stimuli when making decisions about a new stimulus. Recently [1] it was shown that once some reliable knowledge is available for one stimulus, incorporating that knowledge into learning the responses to other stimuli, even if they are largely unrelated, can be very effective. Many psychologists agree that applying successful solutions for old problems to new and often unrelated problems is a useful learning strategy [9] [11] [3]. Although the experiments reported in this paper do not attempt to replicate human behavior at any level, biologically and psychologically inspired mechanisms and methods can often provide useful insights and hints for AI methods (Heckman, 2004).

The research reported in this paper deals with a selection policy for CLAs, called **Tactic-Based Learning** (TBL), which accelerates learning by applying knowledge about one well-learned situation to another. Although many machine learning algorithms can achieve excellent results by identifying similar feature vectors (explicit domain generalization), they all require postulating a sensible and computable distance metric. For example, the *k*-Nearest Neighbor algorithm [7] [8] computes similarity using the Euclidean distance between vectors in an ordered *n*-dimensional space. On the other hand, although case-based reasoning [12] allows feature vectors to be categorical, a distance metric of some kind must be postulated to identify similar cases.

For many problem domains, it is not possible to postulate a meaningful distance metric. For example, in Natural Language Processing there is no direct way to compute the distance between the meanings of words, so other methods must be devised [10]. TBL, however, does not compare feature factors at all, and is thus applicable to a wide problem domain.

2. A note about a change in terminology

After attending the AIPR 2007 workshop itself, the authors have decided to incorporate the very useful feedback they received and make some changes to the terminology. The chart below gives the list of old terms specific to the selection policy and their corresponding new terms. Readers who are unfamiliar with mentoring/Tactic-Based Learning may skip this section.

old term	new term
mentoring	tactic-based learning
mentor	tactic
unguided stimulant	undirected stimulant
student stimulant	directed stimulant
independent stimulant	independent stimulant
supporter stimulant	supporter stimulant
dependence threshold	dependence threshold
independence threshold	independence threshold
withdrawal threshold	recantation threshold
election threshold	support threshold
mentored CLA	tactic-based CLA
unmentored CLA	standard CLA

Table 1 old and new terms for the mentoring/tactic-based learning technique

3. Background

3.1. Collective Learning Systems

In a Collective Learning System (**CLS**) a Collective Learning Automaton (**CLA**) learns how to respond to stimuli appropriately using the algedonic cycle [2], as illustrated in Figure 1. The CLA is embedded in an Environment that sends a stream of stimuli to the CLA and periodically issues evaluations of the CLA's responses to these stimuli. A **stimulus** is a vector of several features that describes some state of the Environment. The CLA uses a State Transition Matrix (**STM**) to store each unique stimuli that has been received, along with its occurrence count (sample size) and an estimate of the probability that each possible response is valid for this stimulus. For each stimulus that is received, the CLA uses these probabilities to select a **response**, which is then sent to the Environment. These selection probabilities are updated based on periodic evaluations issued to the CLA by the Environment at the end of a **stage,** which is a sequence of responses by the CLA.

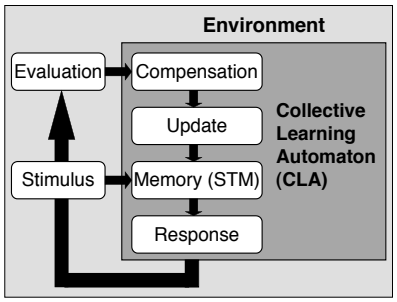

Figure 1: Algedonic cycle of a CLS

For a given stimulus the **Standard CLA** (a CLA that does not use TBL) selects the response with the highest statistical confidence if the confidence is sufficiently high; otherwise, a response is selected at random. All responses are sent to the Environment, and at the end of each stage, the Environment evaluates their collective performance. This **evaluation** is issued to the CLA, where the **compensation** function converts the evaluation into an **update**. The update is applied to all the elements of the probability vectors in the STM that were used to generate the CLA's responses since the last evaluation (the **history** of the stage) [4].

The standard difference of two proportions is used to compute the statistical confidence of each response for every stimulus, which is called the **selection confidence** of a response.

3.2. Tactic-Based Learning

Tactic-based Learning is an algorithm that overrides the **standard selection policy** used by a Standard CLA. A **Tactic-Based CLA** follows the standard selection policy until one stimulus is sufficiently well trained to elect its primary response as a **tactic**. A stimulus supports a tactic when its selection confidence is very high. Stimuli that are using a tactic (**directed stimuli**) simply use this response, assuming it is better than a random response. However, each directed stimulus tracks the effectiveness of the tactic and uses it only as long as it remains effective (an average compensation ≥ 1). When a new tactic becomes available, all stimuli that do not yet have an effective tactic will try it.

The lifecycle of a hypothetical stimulus in a Tactic-Based CLA is described in Figure 2. When there are no tactics in a CLA, all stimuli follow the standard selection policy and are called **undirected stimuli**. As soon as the first tactic appears, all undirected stimuli will investigate it. When a stimulus selects a tactic, it becomes a directed stimulus of that tactic. As long as a tactic remains effective for a directed stimulus, the directed stimulus will continue to use the tactic's responses. However, if a tactic proves ineffective (a parameter of the algorithm), the directed stimulus drops this tactic and looks for another. If no other effective tactics are available, the stimulus reverts to the standard selection policy and becomes an undirected stimulus. After a directed stimulus has attained a specified selection confidence, it becomes an independent stimulus and reverts to the standard selection policy. Dropping the tactic allows the independent stimulus to explore its response range. Exploration is useful because it helps avoid settling in a local maximum. An independent stimulus will either lose confidence in its response and revert to being a directed stimulus, or will become confident enough to become a supporter of a tactic itself. An independent stimulus is allowed some latitude, and it will only revert to being a directed stimulus if its selection confidence falls below the **dependence threshold**.

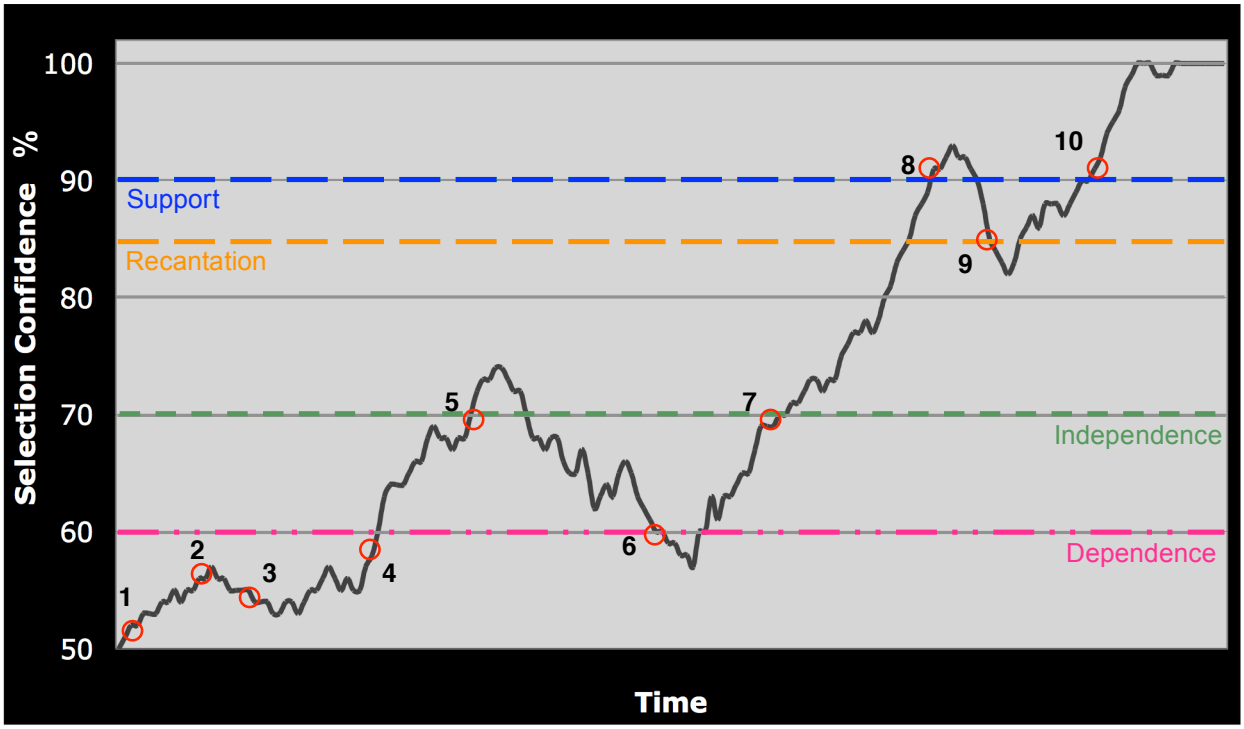

Figure 2 The Lifecycle a Stimulus ϕ: (1) There are no tactics available, ϕ is undirected **(2)** the first tactic appears and ϕ becomes directed **(3)** the first tactic is not effective and so ϕ abandons it and returns to being undirected **(4)** a second tactic appears and ϕ becomes directed again **(5) ϕ's** selection confidence crosses the Independence threshold, ϕ explores its response range and does not use its tactic **(6)** ϕ loses confidence and crosses the dependence threshold and becomes dependent on its tactic again **(7)** ϕ becomes independent again **(8)** ϕ's selection confidence crosses the support threshold and ϕ becomes a supporter. ϕ's response with the highest probability become a new tactic, if no other stimulus supports already supports it **(9)** ϕ loses selection confidence and recants its support of its tactic. If no other stimuli support this tactic, the tactic will no longer be available **(10)** ϕ gains selection confidence and supports a tactic

In the event that a supporter stimulus loses confidence in its response, the supporter stimulus **recants**. If the tactic no longer has any supporters, it will no longer be available for use, and any directed stimuli using it will become undirected stimuli.

4. Objective & Solution Method

The effectiveness of TBL has been shown in previous work [1] for static learning problems, that is, learning tasks that did not change over time. The research presented in this paper demonstrates that TBL can help a CLA recover from a change in the learning task faster than a Standard CLA. In order to observe the effectiveness of TBL to speed recovery, the following simple, but representative, problem was devised. A CLA is trained on the problem in Figure 3 until it is completely confident. Then, some of the stimuli are given new correct responses (see Figure 4). The CLA continues to train until it has recovered its

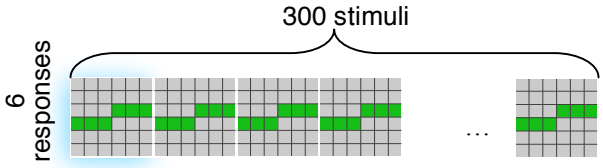

Figure 3. Initial Problem State a CLA is first trained to solve a 300 by 6 classification problem. The green cells are the correct responses

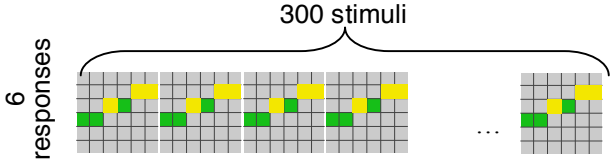

Figure 4. Secondary Problem State After a CLA has trained on the initial state, some of the stimuli are given new correct responses. The green squares are correct responses that have not changed. The yellow squares are new correct responses

confidence or it reaches the maximum training time of 100,000 contests.

5. Factors

There were three factors in these experiments: (1) the TBL thresholds that govern when a stimulus can use or support a tactic (2) the percentage of change in the learning task, that is, what percentage of the stimuli had to be relearned and (3) the collection length (the number of responses that a CLA makes between evaluations.

The TBL thresholds (support, recantation, independence, and dependence) were all tested at the following settings (%) {50, 75, 90, 95, 99} under the following restrictions:

- Independence threshold ≤ Support threshold
- Dependence threshold ≤ Independence threshold
- Recantation threshold ≤ Support threshold

The total number of experiments for each possible threshold setting is 137.

These experiments included three settings for the percentage of change in the problem: a 50% change when the problem started with two correct answers and added a third (see Figures 2 and 3), a 66% change (3 correct responses to 6), and an 87% change (6 correct responses to 1).

The collection lengths included in these experiments were {1, 2, 4, 6, 12}. A collection length of 1 is trivial because the problem becomes a simple process of elimination. A collection length of 12 is considerably more difficult. As the collection length gets longer, it becomes harder and harder to tell which stimuli chose correct responses and which did not.

6. Results

Due to space restrictions, the results from the most difficult factor settings are presented. These results are for a collection length of 12 (the longest collection length tested) and an 87% change in the problem. Since work has been presented previously about the effectiveness of TBL on learning, the results of the initial phases of learning are omitted. Only results for the second phase are presented. These results show the recovery after a change in the problem. If requested, the rest of the results can be provided.

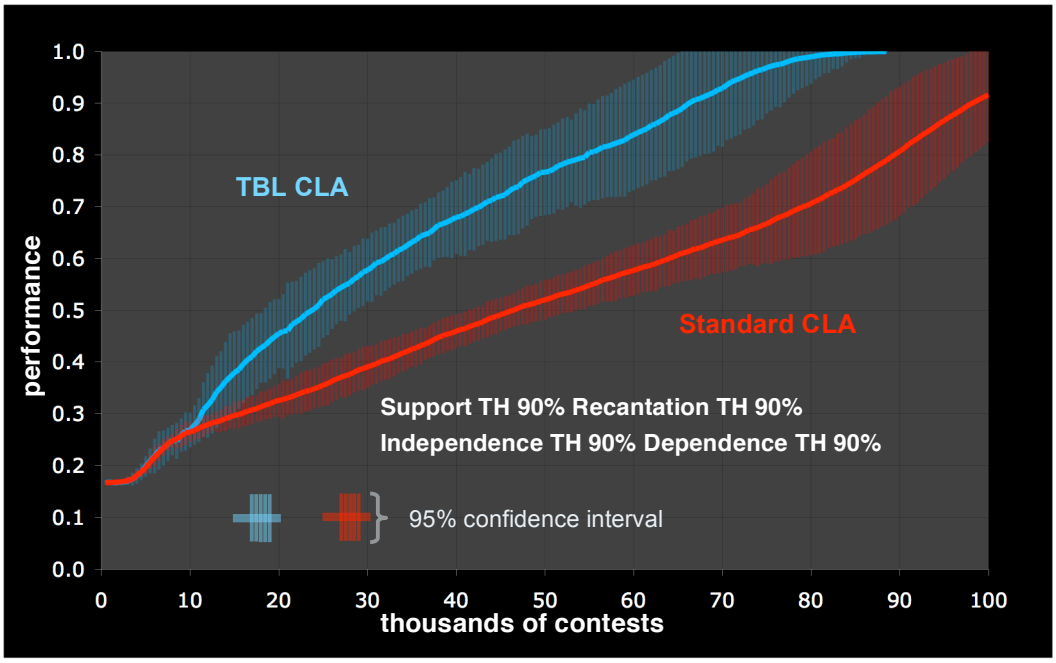

Figure 4. These threshold settings represent one of the most favorable results (on the most difficult set of factors). These results can be explained by the fact that the stimuli are forced to become directed (use a tactic) immediately. They do not spend any time being independent. This is very helpful for recovery in situations where a stimulus has only one correct response.

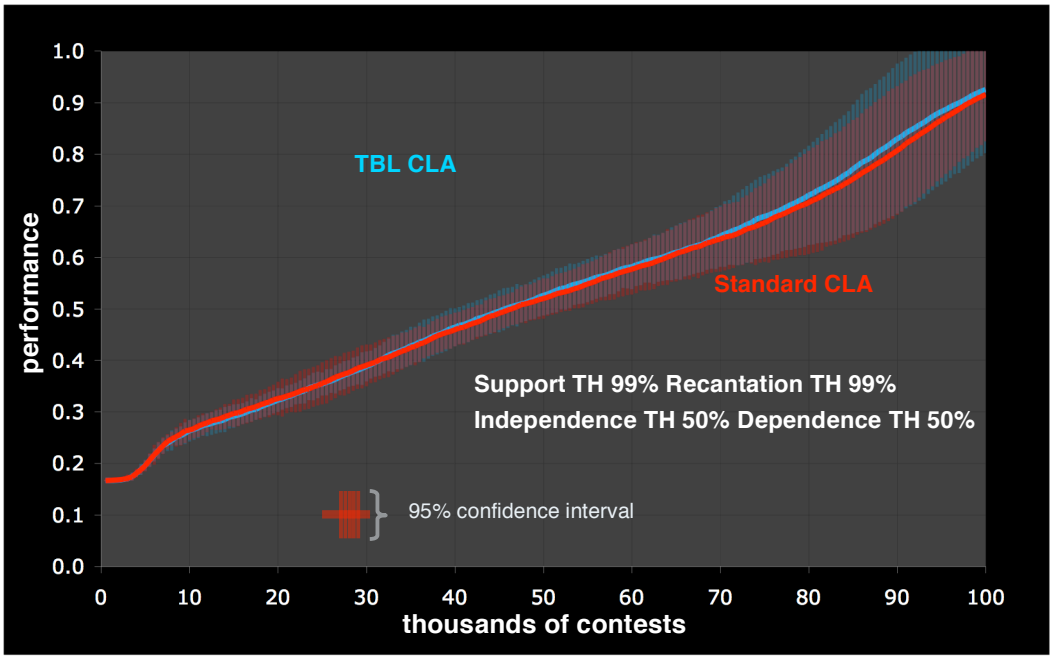

Figure 5. The thresholds can be set so that there is no difference between TBL and the Standard Selection Policy. In this case, a stimuli recants its tactic fairly early in recovery; because the independence TH is at 50% (selection confidences can not be lower than 50%), any stimuli that recants uses the Standard Selection Policy for the rest of its existence (99% selection confidence is sufficient for a stimulus to consistently choose its primary response).

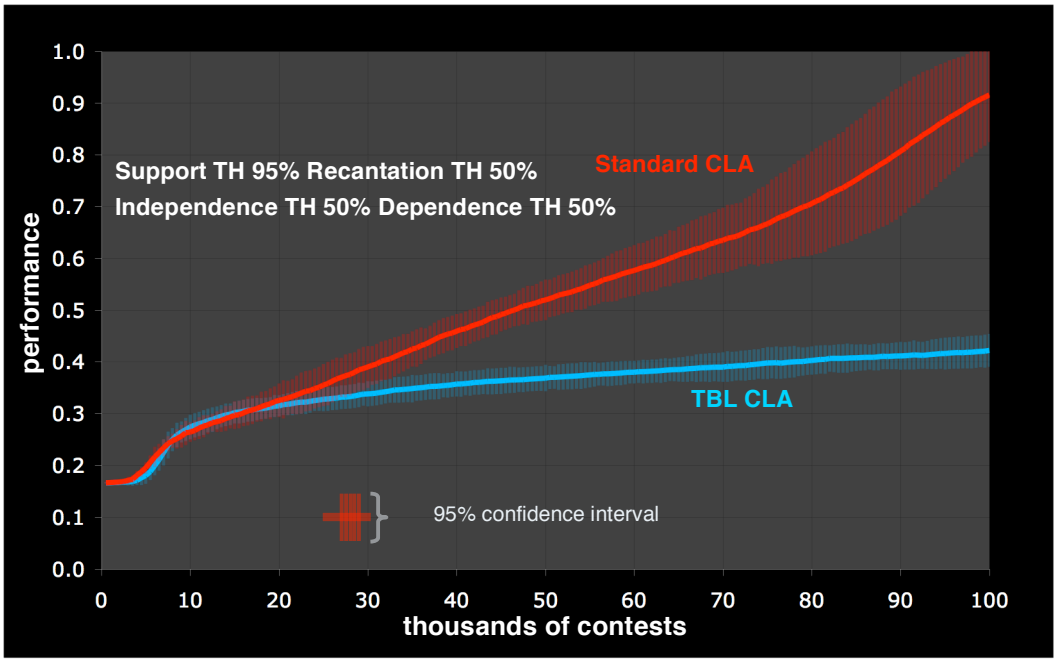

Figure 6. In this case, the stimuli *never* recant their support of a tactic. This is a problem because the compensation policy, which determines how to interpret the evaluation from the environment, has high expectations of stimuli that support and use tactics. In most cases, this policy helps TBL CLAs avoid local maxima, but if all the stimuli are supporters, then nothing except a completely correct evaluation will generate a positive compensation. At a collection length of 12, this is a very rare occurrence, and so the TBL CLA is receiving almost no positive compensations

7. Conclusions and Future Work

Tactic-Based Learning has been shown to speed recovery when the environment undergoes a sudden and dramatic change. In order to acheive this advantage, the thresholds for using a tactic must be chosen carefully, as ill-placed thresholds can severly hinder recovery.

This work leaves some intriguing questions which the authors hope to address in the future.

- **How does TBL affect recovery when stimuli may have more than one correct response?** These experiments focused solely on learning tasks that involved a single correct choice for each stimulus, but there are many situations in which more than one response is acceptable.

- **How much does TBL aid recovery when the changes to the learning problem are introduced gradually?** In these experiments the problem changed all at once, but there are many stituations, especially in real-world environments, that change slowly. It would be interesting to investigate if TBL is still effective under these conditions.

- **Is TBL effective when the changes to the learning task occur regardless of the learner's readiness?** For this research the CLA was allowed to train as long as necessary to become confident and accurate on the initial phase of the problem (or until it reached the time limit). While there are many learning tasks that can allow the learner to reach mastery, many tasks must change before then.

- **What other learning pathologies can be observed in a TBL CLA, and can these be mapped to known psychological phenomena?** In order to more fully understand the learning process, it is important to look at how and why learning fails.

8. References

[1] A. Armstrong, P. Bock, "Mentoring: an Intelligently Biased Selection Policy for Collective Learning Automata", *Proceeding of the 2005 Conference on Intelligent Engineering Systems through Artificial Neural Networks (ANNIE '05),* Volume 15, ASME Press, New York, 2005, pp. 121-130.

[2] S. Beer, Decision and Control: The Meaning of Operational Research and Management Cybernetics, West Sussex, England, 1966.

[3] L., Berk, *Child Development*, Allyn & Bacon, Boston, MA, 2003.

[4] P. Bock, *The Emergence of Artificial Cognition: An Introduction to Collective Learning*, World Scientific, New Jersey, 1993.

[5] P. Bock, "Observation of the Properties of a Collective Learning Stochastic Automaton", *Proceedings of the International Information Sciences Symposium*, Patras, Greece, 1976.

[6] K. Heckman, *An assessment of the effect of personality and forward context on the expertise of a collective learning system using the FIVE-FACTOR model of personality*. Doctoral Dissertation, The George Washington University Press, Washington, DC, 2004.

[7] T. Mitchell, *Machine Learning*. McGraw-Hill, New York, NY, 1997.

[8] A.W. Moore, M.S. Lee, "Efficient algorithms for minimizing cross validation error", *Proceedings of the 11th International Conference on Machine Learning*. Morgan Kauffman, San Francisco, CA, 1994.

[9] J. Piaget, *The Origins of intelligence in children*, Penguin, Harmondsworth, UK, 1977 (Originally published 1936).

[10] D. Portnoy, P. Bock, "Unsupervised Fuzzy-Membership Estimation of Terms in Semantic and Syntactic Lexical Classes", *Proceedings of the 33rd IEEE Applied Imagery Pattern Recognition Workshop*. Los Alamitos, CA, 2005.

[11] M. Pulaski, *Understanding Piaget: An Introduction to Children's Cognitive Development*, Harper & Row, New York, NY, 1980.

[12] K. Sycara, R. Guttal, J. Koning, S. Narasimhan, D. Navinchandra, "CADET: A case-based synthesis tool for engineering design". *International Journal of Expert Systems*, 4(2), 157-188, 1992.