[Yao, 2006] Y.Y. Yao Granular computing for data mining // Proceedings of SPIE Conference on Data Mining, Intrusion Detection, Information Assurance and Data Networks Security / B.V. Dasarathy (Ed.), Kissimmee, Florida, USA. 2006. pp. 1-12 (624105).

[Zadeh, 1997] L.A. Zadeh Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic // Fuzzy Sets Systems. 1997. Vol. 19. pp. 111-127.

Authors' Information

Kagramanyan Alexander – Kharkov National University, Svobody sq., 4, Kharkov, Ukraine, 61077

Mashtalir Vladimir – Kharkov National University of Radio Electronics, Lenina ave., 14, Kharkov, Ukraine, 61166, mashtalir@kture.kharkov.ua

Shlyakhov Vladislav – Kharkov National University of Radio Electronics, Lenina ave., 14, Kharkov, Ukraine 61166

SOLVING A DIRECT MARKETING PROBLEM BY THREE TYPES OF ARTMAP NEURAL NETWORKS

Anatoli Nachev

Abstract: An important task for a direct mailing company is to detect potential customers in order to avoid unnecessary and unwanted mailing. This paper describes a non-linear method to predict profiles of potential customers using dARTMAP, ARTMAP-IC, and Fuzzy ARTMAP neural networks. The paper discusses advantages of the proposed approaches over similar techniques based on MLP neural networks.

Keywords: ARTMAP, neural networks, data mining

ACM Classification Keywords: F.1.1 Models of Computation - neural networks, H.2.8 Database Applications - data mining

Introduction

Many companies use direct mailing to potential customers, or 'junk mail', to market a product or service. This can be an effective marketing approach, however much of this junk mail is really of no interest to the majority of people that receive it. The task how to predict the profiles of potential customers for a product, given information about the clients and a test sample of customers possessing the particular product is a well-known data mining problem from the world of direct marketing. The prediction task discussed in this paper, or the underlying problem, is to find a subset of customers with a probability of having a caravan insurance policy above some boundary probability. Those customers can be targeted by mailing promotional materials. The boundary of the targeted group depends on the cost and benefits such as of the costs of mailing and benefits of selling insurance policies.

The dataset used to test the proposed approach is based on real world business data [Van Der Putten, 2000]. It is a block of very detailed survey information on the people, some of whom bought and plan to buy a caravan insurance policy. The people were asked to answer 85 questions, each of which can be regarded as one feature in the classification. The block of data consists of 3 parts. The first is training data, which contains a number of survey responses, some of which come from caravan policy holders. The second part is testing data, and it contains answers from potential caravan insurance policy buyers. The last part is the true data that shows who of those potential buyers actually bought the policy at last. The maximum number of policy owners that could be found is 238. If a random selection is applied, average results provide 42 policy owners, or a hit rate (percentage of real policy buyers out of all predictions made) is about 6%.

A lot of techniques have been used to predict which customers are likely to respond or purchase a product, both linear and non-linear. Methods include: standard statistics [Van Der Putten, 2000], backpropagation MLP neural networks [Brierly, 2000], [Crocoll, 2000], [Shtovba et al., 2000], self-organizing maps (SOMs) [Vesanto et al., 2000], genetic programming, C4.5, CART, and other decision tree induction algorithms, fuzzy clustering and rule discovery, support vector machines (SVMs), logistic regression, boosting and bagging, all described in [Van Der Putten, 2000]. The best predictive technique reported in [Elkan, 2001] and [Van Der Putten, 2000] is the Naive Bayesian learning. It has been tested on 800 predictions and gives a hit rate about 15.2%. Predictors based on the backpropagation MLP networks show accuracy rate about 71% and hit rate about 13% as reported in [Brierly, 2000], [Candocia, 2004], [Crocoll, 2000], and [Van Der Putten, 2000].

This paper discusses three non-linear approaches based on dARTMAP, ARTMAP-IC, and Fuzzy ARTMAP neural networks.

Section 1 outlines the prediction task and introduces the reader to the discussed domain.

Section 2 outlines the main characteristics and function of predictors based on the three ARTMAP models.

Section 3 discusses the preprocessing steps needed to prepare an input dataset for the three ARTMAP network.

Section 4 describes experiments conducted in order to solve the prediction task and results from those experiments.

Predictors Based on dARTMAP, ATRMAP-IC, and Fuzzy ARTMAP Neural Networks

Adaptive Resonance Theory (ART) began with an analysis of human cognitive information processing [Grossberg, 1976]. Fundamental computational design goals have always included memory stability with fast or slow learning in an open and evolving input environment. As a real-time model of dynamic processes, an ART network is characterized by a system of ordinary differential equations, which are approximated by an algorithm for implementation purposes [Grossberg, 1980].

ART is a family of neural networks for fast learning, pattern recognition, and prediction, including both unsupervised: ART1, ART2, ART2-A, ART3, Fuzzy ART, Distributed ART; and supervised: ARTMAP, ARTMAP-IC, Fuzzy ARTMAP, ART-EMAP, ARTMAP-FTR, dARTMAP, and Default ARTMAP systems.

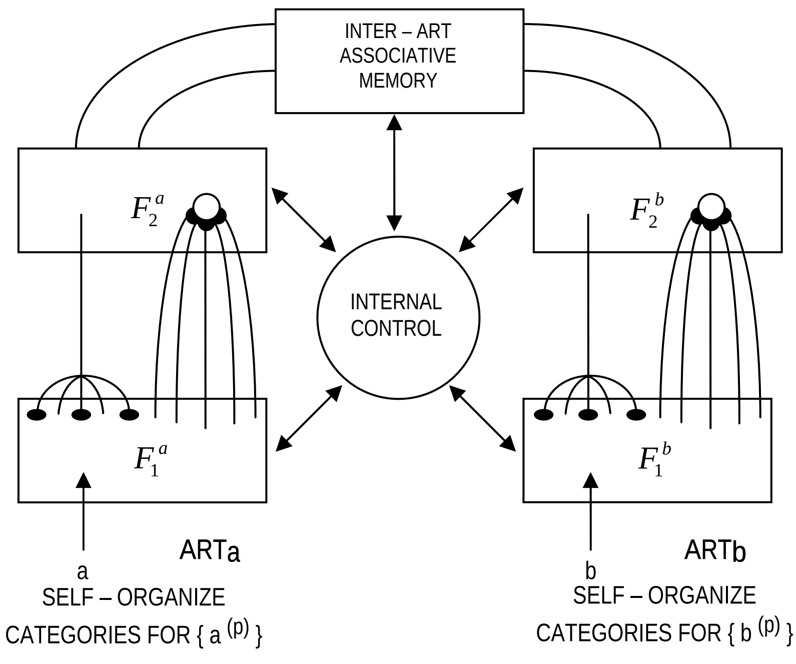

Figure 1. Components of an ARTMAP system.

ARTMAP neural networks develop stable recognition codes in real time in response to arbitrary sequences of input patterns. They were designed to solve the stability-plasticity dilemma that every intelligent machine learning system has to face: how to keep learning from new events without forgetting previously learned information.

ARTMAP networks consist of two ART1 networks, ARTa and ARTb, bridged via an inter-ART module, as shown on Figure 1. An ART module has three layers: the input layer (F0), the comparison layer (F1), and the recognition layer (F2) with m , m and n neurons, respectively. The neurons, or nodes, in the F2 layer represent input categories. The F1 and F2 layers interact with each other through weighted bottom-up and top-down connections, which are modified when the network learns. There are additional gain control signals in the network that regulate its operation.

Distributed ARTMAP (dARTMAP)

A key feature of the dARTMAP system is that it contains distributed ART (dART) modules instead of ART1 [Carpenter et al., 1998a]. A dART module combines the computational advantages of ART1 and MLP systems [Carpenter et al., 1996]. Properties include code stability when learning is fast and on-line, memory compression when inputs are noisy and unconstrained. The coding field of a dARTMAP supervised system is analogous to the hidden layer of a multi-layer perceptron (MLP), where distributed activation helps the network achieve memory compression and generalization. Each dARTMAP input first activates a distributed code. If this code produces a correct prediction, learning proceeds in the distributed coding mode. If the prediction is incorrect, the network resets the active code via match tracking feedback. In dARTMAP networks, the reset process triggers a search for a category node that can successfully code the current input. It also places the system in a winner-takes-all (WTA) coding mode for the duration of the search. In WTA mode, dARTMAP can add nodes incrementally as needed. When a coding node is added to the network, it becomes permanently associated with the output class that is active at the time. From then on, the network predicts this class whenever the same coding node is chosen in WTA mode.

ARTMAP-IC

ARTMAP- IC adds to the basic ARTMAP system new capabilities designed to solve the problem with inconsistent cases, which arises in prediction, where identical input vectors correspond to cases with different outcomes [Carpenter et al., 1998b]. It modifies the ARTMAP search algorithm to allow the network to encode inconsistent cases (IC) by involving an instance counting procedure and a new match tracking algorithm that consistently improve both predictive accuracy and code compression, compared to the basic ARTMAP networks. These added capabilities also allow ARTMAP-IC to encode predictions of inconsistent cases in the training set, giving good test set performance on various problems.

Fuzzy ARTMAP

Fuzzy ARTMAP, introduced in [Carpenter et al., 1992], is a natural extension to ARTMAP that uses Fuzzy ART instead of ART1 modules. Fuzzy ARTMAP is completely equivalent to ARTMAP, when the input domain is the Hamming cube {0,1}. It is capable of forming associative maps between clusters of its input and output domains in a supervised manner. Each module features its own set of parameters, with values assigned independently. ARTa is responsible clustering the input feature space; ARTb – for the output feature space. The inter-ART's role is to establish the correct association between input and output categories (cluster associations). The Fuzzy ARTMAP networks have been found useful in pattern recognition, because classification may be viewed as a many-to-one mapping task that entails clustering of the input space and then association of the produced clusters with a limited number of class labels (output clusters that encode a single class label).

The three ARTMAP modifications discussed above are designed to guarantee stable memories even with fast online learning. In contrast, when multi-layer perceptron (MLP) neural networks are used for classification problems, they employ slow off-line learning in order to avoid catastrophic forgetting in an open input environment, which limits adaptation for each input and so requires multiple presentations (epochs) of the training set. With fast learning, MLP memories suffer catastrophic forgetting.

Input Data

Input data for the three ARTMAP neural network simulators are available from the data mining company Sentient Machine Research [Van Der Putten, 2000].

The train dataset contains 5822 customer records. Each record consists of 86 attributes containing socio-demographic data represented by attributes 1-43 and product ownership attributes 44-86. The socio-demographic data is derived from zip codes. All customers living in areas with the same zip code have the same socio-demographic attributes. Attribute 86, "CARAVAN: Number of mobile home policies", is the target variable.

Evaluation dataset for validation of the prediction model consists of 4000 customer records. It has the same format as the training dataset, only the target is missing. Targets for the evaluation set have been provided by a separate file.

An important feature of the datasets is that the values of all numerical attributes were made discrete in advance. For example, instead of a real-valued feature giving the precise monetary amount that a customer pays for car insurance, the datasets include only a discrete-valued feature that categorizes this amount into one of seven different discrete levels.

Each of the three ARTMAP neural networks discussed here requires data samples, or input patterns, to be presented as M-component vectors of floating point numbers in the range [0, 1]. Therefore both train and evaluation datasets require normalization or mapping the original values into that range. For the purposes of normalization, each attribute was treated as an independent variable and submitted to a linear transformation:

$$\tilde{x}_i = \frac{(x_i - x_i^{\min})}{(x_i^{\max} - x_i^{\min})},$$

where $\tilde{x}_i$ is the normalized variable component; x_i is the original variable component; $x_i^{\max}$ is the max variable component; and $x_i^{\min}$ is the min variable component..

An effective solution of the prediction task should involve a selected subset of attributes, rather than the whole set, as the discriminatory power of a selection would be higher than one of the whole set. The selection itself is critical for a successful prediction. [Van Der Putten, 2000] reports a variety of techniques that rank importance and sensitivity of the attributes in the light of the prediction task, such as greedy feature selection algorithm, statistics, stepwise procedures, evolutionary algorithms, chi-analysis, etc.. For the purposes of the experiments reported here, those rank results were taken into consideration. In addition to that many empirical experiments and simulations were conducted in order to explore how different subsets of attributes influence the predictiveness of the three ARTMAP models. Experimental results show that the highest predictive rate for each of the tree models can be achieved by a set of the following attributes (sequence numbers correspond to the original notation):

- dARTMAP: {43, 47, 59}
- ARTMAP-IC: {43, 47, 59}
- Fuzzy ARTMAP: {1, 5, 12, 16, 18, 25, 30, 32, 34, 37, 42, 43, 44, 47, 59, 61, 65, 68, 80, 82, 85}.

Table 1 describes the attribute meanings.

The three attributes with maximal discriminatory power are:

- Purchasing power class (attribute #43, MKOOPKLA).
- Contribution car policies (attribute #47, PPERSAUT).
- Contribution fire policies (attribute #59, PBRAND).

No	*Attribute Name and Description*	*No*	*Attribute Name and Description*
1	*MOSTYPE Customer Subtype*	*43*	*MKOOPKLA Purchasing power class*
5	*MOSHOOFD Customer main type*	*44*	*PWAPART Contribution private third party insurance*
12	*MRELOV Other relation*	*47*	*PPERSAUT Contribution car policies*
16	*MOPLHOOG High level education*	*59*	*PBRAND Contribution fire policies*
18	*MOPLLAAG Lower level education*	*65*	*AWAPART Number of private third party insurance 1 - 12*
25	*MSKA Social class A*	*68*	*APERSAUT Number of car policies*
30	*MHHUUR Rented house*	*80*	*ABRAND Number of fire policies*
34	*MAUT0 No car*	*82*	*APLEZIER Number of boat policies*
37	*MINKM30 Income < 30.000*	*85*	*ABYSTAND Number of social security*
42	*MINKGEM Average income*		

Table 1. Selected attributes from train and evaluation datasets.

These three attributes describe people who are very likely to hold or possibly be in the market for a caravan insurance policy because such people are likely to be:

- People having a high level of purchasing power. Apart from 'Purchasing Power Class', all socio-demographic attributes, including customer segmentations by lifestyle, income, etc., often do not add any predictive power when behavioral data is available. People with high purchasing power are not necessarily enthusiastic about insuring their property, but they do have quite enough wealth to own a caravan, even if using it were not their prime hobby. Typical customers have high, or at least medium, education, status, social class, and income levels. For the feature selection, all demographic attributes were discarded, except attribute 43, 'MKOOPKLA Purchasing power class'
- Car owners with high contribution to car policy purchases. Those who do not have a car are unlikely to own a caravan, as they generally require to be towed. Car owners can be readily identified as those having existing car insurance policies. The amount spent on policies is also important. People who spend more on car insurance are most likely to be caravan policy buyers, and the more they spend, the more likely a buyer they are.
- People having fire policy with high level of contribution. This may indicate that the fire insurance is for a caravan. The level of the fire insurance cover that is most likely to be indicative of a caravan policy is level 4.

Intuitively, these three predictors identify customers who have a car and are wealthier than average, and who in general carry more insurance coverage than average. It is not surprising that these are the people who are most likely to have caravan insurance.

Experimental results however show that the Fuzzy ARTMAP model requires 13 more attributes to achieve its best predictiveness, as mentioned above. These extra attributes have less discriminatory power, but in combination with the major three, make the Fuzzy ARTMAP the best performer among the three neural network models.

Experimental Results

A number of experiments were conducted using simulators of dARTMAP, ARTMAP-IC, and Fuzzy ARTMAP neural networks. The goal of the experiments was to identify how order in which attributes and input patters are submitted influences the predictiveness; what the optimal network parameters are; and how network parameters affect the train and test time and memory consumption.

To maximize use of the datasets and to avoid bias in the selection of the training and test sets, a cross-validation technique was applied. Cross-validation creates N (N=5) copies of a classifier and tests each on 1/N of the evaluation dataset, after training it on 1/N-th of the training set. In other words, each classifier makes predictions for its 1/N-th of the data, yielding predictions for the whole set.

Results from the experiments showed that both dARTMAP and ARTMAP-IC models are sensitive to the order in which attributes appear in the training set. Out of six permutations of the attributes 43, 47, and 59, dARTMAP gave a satisfactory level of prediction for three permutations: {43, 47, 59}, {43, 59, 47}, and {59, 43, 47}; ARTMAP-IC gave satisfactory results for the four permutations: {43, 47, 59}, {43, 59, 47}, {59, 43, 47}, and {47, 43, 59}. Experiments showed that the Fuzzy ARTMAP is not sensitive to the order in which attributes appear to the input. To see how sequence of input patters affect the predictability, the networks was trained by eight different orders of the training sets: original order; real buyer entries shifted to the beginning; to the end; and five different randomly chosen orders. Experiments showed that the dARTMAP and ARTMAP-IC yield unsatisfactory prediction outputs for the shifted orders and do not distinguish the rest of the orders, yielding satisfactory results. The Fuzzy ARTMAP shows satisfactory results for all the orders.

To see how the network parameters influence the predictiveness, simulations with a full range of parameter values were conducted. Results show that an acceptable level of predictiveness for the three models can be achieved by the following values of network parameters: $\rho_{test} = 0$, $\alpha = 0.01$, $\varepsilon = -0.001$, and $p = 1.0$. The parameter $\beta = 1.0$ ensures best performance for the ARTMAP-IC and dARTMAP models, but Fuzzy ARTMAP performs best when $\beta = 0.968$.

The vigilance parameter ρ (Rhobar) affects the performance of the three models by tuning the details and granularity of the clusters, thus changing accuracy of predictions and hit rate. The parameter was set to various values between $0.915 \leq \rho \leq 0.955$ with step of increment 0.005. Figure 2 shows accuracy rate for each of the networks. Figures 3 shows total positive predictions, both correct and incorrect, which determines the boundary of the targeted group of customers, yet the expenses that should be made for direct marketing. Fuzzy ARTMAP has best performance when the size of the targeted group of is about 160; ARTMAP-IC outputs a group of 63 customers, whilst dARTMAP achieves a good prediction rate with a relatively small group of about 30 customers. Figure 4 shows the number of correct predictions. Given that the three models output different number of potential buyers, as per Figure 3, the number of correct predictions is influenced by the number of the total number of potential buyers, rather than a better prediction. Figure 5 shows how the vigilance parameter affects the hit rate. Maximal hit rates are: 17.96% for the Fuzzy ARTMAP; 14.29% for the ARTMAP-IC, and 30% for the dARTMAP. These hit rates exceed the reported 13% of the MLP networks used for the same prediction task. This result exceeds the best hit rate of 15% reported in [Elkan, 2001] and [Van Der Putten, 2000], but direct comparison of results would not be accurate, as the two cases are based on different size of the targeted group. For boundaries where a scale of a direct marketing is comparable to the scale adopted by the experiments described above, the three neural network models could be considered as good performing and applicable.

Training time is another advantage of the three ARTMAP models, which should be pointed out. With the dataset discussed here, the three neural networks get trained for about 5 seconds in contrast to the MPL networks that require about 35 minutes [Shtovba et al., 2000]. MPL networks, however, outperform ARTMAP in the test time, but both ARTMAP and MLP respond for less that a second, which is response in a real time.

Another advantage of the three ARTMAP models is that for implementation of the long-term memory (LTM) the simulators consume a little RAM - about 4.9 KB.

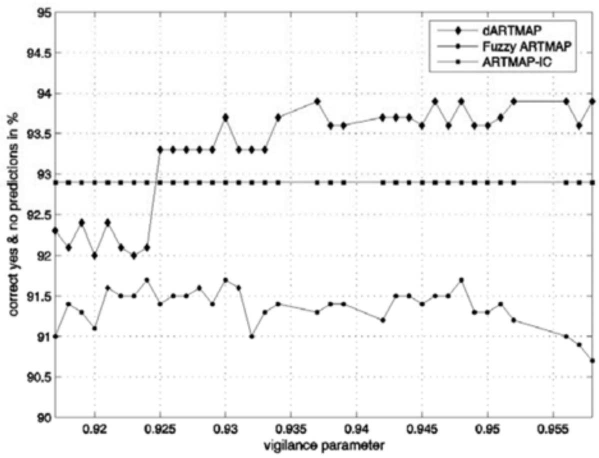

Figure 2. Correct yes and no predictions.

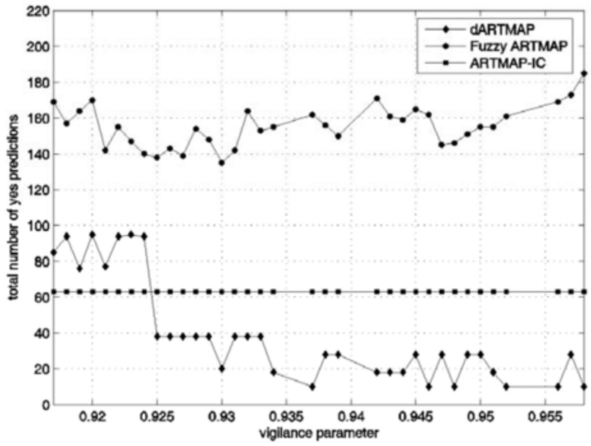

Figure 3. Total yes predictions (both correct and wrong).

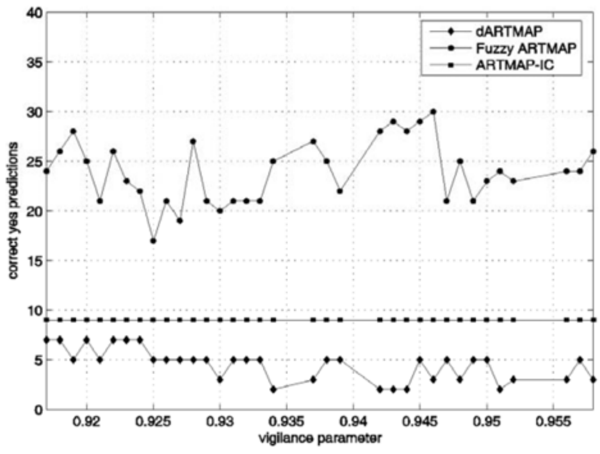

Figure 4. Correct yes predictions.

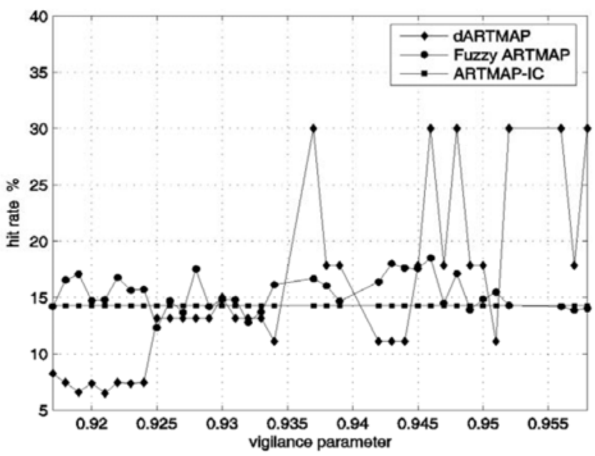

Figure 5. Hit rate.

Conclusion

This paper describes three non-linear methods to predict profiles of potential customers of a product using dARTMAP, ARTMAP-IC, and Fuzzy ARTMAP neural networks, given a test sample of customers possessing the particular product. Such a prediction can be used for direct marketing purposes. The proposed methods require a pre-processing of the test sample data in order to prepare specific input for simulators based on the ARTMAP paradigm. The theoretical concepts and conducted experiments lead to the following conclusions:

- The three ARTMAP models discussed in the paper show higher hit rates than those obtained by other non-linear approaches based on MLP neural networks.
- The Fuzzy ARTMAP model provides stable prediction, not affected by the order of attributes presented to the model, nor the order of input patterns. In contrast, the dARTMAP and ARTMAP-IC are sensitive to both orders and provide satisfactory results in some cases only.
- Predictors based on the three models discussed here have a very short training period, in contrast to the MPL neural networks. The three models also consume a negligible small amount of memory, which makes them applicable in large scale prediction tasks.

All conclusions above feature the Fuzzy ARTMAP model as most stable and suitable for business applications like those discussed here.

Bibliography

Brierly, P. (2000) 'Characteristics of caravan insurance policy owners', [online] available at http://www.liacs.nl/~putten/library/cc2000/ brierl~1.pdf, 2000

Candocia, F. (2004) 'EEL 6825 Pattern Recognition', [online] available at http://www.cise.ufl. edu/~bfeng/eel6825/ 6825_pattern.htm, 2004

Carpenter, G.A. (1996). 'Distributed ART networks for learning, recognition, and prediction.' Proceedings of the World Congress on Neural Networks (WCNN'96), pp. 333–344.

Carpenter, G.A., Milenova, B., & Noeske, B. (1998a). 'dARTMAP: A neural network for fast distributed supervised learning.' Neural Networks, 11, 793-813. Technical Report CAS/CNS TR-97-026, Boston, MA: Boston University.

Carpenter, G. A. and Markuzon, N. (1998b), 'ARTMAP-IC and Medical Diagnosis: Instance Counting and Inconsistent Cases', Neural Networks, 11:2, 323-336.

Crocoll, W. (2000) 'Artificial Neural Network Portion of Coil Study', [online] available at http://www.liacs.nl/~putten/library/cc2000/crocol~1.pdf

Elkan, C. (2001) 'Magical Thinking in Data Mining: Lessons From CoIL Challenge 2000.' Proceedings of the Seventh International Conference on Knowledge Discovery and Data Mining (KDD'01), pp. 426-431.

Grossberg, S. (1976) Adaptive pattern classification and universal recoding. II: Feedback, expectation, olfaction, and illusions. Biological Cybernetics, 23, 187–202, 1976.

Grossberg S. (1980). 'How does a brain build a cognitive code?', Psychological Review, 87, 1–51., 1980

Shtovba, S. and Mashnitskiy, Y. (2000) 'The Backpropagation Multilayer Feedforward Neural Network Based Competition Task Solution', [online] available at http://www.liacs.nl/ ~putten/library/cc2000/shtob~1.pdf, 2000

Van Der Putten, P. and Van Someren M. (eds).(2000) CoIL Challenge 2000: The Insurance Company Case. Published by Sentient Machine Research, Amsterdam. Also a Leiden Institute of Advanced Computer Science Technical Report 2000-09.

Vesanto, J. and Sinkonen, J. (2000) Submission for the Coil Chalange 2000, [online] available at http://www.liacs.nl/~putten/ library/cc2000/vesant~1.pdf, 2000

Author's Information

Anatoli Nachev – Information Systems, Dept. of A&F, NUI Galway, Ireland; e-mail: anatoli.nachev@nuigalway.ie