DATA MINING WITH FUZZY ARTMAP NEURAL NETWORKS: PREDICTION OF PROFILES OF POTENTIAL CUSTOMERS

Anatoly Nachev

Abstract: The task how to predict profiles of potential customers for a product is important for a direct mailing company. A good prediction allows the company to detect potential customers and to avoid unnecessary and unwanted mailing. This paper describes a non-linear methodology to predict profiles of potential customers using Fuzzy ARTMAP neural networks. The paper discusses advantages of the proposed approach over similar techniques.

Keywords: Fuzzy ARTMAP, neural networks, data mining

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

Introduction

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. Direct mailings to a company's potential customers, or 'junk mail' to many can be a very effective way for to market a product or service. However, much of this junk mail is really of no interest to the majority of people that receive it.

A lot of techniques, both linear and non-linear, have been used to predict those customers who are likely to respond or purchase a product, for example statistical methods such as linear regression, decision trees, MPL neural networks (NNs), etc.

The prediction task discussed in this paper, in other words 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 depends on the cost and benefits such as of the costs of mailing and benefits of selling insurance policies. The dataset used for experiments 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) of 6%.

A wide variety of methodological approaches were used to solve this prediction task. 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 proposes a non-linear approach based on Fuzzy ARTMAP neural networks to solve the prediction task outlined above.

Section 1 outlines the prediction task and variety of approaches to solve it.

Section 2 discusses the main characteristics of a predictor based on the Fuzzy ARTMAP model and outlines how it functions.

Section 3 describes the preprocessing steps needed to prepare an input dataset for a Fuzzy ARTMAP network.

Section 4 describes experiments conducted to solve the prediction task by a Fuzzy ARTMAP simulator and discusses experimental results.

Predictors Based on ART Neural Networks

Adaptive Resonance Theory (ART) began with an analysis of human cognitive information processing [Grossberg, 1976]. Fundamental computational design goals have therefore 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, Fuzzy ARTMAP, ART-EMAP, ARTMAP-FTR, Distributed ARTMAP, and Default ARTMAP systems.

ARTMAP architectures are neural networks that 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 were designed to accept binary input patterns [Carpenter et al., 1991]. 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.

Fuzzy ARTMAP, introduced in [Carpenter et al., 1992], is a natural extension to ARTMAP. 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, whose values can be assigned independently. ARTa is clustering the input feature space and ARTb 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).

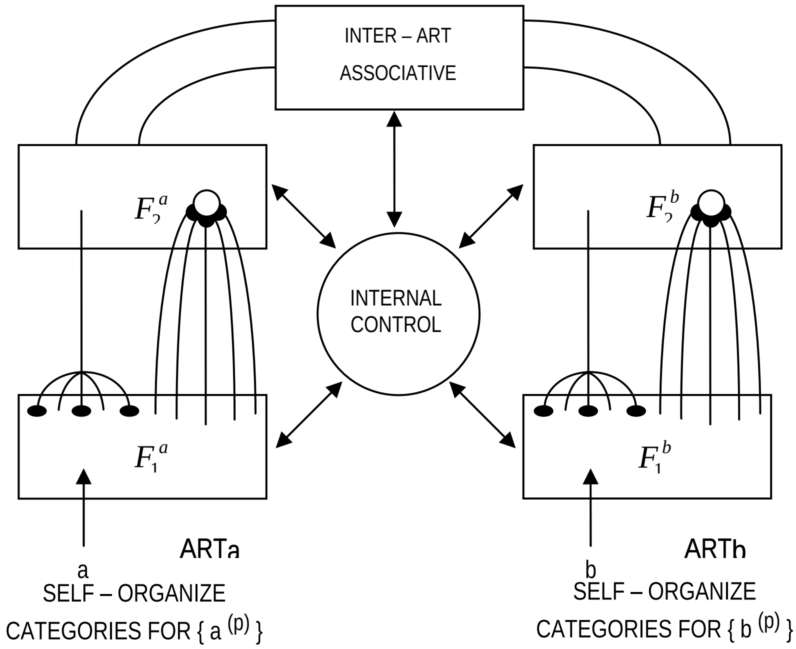

Figure 1. Components of an ARTMAP system.

Many applications of supervised learning systems such as Fuzzy ARTMAP are classification problems, where the trained system tries to predict a correct category given a test set input vector.

From another hand, when multi-layer perceptron (MLP) NNs have been used for classification problems, they employ slow off-line learning 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.

Features of a fast-learn system, such as its ability to encode significant rare cases and to learn quickly in the field, may be essential for the given application domain.

Data Pre-processing

The dataset used for simulations is owned and supplied by the data mining company Sentient Machine Research [Van Der Putten, 2000]. It is subdivided into two parts: a train dataset of 5822 customer records and an evaluation dataset of 4000 customer records

Each record consists of 86 attributes containing socio-demographic data represented by attributes (numbers 1-43) and product ownership attributes (numbers 44-86). The socio-demographic data is derived from zip codes of addresses. 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 that shows if a customer hold an insurance policy or not.

The evaluation dataset is used for validation of the prediction model. It has the same format as the training data set, but only the target attribute 85 is missing. Targets for the evaluation set have been provided by a separate file.

The prediction task can be solved involving a subset of selected attributes, or features, and their selection is critical for a successful prediction. [Van Der Putten, 2000] reports a variety of selection techniques that rank similarly importance and sensitivity of the attributes in the light of the prediction task. For the purposes of the experiments reported here all important attributes have been taken on board. Many simulations were conducted to explore how these attributes influence the predictiveness of the Fuzzy ARTMAP model both individually and in groups. Results show that the highest predictive rate can be achieved by a set of the following features: (numbers correspond to the original dataset notations):

S = {1, 5, 12, 16, 18, 25, 30, 32, 34, 37, 42, 43, 44, 47, 59, 61, 65, 68, 80, 82, 85}. See full feature description in Table 1.

No	Feature Name and Description	No	Feature 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 features from train and evaluation datasets.

The feature set can be generally interpreted as identifying customers who are:

1. 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.

2. 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.

3. People having a high level of purchasing power. 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.

Experiments

A number of experiments were conducted using a simulator of the Fuzzy ARTMAP model. All experiments explored how this model solves the prediction task paying attention to the following critical factors:

- Sensitiveness to the order in which features and input patters are submitted. This is due to the fact that some ART models commit LTM nodes differently in different orders.
- Best tuning of the network and using optimal values of the network parameters.
- How network parameters affect the train time, 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 created N copies of the classifier and tested 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. Cross-validation was applied using N=5.

The first group of experiments showed that the Fuzzy ARTMAP is not sensitive to the order in which features are ordered. The experiments also explored how order of input patterns influences the predictiveness by submitting various randomly generated sequences of input patterns. Results reveal that variations in results are slightingly small and can be ignored.

A series of simulations were conducted to explore how network parameters affect the predictiveness of the model. Experiments show that best results can be achieved by the default values of most of the network parameters, namely: $\rho_{test} = 0$, $\alpha = 0.01$, $\varepsilon = -0.001$, and $p = 1.0$. Default value of the parameter $\beta = 1.0$ however does not ensure best performance. Results show that all values in the interval $0.95 \le \beta \le 0.97$ provide a better performance with a maximum of correct predictions at $\beta = 0.968$. Confusion matrix is shown in Table 1.

		Predicted		
		No	Yes	
Actual	No	3625	137	3762
	Yes	211	30	238
		3852	167	4000

Table 1. Confusion matrix of prediction made by Fuzzy ARTMAP with feature set S and parameters $\beta = 0.968$; $\rho = 0.94$

The vigilance parameter ρ (Rhobar) affects the Fuzzy ATRMAP performance 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 $0.915 \le \rho \le 0.955$ with step of increment 0.005. Figure 2 shows network performance and the best accuracy rate 91.4% obtained at $\rho = 0.94$. Figures 3 and 4 reveal number of total positive predictions and correct positive predictions respectively. Figure 5 shows how vigilance affects the hit rate. In most of the parameter values the model provides a hit rate above 15% with best rate nearly 18% (17.96%) at $\rho = 0.94$. 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 they are based on different boundaries of the prediction task. For boundaries where the scale of a direct mailing is comparable with the scale adopted by the experiments reported here, a Fuzzy ARTMAP predictor outperforms the other techniques in terms of hit rate.

Training time is another advantage of the Fuzzy ARTMAP model that should be pointed out. With this prediction task the Fuzzy ARTMAP neural networks train for about 5 seconds in contrast to the MPL networks that require about 35 minutes as reported in [Shtovba et al., 2000]. From another hand the MPL networks outperforms Fuzzy ARTMAP in the test time, but both models respond for less that a second, which them makes equally effective and working in a real time.

All simulations show that the LTM memory used by the Fuzzy ARTMAP model requires about 4.9 KB RAM, which makes predictors based on this model efficient with large scale prediction tasks.

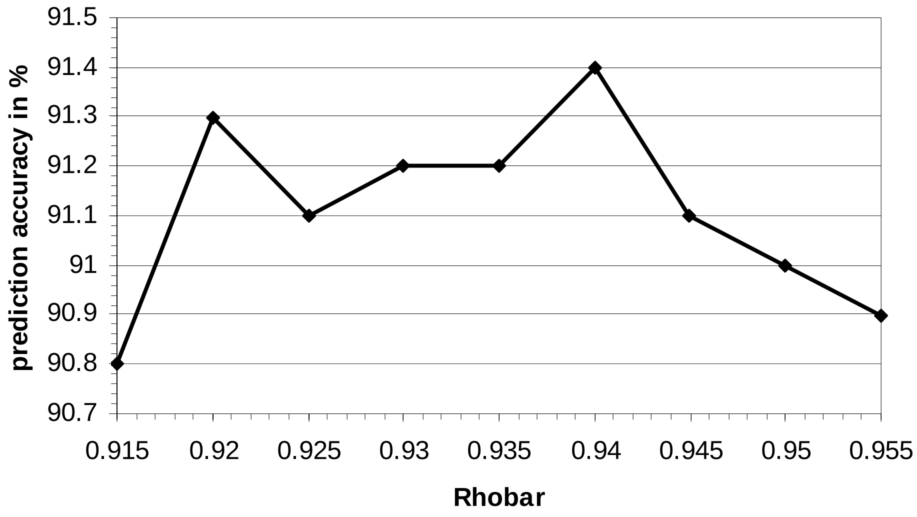

Figure 2. Prediction accuracy (both positive and negative) of Fuzzy ARTMAP neural network.

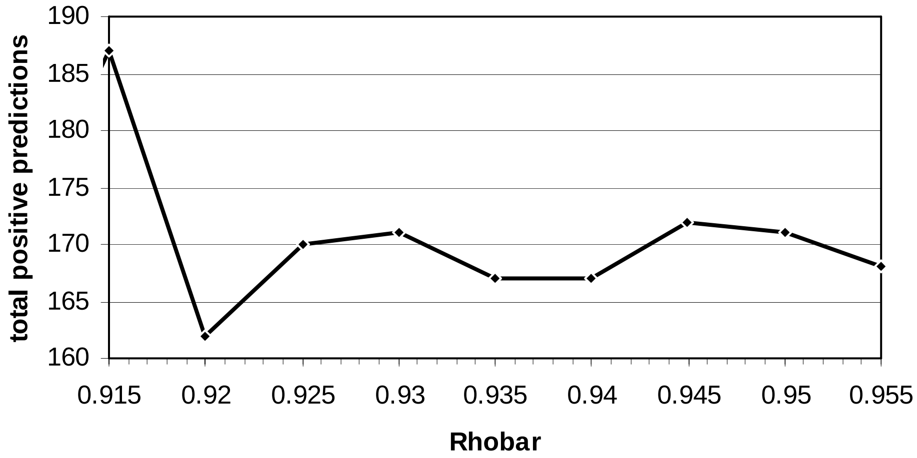

Figure 3. Total positive predictions (both correct and wrong) by Fuzzy ARTMAP neural network.

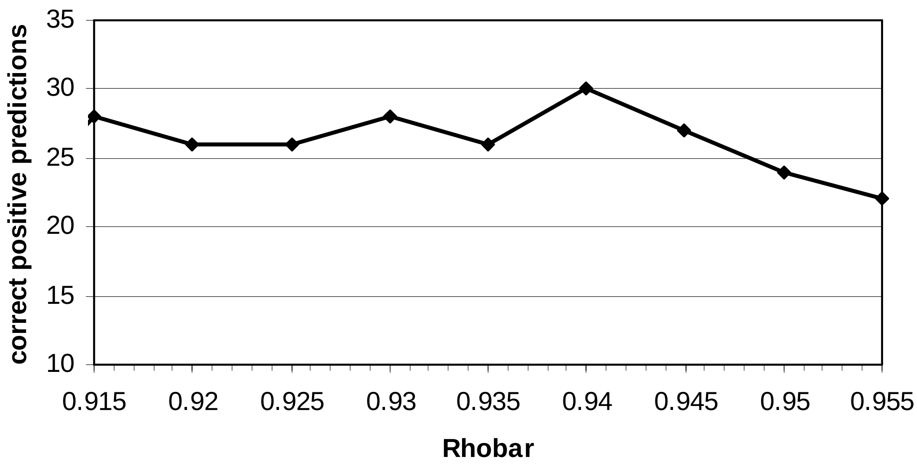

Figure 4. Correct positive predictions by Fuzzy ARTMAP neural network.

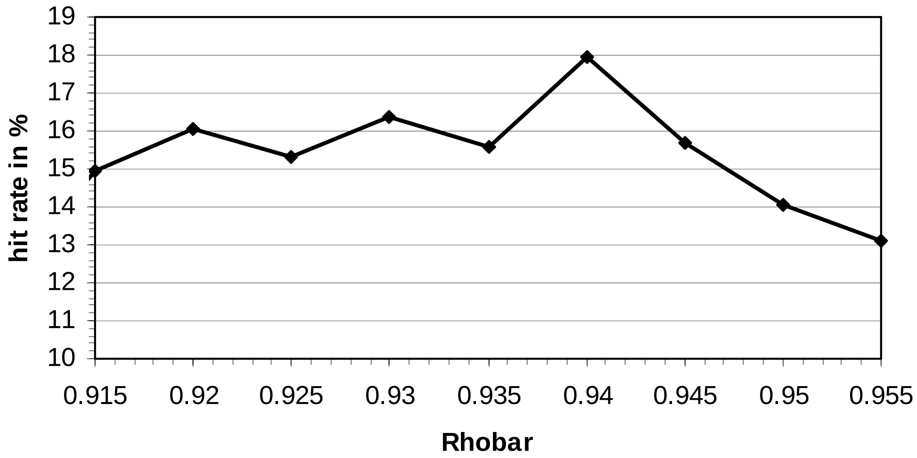

Figure 5. Prediction hit rate by Fuzzy ARTMAP neural network

Conclusion

This paper proposes a non-linear approach, based on Fuzzy ARTMAP neural networks, for solving a prediction task to identify potential buyers of insurance policies. Solution requires an initial processing of the data set to prepare input for a Fuzzy ARTMAP simulator. All conducted experiments lead to the following conclusions:

- Fuzzy ARTMAP model outperforms other predictive techniques, including similar non-linear approaches based on MLP neural networks.
- The model provides stable predictive abilities regardless of order of features submitted and order of input patterns.
- Predictors based on Fuzzy ARTMAP neural networks have a very short training period, in contrast to the MPL neural networks, and small resource consumption, which makes them applicable for large scale prediction tasks.

All conclusions above feature the Fuzzy ARTMAP model as an effective data mining tool for problem areas similar to the prediction task discussed here.

Bibliography

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

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

[Carpenter et al., 1991] Carpenter, G.A., Grossberg, S., & Reynolds, J.H. (1991). ARTMAP: Supervised real–time learning and classification of nonstationary data by a self–organizing neural network. Neural Networks, 4, 565–588.

[Carpenter et al., 1992], Carpenter, G.A., Grossberg, S., Markuzon, N., Reynolds, J.H., & Rosen, D.B., Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps, IEEE Transactions on Neural Networks, 3, 698-713

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

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

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

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

[Shtovba et al., 2000] S. Shtovba, Y. Mashnitskiy, The Backpropagation Multilayer Feedforward Neural Network Based Competition Task Solution, Available at http://www.liacs.nl/ ~putten/library/cc2000/shtob~1.pdf , 2000

[Van Der Putten, 2000] P. Van Der Putten and M. van Someren (eds). 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. June 22, 2000.

[Vesanto et al., 2000] J. Vesanto, J. Sinkonen, Submission for the Coil Chalange 2000, 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