

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Benchmarking the Data Mining Algorithms with Adaptive Neuro-Fuzzy Inference System in GSM Churn Management

Adem Karahoca, Dilek Karahoca and Nizamettin Aydın
*Software Engineering Department, Bahcesehir University
 Turkey*

1. Introduction

Turkey has started to distribute Global Services of Mobile (GSM) 900 licences in 1998. Turkcell and Telsim have been the first players in the GSM market and they bought licenses respectively. In 2000, GSM 1800 licenses were bought by ARIA and AYCELL respectively. After then, GSM market has saturated and customers started to switch to other operators to obtain cheap services, number mobility between GSM operators, and availability of 3G services.

One of the major problems of GSM operators has been churning customers. Churning means that subscribers may move from one operator to another operator for some reasons such as the cost of services, corporate capability, credibility, customer communication, customer services, roaming and coverage, call quality, billing and cost of roaming (Mozer et al., 2000). Therefore churn management becomes an important issue for the GSM operators to deal with. Churn management includes monitoring the aim of the subscribers, behaviours of subscribers, and offering new alternative campaigns to improve expectations and satisfactions of subscribers. Quality metrics can be used to determine indicators to identify inefficiency problems. Metrics of churn management are related to network services, operations, and customer services.

When subscribers are clustered or predicted for the arrangement of the campaigns, telecom operators should have focused on demographic data, billing data, contract situations, number of calls, locations, tariffs, and credit ratings of the subscribers (Yu et al., 2005).

Predictions of customer behaviour, customer value, customer satisfaction, and customer loyalty are examples of some of the information that can be extracted from the data stored in a company's data warehouses (Hadden et al., 2005).

It is well known that the cost of retaining a subscriber is much cheaper than gaining a new subscriber from another GSM operator (Mozer et al., 2000). When the unhappy subscribers are predicted before the churn, operators may retain subscribers by new offerings. In this situation in order to implement efficient campaigns, subscribers have to be segmented into classes such as loyal, hopeless, and lost. This segmentation has advantages to define the customer intentions. Many segmentation methods have been applied in the literature. Thus,

Source: Data Mining and Knowledge Discovery in Real Life Applications, Book edited by: Julio Ponce and Adem Karahoca, ISBN 978-3-902613-53-0, pp. 438, February 2009, I-Tech, Vienna, Austria

churn management can be supported with data mining modelling tools to predict hopeless subscribers before the churn. We can follow the hopeless groups by clustering the profiles of the customer behaviours. Also, we can benefit from the prediction advantages of the data mining algorithms.

Data mining tools have been used to analyze many profile-related variables, including those related to demographics, seasonality, service periods, competing offers, and usage patterns. Leading indicators of churn potentially include late payments, numerous customer service calls, and declining use of services.

In data mining, one can choose a high or low degree of granularity in defining variables. By grouping variables to characterize different types of customers, the analyst can define a customer segment. A particular variable may show up in more than one segment. It is essential that data mining results extend beyond obvious information. Off-the-shelf data mining solutions may provide little new information and thus serve merely to predict the obvious. Tailored data mining solutions can provide far more useful information to the carrier (Gerpott et al., 2001).

Hung et al., 2006 proposed a data mining solution with decision trees and neural networks to predict churners to assess the model performances by LIFT (a measure of the performance of a model at segmenting the population) and hit ratio. Data mining approaches are considered to predict customer behaviours by CDRs (call detail records) and demographics data in (Wei et al., 2002; Yan et al., 2005; Karahoca, 2004).

In this research, main motivation is investigating the best data mining model(s) for churning, according to the measure of the performances of the model(s). We have utilized data sets which are obtained from a Turkish mobile telecom operator's data warehouse system to analyze the churn activities.

2. Material and methods

Loyal 24900 GSM subscribers were randomly selected from data warehouses of GSM operators located in Turkey. Hopeless 7600 candidate churners were filtered from databases during a period of a year. In real life situations, usually 3 to 6 percent annual churn rates are observed. Because of computational complexity, if all parameters within subscriber records were used in predicting churn, data mining methods could not estimate the churners. Therefore we selected 31% hopeless churners for our dataset and discarded the most of the loyal subscribers from the dataset.

In pattern recognition applications, the usual way to create input data for the model is through feature extraction. In feature extraction, descriptors or statistics of the domain are calculated from raw data. Usually, this process involves in some form of data aggregation.

The unit of aggregation in time is one day. The feature mapping transforms the transaction data ordered in time to static variables residing in feature space. The features used reflect the daily usage of an account. Number of calls and summed length of calls to describe the daily usage of a mobile phone were used. National and international calls were regarded as different categories. Calls made during business hours, evening hours and night hours were also aggregated to create different features. The parameters listed in Table 1, were taken into account to detect the churn intentions.

Abbreviation	Meaning
City	Subscribers' city
Age	Subscribers' age
Occupation	Occupation code of Subscriber
Home	Home city of Subscriber
Month-income	Monthly income of Subscriber
Credit-limit	Credit Limit of Subscriber
Avglen-call-month	Average length of calls in last month
Avg-len-call4-6month	Average length of calls in last 4-6 months
Avg-len-sms-month	Number of SMS in last month
Avglensms4-6month	Number of SMS in last 4-6 months
Tariff	Subscriber tariff
Marriage	Marital status of the subscriber
Child	Number of children of the subscriber
Gender	Gender of the subscriber
Month-Expense	Monthly expense of subscriber
Sublen	Length of service duration
CustSegment	Customer segment for subscriber
AvgLenCall3month	Average length of calls in last 3 months
TotalSpent	Total expenditure of the subscriber
AvgLenSms3month	Number of sent SMS in last 3 months
GsmLineStatus	Status of subscriber line
Output	Churn status of subscriber

Table 1. Extracted features

Summarized attributes that are augmented by different data sources are given in Table 1. These attributes are used as input parameters for churn detection process. These attributes are expected to have higher impact on the outcome (whether churning or not). In order to reduce computational complexity of the analysis, some of the fields in the set of variables of corporate data warehouse are ignored. The attributes with highest Spearman's Rho values are listed in Table 2. The factors are assumed to have the highest contribution to the ultimate decision about the subscriber.

Attribute Name	Spearman's Rho	Input Code
Month-Expense	0.4804	in1
Age	0.4154	in2
Marriage	0.3533	in3
TotalSpent	0.2847	in4
Month-income	0.2732	in5
CustSegment	0.2477	in6
Sublen	0.2304	in7

Table 2. Ranked Attributes

A number of different methods were considered to predict churners from subscribers. In this section the brief definitions of data mining methods are given. The methods that described in this section are general methods that are used for modelling the data. These methods are *JRip*, *PART*, *Ridor*, *OneR*, *Nnge*, *Decision Table*, *Conjunction Rules*, *AD Trees*, *IB1*, *Bayesian networks* and *ANFIS*. Except *ANFIS*, all the methods are executed in WEKA (Waikato Environment for Knowledge Analysis) data mining software [Frank & Witten, 2005].

2.1 JRip method

JRip implements a propositional rule learner, “Repeated Incremental Pruning to Produce Error Reduction” (RIPPER), as proposed by [Cohen, 1995]. JRip is a rule learner. In principle it is similar to the commercial rule learner RIPPER. The RIPPER rule learning algorithm is an extended version of learning algorithm IREP (Incremental Reduced Error Pruning). It constructs a rule set in which all positive examples are covered, and its algorithm performs efficiently on large, noisy datasets. Before building a rule, the current set of training examples are partitioned into two subsets, a growing set (usually 2/3) and a pruning set (usually 1/3). The rule is constructed from examples in the growing set. The rule set begins with an empty rule set and rules are added incrementally to the rule set until no negative examples are covered. After growing a rule from the growing set, condition is deleted from the rule in order to improve the performance of the rule set on the pruning examples. To prune a rule, RIPPER considers only a final sequence of conditions from the rule, and selects the deletion that maximizes the function [Frank & Witten, 2005].

2.2 PART method

The PART algorithm combines two common data mining strategies; the divide-and-conquer strategy for decision tree learning and the separate-and-conquer strategy for rule learning. The divide-and-conquer approach selects an attribute to place at the root node and “divides” the tree by making branches for each possible value of the attribute. The process then continues recursively for each branch, using only those instances that reach the branch. The separate-and-conquer strategy is employed to build rules. A rule is derived from the branch of the decision tree explaining the most cases in the dataset, instances covered by the rule are removed, and the algorithm continues creating rules recursively for the remaining instances until none are left. The PART implementation differs from standard approaches in that a pruned decision tree is built for the current set of instances, the leaf with the largest coverage is made into a rule and the tree is discarded. By building and discarding decision trees to create a rule rather than building a tree incrementally by adding conjunctions one at a time avoids a tendency to over prune. This is a characteristic problem of the basic separate and conquer rule learner.

The key idea is to build a partial decision tree instead of a fully explored one. A partial decision tree is an ordinary decision tree that contains branches to undefined sub trees. To generate such a tree, we integrate the construction and pruning operations in order to find a stable subtree that can be simplified no further. Once this subtree has been found tree building ceases and a single rule is read. The tree building algorithm splits a set of examples recursively into a partial tree. The first step chooses a test and divides the examples into subsets. PART makes this choice in exactly the same way as C4.5. [Frank & Witten, 2005].

2.3 Ridor method

Ridor generates the default rule first and then the exceptions for the default rule with the least (weighted) error rate. Later, it generates the best exception rules for each exception and iterates until no exceptions are left. Thus it performs a tree-like expansion of exceptions and the leaf has only default rules but no exceptions. The exceptions are a set of rules that predict the improper instances in default rules [Gaines & Compton, 1995].

2.4 OneR method

OneR, generates a one-level decision tree, that is expressed in the form of a set of rules that all test one particular attribute. 1R is a simple, cheap method that often comes up with quite good rules for characterizing the structure in data [Frank & Witten, 2005]. It turns out that simple rules frequently achieve surprisingly high accuracy [Holte, 1993].

2.5 Nnge

Nearest-neighbor-like algorithm is using for non-nested generalized exemplars Nnge which are hyper-rectangles that can be viewed as if-then rules [Martin, 1995]. In this method, we can set the number of attempts for generalization and the number of folder for mutual information.

2.6 Decision tables

As stated by Kohavi, decision tables are one of the possible simplest hypothesis spaces, and usually they are easy to understand. A decision table is an organizational or programming tool for the representation of discrete functions. It can be viewed as a matrix where the upper rows specify sets of *conditions* and the lower ones sets of *actions* to be taken when the corresponding conditions are satisfied; thus each column, called a *rule*, describes a procedure of the type “if conditions, then actions”.

The performance of this method is quite good on some datasets with continuous features, indicating that many datasets used in machine learning may not require these features, or these features may have few values [Kohavi, 1995].

2.7 Conjunctive rules

This method implements a single conjunctive rule learner that can predict for numeric and nominal class labels. A rule consists of antecedents “AND”ed together and the consequent (class value) for the classification/regression. In this case, the consequent is the distribution of the available classes (or mean for a numeric value) in the dataset.

2.8 AD trees

AD Trees can be used for generating an alternating decision (AD) trees. The number of boosting iterations needs to be manually tuned to suit the dataset and the desired complexity/accuracy tradeoffs. Induction of the trees has been optimized, and heuristic search methods have been introduced to speed learning [Freund & Mason, 1999].

2.9 Nearest neighbour Instance Based learner (IB1)

IBk is an implementation of the k-nearest-neighbors classifier that employs the distance metric. By default, it uses just one nearest neighbor ($k=1$), but the number can be specified [Frank & Witten, 2005].

2.10 Bayesian networks

Graphical models such as Bayesian networks supply a general framework for dealing with uncertainty in a probabilistic setting and thus are well suited to tackle the problem of churn management. Bayesian Networks was coined by Pearl (1985).

Graphical models such as Bayesian networks supply a general framework for dealing with uncertainty in a probabilistic setting and thus are well suited to tackle the problem of churn management. Every graph of a Bayesian network codes a class of probability distributions. The nodes of that graph comply with the variables of the problem domain. Arrows between nodes denote allowed (causal) relations between the variables. These dependencies are quantified by conditional distributions for every node given its parents.

2.11 ANFIS

A Fuzzy Logic System (FLS) can be seen as a non-linear mapping from the input space to the output space. The mapping mechanism is based on the conversion of inputs from numerical domain to fuzzy domain with the use of fuzzy sets and fuzzifiers, and then applying fuzzy rules and fuzzy inference engine to perform the necessary operations in the fuzzy domain [Jang,1992 ; Jang,1993]. The result is transformed back to the arithmetical domain using defuzzifiers. The ANFIS approach uses Gaussian functions for fuzzy sets and linear functions for the rule outputs. The parameters of the network are the mean and standard deviation of the membership functions (antecedent parameters) and the coefficients of the output linear functions (consequent parameters). The ANFIS learning algorithm is used to obtain these parameters. This learning algorithm is a hybrid algorithm consisting of the gradient descent and the least-squares estimate. Using this hybrid algorithm, the rule parameters are recursively updated until an acceptable error is reached. Iterations have two steps, one forward and one backward. In the forward pass, the antecedent parameters are fixed, and the consequent parameters are obtained using the linear least-squares estimate. In the backward pass, the consequent parameters are fixed, and the output error is back-propagated through this network, and the antecedent parameters are accordingly updated using the gradient descent method.. Takagi and Sugeno's fuzzy if-then rules are used in the model. The output of each rule is a linear combination of input variables and a constant term. The final output is the weighted average of each rule's output. The basic learning rule of the proposed network is based on the gradient descent and the chain rule [Werbos, 1974]. In the designing of ANFIS model, the number of membership functions, the number of fuzzy rules, and the number of training epochs are important factors to be considered. If they were not selected appropriately, the system will over-fit the data or will not be able to fit the data. Adjusting mechanism works using a hybrid algorithm combining the least squares method and the gradient descent method with a mean square error method. The aim of the training process is to minimize the training error between the ANFIS output and the actual objective. This allows a fuzzy system to train its features from the data it observes, and implements these features in the system rules. As a Type III Fuzzy Control System, ANFIS has the following layers as represented in Figure 1.

Layer 0: It consists of plain input variable set.

Layer 1: Every node in this layer is a square node with a node function as given in Eq. (1);

$$\mu_{A_i}(x) = \frac{1}{1 + \left[\left(\frac{x - c_i}{a_i} \right)^2 \right]^{b_i}}$$

(1)

where A is a generalized bell fuzzy set defined by the parameters {a,b,c} , where c is the middle point, b is the slope and a is the deviation.

Layer 2: The function is a T-norm operator that performs the firing strength of the rule, e.g., fuzzy AND and OR. The simplest implementation just calculates the product of all incoming signals.

Layer 3: Every node in this layer is fixed and determines a normalized firing strength. It calculates the ratio of the *j*th rule’s firing strength to the sum of all rules firing strength.

Layer 4: The nodes in this layer are adaptive and are connected with the input nodes (of layer 0) and the preceding node of layer 3. The result is the weighted output of the rule j.

Layer 5: This layer consists of one single node which computes the overall output as the summation of all incoming signals.

In this research, the ANFIS model was used for churn data identification. As mentioned before, according to the feature extraction process, 7 inputs are fed into ANFIS model and one variable output is obtained at the end. The last node (rightmost one) calculates the summation of all outputs [Riverol & Sanctis, 2005].

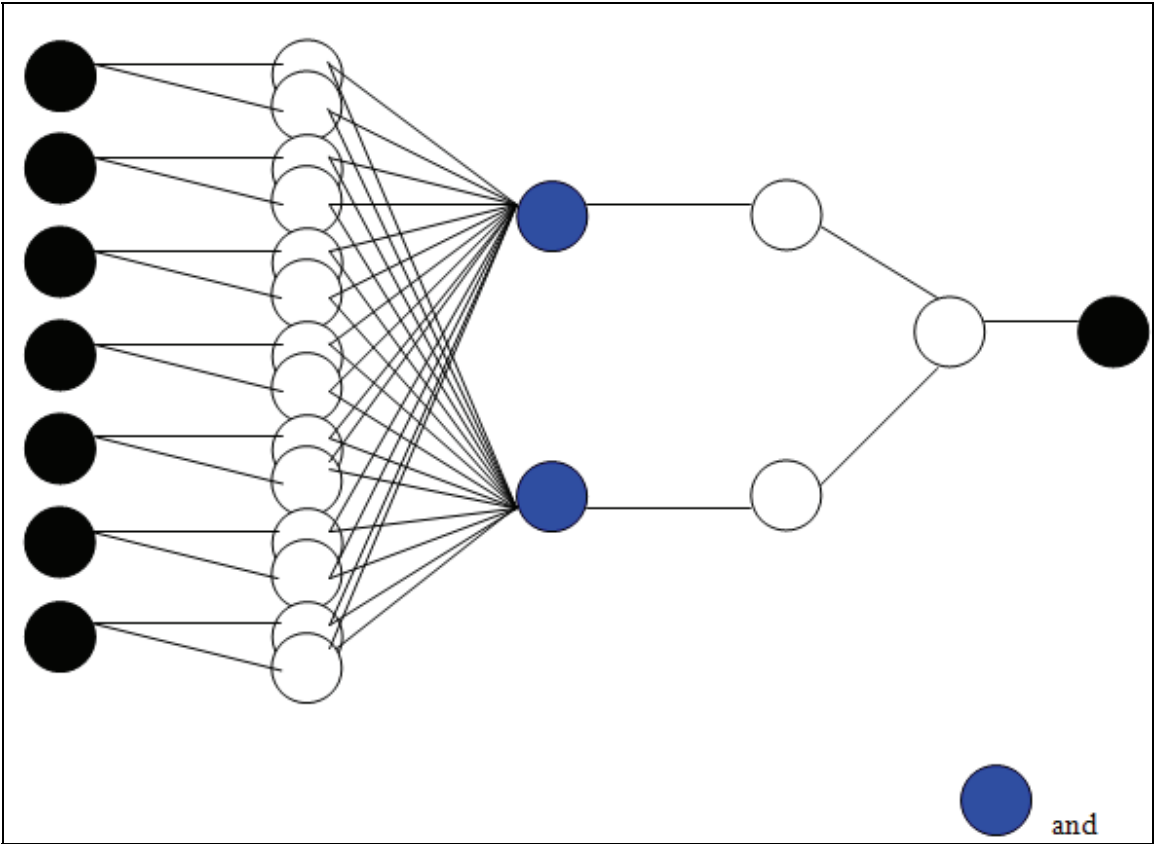


Fig. 1. ANFIS model of fuzzy inference

3. Findings

Benchmarking the performance of the data mining methods’ efficiency can be calculated by confusion matrix with the following terms [Han & Kamber, 2000]:

- 1. True positive (TP) corresponding to the number of positive examples correctly predicted by the classification model.
- 2. False negative (FN) corresponding to the number of positive examples wrongly predicted as negative by the classification model.
- 3. False positive (FP) corresponding to the number of negative examples wrongly predicted as positive by the classification model.
- 4. True negative (TN) corresponding to the number of negative examples correctly predicted by the classification model.

The true positive rate (TPR) or sensitivity is defined as the fraction of positive examples predicted correctly by the model as seen in Eq.(2).

$$TPR = TP / (TP + FN) \tag{2}$$

Similarly, the true negative rate (TNR) or specificity is defined as the fraction of negative examples predicted correctly by the model as seen in Eq.(3).

$$TNR = TN / (TN + FP) \tag{3}$$

Sensitivity is the probability that the test results indicate churn behaviour given that no churn behaviour is present. This is also known as the true positive rate.

Specificity is the probability that the test results do not indicate churn behaviour even though churn behaviour is present. This is also known as the true negative rate.

	Training data	
Method	Sensitivity	Specificity
Jrip	0.85	0.88
Ridor	0.90	0.91
PART	0.60	0.93
OneR	0.79	0.85
Nnge	0.68	0.89
Decision Table	0.85	0.84
Conj.R.	0.66	0.75
AD tree	0.75	0.85
IB1	0.77	0.89
BayesNet	0.98	0.95
ANFIS	0.86	0.85

Table 3. Training Results for the Methods Used

Correctness is the percentage of correctly classified instances. RMS denotes the root mean square error for the given dataset and method of classification. Precision is the reliability of the test (F-score).

RMS, prediction and correctness values indicates important variations. Roughly JRIP and Decision Table methods have the minimal errors and high precisions as shown in Table 3 for training. But in testing phase ANFIS has highest values as listed in Tables 4 and 5.

RMSE (root mean squared error) values of the methods vary between 0.38 and 0.72, where precision is between 0.64 and 0.81. RMS of errors is often a good indicator of reliability of methods. Decision table and rule based methods tend to have higher sensitivity and specificity. While a number of methods show perfect specificity, the ANFIS has the highest sensitivity.

Testing data		
Method	Sensitivity	Specificity
JRIP	0.49	0.70
Ridor	0.78	0.78
PART	0.58	0.64
OneR	0.65	0.51
Nnge	0.58	0.57
Decision Table	0.75	0.73
Conj.R.	0.65	0.71
AD tree	0.74	0.77
IB1	0.72	0.65
Bayes Net	0.75	0.76
ANFIS	0.85	0.88

Table 4. Testing Results for the Methods Used

Testing data		
Method	Precision	Correctness
JRIP	0.64	0.43
Ridor	0.72	0.66
PART	0.69	0.51
OneR	0.70	0.55
Nnge	0.66	0.48
Decision Table	0.72	0.71
Conj.R.	0.70	0.55
AD tree	0.71	0.60
IB1	0.74	0.61
Bayes Net	0.72	0.71
ANFIS	0.81	0.80

Table 5. Testing Results for the Methods Used

Three types of Fuzzy models are most common; the Mamdani fuzzy model, the Sugeno fuzzy model, and the Tsukamoto fuzzy model. We preferred to use Sugeno-type fuzzy model for computational efficiency.

Sub-clustering method is used in this model. Sub clustering is especially useful in real time applications with unexpectedly high performance computation. The range of influence is 0.5, squash factor is 1.25, accept ratio is 0.5; rejection ratio is 0.15 for this training model. Within this range, the system has shown a considerable performance.

As seen on Figure 2, test results indicate that, ANFIS is a pretty good means to determine churning users in a GSM network. Vertical axis denotes the test output, whereas horizontal axis shows the index of the testing data instances.

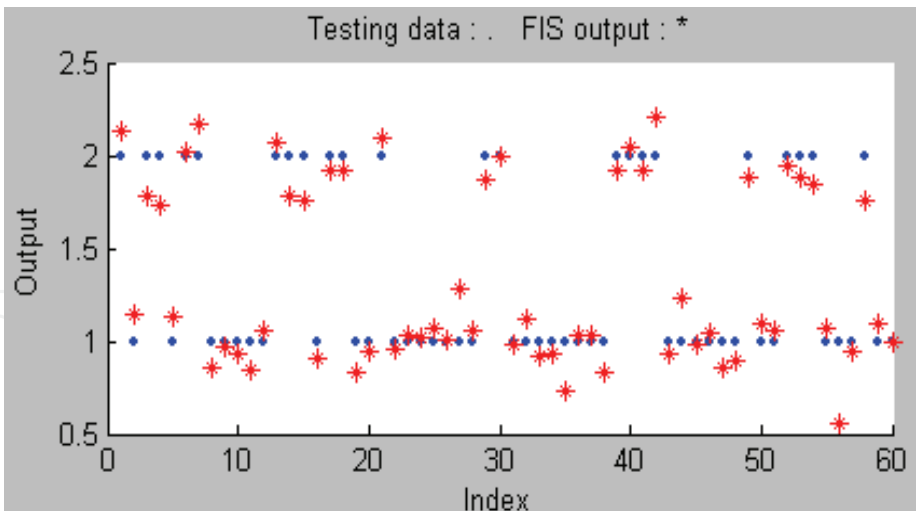


Fig. 2. ANFIS classification of testing data

Figure 2 and Figure 3, show plot of input factors for fuzzy inference and the output results in the conditions. The horizontal axis has extracted attributes from Table 2. The fuzzy inference diagram is the composite of all the factor diagrams. It simultaneously displays all parts of the fuzzy inference process. Information flows through the fuzzy inference diagram that is sequential.

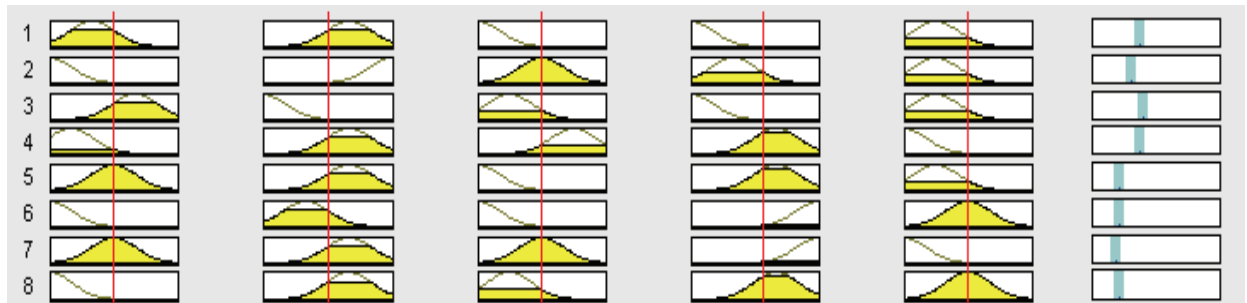


Fig. 3. Fuzzy Inference Diagram

ANFIS creates membership functions for each input variables. The graph shows Marital Status, Age, Monthly expense and Customer Segment variables membership functions. In these properties, changes of the ultimate (after training) generalized membership functions with respect to the initial (before training) generalized membership functions of the input parameters were examined. Whenever an input factor has an effect over average, it shows considerable deviation from the original curve. We can infer from the membership functions that, these properties has considerable effect on the final decision of churn analysis since they have significant change in their shapes. In Figures 4 to 7, vertical axis is the value of the membership function; horizontal axis denotes the value of input factor. Marital status is an important indicator for churn management; it shows considerable deviation from the original Gaussian curve as seen in Figure 4, during the iterative process. Figure 5 shows the initial and final membership functions. As expected, age group found to be an important indicator to identify churn. In network, monthly expense is another factor affecting the final model most. Resultant membership function is shown in Figure 6.

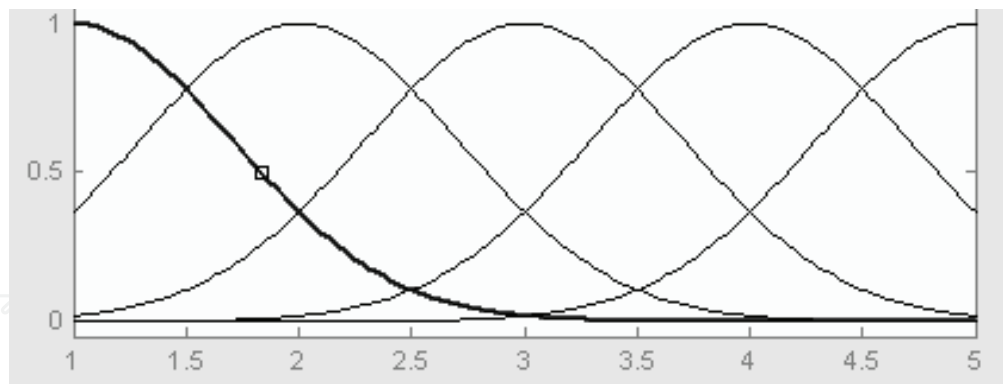


Fig. 4. Membership function for Marital Status

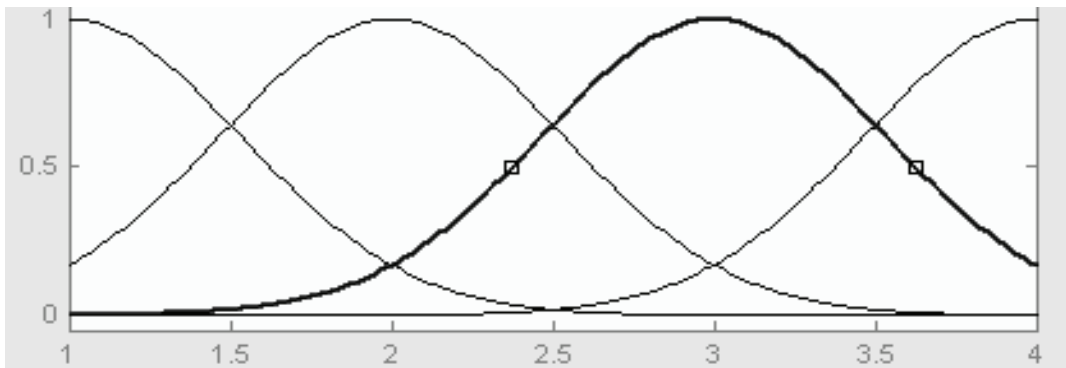


Fig. 5. Membership function for Age Group

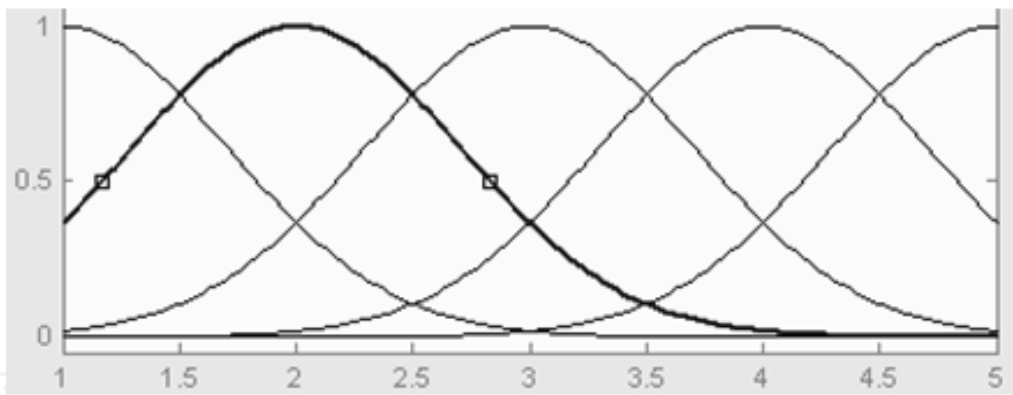


Fig. 6. Membership function for Monthly Expense

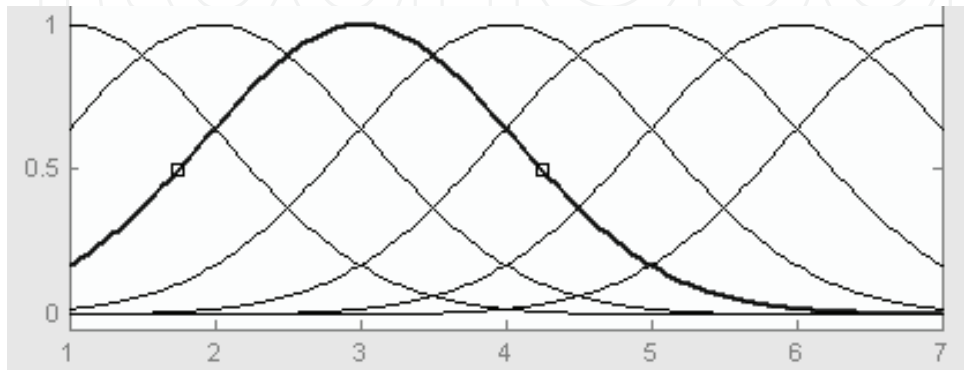


Fig. 7. Membership function for Customer Segment

Subscriber’s customer segment also critically affects the model. As seen on Figure 7, deviation from original curve is significant. These attributes represented in Figures 4 to 7 has the highest effect on final classification, the process has changed the membership functions significantly giving the values more emphasis for the final decision.

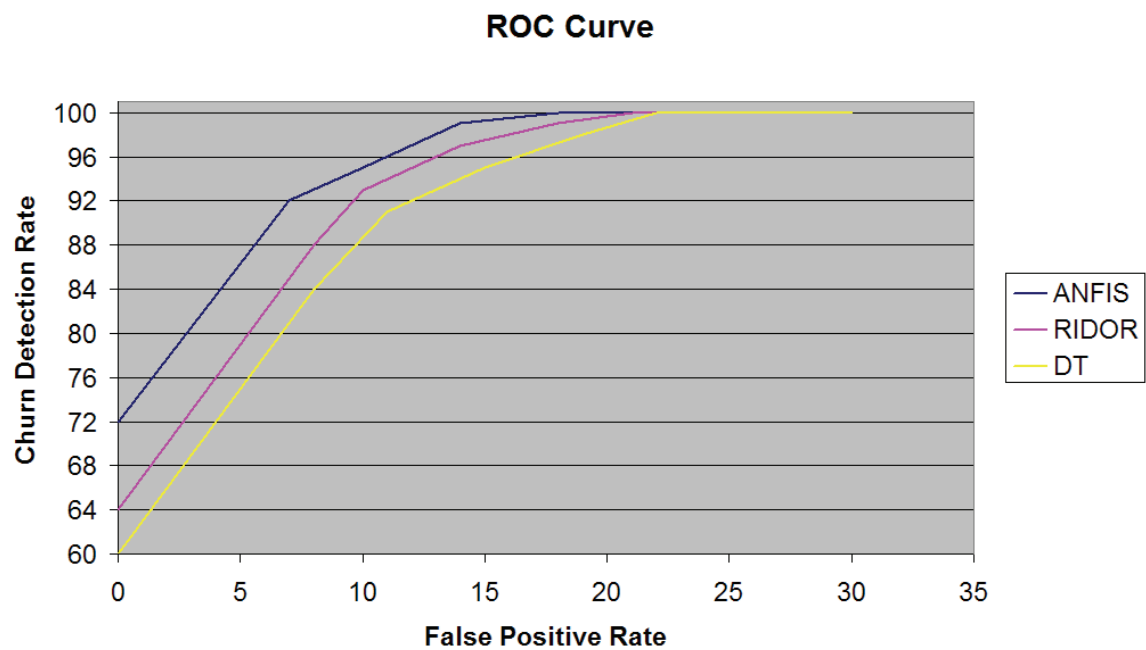


Fig. 8. Reciever Operating Characteristics Curve for Anfis, Ridor and decision trees

Figure 8 illustrates the ROC curve for the best three methods, namely ANFIS, RIDOR and Decision Trees. The ANFIS method is far more accurate where the smaller false positive rate is critical. In this situation where preventing churn is costly, we would like to have a low false positive ratio to avoid unnecessary customer relationship management (CRM) costs.

4. Conclusions

The proposed integrated diagnostic system for the churn management application presented is based on a multiple adaptive neuro-fuzzy inference system. Use of a series of ANFIS units greatly reduces the scale and complexity of the system and speeds up the training of the network. The system is applicable to a range of telecom applications where continuous monitoring and management is required. Unlike other techniques discussed in this study, the addition of extra units (or rules) will neither affect the rest of the network nor increase the complexity of the network.

As mentioned in Section 2, rule based models and decision tree derivatives have high level of precision, however they demonstrate poor robustness when the dataset is changed. In order to provide adaptability of the classification technique, neural network based alteration of fuzzy inference system parameters is necessary. The results prove that, ANFIS method combines both precision of fuzzy based classification system and adaptability (back propagation) feature of neural networks in classification of data.

One disadvantage of the ANFIS method is that the complexity of the algorithm is high when there are more than a number of inputs fed into the system. However, when the system

reaches an optimal configuration of membership functions, it can be used efficiently against large datasets.

Based on the accuracy of the results of the study, it can be stated that the ANFIS models can be used as an alternative to current CRM churn management mechanism (detection techniques currently in use). This approach can be applied to many telecom networks or other industries, since it is once trained, it can then be used during operation to provide instant detection results to the task.

5. Acknowledgement

Authors thanks to Mert Şanver for his helps and works.

6. References

- Cohen, W. (1995). Fast effective rule induction. *Proceedings of the 12th International Conference on Machine Learning*, pp.115-123 Lake Tahoe, CA, 1995.
- Frank, E. & Witten, I. H. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, 0-12-088407-0, San Francisco.
- Freund, Y. & Mason, L. (1999). The alternating decision tree learning algorithm. In *Proceedings of the Sixteenth International Conference on Machine Learning*. Pp. 124-133, Bled, Slovenia, 1999.
- Gaines, B.R. & Compton, P. (1995). Induction of ripple-down rules applied to modelling large databases. *Journal of Intelligent Information Systems* Vol.5, No.3, pp.211-228.
- Gerpott, T.J.; Rams, W. & Schindler, A. (2001). Customer retention loyalty and satisfaction in the German mobile cellular telecommunications market, *Telecommunications Policy*, Vol.25, No .10-11, pp.885-906.
- Hadden, J.; Tiwari, A.; Roy, R. & Ruta, D. (2005). Computer assisted customer churn management: state of the art and future trends. *Journal of Computers and Operations Research*, Vol.34,No .10, pp.2902-2917.
- Han, J. & Kamber, M. (2006). *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 978-1-55860-901-3, San Francisco.
- Hung, S-Y.; Yen, D. C. & Wang, H. Y. (2006). Applying data mining to telecom churn management. *Journal of Expert Systems with Applications*, Vol.31,No.3, pp.515-524.
- Holte, R.C. (1993). Very simple classification rules perform well on most commonly used datasets, *Machine Learning*Vol.11, pp.63-91.
- Jang, J-SR. (1992). Self-learning fuzzy controllers based on temporal back propagation, *IEEE Trans Neural Networks*, Vol.3, No.5, pp.714-723.
- Jang, J-SR. (1993). ANFIS: adaptive-network-based fuzzy inference system, *IEEE Trans. Syst. Man Cybernet*, Vol.23, No.3, pp.665-685.
- Karahoca, A. (2004). Data Mining via Cellular Neural Networks in the GSM sector, *Proceedings of 8th IASTED International Conference: Software Engineering Applications*, pp.19-24, Cambridge, MA. 2004.
- Kohavi, R. (1995). The power of decision tables, In *Proceedings of European Conference on Machine Learning*, pp.174-189 1995.
- Martin, B. (1995). Instance-based learning: Nearest Neighbor with generalization. Unpublished Master Thesis, University of Waikato, Hamilton, New Zealand.

- Mozier, M.C.; Wolniewicz, R.; Grimes, D.B.; Johnson, E. & Kaushansky, H.(2000). Predicting Subscriber Dissatisfaction and Improving Retention in the Wireless Telecommunications Industry, *IEEE Transactions on Neural Networks*, Vol.11, No.3, pp.690-696.
- Pearl, J. (1985). Bayesian Networks: A Model of Self-Activated Memory for Evidential Reasoning. In *Proceedings of the 7th Conference of the Cognitive Science Society*, pp.329-334, University of California, Irvine, CA, 1985, August 15-17.
- Riverol, C. & Sanctis, C. D. (2005). Improving Adaptive-network-based Fuzzy Inference Systems (ANFIS): A Practical Approach, *Asian Journal of Information Technology*, Vol.4, No.12, pp.1208-1212.
- Wei, C. P. & Chiu I-T. (2002). Turning telecommunications call details to churn prediction: a data mining approach, *Journal of Expert Systems with Applications*, Vol.23, pp.103-112.
- Werbos, P. (1974). Beyond regression, new tools for prediction and analysis in the behavioural sciences. Unpublished PhD Thesis, Harvard University.
- Yan, L.; Fassion, M. & Baldasare, P.(2005). Predicting customer behavior via calling links, *Proceedings of Int. Joint Conference on Neural Networks*, pp.2555-2560, Montreal, Canada, 2005.
- Yu, W.; Jutla, D. N. & Sivakumar, S. C. (2005). A churn management alignment model for managers in mobile telecom, *Proceedings of the 3rd annual communication networks and services research conferences*, pp.48-53.

IntechOpen



Data Mining and Knowledge Discovery in Real Life Applications

Edited by Julio Ponce and Adem Karahoca

ISBN 978-3-902613-53-0

Hard cover, 436 pages

Publisher I-Tech Education and Publishing

Published online 01, January, 2009

Published in print edition January, 2009

This book presents four different ways of theoretical and practical advances and applications of data mining in different promising areas like Industrialist, Biological, and Social. Twenty six chapters cover different special topics with proposed novel ideas. Each chapter gives an overview of the subjects and some of the chapters have cases with offered data mining solutions. We hope that this book will be a useful aid in showing a right way for the students, researchers and practitioners in their studies.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Adem Karahoca, Dilek Karahoca and Nizamettin Aydın (2009). Benchmarking the Data Mining Algorithms with Adaptive Neuro-Fuzzy Inference System in GSM Churn Management, Data Mining and Knowledge Discovery in Real Life Applications, Julio Ponce and Adem Karahoca (Ed.), ISBN: 978-3-902613-53-0, InTech, Available from:

http://www.intechopen.com/books/data_mining_and_knowledge_discovery_in_real_life_applications/benchmarking_the_data_mining_algorithms_with_adaptive_neuro-fuzzy_inference_system_in_gsm_churn_management

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2009 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](https://creativecommons.org/licenses/by-nc-sa/3.0/), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen