Predicting Behaviors of Insurance Costumers by Using the Genetic Algorithm

Fatemeh Bagheri[1,*], Majid Ziaratban[1,+], Mohammad Jafar Tarokh[2]

[1] Engineering faculty, Golestan University, Gorgan, Iran
*f.bagheri@gu.ac.ir
+m.ziaratban@gu.ac.ir
[2]Industrial Engineering faculty, K. N. Toosi University of Technology, Tehran, Iran
mjtarokh@kntu.ac.ir

Abstract

Improving customers' satisfaction, retaining loyal customers, and acquiring new customers are the most important goals of companies. Because of existing competition conditions among insurance industries, especially in optional insurances such as vehicle body insurance, implementing policies which have benefits for both customers and insurance companies can increase the corporation profitability and customers' satisfaction. In this paper, general policies of vehicle body insurance in Iran are investigated. In this way, a non-linear function (C_j) is proposed to separate profitable and harmful customers. This function is based on some insurance features of customers such as age of a vehicle, vehicle price, vehicle type, and number of years without accident. Customers are categorized based on the calculated values of C_j in which all parameters are optimally set. The parameters have to be set in the training phase to calculate reliable C_j values for customers. Parameters of C_j are set so that in a subset of the customers of the training set for which the value of C_j is greater than a threshold (Th), the number of profitable customers and disadvantageous customers should be maximized and minimized, respectively. A customer with positive M (monetory) value is a profitable customer. To optimize this multivariable problem with large number of parameters, the genetic algorithm is used. After optimizing the parameters in the training phase by using GA, the behaviors of the customers in the test set are predicted. In the prediction phase, it was estimated that a costumer is profitable or not, even he had accidents and received recompenses from the company. If a customer is predicted as profitable, the company allows him to enjoy some discounts. In the current Iranian insurance policies, all discounts are canceled for next insurance year if a customer has accidents and gives recompense from the company. By implementing the proposed policy, the customers' satisfactions are increased and also more new customers are acquired from other insurance companies.

Keywords: Behavior prediction, insurance customers, genetic algorithm.

1. Introduction

Generally, customers are divided into different groups based on their characteristics and behaviors. Building models and decision support tools to increase the investment on marketing activities is the interesting object of companies. The companies spending considerable amount of their resources to achieve this object [1]. Different groups of customers have different characteristics and behaviors. Some of them are likely and others are unlikely to be profitable. Companies yearn to retain valuable customers and spend resources for them. This purpose is achievable by using customer relationship management (CRM) decision support tools, which help companies to differentiate between different groups of customers [2].

These tools help companies in different stages such as acquiring and retaining customers. CRM is one of the best ways to analyze customers' information and has been implemented in most of international companies. These wide implementations demonstrate two realities: universality of the CRM, and its successful results.

Four main targets of CRM are collecting useful information about customers, delivering the collected data to the management departments of organizations, discovering customers' values, and implementing new policies to increase the customers' loyalty [3]. CRM helps the organizations in several scopes such as finding, acquiring and retaining the customers.

Managers are interested to discover hidden, reliable, and real knowledge from their big database and use this knowledge to increase their companies' profit. Companies use various tools to implement their CRM activities. Data mining is one of the prevalent and powerful CRM analyzing tools. The most important purposes of using the data mining methods are cost reduction, improving customer satisfaction and market share, and improving organizational effectiveness [4]. The data mining includes two groups of techniques: rule-based and basic techniques. Fuzzy logic, decision tree, genetic algorithm, inductive learning system, and statistical methods are some rule-based methods. Clustering, classification, regression analysis, association rules, and sequence analysis are the methods that are in the basic scope [5].

Data mining is one of possible ways to find hidden knowledge in large databases. Managers are enthusiastic to discover hidden knowledge to increase their profits. Furthermore, Companies have to analyze their customers' data to retain them .Decreasing costs, increasing customers' satisfaction, increasing market share of the company, and improving profit level are the most important goals in using data mining methods as a CRM analyzing tool.

There are many articles under the heading of segmentation modeling which have used the statistical methods and data mining approaches to classify different groups of customers [2]. Also, some researches were done in the field of the insurance strategies and its customers.

In a research, Blows et al. examined the association between car crash injury and uninsured motor vehicles. Their study provides new evidence, which is uninsured motorists compared to insured motorists, have higher risks, which being injured or killed in car crashes [6]. Šubelj et al. proposed a novel expert system approach to detect the groups of collaborating automobile insurance fraudsters with networks [7].

Liao et al. proposed an approach to segment market and demand chain analysis on life insurance market for new product development and marketing. They used the a priori algorithm and clustering analysis for data mining to extract knowledge [8].

2. Related works

At the beginning of the present century, amount of premiums that have been issued by the insurance companies is 2.5 trillion dollars. Various factors such as changing speed in job environment and market status risks effect on durability and growing of insurance industry. One of the most important factors is technology which includes intelligent computing and other computing techniques. Studies on combining these techniques with insurance-based programs are performing around the world.

The insurance industry is one of the most important industries in all countries. Several researchers used statistical methods in insurance industry related applications. Lately, due to rapid advancement of technology, most of researchers utilized artificial intelligent techniques (soft computing) to solve problems in insurance domain [9]. Shapiro has reported that three techniques, neural networks, fuzzy logic, and genetic algorithm have been used in most related researches in insurance [10].

Due to current unprecedented economic recession, industries managers are going to use computer modelling in markets such as in exchange markets. They used genetic algorithm in financial strategies and investment to limit global reserve assets. In recent years, companies have used the genetic algorithm in their financial predictions [11].

Genetic algorithm is an optimization method to find the best solution for multi-variable problems in very large search spaces. This algorithm is a special type of evolutionary algorithms and uses some biologic techniques such as inheritance and mutation. Genetics is a science about inheritance and biological transition from one generation to the next. Chromosomes and genes are main factors in biological transition in organisms. Superior and stronger genes and chromosomes remain and weaker genes are lost. In other words, the best and most reasonable creatures remain. This algorithm is based on the Darwinian evolution (Survival of the fittest) that says: "Weaker creatures are destroyed and stronger ones are remained". Indeed, the law of natural selection for survival says: more possibility of adaptability causes more survival and reproduction rate. This rule is based on the mutation and combining chromosome strings.

Genetic algorithms use the principles of Darwinian natural selection to find an optimal function for prediction or pattern matching. Genetic algorithms are often a good way for prediction applications based on random solutions. Genetic algorithm (GA) is a programming technique which uses genetic evolution as a problem solving model. The input of GA is a problem that must be solved and the solutions are encoded as a pattern which is called fitness function. Genetic algorithm evaluates each candidate solution. These algorithms is used in various fields such as optimization, identification and control systems, image processing and fusion problems, determining the topology, and training systems which are based on artificial neural networks and decision rules.

Cachón and Vázquez proposed a genetic algorithm (GA) to automatically adjust parameters, removing the need of manual tuning and increasing efficiency for solving pattern recognition problems [12].

Uysal and Gunal offered a method to obtain better representation of documents in text classification by proposing genetic algorithm oriented latent semantic features (GALSF) [13]. Lin et al. proposed a genetic algorithm feature selection (GAFS) for image retrieval systems and image classification [14].

Homayouni et al. used genetic algorithm for optimization of integrated scheduling of cranes, vehicles, and storage platforms at automated container terminals [15].

Tosun proposed a new recombination operator for the genetic algorithm solution of the quadratic assignment problem [16].

Li and Wen developed a novel method of virtual sample generation which considers the integrated effects and constraints of data attributes to improve the learning robustness [17].

Zarifi and Ghalehjogh proposed a new evolutionary approach for neural spike detection based on genetic algorithm [18]. Pourvaziri and Naderi offered an effective novel hybrid multi-population genetic algorithm [19]. Smail et al. compared the performances of Particle Swarm Optimization and Genetic Algorithm to develop a methodology for wiring network diagnosis allowing the detection, localization and characterization of faults [20]. Kalita and Datta solved the bi-objective corridor allocation problem using a permutation-based genetic algorithm [21].

Maghsoudi et al. proposed a new model for improving k-mean clustering algorithm based on genetic algorithm [22]. Nazif considered a single machine family scheduling problem where jobs were partitioned into families and setup was required between these families. The goal of his research was to find an optimal schedule that minimizes the total weighted completion time of the given jobs in the presence of the sequence independent family setup times by using genetic algorithm [23].

Vahidi and Mirpour proposed a new algorithm for data clustering by using genetic algorithm. Their proposed clustering method was based ona grouping genetic algorithm (GGA) [24].

Eslami et al. presented a solution for optimizing energy consumption in the sensors of wireless network by using a distributed genetic algorithm and solving the famous problem of graph coloration [25].

The genetic algorithm has been used in insurance industry related applications. Sanz et al. proposed an approach to predict insolvency of non-life insurance companies based on the application of genetic programming (GP). They believe despite the emergent statistical learning methods like neural networks or SVMs provide a successful approach in terms of error rate, but their black-box characteristic make the obtained results difficult to be interpreted and discussed [26].

In addition to genetic algorithm, some other algorithms are used in the applications related to the insurance industry. Guillén et al. investigate time-varying effects in the analysis of insurance customer loyalty. Their results showed that the kind of contracts held by customers and the concurrence of an external competitor have strongly influence on customers loyalty; but those factors become much less significant some months later [27].

Guelman et al. proposed a survey of personalized treatment models for pricing strategies in insurance. They show how models of personalized treatment learning can be used to select the policy holders that should be targeted in marketing strategies of a company [28].

Lin offered a tool based on the back propagation neural network (BPNN) model for the underwriter to determine the proper premium rate of in-between risks [29].

3. Motivation

In Iran, if a customer of an insurance company does not have any accident in his first insurance year, the company will discount to him for the following year. Discount increases every year until he does not have any accident. Maximum discount is 70% of the amount of the premium. Upon the occurrence of an accident and the customer demand for recompense from the insurance company, all of his discounts will become zero and the customer have to pay the whole amount without discount on his contract for the next year. Large insurance companies can improve their customers' satisfaction and loyalty by slight variations in their policies. Acquiring new customers is one of the most important goals of companies. Customers'

satisfaction improvement and retaining customers are two effective and powerful advertising to acquire new customers. This paper tries to modify some insurance policies and laws so that the customers who have accidents with low recompenses do not worry about losing all their discounts for future years. Because of existence of competition conditions among insurance industries, these modifications will cause acquiring more new customers.

Due to current insurance rules, knowing of customers is not accurate. In other words, real amount of profit or detriment for customers is not properly specified. The customers which have accident and receive recompense from a company are divided into two categories. One group includes customers that have received recompenses for their accidents lower than their premium paid to the company. These customers are profitable for the company. Customers in another group have received recompenses for their accidents more than their premium. Therefore, they are not profitable customers.

In current insurance policies, a company has similar behavior with customers who had accidents (both profitable and harmful) and all their discounts will be canceled for next year.

The goal of the paper is to predict insurance behaviors in next year for customers who had have accidents and received recompenses from the insurance company. In other words, we want to predict that the amount of recompenses which will pay to a customer is lower than the premium paid by him, or not. After prediction the customer's behavior (profitable or harmful), the company can consider new policies for discounts allocation. If the amount of recompenses which will receive from insurance company is predicted to be lower than the amount of premium, the customer is predicted as a profitable customer and company can consider some discount for him. While due to the current insurance policies, this profitable customer loses all discounts, because he had an accident. Implementation of the policies with win-win interactions between customers and the insurance company improves customers' satisfaction and keeps more customers in the loyal customers group. Obviously, increasing the number of these profitable customers even by considering discounts for them will increase benefits of the company.

Furthermore, the prediction can be implemented for customers from other companies who had accidents in last year and received recompenses from their corresponding insurance companies and they want to contract with our company for the first time. The prediction is implemented based on the insurance history of a customer in his earlier insurance company. The prediction for a new customer means that if a new customer has accidents in next year, he will be profitable customer for our company at the end of the next year or not. In other words, if a new customer is predicted as a profitable customer in the next year, he can enjoy our company discounts at the beginning of the insurance contract. Otherwise, if he extends the contract in his previous insurance company, all discounts are canceled if he had accidents in last year.

4. Methodology

4-1. Data preprocessing and definitions

For evaluating the proposed method, an actual database of an Iranian insurance company was used. Initial data includes information about customers (such as age, gender, etc.), information of the insured vehicle (such as age of vehicle, number of cylinders, the city that the vehicle is driven in, etc.), the number of accidents which occurred in the insurance time, and other information.

The company that we used its database uses its own data just for reporting and archiving. Therefore, this database had poor organization for data mining use. Thus, the preprocessing phase was very time

consuming. The gotten experience is very helpful to improving the company's database structure for extra using.

We requested the insurance company to give us three past years' data. The number of records in the dataset is 34068 records for 26738 customers. All mistakes in the dataset have been corrected. Most frequent mistakes had occurred in filling the fields, such as filling some fields in both English and Farsi, and assigning two different names to a specific model of vehicle.

The age of vehicle (S), type of vehicle (N), number of earlier consequently years that the customer did not have any accident (T), and price of vehicle (G) are four main factors which are used for prediction.

Age of vehicles indicates the past years after the vehicle production. In this paper, the vehicles over 20 years old are inserted into the group of the ones with 20 years old. Therefore, 20 age groups are considered. Totally, 41 various vehicle types are in our dataset. The vehicles with high similarities are inserted in the same vehicle type group. Types of vehicles and the frequency of each one in our dataset are given in Table 1.

The number of years without accident is equal to the number of the earlier consequent years that a customer did not request recompenses from the company. The maximum value of this feature is 5 in the dataset. Hence, 6 groups are considered for the feature (number of years without accident). Since the price of vehicles is a continuous variable and has a very wide range, 10 uncorrelated intervals are considered for price of vehicles as shown in Table 2.

4-2. Proposed model

To determine the value of each customer a fitness function is defined as follows:

$$C_j = G_{i_G(j)}{}^{p_G} \times T_{i_T(j)}{}^{p_T} \times S_{i_s(j)}{}^{p_S} \times N_{i_N(j)}{}^{p_N} \tag{1}$$

C_j is a combination of some customer attributes. $i_S(j)$, $i_N(j)$, $i_T(j)$, and $i_G(j)$ determine the indices of the four customers' features: the age of vehicle, type of vehicle, the number of years without accident, and the group of vehicle price, respectively for j-th customer. S, N, T, and G are sets of coefficient corresponding to these four attributes. S_{i_S} is the i_S-th coefficient of the S. To improve the flexibility of calculation of C_j, four exponents p_S, p_N, p_T, and p_G are used in Eq.1. To determine which customer is profitable for the company, an attribute called Monetory (*M*) is used:

M = *Premium paid to the company - Total recompenses received from the company* $\tag{2}$

Table 1: Various groups of vehicle types

Vehicle type name	Type index	Number of vehicles	Vehicle type name	Type index	Number of vehicles
Autobus	1	11	Motorcycle	22	20
BMW	2	60	MVM 110	23	70
Benz	3	50	Nissan	24	47
C5	4	5	Opel	25	43
Delica	5	107	Patrol	26	32
Daewoo	6	313	Peugeot 206	27	3709
Daewoo Matiz	7	109	Peugeot 405	28	3360
Gol	8	114	Peugeot Pars	29	1386
Golf	9	4	Peugeot RD	30	1570
Hyundai	10	471	Paykan	31	873
Kamion	11	270	Pride	32	7624
Keshandeh	12	135	Proton	33	48
Kamiontereylikeshande	13	268	Renault	34	27
Kamionet	14	276	Rio	35	521
Kia	15	23	Samand	36	1766
L90	16	388	Sepand PK	37	254
Maxima	17	114	Suzuki	38	53
Mazda	18	491	Toyota	39	402
Minibus	19	10	Vanet	40	904
Mitsubishi	20	95	Xantia	41	617
Motefaregheh	21	98			

Table 2: Intervals of vehicle prices

Interval index	Lower boundary (Rials)	Upper boundary (Rials)	Number of vehicles
1	0	54,000,000	2717
2	54,000,000	63,000,000	2678
3	63,000,000	70,000,000	2925
4	70,000,000	80,000,000	3003
5	80,000,000	90,000,000	2831
6	90,000,000	100,000,000	2091
7	100,000,000	120,000,000	3495
8	120,000,000	135,000,000	1759
9	135,000,000	190,000,000	2684
10	190,000,000	-	2546

To have reliable prediction, a threshold (Th) is considered for C_j. If the value of C_j is higher than Th, the model can reliably predict that the customer is profitable or not. In the group of customers for whom the C_j values are greater than Th, the profitable or harmful ones are identified based on their monetary values. If M is higher than zero, it means the customer is profitable and hence, the number of profitable customers (n_{pos}) increases. Otherwise the customer is harmful and therefore, the number of harmful customers (n_{neg}) increases.

$$If \ C_j > Th \Rightarrow \begin{cases} if \ \ M > 0 \Rightarrow n_{pos} = n_{pos} + 1 \\ if \ \ M \leq 0 \Rightarrow n_{neg} = n_{neg} + 1 \end{cases}$$
(3)

Actually, we want to predict that which customer is profitable for next year. In other words, we try to prepare a final list including only the profitable customers. In this way, we use the information of customers in all years except of the final year. Indeed, we try to predict their insurance behavior in the final year. Hence, their real behaviors in the final year are used to evaluate our predictions. As mentioned earlier, to have more reliable predictions, the value of C_j must be greater than Th. Therefore, in the optimization phase, the values of all parameters in Eq.1 and also the value of Th have to be set so that a customer, for whom the C_j is higher than Th, will be profitable. Hence, the fitness function for the genetic algorithm is defined in Eq.4. The goal of the GA is to minimize F_{obj}.

$$F_{obj} = \frac{n_{neg}}{n_{pos} + n_{neg}} * 100$$
(4)

where n_{pos} and n_{neg} are respectively the numbers of profitable and harmful customers for whom the value of C_j is greater than Th. $n_{pos} + n_{neg}$ determines the number of customers in the final list (customers with Cj higher than Th). Minimizing the fitness function means the GA tries to adjust the parameters to set the number of harmful customers in the final list as low as possible. Experiments show that the genetic algorithm acts very strictly by using this fitness function so that the number of harmful customers will be zero in the final list. To achieve this result, the GA has to set very big value for the threshold Th than the computed C_j values. Therefore, the number of profitable customers will be very small. In other words, the final list includes very small number of customers (less than 1% of all customers). On the other hand, our purpose is to find a significant number of customers to give them some additional discounts to improve more customers' satisfaction. Hence, the GA optimization in the training phase is performed by using the fitness function in Eq.4 by considering a lower boundary for the number of customers in the final list. In other words, in the training phase, the GA optimizes the parameters so that the F_{obj} is minimized and also at least N_{min} customers of the training set have C_j values greater than Th. By using the optimized parameters in the prediction phase, we have a significant number of test customers in the final list.

For very large values of N_{min}, more costumers will be predicted as profitable and hence the final list includes very large number of customers. But in these cases, since the threshold value Th has been set with small values than the computed C_j values in the optimization phase, more harmful customers will pass the threshold and placed in the final list. Therefore, the prediction accuracy will be dramatically reduced for very large values of N_{min}.

As noted, the goal of the GA is to adjust several parameters to minimize the number of harmful test customer and maximize the profitable ones in the final list. The problem parameters which have to be set

by GA in the training phase are: S (set of 20 coefficients), N (set of 41 coefficients), T (set of 6 coefficients), G (set of 10 coefficients), p_S, p_N, p_T, p_G, and Th.

Optimizing these large numbers of parameters (totally 82 parameters) and determining their best values is very complex and time consuming in full search approaches. Therefore, the genetic algorithm is used for the optimization.

5. Experiments structures and results

To calculate optimized values for these 82 parameters by using GA, population is considered to include 50 chromosomes in each generation. Maximum number of generations is limited to 1000. Two elite chromosomes completely remain for next generation. 60% out of 48 remaining chromosomes for next generation reproduced by crossover and the remaining is produced by uniform mutation with 40% rate. The mutation rate is considered to be a large value to avoid GA sticking in local minimums and to have more complete search in parameters space.

The optimization was performed based on information of 20000 customers in the training set. Total number of customers is 26738. Hence, test set includes 6738 customers. In both training and test sets, about 25% of customers are harmful and the rests are profitable in their final insurance year.

5-1. Experimental results

Prediction accuracies in terms of N_{min} for training and test sets are shown in Fig. 1. Fig. 2 shows the ratios of customers predicted as profitable in terms of N_{min}. Furthermore, for the test set, the prediction accuracy versus the numbers of customers in the final list is shown in Fig. 3.

Fig. 2 show that in both training and test sets, the ratio of customers that are selected in the final list is about equal to the corresponding value of N_{min}.

From Fig.1, by increasing N_{min}, the prediction accuracy reduces in training and test sets. The reason is that for a large N_{min} value, more harmful customers having similar features to the ones of profitable customers are allowed to be in the final list. If the prediction accuracy is more important than the number of customers in the final list for the company, the optimized parameters based on lower values of N_{min} can be selected. Vice versa, if the company wants to increase more customers' satisfaction, it means the number of customers in the final list is more important than the prediction accuracy. Hence, the optimized parameters based on higher values of N_{min} can be selected. In this paper, we assume that the company wants to give discounts to about 10% of the customers which had accidents in past years. Hence, the optimized parameters based on the N_{min} equal to 10% are selected.

By using the selected optimized parameters, 2010 out of 20000 training customers have Cj values higher than Th. From Fig.3, 745 out of 6738 test customers are selected to enjoy discounts. Variations of the corresponding fitness function by increasing the generation number is shown in Fig. 4.
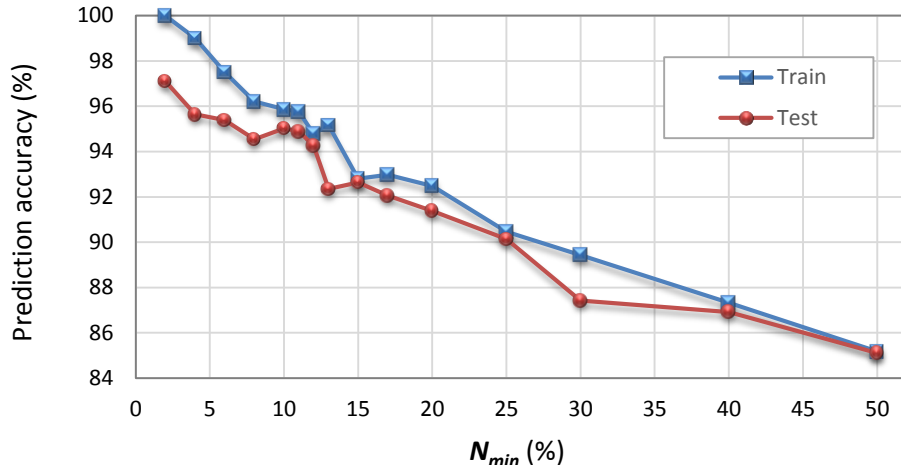
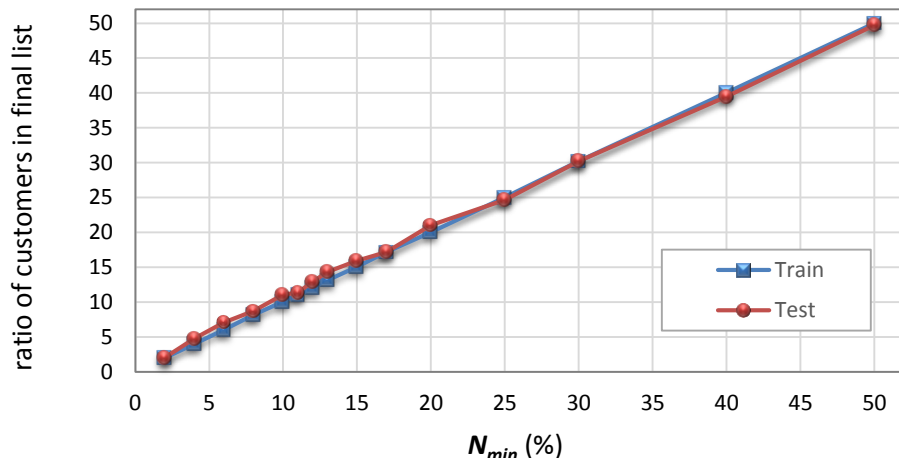Fig. 1: Prediction accuracies in terms of N_{min}

Fig. 2: Ratios of customers predicted as profitable in terms of N_{min}
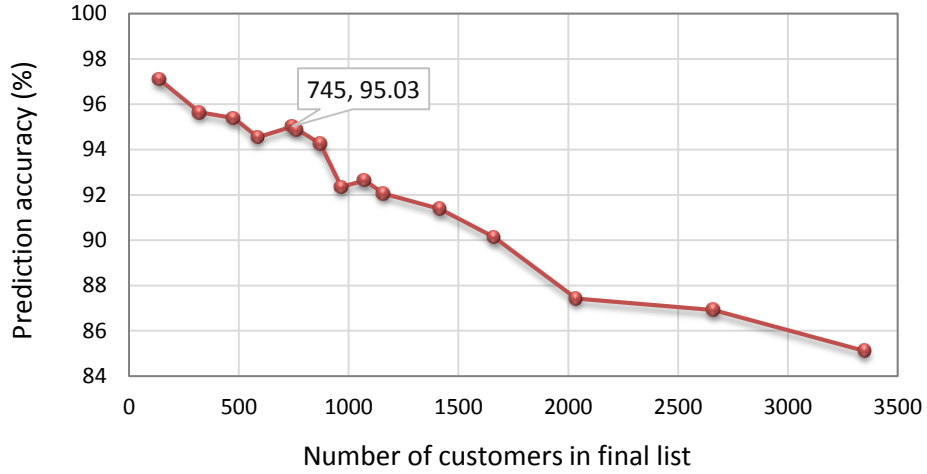
Fig. 3: Prediction accuracy versus the numbers of customers in the final list for the test set
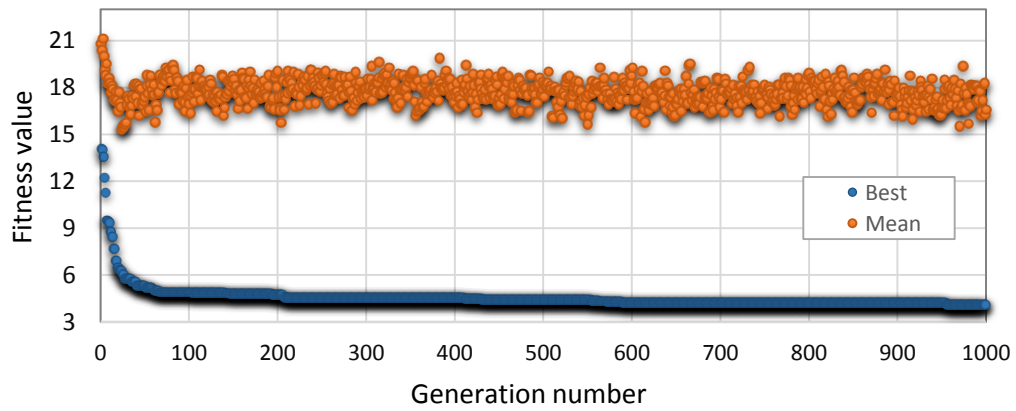

Fig. 4: Variations of the best and mean values of the fitness function by increasing the generation number

The value of *Th* was computed equal to 0.5101. Optimized values for p_G, p_T, p_S, and p_N were obtained equal to 0.6893, -1.3777, 0.4972, and 0.3318, respectively. Figures 5, 6, 7, and 8 show optimized values calculated by GA for parameters corresponding to *G*, *T*, *S*, and *N*, respectively.

Fig. 5: Optimized values of G (vehicle price) coefficients

Fig. 6: Optimized values of T (number of years without accident) coefficients
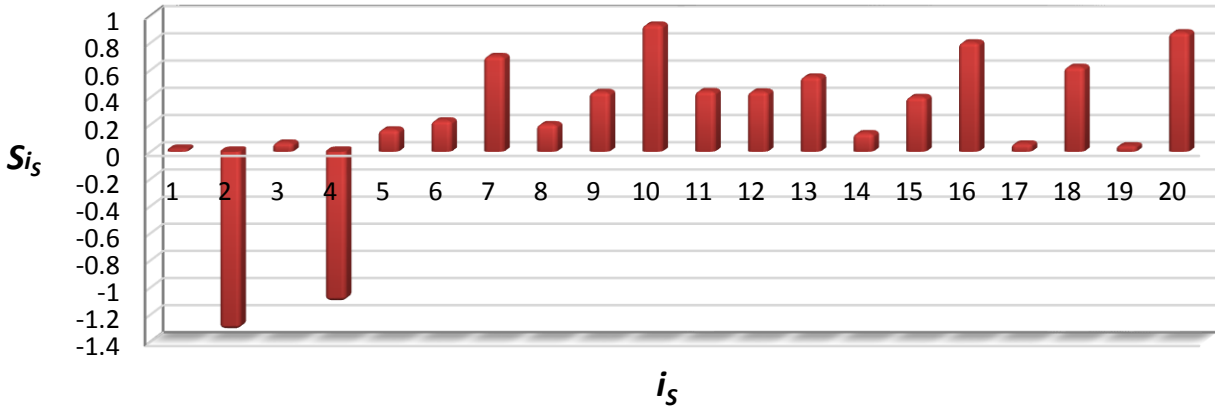

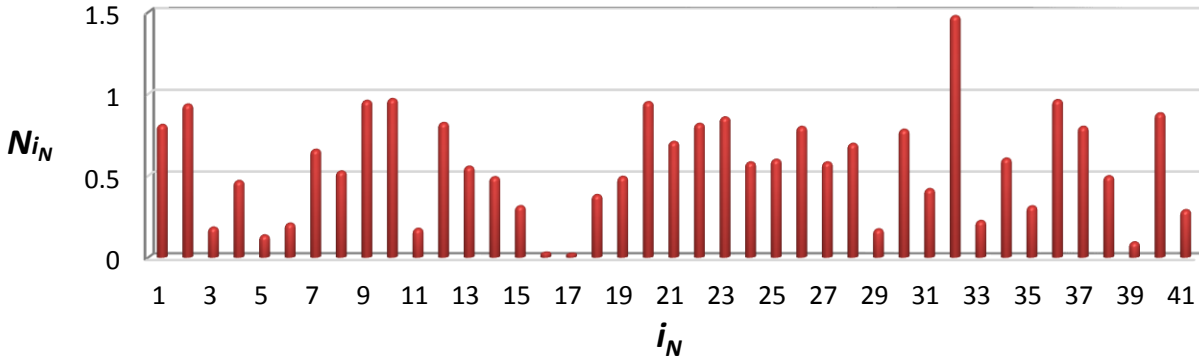Fig. 7: Optimized values of S (age of vehicle) coefficients

Fig. 8: Optimized values of N (type of vehicle) coefficients

Fig. 9, 10, 11, and 12 show the graphs of $G_{i_G}{}^{p_G}$, $T_{i_T}{}^{p_T}$, $S_{i_s}{}^{p_S}$, and $N_{i_N}{}^{p_N}$, respectively. From these figures, the importance of each feature of a customer can be deduced.

5-2. Result analysis

In Fig. 10, the value of $T_{i_T}{}^{p_T}$ increases by increasing the value of i_T. It means that the customers with higher number of years without accident are more profitable. The value corresponding to $i_T = 5$ is very great. It demonstrates that a customer, who did not have any accident in 5 early years, is certainly a profitable customer in next year; because, even if the other three features of the customer is the minimum values ($i_G = 7$, $i_S = 1$, and $i_N = 17$), we have:

$C_j = G_7{}^{p_G} \times T_5{}^{p_T} \times S_1{}^{p_S} \times N_{17}{}^{p_N} = 0.2 \times 151.24 \times 0.15 \times 0.29 = 1.316$

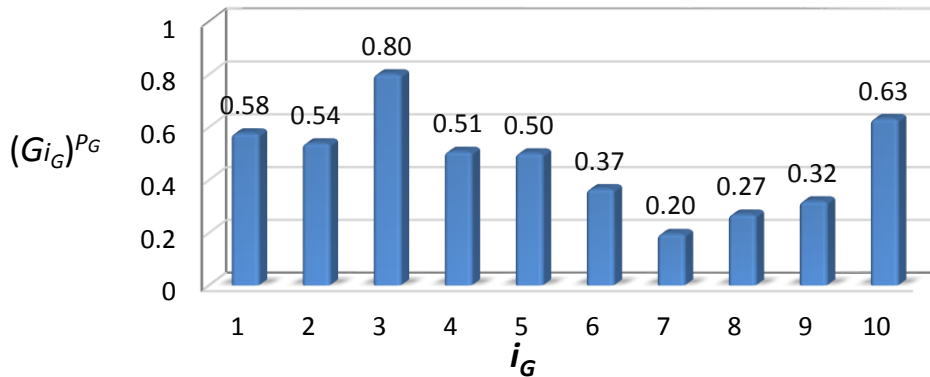Since the calculated C_j is greater than $Th = 0.5101$, the customer is considered as profitable.

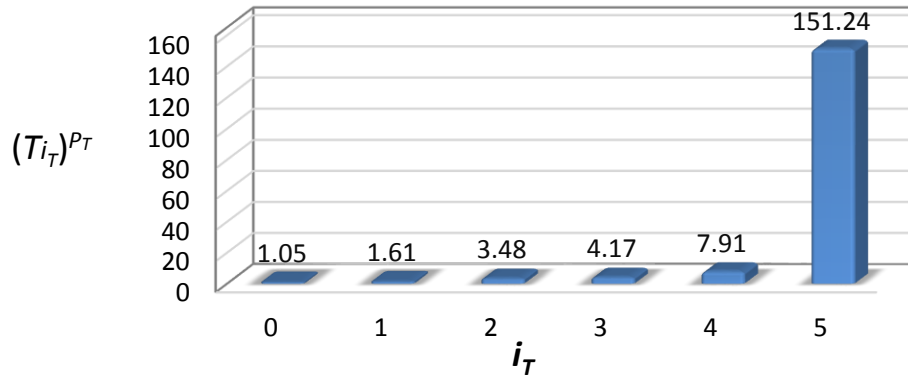Fig. 9: Optimized values of $G_{i_G}{}^{p_G}$
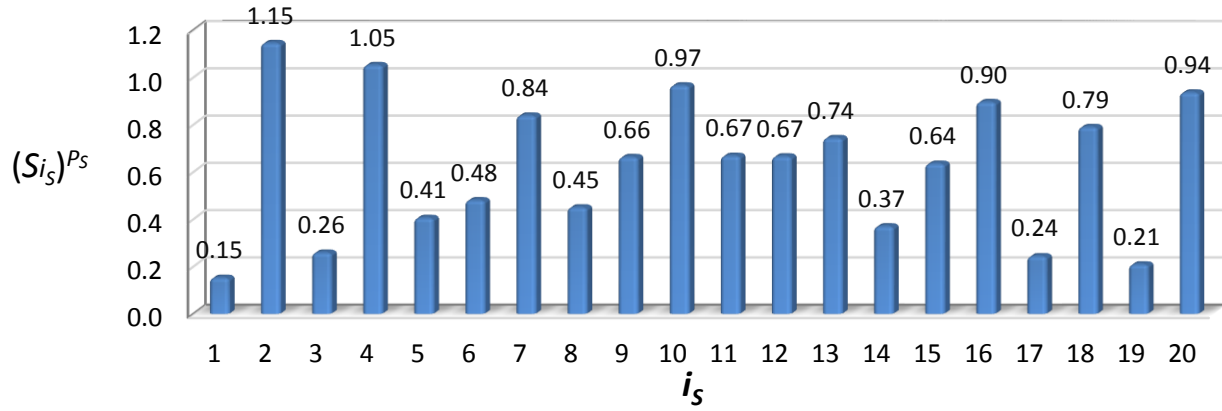
Fig. 10: Optimized values of $T_{i_T}{}^{p_T}$

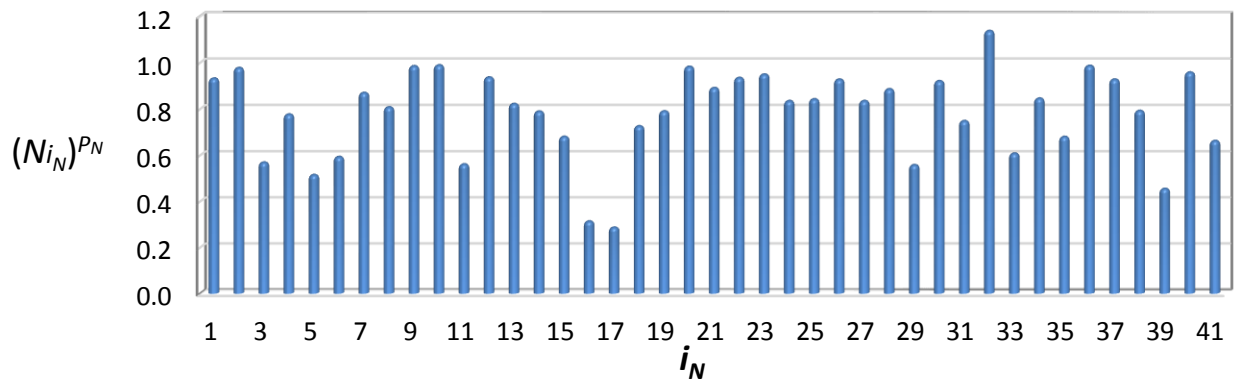
Fig. 11: Optimized values of $S_{i_S}{}^{p_S}$

Fig. 12: Optimized values of $N_{i_N}{}^{p_N}$

To evaluate the prediction model, optimized parameters are used to predict behaviors of customers in the test set. If the value of C_j for a customer is higher than $Th = 0.5101$, the customer is considered as a profitable customer. Because in the training phase, the optimization was performed so that in the set of customers, for whom the value of C_j is higher than Th, the numbers of profitable and harmful customers were maximized and minimized, respectively.

Monetory determines that a customer is profitable or harmful. Therefore, the accuracy of prediction can be evaluated based on M values. In other words, if C_j for a customer in the test set is higher than Th and his M value is positive, the prediction is correct. Obviously, if the M value is negative, the customer is harmful and the prediction is incorrect.

By using the proposed model, the prediction was performed for 6738 customers in the test set. The C_j values for 745 test customers were obtained greater than Th and hence they placed in the final list. In other words, 11.06% of customers in the test set predicted as profitable customers. Although, they have accidents, they enjoy discounts. It does not mean that the remaining 88.94% of customers are harmful. It means that the model is not able to reliably predict these customers' behavior and hence, decisions are not made for them. The prediction accuracy was obtained equal to 95.03% for the customers in the final list. In other words, only 4.97% of customers in the final list are incorrectly considered as profitable customers.

6. Conclusion

In this paper, a model is proposed by which profitable customers can be predicted. In this model, the customers whom predicted as profitable enjoy some discounts for their new insurance contract, although they had accidents. The reason is that our model predicts these customers are profitable even if they have accidents in next year. In addition, for new customers from other companies, the prediction can be performed based on their insurance histories. The new customer can enjoy discounts if he is predicted as profitable customer even having any accidents in past years. Otherwise, if he extends his contract in his old company, he is divested from all discounts of his company.

According to experimental results, 745 customers are predicted as profitable. 95.03% out of all predictions in the final list was correct. In other words, only 4.97% equal to 37 out of 745 customers were incorrectly predicted as profitable customers and the remaining 708 customers were correctly predicted. On the other hand, 5138 out of 6738 customers in test set are really profitable. Hence, 4430 profitable customers were not detected in the proposed model; because the reliability for these customers was not enough. In other words, there are several harmful customers with features similar to these 4430 profitable customers. Therefore, if the model takes easier to detect more profitable customers, more harmful customers are incorrectly placed in the final list and therefore, the prediction accuracy reduces dramatically. However, it does not mean that some detriments are inflicted to these 4430 customers. They only have not enjoyed additional discounts.

References

[1] Gui, G., Wong, M.L., "*Implementing neural networks for decision support in direct marketing*", International Journal of Market Research 46 (2) (2004), 235–254.

[2] Soopramanien, D., Juan, L. H., "*The importance of understanding the exchange context when developing a decision support tool to target prospective customers of business insurance*", Journal of Retailing and Consumer Services 17 (2010) 306–312.

[3] Helfert, M., Heinrich, B., "*Analyzing Data Quality Investments in CRM- A model based Approach*", School of computing, (2003).

[4] Bayam, E., Liebowitz, J., Agresti, W., "*Older drivers and accidents: A meta analysis and data mining application on traffic accident data*", Expert Systems with Applications 29 (2005) 598–629.

[5] Witten, I.H., Frank, E., "*Data mining: Practical machine learning tools and techniques*" (2nd ed.), USA: Morgan Kaufmann Publishers (2005).

[6] Blows S., Ivers, R.Q., Connor, J., Ameratunga, S., Norton, R., "*Car insurance and the risk of car crash injury*", Accident Analysis and Prevention 35 (2003) 987–990.

[7] Šubelj, L., Furlan, S., Bajecl, M., "*An expert system for detecting automobile insurance fraud using social network analysis*", Expert Systems with Applications 38 (2011) 1039–1052.

[8] Liao, S.H., Chen, Y., Tseng, Y.Y, "*Mining demand chain knowledge of life insurance market for new product development*", Expert Systems with Applications 36 (2009) 9422–9437.

[9] JU LIN, Y., HUANG, C.S., LIN, C.C., "*Determination of Insurance Policy Using Neural Networks and Simplified Models with Factor Analysis Technique*" Wseas Transactions On Information Science & Applications, Issue 10, Volume 5, October (2008), 1415-1425.

[10] A.F. Shapiro, *"The merging of neural networks, fuzzy logic, and genetic algorithms"*, Insurance: Mathematics and Economics, Vol. 31, 2002, pp.115-131.

[11] Lotfi, H., Latifi, M., "*classification the Tehran bourse market data by using integrating neural networks and colonial competitive algorithm",* Fifteenth Annual International Conference of Computer Society of Iran.

[12] Cachón, A., Vázquez, R. A., *"Tuning the parameters of an integrate and fire neuron via a genetic algorithm for solving pattern recognition problems"*, Neurocomputing, Available online 30 July (2014).

[13] Uysal, A. K., Gunal, S., "*Text classification using genetic algorithm oriented latent semantic features*", Expert Systems with Applications, Volume 41, Issue 13, 1 October (2014), pp 5938–5947.

[14] Lin, C. H., Chen, H. Y., Wu. Y. S., *"Study of image retrieval and classification based on adaptive features using genetic algorithm feature selection"* Expert Systems with Applications, Volume 41, Issue 15, 1 November 2014, Pages 6611–6621.

[15] Homayouni, S. M., Tang, S.H., Motlagh, O., "*A genetic algorithm for optimization of integrated scheduling of cranes, vehicles, and storage platforms at automated container terminals*", Journal of Computational and Applied Mathematics, Volume 270, November (2014), pp 545–556.

[16] Tosun, U., *"A New Recombination Operator for the Genetic Algorithm Solution of the Quadratic Assignment Problem"*, Procedia Computer Science, Vol. 32 (2014), pp 29–36.

[17] Li, D. C., Wen, I. H*., "A genetic algorithm-based virtual sample generation technique to improve small data set learning"*, Neurocomputing, Vol. 143, 2 November (2014), pp 222–230.

[18] Zarifi, M. H., Ghalehjogh, N. K., *"A new evolutionary approach for neural spike detection based on genetic algorithm"*, Expert Systems with Applications, Available online 4 August (2014).

[19] Pourvaziri, H., Naderi, B., *"A hybrid multi-population genetic algorithm for the dynamic facility layout problem"*, Applied Soft Computing, Available online 22 July (2014).

[20] Smail, M.K., Bouchekara, H.R.E.H., Pichon, L., Boudjefdjouf, H., Mehasni, R., *"Diagnosis of wiring networks using Particle Swarm Optimization and Genetic Algorithms"*, Computers & Electrical Engineering, Available online 11 August (2014).

[21] Kalita, Z., Datta, D., "*Solving the bi-objective corridor allocation problem using a permutation-based genetic algorithm*", Computers & Operations Research, Vol. 52, Part A, December (2014), pp 123–134.

[22] Maghsoudi, R, Ghorbannia Delavar, A, Hoseyny, S, Asgari, R, Heidari, Y, *"Representing the New Model for Improving K-Means Clustering Algorithm based on Genetic Algorithm*", The Journal of Mathematics and Computer Science Vol .2 No.2 (2011) 329-336.

[23] Nazif, N, "*A Genetic Algorithm for Solving Scheduling Problem*", The Journal of Mathematics and Computer Science Vol .5 No.2 (2012) 91-96.

[24] Vahidi, J, Mirpour, S, "*Introduce a New Algorithm for Data Clustering by Genetic Algorithm*" Journal of mathematics and computer Science 10 (2014) 144 – 156.

[25] Eslami, M, Vahidi, J, Askarzadeh, M, "*Designing and Implementing a Distributed Genetic Algorithm for Optimizing Work Modes in Wireless Sensor Network*", Journal of mathematics and computer Science 11 (2014) 291-299.

[26] Sanz, S. S., Villacañas, J. L. F., Vargas, M. J. S, Calzón, C. B., *"Genetic programming for the prediction of insolvency in non-life insurance companies"*, Computers & Operations Research, Volume 32, Issue 4, April (2005), pp 749–765.

[27] Guillén, M., Nielsen, J. P., Scheike, T. H., Pérez-Marín, A. M., *"Time-varying effects in the analysis of customer loyalty: A case study in insurance"*, Expert Systems with Applications Vol. 39, Issue 3, 15 February (2012), Pages 3551–3558.

[28] Guelman, L., Guillén, M., Pérez-Marín, A. M., *"A survey of personalized treatment models for pricing strategies in insurance"*, insurance: Mathematics and Economics Vol. 58, September (2014), pp 68–76.

[29] Lin, C., *"Using neural networks as a support tool in the decision making for insurance industry"*, Expert Systems with Applications, Vol. 36, Issue 3, Part 2, April (2009), pp 6914–6917.