RESEARCH
www.commprac.com
ISSN 1462 2815

PREDICTING CHURN IN TELECOM SECTOR USING A POPULATION-BASED INCREMENTAL ANN LEARNING ALGORITHM

Dr. K Uthra Devi [1], Dr. P. Ramesh [2]*, A Joshua Issac [3], DS Bhupal Naik [4], M. A. Reetha Jeyarani [5] and M. K. Mohamed Faizal [6]

[1] Assistant Professor, Artificial Intelligence and Data Science, Indra Ganesan College of Engineering, Trichy, Tamilnadu, India.
[2] Assistant Professor, Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai, Tamilnadu, India.
[3,5] Assistant Professor, Artificial Intelligence and Data Science, K Ramakrishnan College of Technology, Trichy, Tamilnadu, India.
[4] Assistant Professor, Advanced Computer Science and Engineering, Vignan's Foundation for Science Technology and Research, Guntur, Andhra Pradesh, India.
[6] Assistant Professor, Artificial Intelligence and Data Science, K. Ramakrishnan College of Engineering, Samayapuram - Kariyamanickam Rd, Tamil Nadu, India.
Email: [1]uthra.ud@gmail.com, [2]drrameshp@veltech.edu.in (*Corresponding Author), [3]joshuaissac.ai@krct.ac.in, [4]dsbhupal@gmail.com, [5]reethajeyaranima.ai@krct.ac.in, [6]moh5zal@gmail.com

DOI: 10.5281/zenodo.12783619

Abstract

With global advancement, Information Technology has led to the growth of numerous Service Providers, which, in turn, has resulted in fierce competition between themselves. For Service Providers, the most prevalent obstacle is the handling of customer churn, retention, and satisfaction of customers for successful market sustenance. Customer Relationship Management (CRM) concentrates on boosting, sustaining, and building long term customer associations. CRM relies on the collection of information prior to making decisions. When a customer halts the existing service provider relationship and shifts to another, this is referred to as churn. The overall business profit and image are perturbed by this never-ending motion of churning. Therefore, it is more preferable to stop customers from churning and going for forecasting. In this work, churn prediction in telecom sector is investigated. Artificial Neural Network are used for prediction. To enhance the performance of the ANN, it is required to optimize its structure. From a given solution set, the global optimum can be detected utilizing the probabilistic method of Simulated Annealing (SA). Various optimization problems related to engineering and other areas have been resolved favourably with Population-Based Incremental Learning (PBIL) utilization. For predicting customer churn, this work has proposed a structure optimized Hybrid Simulated Annealing – Population-Based Incremental Learning ANN. Further deep learning techniques were used to improve the Churn prediction.

Keywords: Service Providers, Customer Relationship Management (CRM), Churn, Simulated Annealing (SA), Population-Based Incremental Learning (PBIL), Deep Learning, Rectified Linear Unit (ReLU), Structure Optimized Hybrid Simulated Annealing - Population-Based Incremental Learning Artificial Neural Networks (HSAPBIL ANN).

1. INTRODUCTION

The foundation of Customer Relationship Management (CRM) (Bhat, & Darzi, 2016; Rahimi & Kozak 2017) is Relationship Marketing (RM) principles, which is an emergent field of modern marketing. In the 1990s, the idea of CRM existed in the business field. CRM has garnered much attention from both global businesses and the research community as an academic study. The foundation of this method is the necessity to create new business environments. This, in turn, generates customer relationship management opportunities. In recent times, CRM is a significant emerging business practice employed for the management of interaction between an organization and its customers (existing and prospective). The job of the CRM method is to analyze the customer's historical data with an organization. CRM aids in sales growth by concentrating on customer retention. This will also result in the enhancement of the organization's business relationship with its customers. The objective of this work is to establish how the success of CRM is affected by customer knowledge management, customer orientation, organizational capability, and technology.

In telecom industry, the customer churn signifies that customers have canceled their service from the service provider. The retainment of an organization's existing customers has a key part in preserving the organization's good reputation within the competitive market and also in increasing the organization's overall revenue and image (Nie, Rowe, Zhang, Tian, & Shi 2011). The main challenge of churn prediction is the absence of a single cause for customer churn. This is generally an accumulation of various other causes. It is very complicated to detect these causes, as they are reliant on the organizational services used by the customers and also on the individual views of the customer. In this area of expertise, the key necessity of organizations is early churn prediction, identification of the main reasons for churn, and the prediction of countermeasures to prevent churn. The available organizational data can be employed for all these actions. However, the prediction mechanism is greatly hindered by the data's nature. Tangible or intangible service-oriented areas, product-based businesses, and telecommunication services are churn prediction application areas (Gunther, Tvete, Ass, Sandnes & Borgan 2014).

The Churn prediction gives the telecom sector an understanding of the customers' inclination to drop the organization's services for other competitors. Organizations can use this critical information to focus on these customer categories through marketing promotions to retain customers (Verbeke, Marten, Mues, & Baesens 2011). For an organization, compared to getting new customers, it is more beneficial to retain existing customers due to numerous reasons like the low costs associated with their upkeep.

A metaheuristic is an advanced methodology or heuristic which has been devised to detect or choose a heuristic capable of providing optimization problems with an optimal solution (not best), particularly in problems with the constrained computational ability or with information that is rudimentary or inadequate (Vijaya & Sivasankar, 2019). For the purpose of prediction, the solution's near optimality is a significant disadvantage. Based on observations from the past years, statistical algorithms have got around the near-optimality challenge. For the problem at hand, it can give optimal solutions, that is, the best solution. With the automation of almost every available procedure, there has been a tremendous explosion of data in recent times. Valuable information can be mined from this data as it is rich in information.

Neural network-based hybrid techniques have also been proposed for the prediction of churn (Rere, Fanany, Arymurthy, 2015). For searching the minimum number of excitations, the Simulated Annealing (SA) optimization algorithm is utilized. This method's name is derived from the annealing process in metallurgy, which is reducing a material's defects by controlled cooling. SA algorithm is a part of the heuristic approach and is a popular technique for resolving global optimization problems in a huge search space. During optimization, the goal function (also referred to as fitness, energy) of the optimization algorithm will be maximized (or minimized). The algorithm's convergence is dependent on the problem's coding and its evaluation function.

There is a lot of research in the structural optimization field due to the growing industrial demands for structures that are light in weight, have the best performance, and have a minimal cost. In the preliminary stages, structural optimization (Saka, Hasancebi & Geem 2016) was constrained by the sizing optimization. It is quite challenging to locate structural designs which have a global optimum. These need complex optimization methodologies. There can be numerous locally optimal configurations in standard structural optimization.

Deep learning techniques are now popularly used for performing automatic feature extraction from raw data (feature learning). In deep learning, feature hierarchies are learned, from higher levels to lower level features. As the features are learned automatically, the system is able to map complex functions from input to output. A Convolutional Neural Network (CNN) is a deep learning method using convolution, ReLU and pooling layers.

This work has proposed the structure optimized Hybrid Simulated Annealing – Population-Based Incremental Learning Artificial Neural Network. The remainder of this paper is arranged as follows: related works in literature are presented in Section 2. The employed methods are discussed in Section 3. The experimental results are discussed in Section 4, and the conclusion is made in Section 5.

2. RELATED WORKS

The manner in which an organization's performance is affected by CRM achievement was discussed by Soltani, Zareie, Milani, & Navimipour (2018). Partial Least Squares Structural Equation Modeling (PLS-SEM) was utilized for evaluating these hypotheses. Experimental results have demonstrated that CRM's success is extremely affected by "information technology use" and also that the success of CRM is associated with "customer knowledge management", "organizational capability", and "customer orientation". Subsequently, the discussion was done on the ramifications of the research, its findings, prospective avenues for research, and constraints.

Three hybrid models were studied by Hudaib, Dannoun, Harfoushi, Obiedat, & Faris (2015) to devise an accurate and effective model for the prediction of churn. The proposed models depend on two distinct stages: clustering and prediction. In the clustering stage, there is filtering and grouping of every customer data. In the prediction stage, customer behaviour is predicted. In the first hybrid model, the K-Means algorithm is employed for filtering the data, and the Multilayer Perceptron Artificial Neural Networks (MLP-ANN) is employed for the prediction. In the second hybrid model, hierarchical clustering is employed together with the MLP-ANN. In the third hybrid model, Self-Organizing Maps (SOM) are employed together with the MLP-ANN. Depending on real data and their accuracy, all these hybrid models were devised. In order to determine the proposed models' effectiveness, the values of the churn rate were assessed and later compared with current techniques.

A Particle Swarm Optimization (PSO) method for telecom churn prediction was put forward by Vijaya & Sivasankar (2019). They also proposed three variations of PSO for churn prediction: PSO integrated with a pre-processing mechanism of feature selection, PSO embedded with simulated annealing, and PSO combined with both simulated annealing and feature selection. For analysis of the performance aspects and predictability levels, the proposed classifiers were compared with three hybrid models, Random Forest, Support Vector Machine, K-Nearest neighbour, Naïve Bayes, and decision tree. The utilized performance indicators were precision-recall plots, receiver operating characteristics, F-Measures, Precision, false-positive rate, true negative rate, true positive rate, and Accuracy. Experimental outcomes demonstrated that the metaheuristics' performance was more effective, and they also displayed higher levels of predictability.

A lately devised hypothesis of the Genetic Algorithms (GAs) was examined by Fyfe (1999). In this hypothesis, any generation's GA population is denoted by one vector whose elements are the probabilities of the respective bit positions being equal to 1. The evolution process is denoted by learning the probability vector's elements. This technique is directly connected to the Artificial Neural Network technique of competitive learning. The ANN methods are utilized to broaden this technique's application to produce an order on a sub-population set, to multi-modal problems, to multi-objective criteria, and non-static problems.

For handling PBIL and certain effective local search schemes, a hybrid Evolutionary Algorithm (EA) was devised by Bureerat (2011). There is a development of a simplistic PBIL with real codes. To the PBIL's major process, there is an integration of the evolutionary direction and approximate gradient operators. This technique has been devised for single objective global optimization. For box-constrained optimization, the proposed hybrid algorithm's search performance was compared with various renowned and recently devised EAs and metaheuristics. Upon comparison, it was found that, with the provided optimization settings, the proposed hybrid optimizer surpasses other EAs' performance. This novel derivative-free algorithm could also sustain EAs' exceptional capabilities.

3. METHODOLOGY

ANN classifiers, which are utilized for the classification of non-churners and churners, have been described in this section. For classifier evaluation, the CRM dataset is employed. For research of customer churn prediction, American telecommunication organizations have provided publicly available CRM dataset. The features are extracted using TF-IDF.

3.1. Structure Optimized Simulated Annealing Artificial Neural Networks

A probabilistic technique for the evaluation of global optimum for problems that cannot be resolved with conventional optimization methods is Simulated Annealing (SA) (Bahrami & Doulati (2016). Just like its name, this method will replicate the metallurgical annealing process where the material's defects are reduced through slow cooling. In SA, the metallurgical annealing's controlled cooling is put into effect as the probability of transitioning towards a worse solution. This probability is directly proportional to the temperature. When the temperature is high, more is the opportunity to shift towards a worse solution. Due to this feature, the SA is able to survey the entire search space. As opposed to the genetic and swarm intelligence algorithms, the SA does not suffer from premature convergence. Also, unlike the population-based EAs, the SA works only on a single probable solution (known as state) and attempts to arrive at a better solution by improving it. This improvement is carried out through the generation of a new successor in the current state's neighbourhood and then transitioning towards it in a probabilistic manner. Suppose S denotes the current state, and S denotes the successor (or the neighbour) generated depending on the current state. The proposed move from S to S occurs depending on a fitness function: If the fitness of S is better than S, then the transition will definitely occur; else, it may probably occur. If the successor's fitness is lower than the current state, the probability of a transition corresponds to their fitness value and the temperature.

The outline of the proposed hybrid optimisation algorithm ANNSA is as below:

Step 1: The SA algorithm will begin at a certain specific "high" temperature with one trial solution vector P (biases and weights of the elected ANN) (Nowakowski, Dorogyy & Doroga-Ivaniuk, 2017) of dimension N. Initially, there is a random generation of P's elements within a feasible upper and lower limit.

Step 2: By the current solution's random perturbation, there is a generation of the new candidate solution.

Step 3: There is a temperature reduction for simulating thermal equilibrium at every temperature.

Step 4: The biggest function values from each temperature reduction's end are compared with the most latest and optimum function value. There is the termination of the algorithm if all these differences are lower than a pre-specified small value.

3.2. Structure Optimized Population-based Incremental Learning Artificial Neural Networks

The Genetic Algorithm's alternate, Population-Based Incremental Learning (PBIL), was initially devised by Baluja. It combines the mechanisms of a GA with simple competitive learning. In contrast to the majority of the conventional Evolutionary Algorithms, for estimation of the binary population, the alleged probability vector is employed by the PBIL. In this technique, through the iterative improvement of the probability vector, the optimization search is accomplished. Unlike the original binary-code PBIL (Jaderberg, Dalibard & Osindero, 2017), despite the development of the PBIL's real-code variations, they are not quite well-known.

The neural network's intermediate layer structure is represented by each matrix column. The neural network's initial and final layers are set as per the input dataset sizes. Therefore, for the identification of a dataset's optimum structure, distinct probability vectors are examined. The proposed PBIL technique is nested in the inner and outer loop. The inner loop is utilized for the identification of the applied structure's training error values. The outer loop is utilized for the optimization of the neural network's structure. Within the outer layer, the global best is considered as the particle with the minimum mean square error value, and each probability vector will have its corresponding binary population. For each iteration, the structures are adjusted by the probability vector. In the inner loop, these structures are again trained with the PBIL. Each network will have an intermediate layer with a maximum of nine nodes and a minimum of 2 nodes.

3.3. Hybrid Simulated Annealing- Population-based Incremental Learning

The Hybrid Simulated Annealing – Population-Based Incremental learning methodology is as below:

- Consider an initial simulated annealing temperature.
- Produce a vector that has the same length as the needed chromosome and which has elements with probabilities of a 1 in the chromosome's respective bit position.

ISSN 1462 2815

Every bit in the vector is initialized to 0.5 when a binary alphabet is employed for the chromosomes.

- A number of samples are generated from the vector where the current probability vector will determine the probability of a 1 in each vector's bit position.
- There will be a halt to the algorithm by relating the biggest function values from each temperature reduction's end with the maximum existing one and the optimum function value. The algorithm will come to an end if all these modifications are lower than a pre-definite small value.
- Identification of the population's fittest chromosome(s).
- Modify the elements of the probability vector such that the probability of a 1 is higher in positions where the fittest chromosomes have a 1.

The procedure is initiated with a probability vector with 0.5 in each element. This procedure will come to a halt when the vector's each element arrives at 0 or 1. A supervised learning technique is applied for updating the elements of the probability vector.

3.4. Proposed Structure Optimized Hybrid Simulated Annealing - Population-based Incremental Learning Artificial Neural Networks

In comparison to the generic Genetic Algorithms, PBIL has been successfully utilized for the resolution of numerous benchmark and real-time problems like the power system stabilizer development by incremental learning. By training incremental learning utilization, pattern variations can later be applied to devise models for responsive stabilizer designs.

The following are the PBIL algorithm's fundamental concepts:

- Set-Up of the Solution Structure. The PBIL algorithm's solution structure is akin to the general Genetic Algorithm. The problem's decision variables are found. Later, they are encoded into adjacent binary sub-strings to produce a complete string known as a solution vector. The string's length is determined by the decision variables' nature. For instance, if the decision variable can have values between 0 and 100, it is adequate to have a seven binary digit sub-string. Binary values (0 or 1) are randomly allocated to the string's (solution vector) digits. Subsequent to the decoding of sub-string and assignation of proper values to the corresponding variables, one probable solution is considered to be contained in this string. As a finite set of such solution vectors constitute a population of probable solutions, the size of the population should be chosen by the analyst.
- Evaluation Function. For each probable solution proposed by the metaheuristic, an evaluation function is essential to assess the quality of each of these solutions. There can be maximization or minimization of the evaluation function's results. Generally, the evaluation function is a mathematical function (Bekker & Olivier, 2008). However, simulation models are employed as evaluation functions for stochastic systems, which are complicated and dynamic.
- Probability Vector. This is an extra structure. The probability vector will have the same number of elements as the solution vectors; however, instead of a binary number, each element will enclose a probability value. A specific element's value demonstrates the probability that a solution vector's given digit has a '1'. The low

ISSN 1462 2815

probability of detecting a '1' in digit i of a solution vector is represented by the low value in element i. The PBIL algorithm's foundation is the probability vector. Although the vector's elements initially enclose probabilities of 0.5 each, during a provided iteration, the population's chosen solution vector will alter each element's content.

- Mutation. This is generally utilized with a 0.02 set probability. The term Random (0 or 1) implies that 0.5 is the probability of choosing either a '0' or a '1'. The next generation is produced, and its solution vectors are produced with the probability vector. In the population, each solution's element will have a random number generated for it. If the random number is more than the probability value in the probability vector's respective element, a '0' is allocated for that element of the solution vector. Else, a '0' is allocated. This is continued for the population's solution vectors.
- Convergence. The probability algorithm's elements will converge to either 1 or 0 if there is the convergence of the algorithm. Irrelevant iterations can be avoided by setting threshold values. For instance, if an element arrives at a value greater than 0.95, it is treated as having converged to 1, and if it arrives at a value lower than 0.05, it is treated as having converged to 0. The algorithm is halted if every element fulfils one of these conditions.

An extremely effective technique for structure and weight selection has been found to be the neural networks' evolutionary structure optimization (Karatzoglou, 2019). A design optimization algorithm generally can be embedded with the approximate model's structure optimization.

The hybrid SA-PBIL ANN methodology is as below:

- The initial simulated annealing temperature is provided.
- The population size N is determined. There is a random generation of the initial population. For encoding the chromosome, the decimal number is utilized. The number of variables that are to be optimized are denoted by each chromosome's gene number. The error precision and maximum iterative number M are specified.
- The population's chromosomes are calculated. The fitness function is evaluated as the objective function's minimum optimization problem is ANN training.

3.5 Proposed PBIL - Wide and Deep Learning Model (WDLM) Network

In this work, an algorithm for the prediction of churn based on the wide and deep learning framework is presented. The Rectified Linear Unit (ReLU) is the most commonly used activation function in deep learning models is used as the classification function. This framework is capable of compounding the benefits of memorization and generalization with less feature engineering, useful for analyzing telecom sector data. Thus, features are automatically extracted from the data and non-linear relationships or context-dependent is captured.

The features were processed by a Wide and Deep Learning Model (WDLM) Network which consists of a wide component and a deep component as shown in figure 1. The wide component is a generalized linear model used for large-scale regression and classification. Here the feature interactions are memorized and adds non-linearity to the generalized linear model. The deep component is a deep neural network which converts the high dimensional features to a low-dimensional dense embedding vector.

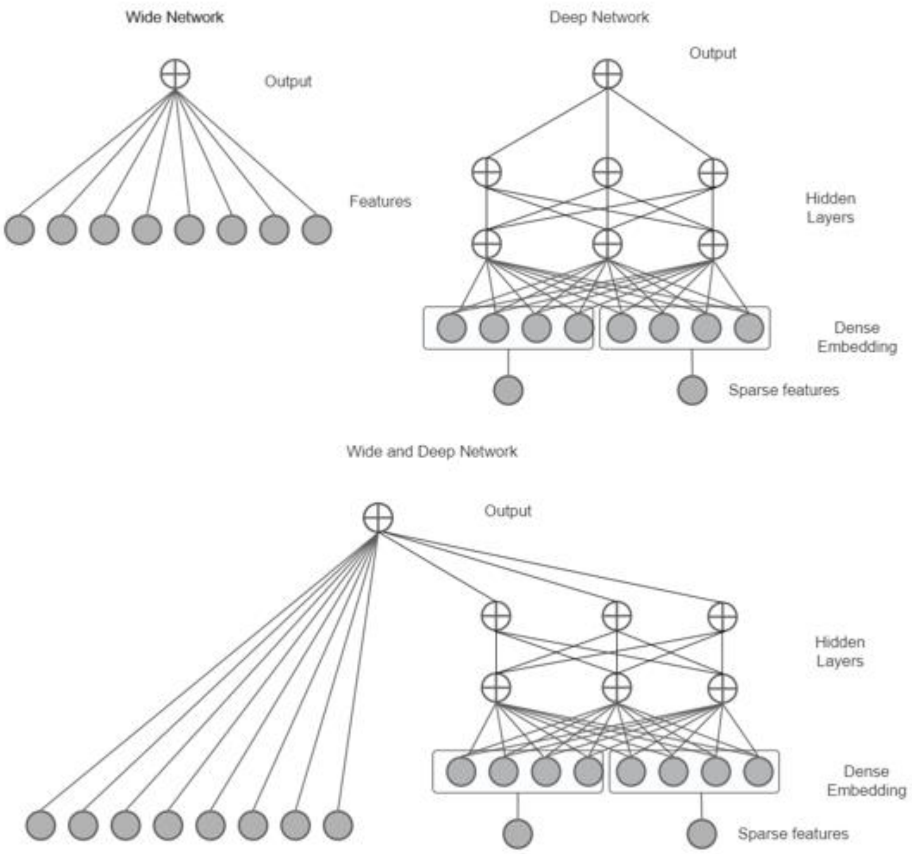

Figure 1: Wide and Deep Learning Network

In the WDLM Network, this component was composed of two types of layers: embedding and hidden layers. *Embedding layers* included two embeddings corresponding to two groups of features: churn and no-churn. All features were put into the wide part which included the crossed features joined with the output of the deep part in the last layer to form a vector. The activation function in other layers was the ReLU. The embedding vectors are generated randomly and during training the values to minimize the final loss function is found. The dense embedding vectors are passed through the network to get the output. The wide and deep network are combined using a weighted sum to obtain the prediction, which is then for joint training. Back-propagation is used for joint training.

In the proposed PBIL-WDLM Network, the structure of the WLDM is optimized using PBIL. The deep component is optimized similar to the steps given in section 3.2. The main steps are:

Initialize probability vector

Generate random Solutions

While termination condition not met

Evaluate Solution

Find best solution

Update probability vector

Mutate probability vector.

4. RESULTS AND DISCUSSION

4.1. Experimental Setup

The experiments, proposed PBIL WDLM Network, WDLM, Structure Optimized Simulated Annealing ANN, Structure Optimized Population-based Incremental Learning ANN, and Structure Optimized Hybrid SA-PBIL ANN methods are evaluated using the CRM dataset. Churn is computed on the basis of the customer leaving the service with 31 to 60 days after the customer was sampled.

4.2. Experimental Results

The experiments were run 15 times, and the averaged results are presented. The average standard deviation was about 2.8%. Table 1 shows the summary of the results. The classification accuracy, recall (for churn & no churn), precision (for churn & no churn), f measure (for churn & no churn) as shown in figures 3 to 6. Figure 7 shows the Fitness Function for Hybrid Population-based Incremental Learning.

Table 1: Summary of Results

	Structure Optimized SA ANN	Structure Optimized PBIL ANN	Structure Optimized Hybrid SA-PBIL ANN	WDLM	PBIL WDLM Network
Classification Accuracy	91.18	92.16	93.85	94.55	95.31
Recall for Churn	0.9429	0.9469	0.95	0.9423	0.9509
Recall for No Churn	0.8381	0.8615	0.9114	0.9531	0.9582
Precision for Churn	0.9325	0.942	0.9622	0.9795	0.9818
Precision for No Churn	0.8608	0.8724	0.8848	0.8743	0.8916
F Measure for Churn	0.9377	0.9444	0.9561	0.9605	0.9661
F Measure for No Churn	0.8493	0.8669	0.8979	0.912001	0.923701

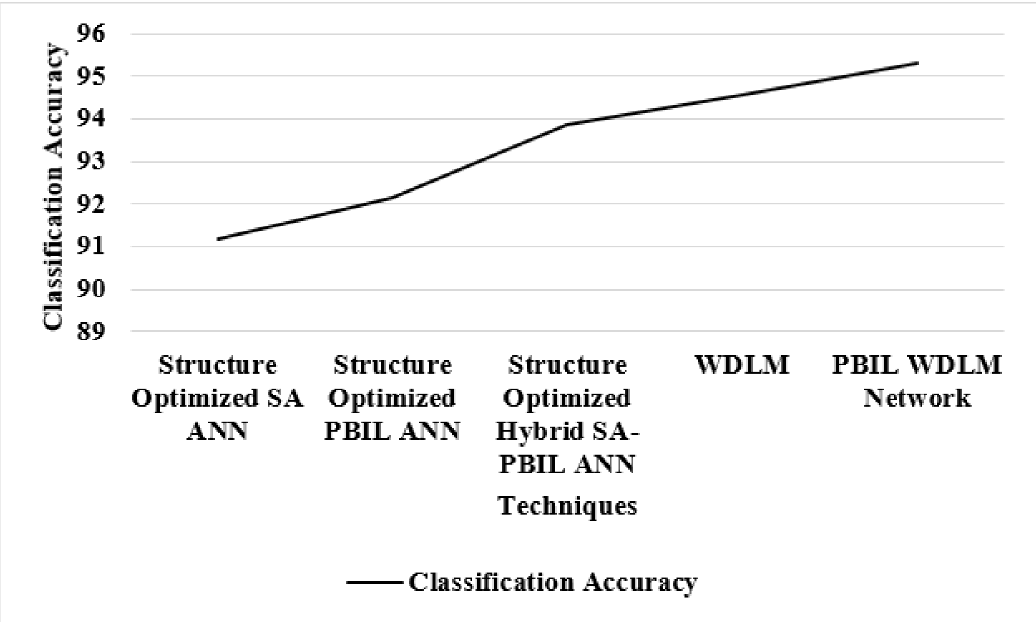

Figure 2: Classification Accuracy for PBIL WDLM Network

From figure 2, it can be observed that the PBIL WDLM network has higher classification accuracy by 4.43% compared to Structure Optimized SA ANN, by 3.36% compared for Structure Optimized PBIL ANN, by 1.54% compared for Structure Optimized Hybrid SA-PBIL ANN and by 0.8% compared for WDLM.

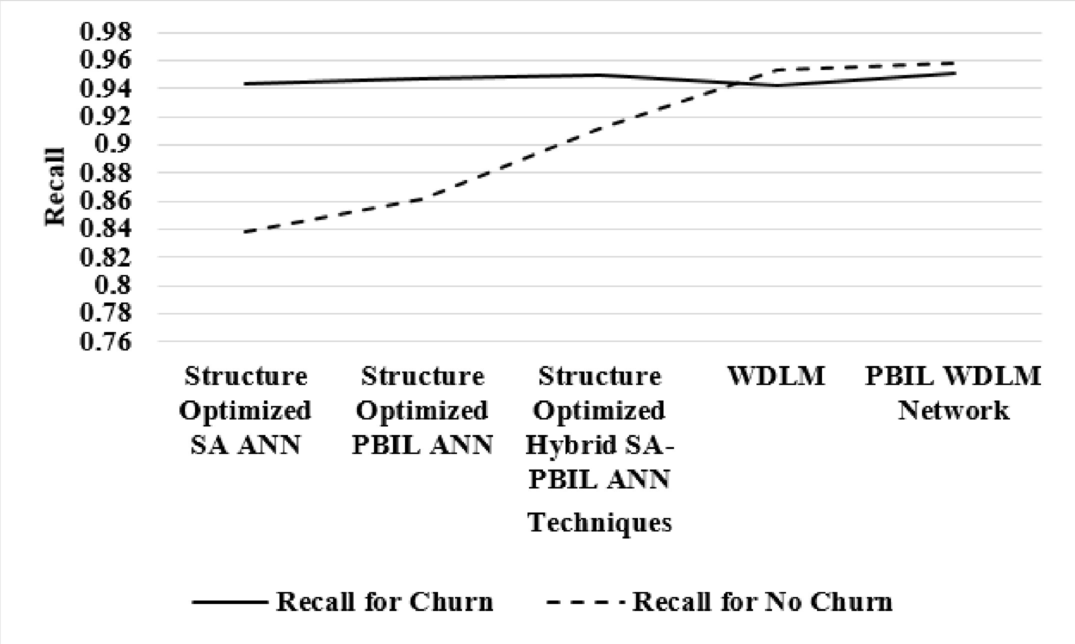

Figure 3: Recall for PBIL WDLM Network

From figure 3, it can be observed that the PBIL WDLM network has higher recall by 0.84% & 13.37% for Structure Optimized SA ANN, by 0.42% & 10.62% for Structure Optimized PBIL ANN, by 0.09% & 5.01% for Structure Optimized Hybrid SA-PBIL ANN and by 0.91% & 0.53% for WDLM when compared with churn and no churn respectively.

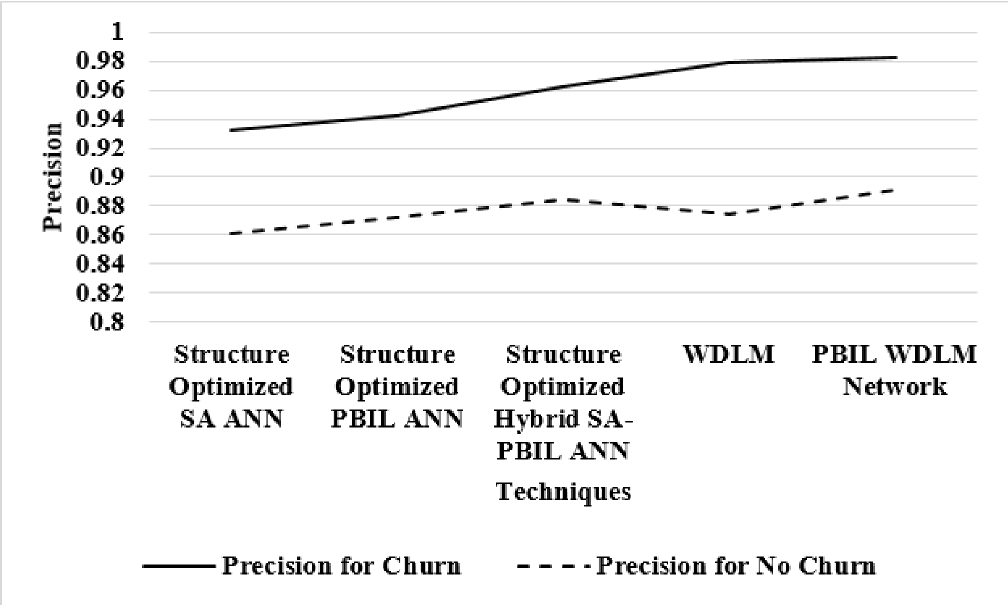

Figure 4: Precision for PBIL WDLM Network

From figure 4, it can be observed that the PBIL WDLM network has higher precision by 5.15% & 3.51% for Structure Optimized SA ANN, by 4.13% & 2.17% for Structure Optimized PBIL ANN, by 2.02% & 0.76% for Structure Optimized Hybrid SA-PBIL ANN and by 0.23% & 1.95% for WDLM when compared with churn and no churn respectively.

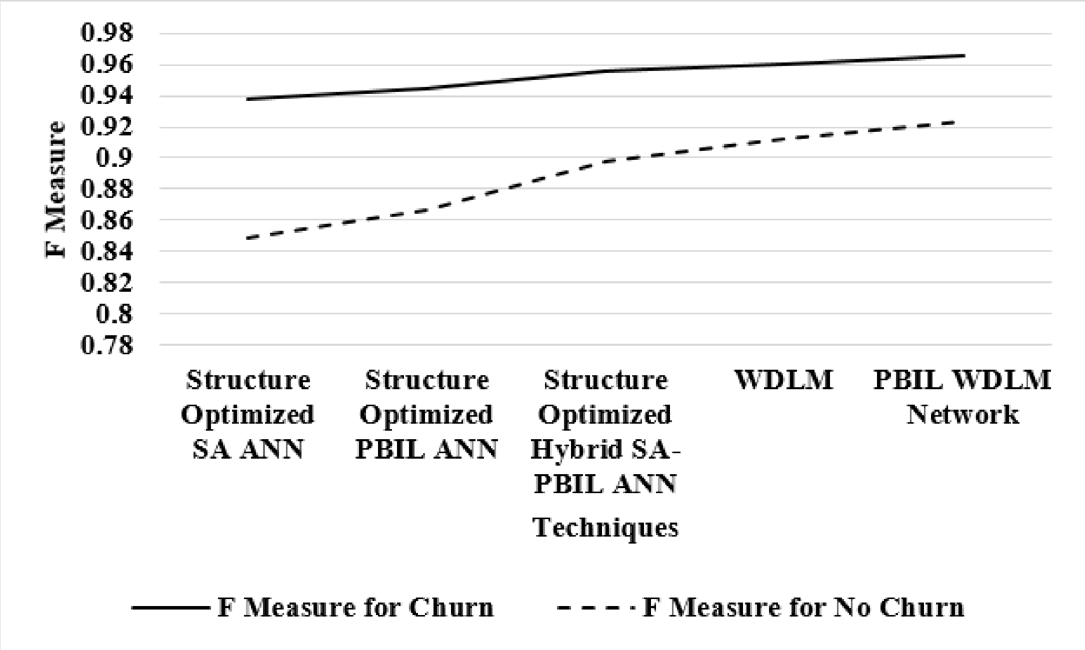

Figure 5: F Measure for PBIL WDLM Network

From figure 5, it can be observed that the PBIL WDLM network has higher f measure by 2.98% & 8.39% for Structure Optimized SA ANN, by 2.27% & 6.34% for Structure Optimized PBIL ANN, by 1.04% & 2.83% for Structure Optimized Hybrid SA-PBIL ANN and by 0.58% & 1.27% for WDLM when compared with churn and no churn respectively.

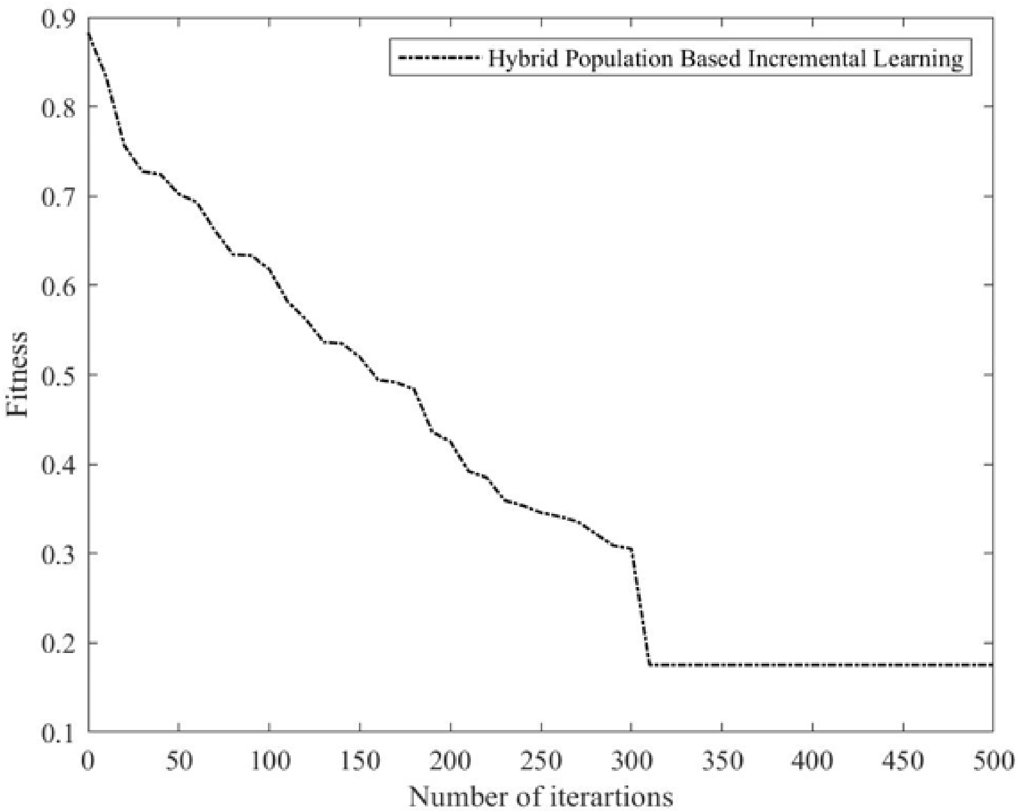

Figure 6: Fitness Function for Hybrid Population-based Incremental Learning

5. CONCLUSION

The churn prediction of customers of telecom sector is investigated in this paper. A distinct training plan is denoted by each customer. ANN are used for predicting the churn. In this work, metaheuristics are used for structure optimizing the ANN. The most effective techniques for resolving most of the hard optimization problems are metaheuristics. Through iterative probability vector improvement, an optimization search is accomplished by PBIL. Global optimal solutions can be detected with the probabilistic optimization method of Simulated Annealing. In optimal control problems, determination of time's control function is done through the exploitation of the ANN's capability to estimate any random nonlinear function. Experimental results demonstrate that the PBIL WDLM network has higher classification accuracy by 4.43% compared to Structure Optimized SA ANN, by 3.36% compared for Structure Optimized PBIL ANN, by 1.54% compared for Structure Optimized Hybrid SA-PBIL ANN and by 0.8% compared for WDLM

References

1) Bahrami S, Doulati F. Prediction of pyrite oxidation in a coal washing waste pile using a hybrid method, coupling artificial neural networks and simulated annealing (ANN/SA). J Clean Product 2016; 137: 1129–37.

2) Bekker, J., & Olivier, Y. (2008). Using the Population-Based Incremental learning algorithm with computer simulation: Some applications. *South African Journal of Industrial Engineering*, *19*(1), 53-72.

3) Bhat, S. A., & Darzi, M. A. (2016). Customer relationship management. *International Journal of Bank Marketing*.

4) Bureerat, S. (2011, January). Hybrid population-based incremental learning using real codes. In *International Conference on Learning and Intelligent Optimization* (pp. 379-391). Springer, Berlin, Heidelberg.

5) Fyfe, C. (1999). Structured population-based incremental learning. *Soft Computing*, *2*(4), 191-198.

6) Gunther, C.C., Tvete, I.F., Aas, K., Sandnes, G.I., Borgan, O.: Modelling and predicting customer churn from an insurance company. Scand. Actuar. J. 2014(1), 58–71 (2014).

7) Hudaib, A., Dannoun, R., Harfoushi, O., Obiedat, R., & Faris, H. (2015). Hybrid data mining models for predicting customer churn. International Journal of Communications, Network and System Sciences, 8(05), 91-96.

8) Jaderberg, M., Dalibard, V., Osindero, S., Czarnecki, W. M., Donahue, J., Razavi, A. ... & Fernando, C. (2017). Population based training of neural networks. *arXiv preprint arXiv:1711.09846*.

9) Karatzoglou, A. (2019, May). Evolutionary optimization on artificial neural networks for predicting the user's future semantic location. In *International Conference on Engineering Applications of Neural Networks* (pp. 379-390). Springer, Cham.

10) Nie, G., Rowe, W., Zhang, L., Tian, Y., Shi, Y.: Credit card churn forecasting by logistic regression and decision tree. Expert Syst. Appl. 38(12), 15273–15285 (2011).

11) Nowakowski, G., Dorogyy, Y., & Doroga-Ivaniuk, O. (2017, September). The realisation of neural network structural optimization algorithm. In *2017 Federated Conference on Computer Science and Information Systems (FedCSIS)* (pp. 1365-1371). IEEE.

12) Rahimi, R., & Kozak, M. (2017). Impact of customer relationship management on customer satisfaction: The case of a Budget Hotel Chain. Journal of Travel & Tourism Marketing, 34(1), 40–51.

13) Rere LMR, Fanany MI, Arymurthy AM. Simulated annealing algorithm for deep learning. Proc Comput Sci 2015;72:37–144

14) Saka, M. P., Hasançebi, O. Ğ. U. Z. H. A. N., & Geem, Z. W. (2016). Metaheuristics in structural optimization and discussions on harmony search algorithm. *Swarm and Evolutionary Computation*, *28*, 88-97.

15) Soltani, Z., Zareie, B., Milani, F. S., & Navimipour, N. J. (2018). The impact of the customer relationship management on the organization performance. *The Journal of High Technology Management Research*, *29*(2), 237-246.

16) Verbeke, W., Martens, D., Mues, C., & Baesens, B. (2011). Building comprehensible customer churn prediction models with advanced rule induction techniques. Expert Systems with Applications, 38(3), 23542364.

17) Vijaya, J., & Sivasankar, E. (2019). An efficient system for customer churn prediction through particle swarm optimization based feature selection model with simulated annealing. *Cluster Computing*, *22*(5), 10757-10768.

18) Nguyen, B. P., Pham, H. N., Tran, H., Nghiem, N., Nguyen, Q. H., Do, T. T., ... & Simpson, C. R. (2019). Predicting the onset of type 2 diabetes using wide and deep learning with electronic health records. Computer methods and programs in biomedicine, 182, 105055.