IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 19, Issue 4, Ver. VII (Jul.-Aug. 2017), PP 30-39
www.iosrjournals.org

Churn Prediction on Huge Telecom Data Using Hybrid Firefly-Particle Swarm Optimization Algorithm Based Classification

*Ammar A. Q. Ahmed[1], Maheswari D.[2]
Research Scholar[1], RathnavelSubramaniam College of Arts&ScienceSulur, Coimbatore, Tamil Nadu, India
Assistant Professor[2], RathnavelSubramaniam College of Arts&ScienceSulur, Coimbatore, Tamil Nadu, India
Corresponding Author: Ammar A. Q. Ahmed

Abstract: *Big data analytics brings in considerable value to decision making and provide more accurate and actionable insights into customer churn in telecom industry. As general meta-heuristics such as Particle Swarm Optimization (PSO) and Firefly Algorithm (FFA) are incapable of handling huge dataeffectively, Spark 2.0 on Hadoop architecture is employed. The parallelized nature of Spark &Hadoop will provide accurate and faster results. FFA and PSO were individually used to predict customer's behavior efficiently but it has issues with highly sparse dataalong with computational complexity and inaccuracy prediction results for the churn dataset. To overcome the above mentioned issues, in the proposed system, Hybrid Firefly with PSO (HFFPSO) based classification is proposed which combines the advantages of PSO and FFA. The main objective of the proposed system is to ensure higher accuracy for large churn dataset using HFFPSO algorithm.This method improves the search efficiency,local well as global through best objective function values. Theoverall performance of the customer churn prediction is improved with the help of the optimized features which is proved through the ROC, PR, accuracy, True Positive Rate (TPR), False Positive Rate (FPR), f-measure, and execution time. Experimental results performed using Spark 2.0 on Hadoop 2.7 confirmed that HFFPSO based classification provides a good, highly scalable churn prediction analysis in the telecommunication industry.*

Keywords: *Churn prediction, Telecom, Meta-heuristics, Spark, Hadoop, PSO, FFA, HFFPSO*

Date of Submission: 18-08-2017 Date of acceptance: 05-09-2017

I. Introduction

Customer churn is a serious issue affecting commercial and manufacturing. Telecommunication businesses frequently lose respected clients and, therefore, revenues to the competition. The telecommunication business has gone via incredible variations over the past few years such as addition of innovative facilities, technical progressions and enlarged competition since deregulation [1]. Hence, customer churn prediction in telecommunication industry is turned into significant for commercial players to defend their trustworthy client base, association growth, and progress its Customer Relationship Management (CRM) [2, 3].

Retaining customers along with great churn risk is one of the strongest challenges in telecommunication business currently [4]. Since large amount of service providers and more intense competition, clients nowadays have a several selections to churn. In telecommunications, the word "churn" tends to the cost of subscribers who switch from one provider to another through a specific time.

As it is more gainful to recollect preceding clients than to continuously charm new clients [5] [6], it is critical to construct an exact churn prediction model for recognizing those clients who are maximum prone to churn. The main objective of churn prediction model is to identify those customers thus the retention approaches are targeted depends on them and the corporation might embellishment through increasing its general income. In [7] represent levels of a common churn prediction model such as data gathering, training, arrangement and forecast. It also defines that recognizing the correct collection grouping of variables has important inspiration in developing the proportion of true predictions.In [8] novel technique to support the industry better lessen the risk of customer churn, and therefore to increase higher benefits. It mainly examines customer churn problem in service industry under the big data surroundings. The usage of business big data for customer churn administration has turn intosignificantsinceconventional techniques are not engineered for the kind of big, dynamic and unstructured data. For instance, ClouderaHadoop introduces some big data use cases for telecommunication companies including customer experience supervision, network optimization, operational analytics and information monetization.

Meta-heuristic global search algorithms like evolutionary methods have been employed to optimize an extensive range of difficult, huge scale methods such as big data, ranging from manufacturing design to re-establishment of natural networks. Meta-heuristics illustrated to be effective for resolving optimization problems. It is proficient to determine an approximate solution within a reasonable time, they have been widely used in solving thedata mining problem in recent years. The major purpose is that for a populationbasedsearch process, various individualscan internment diverse trade-off relationships among the conflicting objectives, say example, in composite structural enterprise optimization. In [9] explored the idea of big data investigation and distinguish some significant data from some sample big data source, such as Twitter twits, using one of companies emerging tool, named as Spark by Apache.Spark helps in-memory computing, that activates it to query data much faster compared to disk based engines such as Hadoop, and also it provides a commonimplementation model that can optimize arbitrary operator graph. Spark is quick andreliable, scalable, and makes sure to transforminformation. It is also not very complex to be deployed and utilized.Spark is advance and highly proficient upgrade to Hadoop aimed at enhancing Hadoop ability of cutting edge analysis. Spark engine functions quite advance and different than Hadoop. Spark engine is developed for in-memory processing as well a disk based processing. This in-memory processing capability makes it much faster than any conventionalinformation processing engine. In the proposed system, the HFFPSO model is used in Spark 2.0 for churn prediction using efficient selection of customer's behavior.

II. Related Works

Jadhav and Pawar [10] an attempt is made to construct a decision support schemethrough data mining expertise for churn prediction in Telecommunication business. It needsgreatly sophisticated customized and superior decision support classification. Huang et al [11] presents novelgroup of attributes for land-line customer churn prediction, consisting two six-month Henley segmentation, precise four-month call details, line information, bill and payment details, account details, demographic profiles, service orders, find fault information, etc. Then the seven classification methods (Logistic Regressions, Linear Classifications, Naive Bayes(NB), Decision Trees(DT), Multilayer Perceptron Neural Networks(MPNNs), Support Vector Machines(SVMs) and the Evolutionary Data Mining Algorithm) are used in customer churn as predictors, depends on the new features. It also defines that recognizing the right grouping of variables has important influence in increasing the percentage of true predictions (TP). Kirui et al [12] discussed a new group of attributes along with the objective of developing the prediction ratios of feasible churners. The features are derived from call information and client profiles and classified as contract-related, call pattern description, and call pattern changes description features.Dalvi et al [13] presents a statistical endurancestudy tool to forecast customer churn depends on evaluationamong decision trees and logistic regression. Selecting the right combination of attributes and fixing the proper threshold values may produce more accurate results of predicting churn customers. The model suggests that data mining techniques can be a promising solution for the customer churn management. Yihui et al [14] presents a feature selection method based on Orientation Ordering Pruning Method (OOPM). It used an indicator system for customer churn prediction of telecom industry. Two effective methods of feature selection and extraction for customer churn prediction are also effective.

Ahmed et al [15] deals with recognizing and predicting churn in the telecom big data. This method provides an effective hybridized firefly approach for churn classification. Hybrid firefly presents the normal firefly algorithm viarevealing very less time latency.Yang et al [16] discussed subtractive clustering based boundary restricted adaptive PSO algorithm for clustering multidimensional statistics. PSO generates superior outcomes in difficult issues alongwith few parameters to adjust giving fast as well as accurate computation results.Karthikeyan et al [17] used a hybrid algorithm of grouping of PSO and Artificial Bee Colony (ABC) algorithm, easy and robust optimization method is utilized in clustering of the benchmark prediction problems for classification purpose. It gives good decision making for specified diabetes database.Ragins et al [18] presents CRM associationnumerouscompensation and profit. These sentiments can comprise trust, liking, and believing in the organization's capability to replyefficiently and rapidly to clients' problem.Umayaparvathi et al [19] discussed a churn classificationmethod for telecom customers. The problem is modeled as a binary prediction problem and thereforenumerous state-of-the-art classifiers are utilized for making churn model. Amin et al [20]provides an intelligent rule-based decision-making method, depends on Rough Set Theory (RST), to minesignificant decision rules associated to consumer churn and non-churn. This algorithm efficiently implements prediction of churn from non-churn consumers, along with prediction of those customers who will churn or may perhaps churn in the near future. The research works in the literature has both merits and demerits in the evaluation of churn prediction. The big data telecom dataset has issues with time complexity and lack of accuracy in the decision making process. To overcome these problems, the hybrid optimization based prediction algorithm is developed in the proposed model.

III. Proposed Methodology

In this research, HFFPSO is introduced to predict the optimized based churn classification result from churn data. Telco big data can make churn prediction much easier from the 3V's perspectives: Volume, Variety, and Velocity. Enterprises have vast amounts of customer behavior data in the era of big data. Most of traditional customer churn predicting models ignore customer segmentation and misclassification cost, which reduces the rationality of model. The purpose of optimization algorithm is focused to increase the classification of churn behavior result optimally. The hybrid meta-heuristic algorithms are proposed to improve the best attribute selection through the local and global optimal solutions. Optimization is aims to discover the more imperative data from a specified group of consumer churn behavior. These features are encloseddata about consumers who left within the last month – the column is called Churn. Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies. Customer account information – how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges. The demographic data of clientele such as femininity, age series, and also verifies if they containassociates and dependents. These features are classified as arithmetical and definite. The attributes are contained 'yes' or 'no' and '0' or '1' etc., so significant attributes of the customeris chosen based on the HFFPSO algorithm.

Firefly algorithm based classification

Firefly optimization method is based on the information that the every firefly attracts to other firefly on the basis of the brightness i.e. firefly alongwith lower brightness is attracted toward firefly with additional brightness and therefore search gap is discoveredeffectively. With the enlargequantity of information generated, the competence of the numericalmethods is now being overshadowed through the large amount of time considered via them to construct solutions [21]. Thus the necessity for metaheuristic approaches is examined in big data churn classification. The small error that appears as an intrinsic part of the solutions presented through metaheuristics has turn into acceptable in most applications whose mainnecessities is to givequicker results. The need for online processing has provoked the usage of metaheuristics to a huge extent. Amplified processing capability from the hardware characteristic has activated superior and additionaleffectual processing, in turn developing the precision of these optimization methods. This algorithm used to identify probable customers for preventing churn.

The firefly algorithm is based on the idealized behavior of the flashing characteristics of fireflies. These flashing characteristics provided as the following three rules:

1- All fireflies are unisex so that one firefly is attracted to other fireflies regardless of their sex.

2- Attractiveness is proportional to their brightness, thus for any two flashing fireflies, the less bright one will move towards the brighter one. The attractiveness is proportional to the brightness and they both decrease as their distance increases. If no one is brighter than a particular firefly, it moves randomly.

3- The brightness or light intensity of a firefly is affected or determined by the landscape of the objective function to be optimized.

For a maximization problem, the brightness can simply be proportional to the objective function. For simplicity, it is assumed that, the attractiveness of a firefly is determined by its brightness or light intensity which in turn is associated with the encoded objective function. However the number of customers is increased and it has insufficient memory space to store all the churn information. It takes long time for overall execution to produce churn prediction result in telecom industry. The disadvantages are, it leads to premature convergence and it reduces the chance to find global optimum. In few cases, it has issue with handling high dimensional dataset since it has lower memory storage. It has also issue with computational complexity and it provides less number of optimal solutions. Thus, the churn prediction accuracy is reduced significantly.

PSO Algorithm based classification

The phenomenon of PSO is that information is optimized through social interaction in the population where idea is not only personal but also social. PSO depends on the rule that every solution is illustrated as a particle in the swarm. Every particle has a position in the search space, which is represented through a vector. Itextendslarge range of statistical and big data depends on models to preciselyclassify purchaser churn. The PSO focused to improve the capability of identifying customer churn behavior. Particles move in the search space to search for the optimal solutions. Therefore, each particle has a velocity and during the movement, each particle updates its position and velocity according to its own experience and that of its neighbors. The best previous position of the particle is recorded as the personal best pbest, and the best position attained through the population so far named asgbest. Depends on pbestand gbest update, PSO is searching for the optimal solutions through updating the velocity and the position of every particle. PSO is initiallymeasured for searching multidimensional continuous spaces. Every feature subset can be assumed as a point in feature space [22]. The optimal point is the subset with least length and highest prediction accuracy. The initial swarm is disseminatedarbitrarily over the explore space, every particle consumes one position. The objective of particles is to fly to the best position. Via passing the instance, their position ismodified through communicating alongwith each other and they seek around the local best and global best position. It should converge on high-quality, probably optimal, positions because they have examinationcapability that equips them to execute feature and determine optimal subsets. Therefore it provides best on churn data providing effective and faster results. The method easily suffers from the partial optimism, which causes the less accuracy on given churn dataset. For satisfactory results, tuning of input parameters and experimenting with various versions of the PSO method is sometimes necessary. From the meta-heuristic algorithms, the issues are handled such as processing time delay and inaccuracy of churn prediction results. The PSO is not ensured optimal position in large churn dataset and hence it leads inaccuracy prediction results. FFA has problem with light intensity thus it provides lower optimal results. This research proposed big data technique to be implemented by Spark 2.0 on Hadoop. HFFPSO is proposed to provide faster and more accurate results.

Hybrid Firefly algorithm with PSO (HFFPSO) based classification

To overcome the limitations of FFA & PSO for churn prediction, HFFPSO algorithm is introduced in which PSO algorithm is combined with FFA. PSO is used with FFA to increase convergence and also to enhance its capability for not falling into the local minimum. It is used to increase the overall efficiency and effectiveness of the churn prediction accuracy by using optimal solutions. A new hybrid model is used to predict customers with high propensity to churn, profiling the reason of churn and examining the gap between the churn decision and the deactivation time. It improves the TPR and FPR for more accuracy by generating the optimal positionsusing best fitness function values in PSO. It is hybrid with FFA to improve the light intensity and brightness for adjusting the firefly behavior. It enhances global search and generates new churn behavior results through determination of best firefly moves. The proposed HFFPSO for churn prediction is shown in Fig 1.

Real fireflies produce a short and rhythmic flash that helps them in attracting (communicating) their mating partners and also serves as protective warning mechanism. FA formulates this flashing behavior with the objective function of the problem to be optimized. The light intensity of each firefly determine its brightness and hence its attractiveness. Attractiveness of the firefly is calculated using equation 1.

$$\beta\,(r) = \beta_0 e^{-\gamma r_{ij}} \tag{1}$$

where$r_{ij} = d(x_i, x_j)$, a Euclidean distance between two data points iand j. In general, $\beta_0 \in [0, 1]$, describes the fitness value o distance at $r = 0$, i.e., when two data points are found at the same point of search space S. The value of $\gamma \in [0, 10]$ determines the variation of fitness value with increasing distance from communicated data points. It is basically the light absorption coefficient and generally $\gamma \in [0, 10]$ could be suggested [26]. The movement of the firefly i in the space which is attracted toward another firefly j is defined by using equation 2.

$$X_i = x_i + \beta_0 e^{-\gamma r_{ij}} + \alpha(\text{rand} - \tfrac{1}{2}) \tag{2}$$

Where α is the randomization parameter in interval [0, 1] and rand is random number generator with numbers uniformly distributed in range [0, 1]. Parameter γ is controls the variation in attractiveness and define convergence.

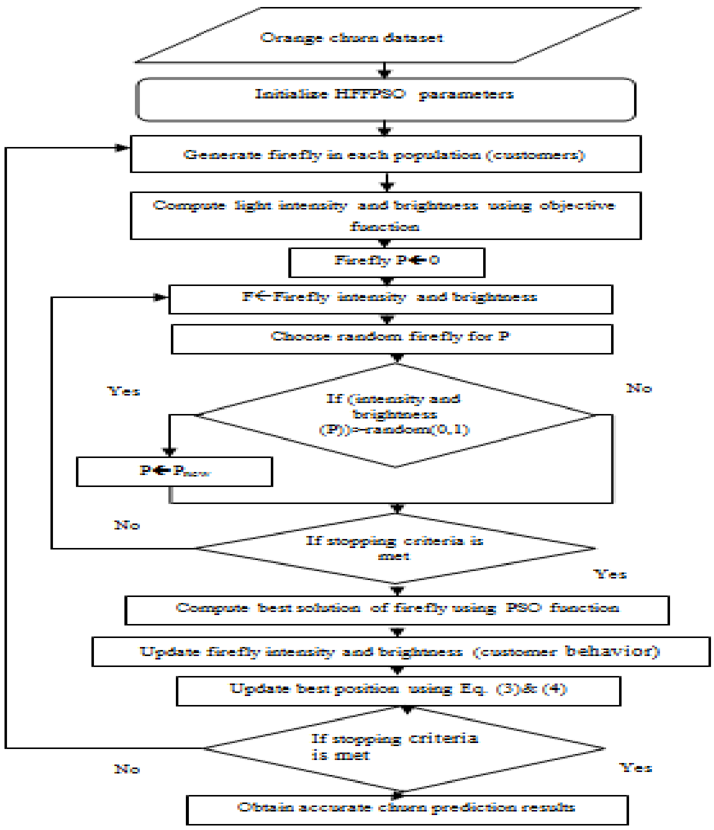

Fig 1 HFFPSO based Classification Architecture

Each firefly-particle updates its own position and velocity according to formula (3) and (4) in every iteration.

$$v_{id}^{k+1} = \omega v_{id}^{k} + c_1\gamma_1\left(p_{id}^{k} - x_{id}^{k}\right) + c_2\gamma_2\left(p_{gd}^{k} - x_{id}^{k}\right) + \alpha\left(\text{rand} - \tfrac{1}{2}\right) \qquad (3)$$

$$x_{id}^{k+1} = \begin{cases} 1 & s\left(v_{id}^{k+1}\right) > rand\ (0,1) \\ 0 & \text{else} \end{cases} \qquad (4)$$

where the $s\left(v_{id}^{k+1}\right)$is the sigmoid function S(vid) = 1/(1 + exp(−v_{id})),i = 1, 2, 3 ... m, m is the number of particles in the swarm, v_{id}^{k} and x_{id}^{k}stand for the velocity and position of the ith particle of the kth iteration, respectively. p_{id}^{k}denotes the previously best position of particle i, p_{gd}^{k} denotes the global best position of the swarm. ω is the inertia weight, c1 and c2 are acceleration constants (the general value of c1 and c2 are in the interval [0 2]), γ1 and γ2 are random numbers in the range [0 1].

Algorithm: HFFPSO

```
Initializing a population with N individuals (churn customers)
Set algorithm factors: Objective function of f(x), from, where x = (x_1,........,x_d)^T
Where x_1,........,x_d is churn customers on large dataset
Generate primary population of fireflies or x_i (i = 1,2,...,n)
x_i is behavior of the new and old customers
Describe light intensity of I_i at x_i via f (x_i)    from equation(p)
While(t < MaxGen) do
        For i = 1 to n(all n fireflies)
        For j = 1 to n(all n fireflies)
        If (I_j > I_i)f_i towards f_j
        End if
        Attractiveness vicissitudes with distance 'r' via Exp[− r2]
        Estimation novel solutions and study light intensity
        End for j
        End for i
Construct new position of firefly (customer behavior) solution
Evaluate the fitness of the new firefly solution which is directly proportional to its brightness
If the fitness value is better than its personal best (pBest)
        Set current value as the new pBest
End if
Choose the particle with the best fitness value of all as gBest
For each particle
        Calculate particle velocity and Update particle position using Eq.(3) & (4)
        Update the best solutions of firefly
        Update the pbest and gbest
End for
End
```

HFFPSO is used toidentify a global optimum of a given objective function. In this approach,the firefly intensities are considered as the objective functionand the requirement for the algorithm is to identify the firefly withmaximum intensity and brightnessto evaluate the churn big data. It identifies the firefly with maximum intensity and brightnessusing best position and particle solutions which is to predict the churn prediction performance on huge data. Initialize the number of churn customers, behavior of new and old customers in given telecom dataset. In FFA, the PSO is combined by best position of PSO which is to increase the brightness and intensity value for providing best churn prediction results. The objective function is calculated using (2) and (3) to update new solutions and it develops the profits in telecom industry through the selection of useful customers. The classifier based HFFPSO model is used to predict the accurate churn behavior information from the large churn orange dataset.

The efficiency of churn prediction model, based on HFFPSO algorithms relies on learning acquired through the orange dataset. The appropriately dataset helps the hybrid model to attain desirable performance. Telecommunication companies archive data by acquiring a lot of information about customers. This HFFPSO approach finds and ranks the most informative instances given churn dataset to develop an optimal prediction result. In this way HFFPSO evolves an optimal prediction that helps in building an effective churn prediction model.

IV. Experimental Result

In this research, the orange dataset is implemented using Spark 2.0 on Hadoop (version 2.7) tool. It's suitable for different users, beginner data mining users and users who prefer scripting interfaces. In this research, large orange dataset is considered to evaluate by using PSO, FF and HFFPSO. The orange dataset contains five main properties such as attribute density, number of records, missing values, number of numerical attributes and number of categorical attributes. The attribute density is 15000, number of records is 50000, missing values 60%, number of numerical attributes is 14740 and number of categorical attributes is 260 shown in Table 1. The comparisons are made in terms of ROC, PR, accuracy, f-measure, true positive rate, false positive rate and execution time.

Table 1: Orange Dataset

Property	Large Orange dataset
Attribute density	15000
No of Records	50000
Missing values	60%
No of Numerical Attributes	14740
No of Categorical Attributes	260

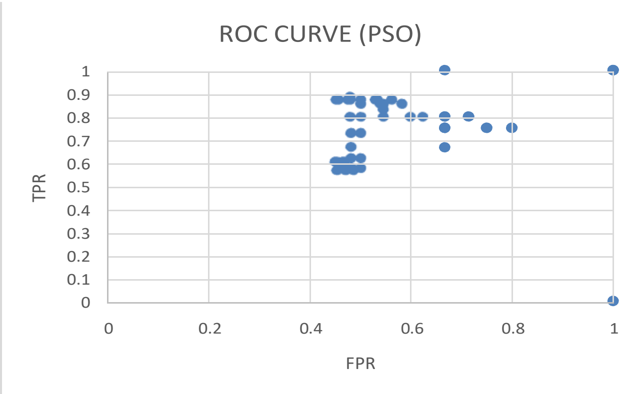

Fig 2 ROC (PSO)

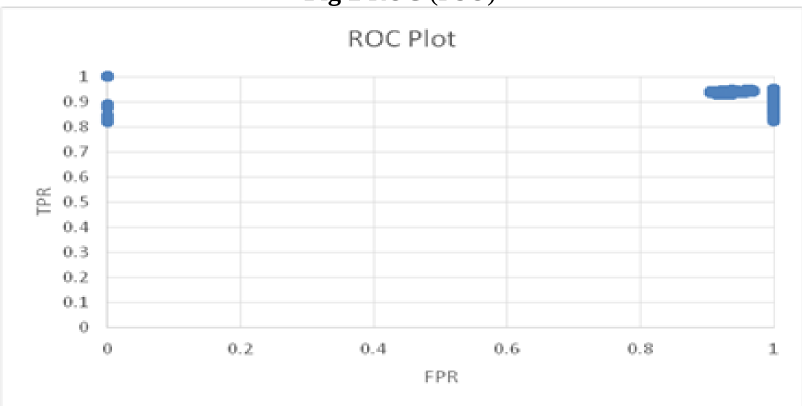

Fig 3 ROC (FFA)

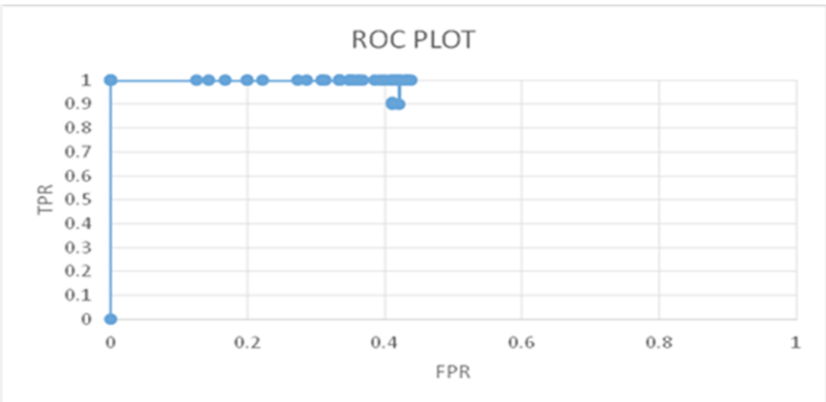

Fig 4 ROC (HFFPSO)

From the above fig 2, 3 and 4, the ROC curve for PSO, FF and HFFPSO is illustrated. The proposed HFFPSO algorithm provides better ROC compare than PSO and FF algorithms. The proposed HFFPSO shows better ROC value than the existing methods which indicates more accurate prediction result in churn dataset.

Precision vs. Recall

Precision is explained as the ratio of the true positives contrary to both true positives and false positives outcomes for imposition and real features. Recall value is calculated on the root of the data retrieval at true positive forecast, false negative.

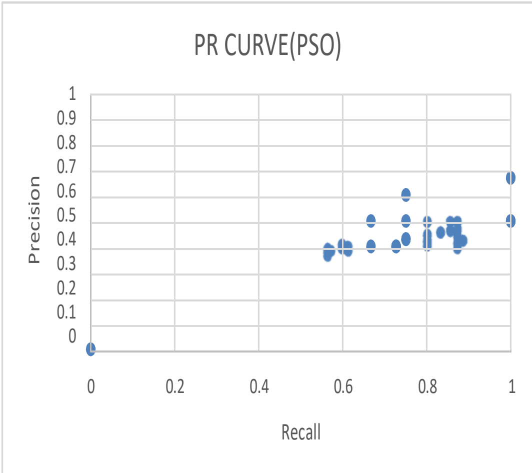

Fig 5 PR Curve- PSO

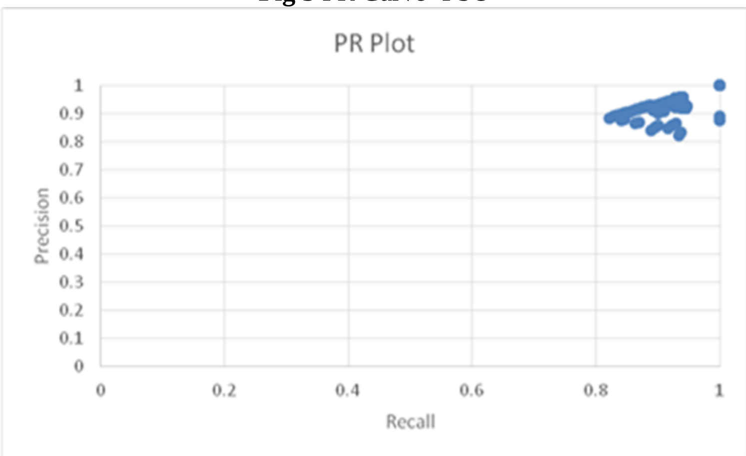

Fig 6 PR Curve- FF

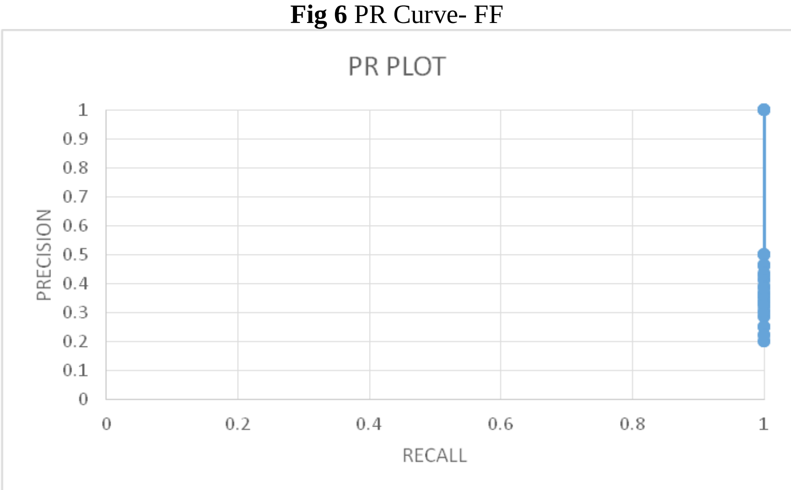

Fig 7 PR Curve- HFFPSO

From the above fig 5,6 and 7, the PR curve forPSO, FF and HFFPSO are illustrated. The proposed HFFPSO algorithm provides better valuescompare than PSO and FF algorithms. The low PR levels attributes to the selection probabilities and the absence of data (sparse nature of data) in the large churn dataset. The proposed HFFPSO shows higher PR than the existing PSO and FFA algorithms.

Accuracy: Accuracy is well-defined as the complete accuracy of the model and is evaluated as the sum of definite classification factors

$$\text{Accuracy} = \frac{\text{True positive+True negative}}{\text{True positive+True negative+False positive+False negative}} \qquad (5)$$

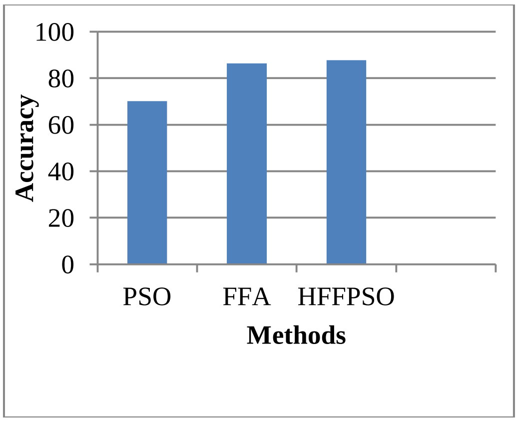

Fig 8 Accuracy

From the above Fig 8, it can be observed that the comparison metric is evaluated using existing and proposed method in terms of accuracy. For x-axis the methods are taken and in y-axis the accuracy value is plotted. The existing method provides lower accuracy whereas the proposed system provides higher accuracy for the given churn dataset. The proposed feature selection HFFPSO algorithm selects best customer churn behaviorfeatures. The result proves that the proposed system attains higher churn prediction results using HFFPSO algorithm. The proposed algorithm shows higher accuracy by 17.67% and 1.41% than the PSO and FFA respectively. It predicts the optimized churn behaviour results by generating the more relevant information. Thus the proposed HFFPSO algorithm is superior to the previous the PSO and FFA methods.

True Positive Rate (TPR)

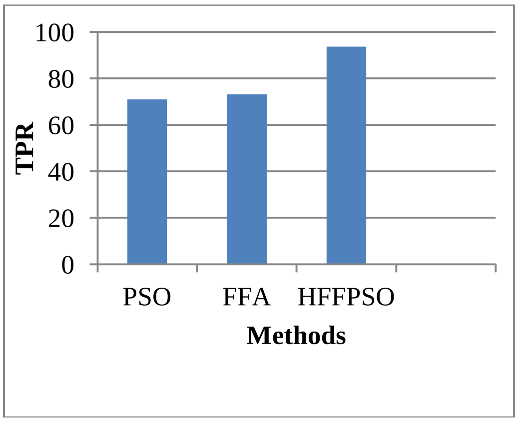

Fig 9 TPR

From the above Fig 9, it can be observed that the comparison metric is evaluated using existing and proposed method in terms of TPR. For x-axis the methods are taken and in y-axis the TPR value is plotted. The existing method provides lower TPR whereas the proposed system provides higher TPR for the given churn dataset.The result proves that the proposed system attains higher churn predictionresults using HFFPSO algorithm. The proposed algorithm shows higher TPR by 22.7% and 20.57% than the PSO and FFA respectively. It predicts the optimized churn behaviour results by generating the true positive features for accurate decision making. Thus the proposed HFFPSO algorithm is superior to the previous the PSO and FFA methods in terms of true customer samples.

False Positive Rate (FPR)

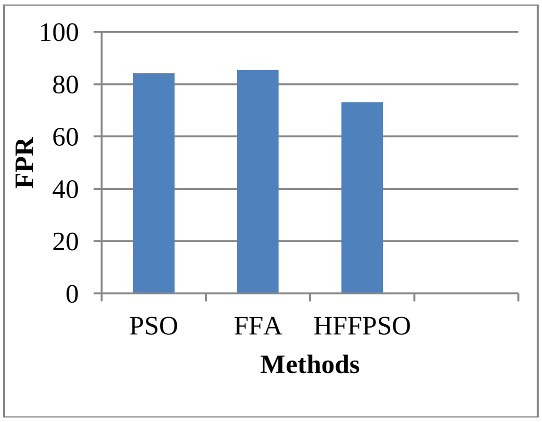

Fig 10 FPR

From the above Fig.10, it can be observed that the comparison metric is evaluated using existing and proposed method in terms of FPR. For x-axis the methods are taken and in y-axis the FPR value is plotted. The proposed algorithm shows lower FPR by 11.08% and 12.34% than the PSO and FFA respectively. It predicts the optimized churn behaviour results by reducing the false information significantly. The proposed HFFPSO algorithm is superior to the previous the PSO and FFA methods.

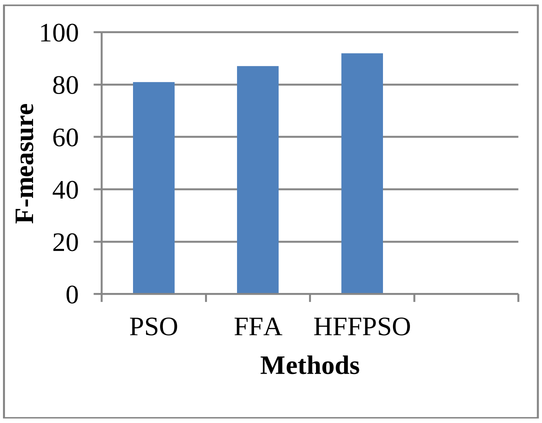

Fig 11 F-measure

From the above Fig 11, it can be observed that the comparison metric is evaluated using existing and proposed method in terms of F-measure. For x-axis the methods are taken and in y-axis the F-measure value is plotted. The proposed HFFPSO prediction model discovers frequent churn prediction and higher f-measure results than existing prediction algorithms of PSO and FFA.The proposed algorithm shows higher FPR by 11% and 5% than the PSO and FFA respectively. It predicts the optimized churn behaviour results by generating the more relevant information. The proposed HFFPSO proves that the better customer churn prediction results.

Execution time

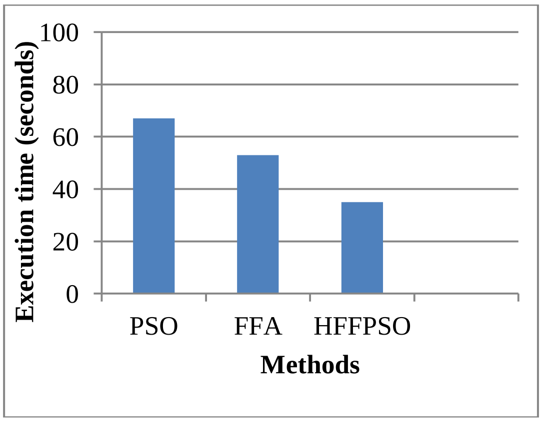

Fig 12Execution time

From the above Fig 12, it can be observed that the comparison metric is evaluated using existing and proposed method in terms of execution time. For x-axis the methods are taken and in y-axis the execution timeisplotted.The proposed HFFPSO algorithm takes less execution timeto predict the customer churn when compared to existing system. The proposed algorithm shows lower time complexityby 32 (sec) and 14 (sec) than the PSO and FFA respectively. It predicts the optimized churn behaviour results by increasing faster convergence through the optimal feature selection.The HFFPSO algorithm is reportedly working efficiently for large churn dataset and it's much faster than PSO and FFA algorithms.

V. Conclusion

Customer churn is one of the mounting issues of today's rapidly growing and competitive telecom sector. Huge size of churn data is evaluated by using the hybrid meta-heuristics algorithms. This research proposedbig data technique to be implemented by Spark 2.0 on Hadoop.HFFPSO is introduced and the prediction levels of the algorithm are to be analyzed. Due to the highly parallelizable nature of Spark and Hadoop, the algorithm can have more processing cycles. Hence the proposed approach HFFPSO improves the accuracy and processing time considerably. The result proves that the proposed accuracy is higher by 17.67% and 1.41% compared to PSO and FFA while also the false positive rate is very low. Thus, it is concluded that the proposed HFFPSO algorithm provides high accuracy of customer churn prediction. In the near-future, large datasets can be effectively processed using this churn prediction model.

Refereences

[1]. Y. Huang, B. Huang, and M. T. Kechadi, "A rule-based method for customer churn prediction in telecommunication services," in Advances in Knowledge Discovery and Data Mining, Springer, 2011, pp. 411–422.
[2]. Idris and A. Khan, "Customer churn prediction for telecommunication: Employing various various features selection techniques and tree based ensemble classifiers," in Multitopic Conference (INMIC), 2012 15th International, 2012, pp. 23–27.
[3]. M. Kaur, K. Singh, and N. Sharma, "Data Mining as a tool to Predict the Churn Behaviour among Indian bank customers," Int. J. Recent Innov. Trends Comput. Commun. vol. 1, no. 9, pp. 720–725, 2013.
[4]. V. L. Miguéis, D. Van den Poel, a. S.Camanho, and J. Falcão e Cunha, "Modeling partial customer churn: On the value of first product-category purchase sequences," Expert Syst. Appl., vol. 39, no. 12, pp. 11250–11256, Sep. 2012.
[5]. Coussement K. and D. V. Poel, "Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques," Expert Syst. Appl., vol. 34, no. 1, pp .313–327, 2008.
[6]. Verbeke W., K. Dejaeger, D. Martens, J. Hur, and B. Baesens, "New insights into churn prediction in the telecommunication sector: A profit driven data mining approach," Eur. J. Oper. Res., vol. 218, no. 1, pp. 211–229, 2012.
[7]. Huang, Bingquan, MohandTaharKechadi, and Brian Buckley. "Customer churn prediction in telecommunications." Expert Systems with Applications 39, no. 1 (2012): 1414-1425.
[8]. Bi, Wenjie, MeiliCai, Mengqi Liu, and Guo Li. "A big data clustering algorithm for mitigating the risk of customer churn." IEEE Transactions on Industrial Informatics 12, no. 3 (2016): 1270-1281.
[9]. Shoro, Abdul Ghaffar, and Tariq Rahim Soomro. "Big data analysis: Apache spark perspective." Global Journal of Computer Science and Technology 15.1 (2015).
[10]. Jadhav, R.J. and Pawar, U.T., "Churn prediction in telecommunication using data mining technology", International Journal of Advanced Computer Science and Applications, 2(2), 2011.
[11]. Huang, B., Kechadi, M.T. and Buckley, B., "Customer churn prediction in telecommunications", Expert Systems with Applications, 39(1), pp.1414-1425, 2012.
[12]. Kirui, C., Hong, L., Cheruiyot, W. and Kirui, H., "Predicting customer churn in mobile telephony industry using probabilistic classifiers in data mining", IJCSI International Journal of Computer Science Issues, 10(2), pp.1694-0814, 2013.
[13]. Dalvi, Preeti K., Siddhi K. Khandge, Ashish Deomore, Aditya Bankar, and V. A. Kanade. "Analysis of customer churn prediction in telecom industry using decision trees and logistic regression." In Colossal Data Analysis and Networking (CDAN), Symposium on, pp. 1-4. IEEE, 2016.
[14]. Yihui, Qiu, and Zhang Chiyu. "Research of indicator system in customer churn prediction for telecom industry." Computer Science & Education (ICCSE), 2016 11th International Conference on. IEEE, 2016
[15]. Ahmed, Ammar AQ, and D. Maheswari. "Churn prediction on huge telecom data using hybrid firefly based classification." Egyptian Informatics Journal(2017).
[16]. Yang, Xin-She. "Flower pollination algorithm for global optimization."Unconventional computation and natural computation. Springer Berlin Heidelberg, 2012. 240-249.
[17]. Karthikeyan, S., and T. Christopher. "A Hybrid Clustering Approach using Artificial Bee Colony (ABC) and Particle Swarm Optimization."International Journal of Computer Applications100.15 (2014).
[18]. Ragins, Edna Johnson, and Alan J. Greco. "Customer relationship management and e-business: more than a software solution." Review of Business 24.1 (2003): 25.
[19]. Umayaparvathi, V., and K. Iyakutti. "Attribute selection and Customer Churn Prediction in telecom industry." Data Mining and Advanced Computing (SAPIENCE), International Conference on. IEEE, 2016
[20]. Amin, Adnan, et al. "Churn prediction in telecommunication industry using rough set approach." New Trends in Computational Collective Intelligence. Springer International Publishing, 2015. 83-95.
[21]. Abdelaziz, A. Y., S. F. Mekhamer, M. A. L. Badr, and M. A. Algabalawy. "The firefly metaheuristic algorithms: Developments and applications." International Electrical Engineering Journal (IEEJ)6, no. 7 (2015): 1945-1952.
[22]. Bird S and Xiaodong Li, "Enhancing the robustness of aspeciation-based PSO," IEEE Congress on Evolutionary Computation, 2006, pp. 843-850.

IOSR Journal of Computer Engineering (IOSR-JCE) is UGC approved Journal with Sl. No. 5019, Journal no. 49102.

Ammar A. Q. Ahmed. "Churn Prediction on Huge Telecom Data Using Hybrid Firefly-Particle Swarm Optimization Algorithm Based Classification." IOSR Journal of Computer Engineering (IOSR-JCE) , vol. 19, no. 4, 2017, pp. 30–39.