Research Article Available online www.ijsrr.org ISSN: 2279–0543

International Journal of Scientific Research and Reviews

Role of Evolutionary Computing Techniques for Knowledge Discovery: A Survey

Mehta Mala[1*] and Chauhan Narendra[2]

[1]Dept. of Information Technology, Sardar Vallabhbhai Patel Institute of Technology (GTU), Vasad-388306, GJ, IN
[2]Dept. of Information Technology, A D Patel Institute of Technology, V.V.Nagar-388120, GUJARAT, INDIA.

ABSTRACT

Evolutionary computing is an exciting development about building, applying and studying algorithms based on Darwin principle "Survival of fittest" and natural selection. Its root lies in evolutionary programming. Genetic Algorithm and Evolutionary strategies. Genetic programming is also considered a part of evolutionary computing. The techniques are used for optimization and efficient search. Data mining techniques have been widely used to mine knowledgeable information from databases. Due to large data collected into database, it is interesting to find out patterns from database, which is useful and understandable as well. The present article provides a survey which focuses on applicability and use of evolutionary computing techniques in field of data mining. Major two application areas which are also considered, is medical data mining and education data mining. An adequate bibliography is also included.

KEYWORDS: Evolutionary Algorithm, Data Mining, Knowledge Discovery, Classification.

***Corresponding author**

Prof. Mala Mehta H.

Dept. of Information Technology, SVIT (GTU),

Vasad-388306, GUJARAT, INDIA.

Email: mhmehta.svit@gmail.com, Mob No – 9879288381

INTRODUCTION

As digital revolution is going on, it has made digitized information easy to store and capture as well. Large information is collected and saved in databases as development in all areas related to computerization is done. Data is stored at very fast rate in today's world but to evaluate data faster techniques are needed and no ad-hoc measures can be used to find any statistical conclusion and findings. Telecommunication networks, manufacturing and production, World Wide Web, Medicine and Education fields are only few examples of domains where large amount of data are stored in databases for analysis purpose. Data is of no special use if it is not processed and analysed properly to find out knowledge form it. In early days, typically manual processes were applied to analyse data. However, such traditional methods failed completely when huge amount of data was gathered. Databases containing number of data in order 10^9 and dimension in order 10^3 are becoming increasingly common. When the scale of data manipulation, exploration and inference goes beyond human capacities, people look to computing technologies for automating the process[1]. Intelligent data analysis methods are a necessity of today's world that discovers useful knowledge. KDD is the term that refers to knowledge discovery in databases. Data mining is considered just a part of this process. KDD focuses on overall process of knowledge discovery from large volumes of data, including the storage and accessing such data, scaling of algorithms, interpretation and visualization of results, modelling and support of overall human machine interaction[1]. Data mining is considered as a form of knowledge discovery only in which specific data sets which is of interest are gathered and studied carefully for analysis to find out useful patterns and discovering knowledge out of them.

Evolutionary computing techniques are an exciting development in Computer Science. It amounts to building, applying and studying algorithms based on the Darwinian principles of natural selection[2]. Evolutionary computing (EC) has its roots in Evolutionary programming, Evolutionary strategies and Genetic algorithms (GA)[3]. In nineties, these branches were considered as a part of EC. Later on, Genetic programming also joined EC. Due to its powerful working Evolutionary Algorithms have emerged from GA and has become most focused part of EC. Researchers found out this whole new branch of algorithms, which can basically aid in solving combinatorial optimization problems known as evolutionary algorithms[4] but its applications are not just limited to pure optimization problems. Ant-Colony optimization[5],

Artificial-Bee Colony optimization[6] and Swarm Intelligence[7] are evolutionary algorithms developed originally for optimization purpose. Evolutionary algorithms are designed by taking inspiration from the

nature and behaviours of different species like birds flocking, fish schooling and bees' working method. Genetic algorithms are designed by taking idea from the human genetics. It follows Charles Darwin's "Survival of the fittest" kind of approach to achieve the optimal solution. Many evolutionary algorithms have proved successful enough in getting desired results. However, to get most optimal answers hybrid methods can also be proposed which can be a balanced mixture of ideas of two evolutionary algorithms[8].

Evolutionary algorithms can be considered as a branch of approximation algorithms where emphasis is given much on proper heuristics rather than static approach. EAs are also considered as a part of soft computing. Evolutionary Algorithms have played crucial role in various fields of data mining as well. In data mining, gathered data can be of huge amount, after pre-processing, data has to be reduced further to operate upon. Here, problem of feature reduction or instance reduction arises which are handled by EAs. In classification and clustering methods, many EAs have been successfully applied to increase the accuracy. Parameter optimization is also a field where EAs can show potential. In data mining, problem of mixed data is faced where different kinds of data are to be handled; here also EAs have been successfully applied.

Remainder of this paper is organized as follows: Section II describes the Evolutionary computing techniques and data mining in detail. Section III presents literature survey on application of evolutionary computing techniques in medical and education data mining. Section IV concludes the paper.

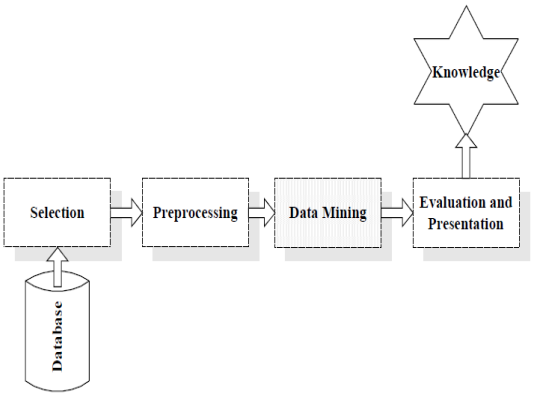

Figure 1: The Knowledge Discovery from Database Process[9]

EVOLUTIONARY COMPUTING TECHNIQUES AND DATA MINING

The term evolutionary computation encompasses a host of methodologies inspired by natural evolution that are used to solve hard problems[10]. Carlos and Moshe[10] explained basic workings of six types of evolutionary algorithms: genetic algorithms, genetic programming, evolution strategies, evolutionary programming, classifier systems and hybrid systems. Basically, evolutionary computation involves study of foundation and applications of computational techniques based on principles of natural evolution. Evolutionary techniques can be perceived as either search methods or optimization methods. Natural evolution has three basic mechanisms: reproduction, mutation and selection. Chromosomes contain genetic information of individual and these mechanisms work on them. Reproduction process introduces new individuals in population. Crossover creates new off springs from two parents combining their genetic material. Mutation introduces small changes into inherited off springs. This whole process follows Charles Darwin's "Survival of the fittest' strategy. It means that only the fittest individuals will survive in the whole process.

In 1975, C.J.Holland first proposed genetic algorithm[3]. Genetic algorithm is an iterative search and optimization method, which works on the human evolution process. Genetic algorithm has taken inspiration from human genetics. Any genetic algorithm has basically four steps to perform. The first step is initial population generation by choosing proper encoding method. After creating the initial population, selection operator chooses the optimal value, based upon the fitness function. Fitness function is application specific objective function which is to be maximized or minimized. Selection operator's job is to select the most optimal population(s) that can go ahead in the evolution and generate best off springs for the future generations. Many different types of selection operators are available in the literature[11]. After selecting the best population, crossover operator operates on the chromosomes that are selected population, to generate new off springs. Crossover operators actually make permutations on the chromosomes that result in new chromosomes. Different crossover operators' efficiencies and effects are different[12]. Mutation operator is included in genetic algorithm to provide population diversity. Mutation operator necessarily gets whole process out of any local minima if realized by the process. Mutation operator is necessary for genetic algorithm so that whole process of getting optimal value does not stick into some valley. Generally, crossover is given more emphasis in traditional genetic algorithm than mutation meaning that mutation is done in fewer amounts on chromosome than crossover operation. Every time, the best fitness value is considered as global minimum or maximum and the process terminates when it reaches to its stopping criteria.

John koza[10] developed a variant of genetic algorithms called genetic programming. Here, instead of encoding possible solutions to a problem as a fixed-length character string, they are encoded as computer programs. Programs here are expressed as parse tress, rather as lines of code. Set of possible internal nodes are called function set, whereas terminal nodes make a set known as terminal set consisting of variables or constants. Evolution in genetic programming proceeds along the general lines of generic evolutionary algorithm, with genetic operators adapted to the tree representation. Reproduction is carried out in both asexual and sexual ways. Asexual reproduction or cloning is similar to elitist selection in GA. Sexual reproduction is crossover that starts out by selecting a random cross over point in each parse tree and then exchanging the sub trees cut from these points, thus producing two offspring trees. Mutation is performed less frequently in genetic programming and is worked by removing a sub tree at a randomly selected point and inserting at that point a new random sub tree. Problem in genetic programming is size of the tress. Cross over operators drastically increases the size of trees and thus create problems for Getting optimal value does not stick into some valley. Generally, crossover is given more emphasis in traditional genetic algorithm than mutation meaning that mutation is done in fewer amounts on chromosome than crossover operation. Every time, the best fitness value is considered as global minimum or maximum and the process terminates when it reaches to its stopping criteria. John koza[10] developed a variant of genetic algorithms called genetic programming. Here, instead of encoding possible solutions to a problem as a fixed-length character string, they are encoded as computer programs. Programs here are expressed as parse tress, rather as lines of code. Set of possible internal nodes are called function set, whereas terminal nodes make a set known as terminal set consisting of variables or constants. Evolution in genetic programming proceeds along the general lines of generic evolutionary algorithm, with genetic operators adapted to the tree representation. Reproduction is carried out in both asexual and sexual ways. Asexual reproduction or cloning is similar to elitist selection in GA. Sexual reproduction is crossover that starts out by selecting a random cross over point in each parse tree and then exchanging the sub trees cut from these points, thus producing two offspring trees. Mutation is worked by removing a sub tree at a randomly selected point and inserting at that point a new random sub tree. Problem in genetic programming is size of the trees. Cross over operators drastically increases the size of trees and thus create problems for memory and speed. Most implementations try to stop trees growing too large and deep but have disadvantage of reducing genetic diversity in larger trees. Evolution strategies are not much studied. They were introduced by Rechenberg and Hans Paul Schweffel. It was proposed as method of solving parameter optimization

problems. Here selection is performed after genetic operators have been applied. Crossover and mutation operators are parts of ES. Two types of crossover operators are: discrete and intermediate. Evolutionary programming (EP) was proposed to develop artificial intelligence. Finite state machines were selected to represent individuals in evolutionary programming as they provide a meaningful representation of behaviour based on interpretation of symbols. EP maintains a population of finite state machines, each one representing particular candidate behaviour for solving problem at hand. The fitness of an individual is calculated by presenting sequentially to FSM the symbols in environment and observing predicted output. Quality of prediction is quantified according to given payoff function. Here no cross over is used but five types of mutation operators are used[10]. Classifiers systems and hybrid systems are also considered as a part of evolutionary computing techniques. However, classifiers systems are considered as evolution-based learning systems. They are considered as a restricted version of classical rule-based systems. It has three main parts: 1) rule and message System performs inference and decides behaviour of system, 2) apportionment of credit system adapts behaviour by credit assignment, and 3) genetic algorithm, which adapts system knowledge by rule discovery. Two main approaches of application of evolutionary techniques in design of rule-based system are: Michigan approach and Pittsburgh approach. In first approach, each individual represents a single rule, and classifier list is represented by whole population. In Pittsburgh approach, genetic algorithm maintains a population of candidate classifier lists, with each individual representing an entire list. In hybrid systems, the attention is drawn to evolutionary-fuzzy hybrid systems more. A prime characteristic of fuzzy logic is its capability of expressing knowledge in human way, allowing rules to be "human-friendly". A fuzzy inference system uses fuzzy logic. System's main components are: 1) fuzzier, 2) inference engine, 3) defuzzifier and 4) knowledge base. Fuzzy rules are described by fuzzy logic. Evolutionary-neural hybrid systems are also considered an important hybrid system which uses neural networks.

Data mining is basically a part of KDD process which is shown in figure 1. Digital revolution has made digitized information easy to capture and fairly inexpensive to store. This has led to many new dimensions open for research. Different fields like education, medical, financial and space engineering gather huge amount of data for processing. Data mining is the process to mine useful patterns out of the data. Data mining gives as output interesting patterns among different variables related to problem in hand. Not only this, data mining is having many different tasks to do like clustering, classification and association rule mining. Classification is the task in which given new object to data mining algorithm

and a target class, algorithm gives yes or no for that meaning that whether that object relates to that class or not. Classification is supervised learning method; it is also known as inductive process. Clustering is unsupervised learning method in which, no target class is given. In clustering, without any supervisor, clusters are created based on similar properties of clusters. Association rule mining is interesting procedure, in which, rule set is created in which deductive rules are generated which shows interesting patterns among data objects. Data mining is most recent area of research and has gain immense popularity among researchers. Data mining combined with evolutionary computing can yield motivating result as evolutionary computing is a field which can be applied with the Darwinian principle of survival of fittest and can find solutions where problem under consideration is considered as hard.

LITERATURE REVIEW

In 2015, T. Santhanam and M.S Padmavathi b proposed an application of k-means and genetic algorithms for dimension reduction by integrating SVM for diabetes diagnosis[13]. Paper starts with a brief introduction about medical mining. It also gives a very brief discussion about GA, K-means and SVM. SVM is a supervised classification method which is considered as an efficient method for classification. Whereas Genetic algorithm, is an evolutionary algorithm, which is based on Darwin's "survival of the fittest strategy", it explores global search space and provides best approximation results by optimization. K-means algorithm is an unsupervised clustering algorithm. Paper further describes a brief literature review. Following by k-means algorithm's detail description. K-means clustering aims at minimizing an objective function known as squared error function. Then authors have described k-means algorithm steps. Next authors have described genetic algorithm in detail. Which says classical model of GA includes three operations: 1) evaluation individual fitness, 2) formation of gene pool by mutation. Support Vector Machine (SVM) is a supervised classification method, it is a classifier which does classification based on hyper planes that separates cases of different class labels in multidimensional space. There are two types of SVM: 1) linear SVM and 2) non-linear SVM. Traditional SVM is based on QP (quadratic programming) but this process is very slow and needs a very large memory. Thus SMO (Sequential Minimal Optimization) is formed to avoid QP routines.

Author has taken Pima Indian Diabetes data set as data source for his work. Which has 8 attributes and 768 instances, then author has proposed his method, in which, working principle of proposed method is described.

Proposed method consists of three steps: 1) data cleaning by replacing null values by mean, 2) instance reduction on output from step 1 by K-means algorithm and 3) applying on output of step 2 GA to reduce features. Finally, SVM is used to classify reduced data set.

In 2013, M.Akhil Jabbar and B.L Deekshatulu Priti Chandra proposed classification of heart disease using k-nearest neighbour and genetic algorithm[14].

Paper starts with brief introduction of data mining, nearest neighbour technique and evolutionary algorithms. Authors also narrated briefly that heart disease is leading cause of death in India. Basic concepts follow introduction section. It includes: 1) K nearest neighbour classifier, 2) genetic algorithm and 3) heart disease. K nearest neighbour is a simple algorithm, which stores all cases and classifies new cases based on simple measure. KNN is of two types; 1) structure less KNN and 2) structure based KNN. Structure less KNN works by dividing data into two parts of training set and testing set. Then author has given K nearest neighbour algorithm. There are generally three cases by which similarities between samples can be measured. Accuracy of K-means algorithm depends on good selection of K. Then GA is described in detail with its applications. One of the applications is in medicine field as well. Flow chart of an evolutionary algorithm is given. Advantages and benefits of genetic algorithm are also narrated. The basic components of GA are also discussed. These are cross over, selection, mutation and fitness functions. Working and flow chart of genetic algorithm is also described. Heart disease and its types are discussed with risk factors of it.

Proposed method is described, which is divided in two parts:

1) First part deals with evaluating attributes using genetic search.
2) Second part deals with building classifier and measuring accuracy of the classifier.

As shown in Figure 2, step 1 to 4 comes under part-1 and steps 5 and 6 come under part-2. Classifiers accuracy is computed as number of samples correctly classified in test data divided by total number of samples in the test data.

Authors have taken various 8 datasets for his work evaluation. These data sets are weather data set, Pima data set, hypothyroid data set, breast cancer data set, liver disorder data set, primary tumor data set, heart stalog data set and lymph data set. Results proved that proposed algorithm is efficient in getting good results.

In 2000, Man Leung Wong, Wai Lam, Kwong Sak Leung Po Shun Ngan and Jack C.Y. Cheng wrote a paper on “Discovering knowledge from medical databases using evolutionary algorithm”[15].

Authors have considered two medical databases for the research. One is fracture database including 6500 records and 8 attributes. Second is scoliosis database, which includes 20 attributes and 500 records. Task of discovering rules from two databases depend on representation of the rule, which is of form "if antecedent then consequence". On the basis of accuracy, rule can be exact, strong or weak. Bayesian network is DAG (directed a cyclic graph), which is a mathematical model and can be used to represent attributes as nodes. To discover rules from database, GGP (Generic genetic programming) is used to represent grammar. A grammar is a representation form of rules. And on this grammar, different GA operators like selection, crossover, mutation and dropping condition is applied. Support and confidence factor are considered as evaluation criteria for rules. Evaluation of these rules is done by support-confidence framework. Fitness function is developed in terms of support and confidence. Support measures the coverage of a rule. Confidence factor is the confidence of the consequent to be true under the condition that the antecedents are also true. Formulas are developed based on support and confidence. A minimum threshold for support is determined. If support is less than this minimum threshold than that confidence factor is ignored. Authors have employed token competition in rule learning approach to search for a set of rules instead of just one rule. Each record in the training set can provide a token. If a rule can match a record, it sets a flag to indicate that token is seized. Weaker rules cannot get token. Fitness score of each individual is modified based on tokens it can seize.

Proposed algorithm

Step 1) load the data set

Step 2) Apply genetic search on the data set

Step 3) attributes are ranked based on their value

Step 4) select the subset of higher ranked attributes

Step 5) Apply (KNN+GA) on the subset of attributes that maximizes classification accuracy

Step 6) calculate accuracy of the classifier, which measures the ability of the classifier to correctly classify unknown sample.

Figure 2: Proposed algorithm[14]

That is why; the authors also modify fitness function. Then authors have narrated learning Bayesian networks from discrete variables based on MDL (minimum description length) and EP (evolutionary programming). Authors have then explained in detail MDL discretization policy. In that also for learning this policy, genetic algorithms are used. Individual representation of chromosomes and genetic operators are discussed.

Experiments are carried out on fracture database and scoliosis database. Results show that some very interesting patterns are found out. This can be very useful to doctors.

In 2013, Quang Hung Do and Jeng-Fung Chen presents Neuro-fuzzy approach in classification of students' academic performance[16]. Accurately predicting students' performance is useful in many different contexts in educational environments. When admission officers review applications, accurate predictions help them to distinguish between suitable and unsuitable candidates for an academic program. The failure to perform and accurate admission decision may result in an unsuitable candidate being admitted to the university. Since quality of educational institutional institution is mainly reflected in research and training, quality of students admitted affects level of institutes. Accurate prediction enables thus managers to improve student academic performance by offering students additional support such as customized assistance and tutoring resource.

Then various classification approaches are discussed which are used to discover knowledge from databases. Support vector machine, naïve Bayes classifier, neural network and decision trees are explained in brief.

Neuro-fuzzy classifier architecture is then described and explained. A neuro-fuzzy classifier is as shown in Figure 3.

A typical fuzzy classification rule Ri demonstrates relation between input feature space and classes. In the NFC, the feature space is partitioned into multiple fuzzy subspaces by fuzzy if-then rules.

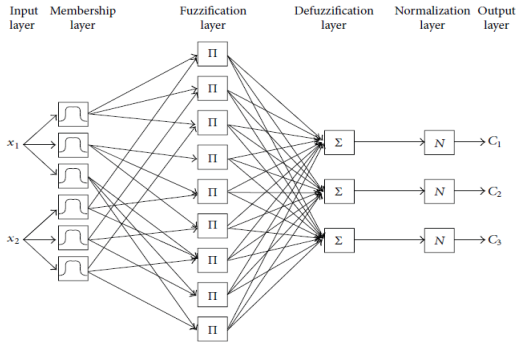

Figure 3: A Neuro-fuzzy classifier[16]

These fuzzy rules can be represented by a network structure. An NFC is a multiplayer feed-forward network consisting of following layers: input, fuzzy membership, fuzzification, defuzzification, normalization and output. In membership layer, membership function of each input is identified. In fuzzification layer, each node generates a signal corresponding to the degree of fulfillment of fuzzy rule for sample x. In defuzzification layer, weighted outputs are calculated each rule affects each class according to their weights. Outputs of network should be normalized. Finally class label of sample is

obtained by picking maximum normalized output value. K-means clustering is applied to obtain initial parameters to form fuzzy if-then rules. K-means clustering method aims to partition the input feature space into a number of clusters in which each data point belongs to cluster with the nearest mean. To optimize the parameters of NFC, SCG algorithm is applied. Cost function is determined from least mean squares of difference between target value and calculated class value. Aim of SCG algorithm is to find the optimal or near optimal parameter for cost function. SCG algorithm is then narrated and explained. Application of proposed method is done on prediction of students' academic performance level. First of all, input and output variables are identified. Input variable were obtained from the admission registration profile. For classification, data set is classified and class labels were identified as "good", "average" and "poor'. Data set was obtained from university of transport technology, Vietnam.

Model was coded and implemented in MATLAB and simulation results were obtained. NFC gives an accuracy of 84.51%, 93.2% and 85.17% for "good", "average" and "poor' classification respectively. This provides 90.03% accuracy for NFC. Results show that NFC can be used to classify students into different groups based on their expected academic performance levels. The model achieved an accuracy of over 90%, which shows that it may be accepted and good enough in comparison of other traditional classifiers.

In 2016, Elham Taherifar and Toraj Banirostam proposed Assessment of Student Feedback from the Training Course and Instructor' Performance through the Combination of Clustering Methods and Decision Tree Algorithms[17]. The main purpose of the application of data mining in the article is to explore the factors affecting the level of satisfaction of students with performance of teaching professor, as well as courses. Data of this article is collected from the UCI website which has been collected in 2013, which corresponds to the data from the survey forms distributed among students of GAZI University in turkey. Authors have the explained basic concepts of performance appraisal system, student evaluation of teaching, educational data mining and the application of data mining in teaching performance appraisal system. Related work is described in section 3. Educational data mining process is explained in detail by authors in which, they have explained basic steps of data mining like preprocessing, data mining and post processing. Proposed model is a hybrid model, which is described in figure 4 below.

Two-step and kohonen algorithms run on 28-item data set. Then using PCA and extracted two components of 28 items, clustering algorithm runs again, and their results are compared. The best method of cluster is chosen, and by running decision tree algorithm on the results of clustering,

satisfaction of students can be predicted. In first step, two-step and kohonen algorithms are run and results are obtained without applying PCA. Performance is evaluated in terms of silhouette value and number of clusters. Then PCA is individually applied to both clustering algorithms and it is found out that two-step method outperforms kohonen in both steps. PCA is proved effective in both algorithms. Then evaluation of clustering results is done by analysis of results of PCA and two-step algorithm and labeling the clusters according to analysis results. Forecasting of student satisfaction is done with Quest decision tree algorithm. Classification rules are obtained by Quest and clusters are obtained by labeling the clusters. Analysis of results of classification clusters are done by which in training set, correct classification is 94.79% and in testing set, classification is 93.77%. Evaluation of work is done be comparing with previous work, in which it is found that proposed method outperforms k-means+ D-tree algorithm in two parameters Accuracy and Error rate.

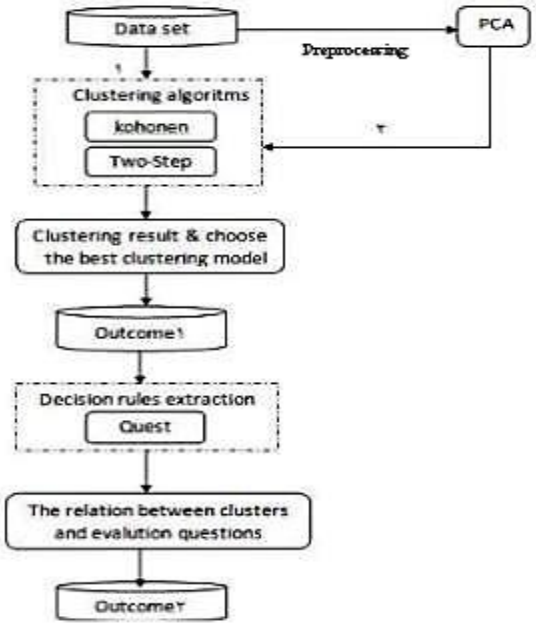

Figure 4: Proposed model by Elham & Toraj[17]

CONCLUSION

Evolutionary computing techniques are powerful techniques that work on basis of Charles Darwin's "survival of fittest" policy. Its applications are immense. But if applied to mining field, it can yield powerful and useful results for researchers. Here in this paper, brief survey of technical papers is given which are concentrating on medical and education field

- In both fields, the usefulness of EC is easily derivable.

- In education data mining, clustering and classification both can be applied for students grouping or students' performance prediction.
- In medical data mining, diabetes, heart disease and many other such diseases can be addressed and novel and helpful results can be found. This paper puts emphasis on the fact that EC is powerful tool to use in mining field.

REFERENCES

1. Mitra S, Pal S K, Mitra P. "*Data Mining in Soft Computing Framework: A Survey"*. IEEE Transactions On Neural Networks. 2002; 13(1):3-14.
2. A. E. Eiben, M.Schoenauer. "*Evolutionary Computation*". Information Processing Letters. Elsevier. 2002; 82(1):1-6.
3. Goldberg D E. "*Genetic Algorithms in search, optimisation & machine learning*" Addison-Wesley Longman Publishing Co: United States; 1989; Available from: URL: https://dl.acm.org/doi/book/10.5555/534133.
4. Mehta M H and Kapadia V V. "*Intelligent Adaptive Mutation Based Genetic Algorithm for Combinatorial Optimization Problems*". IEEE. ICECT. 2012; 3(1):563-66.
5. Fernandes C M, Rosa A C, Ramos V. "*Binary Ant Algorithm*". ACM. 2007; 41-48.
6. Banharnsakun A, Achalakul T, Sirinaovakul B. "*ABC-GSX: A hybrid method for solving the Traveling Salesman Problem*". NaBIC. IEEE. 2010; 7-12.
7. Song M P, Gu G C. "*Research on particle swarm optimisation: a review*". IEEE.2004; 2236-41.
8. Mehta M H and Kapadia V V. "*Intelligent Adaptive Mutation Based Genetic Algorithm With Catfish-PSO effect For Combinatorial Optimization Problems*" IEEE. 2012; 1(1):111-115.
9. Qi Y. "*Intelligent Data Mining Via Evolutionary Computing – A Thesis"*. National University Of Singapore, 2003; Available from: https://core.ac.uk/download/pdf/48625537.pdf
10. Andre's C., Reyes P, Sipper M. "*Evolutionary computation in medicine: an overview*". AIM.ELSEVIER.2000; 19(1):1-23.
11. Razali N M, Geraghty J. "*Genetic Algorithm Performance with different selection strategies in solving TSP*".WCE.2011; 2(1):1134-39.
12. Otman A, Jafar A. "*A comparative study of adaptive crossover operators for genetic algorithms to resolve the travelling salesman problem*". IJCA. 2011; 31(11):49-57.

13. Santhanama T A, Padmavathi M S B. "*Application of K-Means and Genetic Algorithms for Dimension Reduction by Integrating SVM for Diabetes Diagnosis*", PCS. Elsevier. 2015; 47(1):76-83.
14. Jabbar M A, Deekshatulu B L, Chandra P. " *Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm*". PT. Elsevier. 2013; 10(1):85-94.
15. Wong M L, Lam W, Sak K, Po L, Ngan S, Cheng J C Y. "*Discovering Knowledge from Medical Databases Using Evolutionary Algorithm: Learning Rules and Causal Structures for capturing patterns and Causality Relationships*". EIM. IEEE. 2000; 19(4):45-55.
16. Do Q H, Chen J F. "*A Neuro-Fuzzy Approach in the classification of students' Academic Performance".* CIN. Hindawi. 2013; 2013(1):1-7.
17. Taherifar E, Banirostam T. "*Assessment of student feedback from the training course and instructor's performance through the combination of clustering methods and decision tree algorithms".* IJARCSSE. 2016; 6(2):56-64.