INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881; IMPACT FACTOR :6.514(2021); IC VALUE:5.16; ISI VALUE:2.286
Peer Reviewed and Refereed Journal: VOLUME:10, ISSUE:4(3), April:2021
Online Copy Available: www.ijmer.in
Digital certificate of publication:http://ijmer.in/pdf/e-Certificate%20of%20Publication-IJMER.pdf

IMPLEMENTATION OF BIO-INSPIRED ALGORITHMS ON STUDENTS PERFORMANCE PREDICTION IN CLOUD COMPUTING ENVIRONMENT

[1]M. Rajathi and [2]Dr. M. Ramaswami
[1]Research Scholar and [2]Department of Computer Applications
[1&2]Madurai Kamaraj University
Madurai,Tamil Nadu, India

Abstract

The field of Educational Data Mining (EDM) is the most budding area of research which helps to acquire useful knowledge from the existing database of educational academic activities which is used to predict the performance of the students and giving them right guidance to improve themselves for better success in their future. Nowadays enormous amount of data produced by the online educational platforms are handled by the Big Data technologies. The Big data collected, is analyzed in distributed environment to maintain the efficiency, and reduce the computational complexity. In the recent past, the bio-inspired optimization algorithms are acknowledged in machine learning to label the optimal solutions of complex problems in education. In this paper, meta-heuristic technique such as Genetic Algorithm (GA) along with optimization algorithms such as Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO) and their hybrid techniques are applied to analyze the performance of Higher Secondary School(HSC) Students in Madurai District of Tamil Nadu, India. The comparative analysis of GA with hybrid techniques GA-PSO, GA-ACO are analyzed with metrics such as predictive accuracy, speed up ratio, CPU utilization, memory usage, precision, recall and F-Measure in standalone environment and cloud environment. The GA-ACO algorithm shows a better accuracy and GA shows a better speedup ratio and CPU utilization both in local and virtual mode.

Keywords: Big Data, Education Data Mining, GA, Hybrid GA-PSO, GA-ACO, Performance Analysis.

Introduction

In recent years, bio-inspired algorithms are used to solve many real-life complex problems that arise from different fields such as engineering, management, politics, economics, and education. Most of the bio-inspired algorithms are inspired from biological evolution process and swarm behavior. The bio-inspired algorithms are broadly categorized into two categories namely single solution and population based meta-heuristic algorithm. The meta-heuristics maintain the diversity in population and avoid the solutions are being stuck in local optima.Genetic Algorithm (GA) [1], Particle Swarm Optimization (PSO) [2], Ant Colony Optimization (ACO) [3], Spotted Hyena Optimizer (SHO), Emperor Penguin Optimizer (EPO), and Seagull Optimization are some of the more well-known population-based meta-heuristic algorithms (SOA). Among these algorithms, the Genetic Algorithm (GA) is a well-known one that is based on biological evolution. In nature, GA mimics Darwin's theory of survival of the fittest. J.H. Holland proposed the GA in 1992. Many studies have demonstrated the effectiveness of using GA to solve various optimization problems, such as data mining and network traffic control. Indeed, GA has been adopted to solve some problems and was hybridized with other approaches for more enhanced results. Commonly, GA is hybridized with heuristic search approaches such as Particle Swarm Optimization and Ant Colony Optimization. Genetic Algorithm is basically used as a problem-solving strategy in order to provide with an optimal solution [4]. These algorithms are the best way to solve the problem for which limited information is known. It works well in any search space because they form a very general algorithm. The only thing to be known is what the specific situation is where the solution performs very well, and a genetic algorithm will generate a high-quality solution.

Particle swarm optimization (PSO) provides global optimal solution through exploiting the particle's memory and the swarm's memory. PSO becomes one of the evolutionary computation algorithms and swarm intelligence methods due to its ability of adapting in dynamic environments. Ant colony algorithm (ACO) is a heuristic optimization algorithm initially proposed by Italian scholar Dorigo M, etc. in 1991. It is a parallel algorithm in which individuals can cooperate by finding the optimal routes by exchanging and transmitting information continuously. This algorithm can be improved by combining with global optimization algorithms in different environments.

Genetic Algorithm [5] plays a vital role, to handle complex spaces, in many fields such as artificial intelligence, engineering, education, robotic, etc. The genetic processes on the natural evolution principles of populations have been fairly successful in solving problems and produce optimized solution from generation to generation. This is applied in students' data analysis [6] to identify the performance in their curriculum. The performance in higher secondary school education in India is a turning point in the academic lives of all students. School education in India is a two-tier system, the first ten years covering general education followed by two years of higher secondary education. This two-year education, which is also known as Higher Secondary Education, is important because it is a deciding factor for opting desired courses of study in higher education. In fact, the higher secondary education acts as a bridge between school education and the higher learning specializations that are offered by colleges and universities. The performance of the students in the examination relies on mainly three factors namely socio-economic factors, demographic factors and academic

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881; IMPACT FACTOR :6.514(2021); IC VALUE:5.16; ISI VALUE:2.286
Peer Reviewed and Refereed Journal: VOLUME:10, ISSUE:4(3), April:2021
Online Copy Available: www.ijmer.in
Digital certificate of publication:http://ijmer.in/pdf/e-Certificate%20of%20Publication-IJMER.pdf

environment factors [7]. Predicting the performance based on these factors is a challenging task. The prediction will help the students to improve the learning process by identifying the assistance required. The main goal of education has always been better student academic achievement. In past decades, researchers and educators have conducted many researches to determine the factors that influence (positively or negatively) the student achievement in their academic track. It was reported that assessing a student's academic performance would be difficult because student performance is influenced by socioeconomic, psychological, and environmental factors. Based on the average literacy rate in Tamil Nadu, the investigation's geographical scope is limited to Madurai district.

Different types of schools are chosen mainly based on their existence (Urban, Rural, Semi-Urban), Nature of financial existence (Aided, Government, Private), Medium of instruction (Tamil, English), along with sex of the student (Male, Female) and Category of school (Boys, Girls, Co-ed). Nearly 350 school students HSC results are considered for this study and around 38,000 students appear for the exams in a year. This experiment is carried out for three continuous years of result data (2016, 2017 and 2018). Genetic algorithm helps in improving the prediction of students' performance when combined with optimization methods. The results will help the educational institutions to improve the quality of teaching after evaluating the marks achieved by the students in academic career.

Big data analytics is the process of analyzing the vast amount of different data types, to discover the unknown patterns and measurable outcomes. Big data [8] plays a major role in education sector also. It helps in educationist to study the students' data in efficient manner. Big data is getting all kind of data and cloud computing is, what you give to the end user from the collected data. Cloud computing is used in processing big data because traditional processing techniques failed to handle big data. Amazon is one of the best cloud environment providers named as AWS (Amazon Web Services). Genetic algorithms are adaptive to their environments, as this type of method is a platform appearing in the changing environment. Several improvements can be made in order that GAs could be more generally applicable.

In this paper, the result of higher secondary school students of Madurai District is analyzed, and performance of the students is predicted using Genetic algorithm in standalone and also cloud environment. Moreover, Hybrid approaches such as Genetic algorithm with Particle Swarm Optimization and Genetic Algorithm with Ant Colony Optimization are incorporated to improve the efficiency of Genetic algorithm in both the environments.

Literature review

Farissi et al. [9] proposed a Genetic algorithm-based feature selection (GAFS) algorithm for predicting students' performance. Their two-phase experiment was carried out on Kaggle dataset, single classifier without GAFS and single classifier with GAFS. The outcome of their study shows that the proposed GAFS gives a better accuracy when compared to the existing techniques.

Rajalaxmi [11] demonstrated a hybrid GA-BPSO algorithm for predicting the Engineering student's performance. GA when combined with PSO produces a better result for finding the global optimum. The proposed algorithm is validated against Naïve Bayesian and K-Nearest Neighbor algorithms. The results of their study suggest that the Naïve Bayes approach produces a high accuracy with prominent features.

Hasheminejad and Sarvmili [12] designed a prediction model using Particle Swarm Optimization to predict student performance (S3PSO). To predict the students' final outcome, S3PSO extracts the rules that are hidden in the data. When S3PSO is compared to other rule-based algorithms like CART, C4.5, and ID3, they find that S3PSO improves the moodle dataset's fitness value by 31%. Furthermore, S3PSO improves 9 percent of the accuracy value for the moodle dataset when compared to other algorithms such as SVM, NN, KNN, and Nave Bayes.

Sowmiya and Kalaiselvi [13] suggested an algorithm for predicting the instructor's performance using teaching style and student's profile. They used Ant Colony Optimization to optimize the quality of the teacher's content. An ant as intelligent agents improves the efficiency by defining the filtering agent and teaching path agent. Turkey student's evaluation dataset is used for analyzing and ACO outperforms other algorithms such as Particle Swarm Optimization and Cuckoo Search in terms of accuracy.

Cloud computing technologies provide more advantages due to its reliable and scalable services. The main advantage of cloud computing in education sector is it handles wider range of students' data more effectively so that they can be analyzed using bio-inspired algorithms.

Using the Technology Acceptance Model (TAM) as a theoretical foundation, **Zulqurnain Ali et al. [14]** investigated how cloud computing adoption improves student academic performance through personal characteristics and knowledge management paradigm. The current study used a survey approach to recruit 322 university students who are familiar with cloud-based services (G-mail, G-drive, and WhatsApp). By integrating the TAM, it reveals that how knowledge management dimensions and individual characteristics affect cloud computing adoption and student academic performance. Their findings show that perceived usefulness is positively associated with knowledge sharing, learning ability, and knowledge application.

In our work, we incorporate the bio-inspired algorithms, in cloud environment for higher secondary school students' data prediction. The data is collected manually, formed as a big data and implemented in the parallel processing environment.

Bio-inspired algorithms in Big data Analytics

Focusing on the digitization of technical and academic records causes digital libraries to become overburdened. With today's data processing techniques, handling this much data is impossible. Industries and academic institutions are introducing bio-inspired models and algorithms to make big data analytics easier [15]. For analytics, there are three types of bio-inspired algorithms: ecological, swarm-based, and evolutionary algorithms. Genetic algorithm, Simulation Annealing, Cuckoo Search Optimization, and Evolutionary Strategy are examples of ecological algorithms. Artificial Bee Colony, Particle Swarm Optimization, Ant Colony Optimization, and Fish Swarm Optimization are some of the swarm-based algorithms.

Some evolutionary algorithms include invasive weed colony, multi-species optimizer, and biogeography-based optimizer. To predict the performance of higher secondary school students in their final exams, we used the Genetic algorithm, Particle Swarm algorithm, and Ant Colony algorithm. The procedure is carried out in both a standalone and a cloud environment.

Genetic algorithm

Genetic algorithms are one of the tools we can use to apply algorithms to find good, sometimes even optimal, solutions to problems with millions of potential solutions. They use biological processes in software to find answers to problems that have really large search spaces by continuously generating candidate solutions, evaluating how well the solutions fit the desired outcome, and refining the best solutions. When it comes to solving a problem with a genetic algorithm, instead of asking for a specific solution, we can provide the characteristics that the solution must have or the rules that its solution must have in order to be accepted. Genetic algorithms and genetic programming are extremely effective at solving large problems. They do this by taking millions of samples from the search space, making small changes, possibly recombining parts of the best solutions, comparing the resultant fitness to the current best solution, and keeping the best of the two. This process continues until one of the following conditions is met: the known solution is found, a solution that meets all requirements is found, a set number of generations has passed, a set amount of time has passed, etc.

In GA, the search space consists of strings, each one representing the solution to the problem called as chromosomes. The function value of each chromosome is called as fitness value. The population is chromosomes along with its fitness value. Generations are populations generated in the iteration of the GA. Genetic algorithm to find the best solution is as follows:

Procedure:

{

1. [Start] Generate random population of n chromosomes.
2. [fitness] Evaluate the fitness f(x) of each chromosome x in the population.
3. [New Population] Follow the steps below until the new population is complete.
4. [Selection] Select two parent chromosome from the population which has the best fitness value.
5. [Cross Over] With a cross over probability cross over the parents to form new off springs (children).
6. [Mutation] With the mutation probability mutate each off spring by each position in the chromosome.
7. [Accepting] Place off springs in the new population.
8. [Replace] Use newly generated population for the next iteration of the algorithm.
9. [Test] If the end condition is satisfied, stop and return the best solution in current population.
10. [loop] Go to step 2.

}

Selection, mutation, and cross-over operations are all part of the genetic algorithm. In GA, the fitness function is very important. The fitness function specifies the selection criteria for the next generation population. If the problem is one of classification, the fitness function has a component that scores the rule's classification accuracy over a set of training examples. The fitness function is used to assess the procedure's overall performance rather than individual performance. Figure 1 shows a flowchart of the genetic algorithm.

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881; IMPACT FACTOR :6.514(2021); IC VALUE:5.16; ISI VALUE:2.286
Peer Reviewed and Refereed Journal: VOLUME:10, ISSUE:4(3), April:2021
Online Copy Available: www.ijmer.in
Digital certificate of publication:http://ijmer.in/pdf/e-Certificate%20of%20Publication-IJMER.pdf

PSO

PSO is very successful in solving many optimization problems. It is widely used because of its simple conceptual framework. In PSO [16], solutions are represented as particles and the populations of solutions are represented as swarm of particles. The main properties of each particle are position and velocity. Each particle moves to a different position using the velocity. Once a new position is reached, the best position of each particle and swarm are updated. The velocity of each particle is then adjusted based on the experience of the particle. The process is repeated until a stopping condition is reached. The flowchart of PSO process is given in figure 2.

The first step in the PSO process is initialization, which involves creating a swarm of particles. A random position and velocity are assigned to each particle at the start. The fitness value of each particle is updated, and it is compared to the particle's previous best fitness value, as well as the swarm's previous best fitness value, personal best, and global best. If the stopping criteria are not met, the swarm's position and velocity are updated. In the velocity update, the personal best and global best positions, as well as old velocity, are used. The update of velocity and the update of position are two of PSO's most important operations.

The velocity is updated based on three components: the old velocity, experience of an individual particle and experience of the whole swarm. Each term has a weight constant associated with it. For basic PSO algorithm, the number of required constants is three. This might be a significant computational advantage when the population size is large. The updates of velocity and position in PSO also only require a simple arithmetic operation of real numbers.

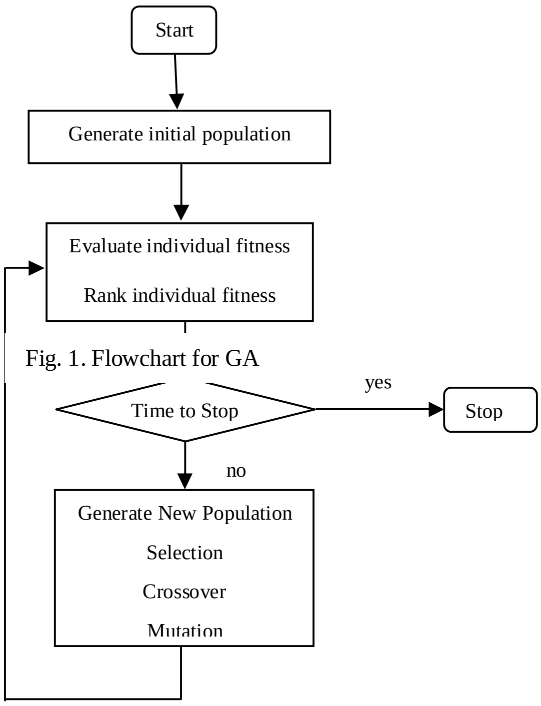

Fig. 1. Flowchart for GA

Hybrid GA-PSO

The hybrid GA-PSO [17] algorithm combines the benefits of PSO's swarm intelligence and the GA's natural selection process. High diversity and low computational cost are the results of hybridization. The GA-PSO algorithm (figure. 3.) begins by generating a random population and then specifies the number of iterations as a parameter. Each solution is a distribution of the entire workflow tasks over the available VMs, and the population represents several solutions to the workflow tasks problem.

The GA algorithm is run with the initialized population for the first half of the defined iterations; that is, if the number of iterations is (n), the GA algorithm will be run (n/2) times. The reason for using (n/2) iteration is to simplify the proposed algorithm, as the GA algorithm's performance is largely determined by the method used to encode solutions into chromosomes and particles, as well as the fitness function's measurement, as well as the population size, or number of iterations.

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881; IMPACT FACTOR :6.514(2021); IC VALUE:5.16; ISI VALUE:2.286
Peer Reviewed and Refereed Journal: VOLUME:10, ISSUE:4(3), April:2021
Online Copy Available: www.ijmer.in
Digital certificate of publication:http://ijmer.in/pdf/e-Certificate%20of%20Publication-IJMER.pdf

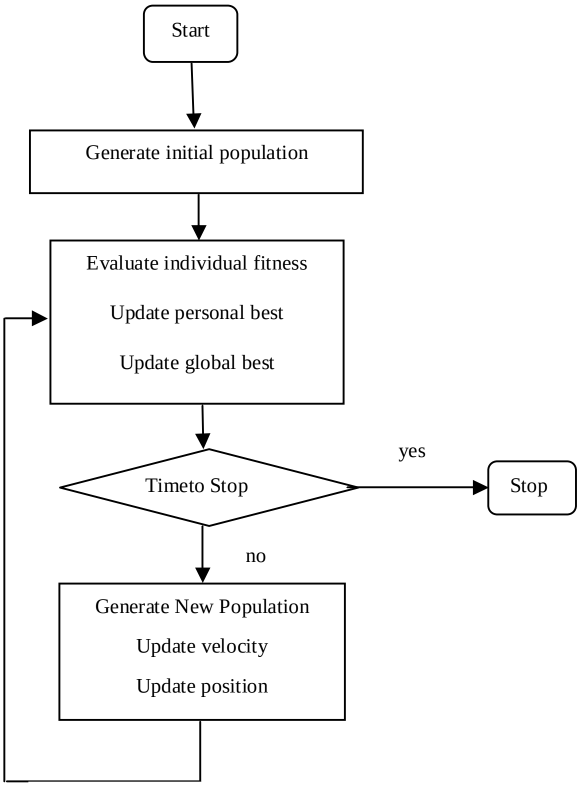

Fig.2. Flowchart for PSO

These parameter values can be tweaked after a few trial runs to see how well the algorithm performs. When the defined number of iterations is divided evenly between the GA and PSO algorithms, the GA-PSO algorithm performed the best in experiments. The solutions are referred to as chromosomes in the GA algorithm, and the chromosomes are gradually improved through the GA operators at each iteration (i.e., selection, crossover, and mutation). At the second half of the defined iterations, the resulting chromosomes are passed to the PSO algorithm. The chromosomes are referred to as particles in the PSO algorithm, and the particles are gradually improved with each iteration. The particle with the lowest fitness value is chosen to represent the problem's solution.

ACO

The ACO [18] is a stochastic optimization method that is based on the social behavior of ant colonies, which try to find the shortest route to food sources. Real ants leave a trail of pheromone behind them, indicating the path they will take. An isolated ant will move at random, but an ant encountering a previously laid pheromone will detect it and decide to follow it with a high probability, bolstering it with more pheromone. The auto-catalytic behavior of a real ant colony is represented by the repetition of the above mechanism, where the more ants follow a trail, the more appealing that trail becomes.

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881; IMPACT FACTOR :6.514(2021); IC VALUE:5.16; ISI VALUE:2.286
Peer Reviewed and Refereed Journal: VOLUME:10, ISSUE:4(3), April:2021
Online Copy Available: www.ijmer.in
Digital certificate of publication:http://ijmer.in/pdf/e-Certificate%20of%20Publication-IJMER.pdf

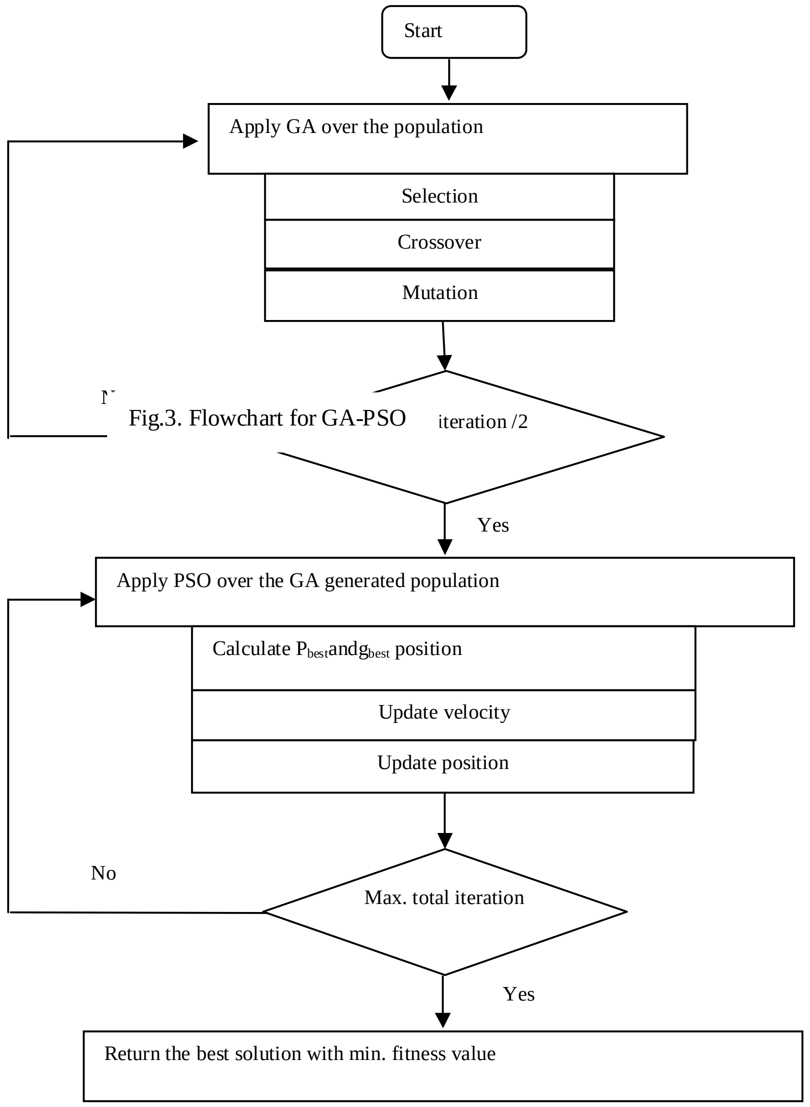

Fig.3. Flowchart for ACO

The idea comes from observing the exploitation of resources of food among ants, in which ants have collectively been able to find the shortest path between to the food. The ACO is implemented as a team of intelligent agents, which simulate the ant's behavior, walking around the graph representing the problem to solve. The flowchart for ACO process is given in figure 4. The problem needs to be represented appropriately, which would allow the ants to incrementally update the solutions through the use of a probabilistic transition rules, based on the amount of pheromone in the trail and other problem specific knowledge.

A heuristic function that measures the quality of components that can be added to the current partial solution and is problem-dependent. A pheromone updating rule set that specifies how to change the pheromone value. Iteratively construct a solution using a probabilistic transition rule based on the value of the heuristic function and the pheromone value.

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881; IMPACT FACTOR :6.514(2021); IC VALUE:5.16; ISI VALUE:2.286
Peer Reviewed and Refereed Journal: VOLUME:10, ISSUE:4(3), April:2021
Online Copy Available: www.ijmer.in
Digital certificate of publication:http://ijmer.in/pdf/e-Certificate%20of%20Publication-IJMER.pdf

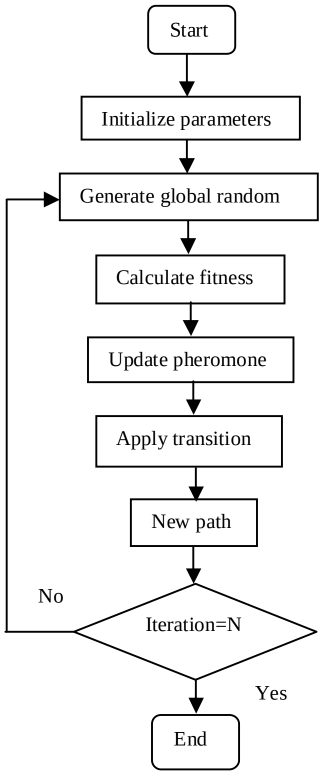

Fig.4. Flowchart for ACO

Hybrid GA-ACO Algorithm

Both GA and ACO [19] are population-based algorithms, but ACO does not require an initial population like GA. ACO is a constructive method in which the ants are guided by the pheromone parameter to find a good solution. The initial pheromone is the same for everyone at first, but after a few iterations, the pheromone levels are updated. As a result, the elements that represent better solutions receive more pheromone than other elements, making them more desirable in the next iteration. The GA solutions are treated as solutions achieved by the ACO in a previous iteration in our hybrid algorithm, and they are used to specify the initial pheromone level of the initial pheromone graph. After that we search for the solution using the ACO algorithm. The flowchart for GA-ACO process is given in figure.5.

Implementation and Results

The dataset used in this study is the results of final exams taken by higher secondary school students in Madurai District of Tamil Nadu. There are 115326 records in the dataset, which has 17 attributes. This information was obtained from the Chief Educational Officer's office. The attributes used in this classification are Year, Taluk name, School location, Type of school, Category of school, Age, Community, Sex, Group, Marks in 6 subjects, Total, and Result. There are two classes for the class variable "result" (Pass, Fail). Labeling the class values is fixed in this two-case problem based on the students' grades (Pass is for students with 40 percent and above and Fail is students for students with below 40 percent). The model is built by converting all predictor variables' categorical values to numeric values. The data was preprocessed before being used in the classification models to ensure that it was ready for analysis.

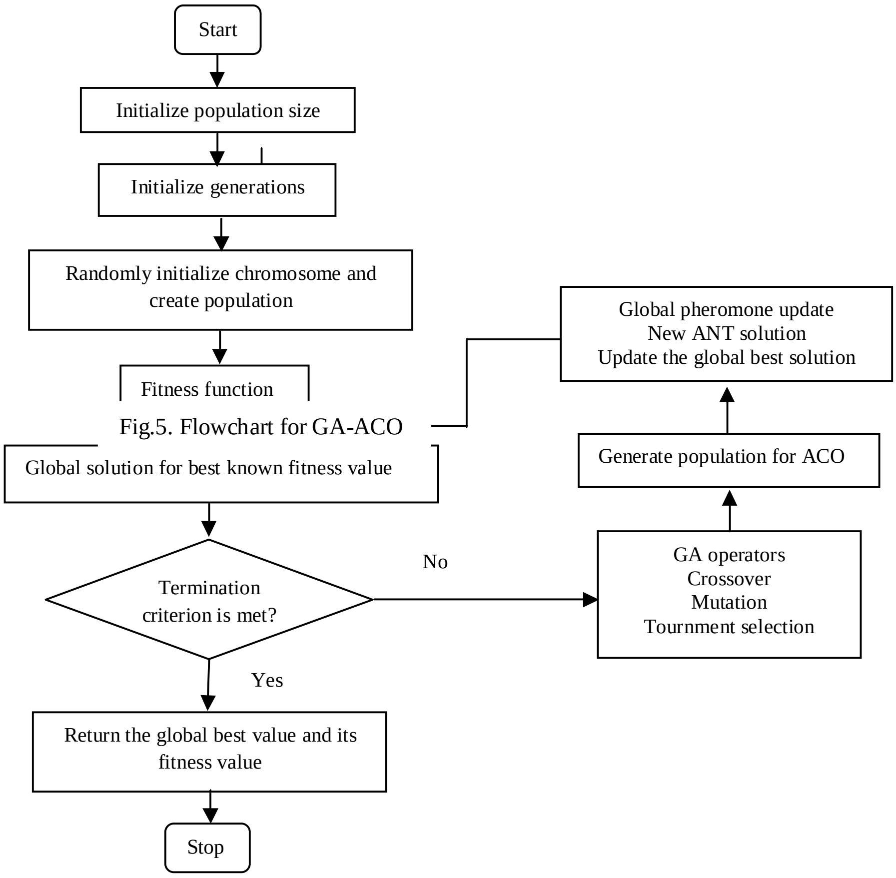

Fig.5. Flowchart for GA-ACO

The dataset had valueless attributes, missing examples, insufficient attribute data types, and other issues that required preprocessing before it could be fed into the analysis phase. The dataset was therefore cleaned, encoded, and the missing values were imputed. In the classification task, the objective is to assign a predefined label or class to a record based on a set of new attributes. An important feature of the classification model is that it is built on a part of the data, known as the training set, which is used to train the models.

After the model is completed, it is used to assign the label to new records where the class attribute is unknown. The classification problem is carried out using bio-inspired algorithms GA, Hybrid GA-PSO, Hybrid GA-ACO both in standalone environment and cloud environment. The cloud computing is achieved through AWS and cloud infrastructure is built around Regions and Availability Zones (AZs) [20]. It consists of many cloud services that we can use in combinations fitted to business or organizational needs. Some of the IAAS by AWS are Amazon EC2, Amazon Elastic Block Store, Auto scaling, Elastic Load Balancing. Amazon EC2 is used for Virtual Machine hosting to provide compute capacity in the cloud.

In our work, we have incorporated Amazon EC2 instance for implementation. Amazon Ubuntu AMI (Amazon Machine Image) based SSD (Solid State Drivers) are created to support the better I/O performance. T2Micro instance is chosen for this work having 1vCPU and 1024Mib of memory. A keypair is generated for this instance and this keypair is used to connect to the virtual machine. The implementation is carried out on both standalone and virtual environments having same configuration.

To evaluate the performance of the hybrid algorithms on the prediction problem in comparison with other problems, we ran extensive experiments on virtual machine on AWS. In GA-PSO, the algorithm starts with 30 random solutions, called population. The mutation rate is as defined as 0.05 at the mutation stage. A single point crossover is defined in the GA phase. In the PSO algorithm, the acceleration coefficients and random numbers are defined to update the velocity and position. The experiments were executed

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881; IMPACT FACTOR :6.514(2021); IC VALUE:5.16; ISI VALUE:2.286
Peer Reviewed and Refereed Journal: VOLUME:10, ISSUE:4(3), April:2021
Online Copy Available: www.ijmer.in
Digital certificate of publication:http://ijmer.in/pdf/e-Certificate%20of%20Publication-IJMER.pdf

nearly 100 iterations to reach the optimal solution. The experiments were carried out 500 times and results are compared. In GA-PSO, the algorithm starts with 30 random solutions, called population. The mutation rate is 0.05- and single-point crossover is defined at GA stage. In ACO algorithm, the numbers of ants are defined to be 20 and 100 ACO generations.

The classification performance of the bio-inspired algorithms GA, GA-PSO and GA-ACO are compared using parameters: predictive accuracy, speedup-ratio, CPU utilization, memory used in each process, precision, recall and F-measure. Predictive accuracy [21] is defined as the percentage of correct predictions for the test data. It can be calculated easily by dividing the number of correct predictions by the number of total predictions.

Speedup ratio [22] is a number that measures the relative performance of two systems processing the same problem. More technically, it is the improvement in speed of execution of a task executed on two similar architectures with different resources. The notion of speedup was established by Amdahl's law, which was particularly focused on parallel processing. However, speedup can be used more generally to show the effect on performance after any resource enhancement. We are testing the effectiveness of a branch predictor on the execution of a program.

CPU utilization [23] is the percentages of allocated EC2 compute units that are currently in use on the instance. This metric identifies the processing power required to run an application on a selected instance. Depending on the instance type, tools in your operating system can show a lower percentage than Cloud Watch when the instance is not allocated a full processor core.

Precision is defined as the fraction of relevant examples (true positives) among all of the examples which were predicted to belong in a certain class.

Recall is defined as the fraction of examples which were predicted to belong to a class with respect to all of the examples that truly belong in the class.

The F-Measure is a way of combining the precision and recall of the model, and it is defined as the harmonic mean of the model's precision and recall. It is used to evaluate binary classification systems, which classify examples into 'positive' or 'negative'.

Table I. Comparison of algorithms in standalone environment

Algorithm	Accuracy	CPU utilization	Memory usage (Mb)	Precision	Recall	F-Measure
GA	0.9816	24.6	1256983	0.96	0.98	0.99
GA-PSO	0.9863	30.6	1494988	0.98	0.99	0.97
GA-ACO	0.9866	31.9	1535526	0.98	0.95	0.96

From the table I, the accuracy of hybrid GA-ACO algorithm is higher than the other two algorithms. The memory occupied by the GA-ACO is high when compared to other algorithms. The CPU utilization is less in GA when compared to other algorithms. The comparison of GA, GA-PSO, GA-ACO in cloud environment is given in the Table II.

Table II. Comparison of algorithms in cloud environment

Algorithm	Accuracy	CPU utilization	Memory Usage (Mb)	Precision	Recall	F- Measure
GA	0.9835	14.6	1196852	0.99	0.96	0.95
GA-PSO	0.9868	20.6	1396852	0.99	0.98	0.97
GA-ACO	0.9891	21.9	1489632	0.95	0.96	0.97

In the cloud, GA-ACO algorithm shows higher accuracy than the other algorithms, just as it does in standalone mode; again, CPU utilization is low with less memory usage in GA. Table III shows the speedup ratio of the algorithms. In our problem, we execute the program for GA with the Ubuntu local machine, which yields an execution time of 40.2 seconds. Next, we execute the program with our Ubuntu cloud machine, which produces an execution time of 36.26 seconds. In both cases the execution workload is the same. Using our speedup formula, we know, our cloud machine has provided a 1.10x speedup over the original. From the table, we find that GA has better speed-up ratio due to the less execution time of the algorithm.

Table III. Comparison of speedup ratio

Algorithm	Execution Time (Local) sec	Execution Time (cloud) sec	Speedup ratio
GA	40.2	36.26	1.10
GA+PSO	60.87	58.59	1.01
GA+ACO	61.39	60.25	1.03

$S=T_{local}/T_{cloud}=40.2/36.26=1.10$ {\displaystyle S_{\text{latency}}={\frac {L_{\text{old}}}{L_{\text{new}}}}={\frac {2.25~\mathrm {s} }{1.50~\mathrm {s} }}=1.5.}

Conclusion

In this paper, bio-inspired algorithms GA, Hybrid GA-PSO, and Hybrid GA-ACO are used to predict student performance in a stand-alone and cloud environment. In both cloud and standalone environments, the algorithms are compared to a variety of intrinsic metrics such as predictive accuracy, speedup ratio, and CPU utilization. AWS's EC2 service is used to implement the cloud. Due to the update of position and velocity in PSO and the pheromone distribution in ACO, the Hybrid algorithms perform better. The GA-ACO algorithm outperforms the other two algorithms in experiments conducted in both standalone and cloud environments. In both environments, the speedup ratio is best for GA, with less CPU utilization.

References

1. Katoch, S., Chauhan, S.S. & Kumar, V. A review on genetic algorithm: past, present, and future. Multimedia Tools Appl (2020).
2. Kennedy J, Eberhart RC (1995), Particle swarm optimization. In: Proceedings of IEEE international conference on neural networks (1995), pp 1942–1948
3. Dorigo M, Birattari M, Stutzle T, Ant colony optimization - artificial ants as a computational intelligence technique. IEEE ComputIntell Mag 1(2006):28–39
4. Saad Alharbi1 and Ibrahim Venkat2, A Genetic Algorithm Based Approach for Solving the Minimum Dominating Set of Queens Problem, Hindawi Journal of Optimization Volume 2017.
5. Lingaraj, Haldurai, A Study on Genetic Algorithm and its Applications. International Journal of Computer Sciences and Engineering, 2016, pp 139-143
6. T.Miranda Lakshmi, A.Martin, V.Prasanna Venkatesan, An Analysis of Students Performance Using Genetic Algorithm, , Journal of Computer Sciences and Applications, 2013, Vol. 1, No. 4, 75-79.
7. RamaswamiMurugesh& R. Bhaskaran, A CHAID Based Performance Prediction Model in Educational Data Mining. International Journal of Computer Science Issues, 7, (2010).
8. Abdullah M. Alghamdi, 2Fahad A. Alghamdi, Enhancing Performance of Educational Data Using Big Data and Hadoop, International Journal of Applied Engineering Research Volume 14, Number 19 (2019) pp. 3814-3819
9. Al Farissi&SamsuryadiSahmin& Halina Dahlan, Genetic Algorithm Based Feature Selection for Predicting Student's Academic Performance, International conference of reliable information and communication,2019, pp-111-117.
10. Xu Zhang, Pan Guo, Hua Zhang and JinYao,Hybrid particle swarm optimization algorithm for process planning, Mathematics 2020, 8, 1745.
11. R R, Rajalaxmi, Feature Selection using Hybrid GA-BPSO for Engineering Students Performance Prediction. International journal of Applied engineering research. 9, 2014, 19945-58.
12. S. M.- H. Hasheminejad and M. Sarvmili, S3PSO: Students' Performance Prediction Based on Particle Swarm Optimization, Journal of AI and Data Mining Vol 7, No 1, 2019, 77-96.
13. J. Sowmiya , K. Kalaiselvi, Predicting Instructor Performance in Higher Education Using Intelligent Agent Systems, EAI Endorsed Transactions on Energy Web 2020, Volume 7 , Issue 30.
14. Zulqurnain Ali &Gongbing, Bi &Mehreen, Aqsa. (2018). Understanding and predicting academic performance through cloud computing adoption: a perspective of technology acceptance model, Journal of Computers in Education, 2018.
15. Gill, Sukhpal Singh &Buyya, Rajkumar, Bio-Inspired Algorithms for Big Data Analytics: A Survey, Taxonomy, and Open Challenges, 2019.
16. Almayan, Hind &Mayyan, Waheeda,Improving accuracy of students' final grade prediction model using PSO,2016, pp 35-39.
17. B.B.Gupta, Ahmad Manasrah M, Hanan Ba Ali, Workflow Scheduling Using Hybrid GA-PSO Algorithm in cloud computing, Wireless Communications and Mobile Computing, Jan 2018.

18. Olympia Roeva , Marcin Paprzycki, StefkaFidanova, Hybrid GA-ACO Algorithm for a Model Parameters Identification Problem, Proceedings of the 2014 Federated Conference on Computer Science and Information Systems pp. 413–420.
19. H. Guangdong, L. Ping and W. Qun, "A Hybrid Metaheuristic ACO-GA with an Application in Sports Competition Scheduling," Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD 2007), Qingdao, China, 2007.
20. R., Tanmay &Borse, Yogita. (2018). Implementation of Cloud computing Service Delivery Models (IAAS, PAAS) by AWS and Microsoft Azure: A Survey. International Journal of Computer Applications. 179. 19-21.
21. Ramaswami. M., “Validating Predictive Performance of Classifier Models for Multiclass Problem in Educational Data Mining” 2014.
22. NewshaArdalani, UrmishThakker, AwsAlbarghouthi, KaruSankaralingam, “A static Analysis-Based Cross-Architecture Performance Prediction Using Machine Learning”, June 2019.
23. KarlMason, MartinDuggan, EndaBarrett, JimDuggan, EndaHowley, “Predicting host CPU utilization in the cloud using evolutionary neural networks”, Future Generation Computer Systems, Volume 86, September 2018, pp. 162-173.
24. G. Vaitheeswaran, and L. Arockiam., “Big Data for Education in Students’ Perspective”, IJCA(0975-8887) and (ICACCTHPA-2014

Filename:	14
Directory:	C:\Users\DELL\Documents
Template:	C:\Users\DELL\AppData\Roaming\Microsoft\Templates\Normal.dotm
Title:	
Subject:	
Author:	Windows User
Keywords:	
Comments:	
Creation Date:	4/16/2021 4:41:00 PM
Change Number:	6
Last Saved On:	4/25/2021 2:01:00 PM
Last Saved By:	Murali Korada
Total Editing Time:	24 Minutes
Last Printed On:	4/29/2021 8:17:00 PM
As of Last Complete Printing	
Number of Pages:	11
Number of Words:	5,499 (approx.)
Number of Characters:	31,346 (approx.)