Journal of Basic and Applied Engineering Research
Print ISSN: 2350-0077; Online ISSN: 2350-0255; Volume 2, Number 2; January-March, 2015, pp. 132-135

http://www.krishisanskriti.org/jbaer.html

Classification Rule Construction Using ABC Algorithm for Business Intelligence

Mrs. R. Sharmila[1] and Ananthi Sheshasaayee[2]

[1]Research Scholar, Manonmaniam Sundaranar University, Tirunelveli Asst. Prof, Department of Applied Science Er. Perumal Manimekalai College of Engineering, Hosur.
[2]Research Supervisor, Manonmaniam Sundaranar University, Tirunelveli Asso. Prof and Head of Computer Science, Quid-e-Millath Govt. College for Women, Chennai
E-mail: [1]sharmi.saravanan@yahoo.in, [2]ananthiseshu@gmail.com

Abstract—*Data mining is "The nontrivial extraction of implicit, previously unknown, and potentially useful information from data." Data mining is an inter-disciplinary field, whose core is at the intersection of machine learning, statistics and databases. A major objective of this work is to evaluate data mining tools in Business applications to develop a tool that can help make timely and accurate decisions. One technique used in data mining is Classification where the desired output is a set of Rules or Statements that characterize the data. Within the rule induction paradigm, the algorithm used is Artificial Bee Colony. Artificial Bee Colony (ABC) is a swarm based meta-heuristic algorithm for optimizing numerical problems which were inspired by the intelligent foraging behavior of honey bees. This proposed work is intended to develop Classification rules to extract data from historical or training data of business data which is developed into patterns relevant for and suitable for quicker analysis, automated processing, thus reducing cost.*

Keywords: *Data mining, Classification, Artificial Bee Colony algorithm, Particle Swarm optimization, foragers.*

1. INTRODUCTION

ABC is a Swarm based meta-heuristic algorithm for optimizing numerical problems. The model consists of three essential components: employed and unemployed foraging bees, and food sources, which is the third component, close to their hive. One technique used in data mining is Classification where the desired output is a set of Rules or Statements that characterize the data. Within the rule induction paradigm, the algorithm used is ABC. To apply ABC, the considered optimization problem is first converted to the problem of finding the best parameter vector which minimizes an objective function. It is the simulation of foraging behavior of honey bees. This proposed work is intended to develop Classification rules to extract data from historical or training data of business data sets which is developed into patterns relevant and suitable for quicker analysis, automated processing.

The minimal model of foraging selection that leads to the emergence of collective intelligence of honey bee swarms consists of three essential components: food sources, employed foragers and unemployed foragers.

- ***Food sources****:* it represents a position of solution of optimization problem, the profitability of food source are expressed as fitness of the solution.
- ***Unemployed foragers****:* there are two types of them, scouts and onlookers. Their main task is exploring and exploiting food source. At the beginning, there are two choices for the unemployed foragers: (i). it becomes a scout–randomly search new food sources around the nest. (ii). It becomes an onlooker–determine the nectar amount of food source after watching the waggle dances of employed bee, and select food source according to profitability.
- ***Employed foragers****:* the honeybees found food source, which also known as the employed bees, are equal to the number of food sources. The employed bees store the food source information and share with others according to a certain probability. The employed bee will become a scout when food source has been exhausted.

2. PROBLEM DESCRIPTION

This proposed work is intended to review Swarm Intelligence (SI), an artificial intelligence technique for machine learning. Swarm intelligence is based on distributive self organized system and is chosen for this paper because of its vast uses and simplicity. The results show that ABC is very competitive in terms of accuracy and that ABC produces significantly simpler (smaller) rule sets, a desirable result in data mining–where the goal is to discover knowledge that is not only accurate but also comprehensible to the user. When we obtain an optimal solution using PSO (Particle Swarm Optimization), PSO is based on a swarm of n individuals called particles. Each particle represents a possible solution to a problem with

N dimensions and its genotype consists of 2*N parameters. The first N of them represents the coordinates of particle position, while the latter N its velocity components in the N dimensional problem space. From the evolutionary point of view, a particle moves with an adaptable velocity within the search space and retains in its own memory the best position it ever reached.

To improve an optimal learning solution of PSO, this paper deals with applying ABC (Artificial Bee Colony). Data mining uses software techniques for finding patterns and regularities in sets of data that are too large to analyze manually. Data mining helps to predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. Data mining is a core step of a broader process of automatic information extraction, called knowledge discovery in databases (KDD) or knowledge discovery.

3. THOUGHTS OF EXPERIMENT

Artificial Bee Colony (ABC) is a heuristic technique suited for search of optimal solutions and based on the concept of swarm. ABC has the ability to effectively face classification of multi-class database instances. The process of knowledge discovery, known as data mining unifies research in fields such as statistics, databases, machine learning and artificial intelligence.

ABC is a Swarm based meta-heuristic algorithm for optimizing numerical problems. The model consists of three essential components: employed and unemployed foraging bees, and food sources, which is the third component, close to their hive. One technique used in data mining is Classification where the desired output is a set of Rules or Statements that characterize the data. Within the rule induction paradigm, the algorithm used is ABC.

To apply ABC, the considered optimization problem is first converted to the problem of finding the best parameter vector which minimizes an objective function. This proposed work is intended to develop Classification rules to extract data from historical or training data of business data sets which is developed into patterns relevant and suitable for quicker analysis, automated processing.

PSO is based on a swarm of n individuals called particles. Each particle represents a possible solution to a problem with N dimensions and its genotype consists of 2*N parameters. The first N of them represents the coordinates of particle position, while the latter N its velocity components in the N dimensional problem space. From the evolutionary point of view, a particle moves with an adaptable velocity within the search space and retains in its own memory the best position it ever reached. But this does not produces simple and best rules as ABC generates.

To improve an optimal learning solution of PSO, this paper deal with applying ABC (Artificial Bee Colony) .The focus of this paper is on supervised learning, more specifically, the classification task of data mining. In classification the knowledge or patterns discovered in the data set can be represented in terms of a set of rules.

A rule consists of an antecedent (a set of attribute-values) and a consequent (class): IF <attrib = value> AND ... AND <attrib = value> THEN<class>.The consequent of the rule is the class that is predicted by that rule. The antecedent consists of a set of terms, where each term is essentially an attribute-value pair.

More precisely, a term is defined by a triple <attribute, operator, value>, where value is a value belonging to the domain of attribute. The operator used in this paper is "=" in the case of categorical/nominal attributes, or "≤" and ">" in the case of continuous attributes. The knowledge being intuitively comprehensible to the user. This is important, because the general goal of data mining is to discover knowledge that is not only accurate, but also comprehensible.

In broad sense, learning is concerned with the algorithms and techniques which allow system to learn. Depending upon how system learns, many categories of algorithms are available including Swarm Intelligence. In Swarm Intelligence, population is made up of agents. These agents interact locally i.e. with each other and to the environment to find the solution but don't have any central authority to control them. So their interactions lead into global behavior of the system. It is obvious that this technique is also inspired by the elements of nature like teamwork of ants, bird flying together, animal moving in heard, foraging behavior of honey bees etc.

III a) Process techniques

- Training Data and Information collection
- Classification Model construction Using ABC Algorithm
- Prediction

Training data and Information Collection

Every transaction in the business industry is often "memorized" for perpetuity. Such transactions are usually time related and can be inter-business deals such as purchases, exchanges, banking, stock, etc., or intra-business operations such as management of in-house wares and assets. Large department stores, for example, thanks to the widespread use of bar codes, store millions of transactions daily representing often terabytes of data. Storage space is not the major problem, as the price of hard disks is continuously dropping, but the effective use of the data in a reasonable time frame for competitive decision-making. It is definitely the most important problem to solve for businesses that struggle to survive in a highly competitive world. The Types of data collected are as follows:

- Operational or transactional data (sales, cost, inventory, payroll, and accounting)
- Non operational data (industry sales, forecast data, and macro economic data),
- Meta data (logical database design or data dictionary definitions).

Classification Model Construction Using ABC

Algorithm

Each tuple/sample is assumed that it belong to a predefined class. The class of a tuple/sample is determined by the class label attribute and the attributes are related to form various conditions.

The training set of tuples/samples is used for model construction. The model is represented as classification rules, decision trees or mathematical Formulae.

The minimal model of foraging selection that leads to the emergence of collective intelligence of honey bee swarms consists of three essential components: food sources, employed foragers and unemployed foragers.

- *Food sources:* it represents a position of solution of optimization problem, the profitability of food source are expressed as fitness of the solution.
- *Unemployed foragers:* there are two types of them, scouts and onlookers. Their main task is exploring and exploiting food source. At the beginning, there are two choices for the unemployed foragers: (i). it becomes a scout–randomly search new food sources around the nest. (ii). It becomes an onlooker–determine the nectar amount of food source after watching the waggle dances of employed bee, and select food source according to profitability.
- *Employed foragers:* the honeybees found food source, which also known as the employed bees, are equal to the number of food sources. The employed bees store the food source information and share with others according to a certain probability. The employed bee will become a scout when food source has been exhausted.

Prediction

Classify future or unknown objects Estimate the accuracy of the model: the known class of a test tuple/sample is compared with the result given by the model accuracy rate = percentage of the tests tuples/samples correctly classified by the model.

Test set should be independent of training set. Accuracy, Speed, Robustness, Scalability, Interpretabilityandsimplicity are the parameters to analyze classification.

The pruned rule set will be used to predict the new data which their classes are unknown. But sometime, one testing data record will be covered by more than one rule for different class. When this happened, the prediction strategy will determine which class should be predicted. There are three main steps for the prediction approach, they have specified as follow:

- Calculate the prediction value for all rules which cover the test data record;
- Accumulate these prediction value according to different possible class;
- Select the class which has the highest prediction value as the final class.

Prediction value = (α × rule fitness value) + (β × rule cover percentage) Where α and β are two weighted parameters associated with rule fitness value and rule cover percentage respectively. The rule cover percentage defines that the proportion of the records which covered by the rule that have the class predicted by the rule (TP). It is calculated by the expression.

Cover percentage =*TP/N* where N is the total number of the records which belong to the predicted class by the rules. The prediction strategy balanced the effect of fitness value and cover percentage for the final predicted class. We need to choose the value of α and β carefully, since they will affect the classification accuracy.

4. PROCESS IMPLEMENTATION

Implementing the application developed in Java is checked by testing using various techniques. First, Planning is the first task in the system implementation. Use the application to different environments in which it should supports Java. This work can be adopted in any business area which expects the accuracy of the decisions taken in the concern. The system should meet the requirements like time and accuracy when it applied for large datasets. The algorithms following below are implemented.

PSO faces classification of multi-class database instances. PSO is based on a swarm of n individuals called particles. Each particle represents a possible solution to a problem with N dimensions and its genotype consists of 2*N parameters. The first N of them represents the coordinates of particle position, while the latter N its velocity components in the Dimensional problem space.

From the evolutionary point of view, a particle moves with an adaptable velocity within the search space and retains in its own memory the best position it ever reached. The process of knowledge discovery, known as data mining unifies research in fields such as statistics, databases, machine learning and artificial intelligence.

IV a) Implementation procedure

The implementation starts when we enhance the existing system into the proposed ones. Here, the existing system is Particle Swarm Optimization Algorithm and proposed is Artificial bee colony Algorithm.

To apply ABC, the considered optimization problem is first converted to the problem of finding the best parameter vector which minimizes an objective function. This proposed work is intended to develop Classification rules to extract data from historical or training data of business data sets which is developed into patterns relevant and suitable for quicker analysis, automated processing.

This system is converted into the proposed system using the enhanced algorithm called ABC which gives good accuracy and efficiency for large data sets when compared to the formers.

5. CONCLUSION

Data mining is an inter-disciplinary field, whose core is at the intersection of machine learning, statistics and databases. A major objective of this work is to evaluate data mining tools in business applications to develop a tool that can help make timely and accurate decisions. One technique used in data mining is Classification where the desired output is a set of Rules or Statements that characterize the data.

Within the rule induction paradigm, the algorithm used is Artificial Bee Colony. Artificial Bee Colony (ABC) is a swarm based meta-heuristic algorithm for optimizing numerical problems which were inspired by the intelligent foraging behavior of honey bees.

This proposed work is intended to develop Classification rules to extract data from historical or training data of business data sets which is developed into patterns suitable for quicker analysis, automated processing, thus reducing cost and helping to provide enhanced care. The project implemented it with benchmark of business data sets.

Thus it can also conclude that, maximum number of iterations provide better success rates Thus rules were developed with rate of accuracy defining the underlying attributes .

REFERENCES

[1] Ali Hadidi, Sina Kazemzadeh Azad, Saied Kazemzadeh Azad, Structural optimization using artificial bee colony algorithm, 2nd International Conference on Engineering Optimization, 2010, September 6–9, Lisbon, Portugal.

[2] "Artificial bee colony (ABC) algorithm homepage". Mf.erciyes.edu.tr. Retrieved 2012-02-19.B.Basturk, Dervis Karaboga, An Artificial Bee Colony (ABC) Algorithm for Numeric function Optimization, IEEE Swarm Intelligence Symposium 2006, May 12–14, 2006, Indianapolis, Indiana, USA.

[3] D. Karaboga, An Idea Based On Honey Bee Swarm for Numerical Optimization, Technical Report-TR06,Erciyes University, Engineering Faculty, Computer Engineering Department 2005.

[4] D. Karaboga, B. Basturk, A Powerful And Efficient Algorithm For Numerical Function Optimization: Artificial Bee Colony (ABC) Algorithm, Journal of Global Optimization, Volume:39 , Issue:3 ,pp: 459–471, Springer Netherlands, 2007. doi:10.1007/s10898-007-9149-x

[5] D. Karaboga, B. Basturk, On The Performance Of Artificial Bee Colony (ABC) Algorithm, Applied Soft Computing, Volume 8, Issue 1, January 2008, Pages 687–697. doi:10.1016/j.asoc.2007.05.007

[6] D. Karaboga, B. Basturk, Artificial Bee Colony (ABC) Optimization Algorithm for Solving Constrained Optimization Problems, LNCS: Advances in Soft Computing: Foundations of Fuzzy Logic and Soft Computing, Vol: 4529/2007, pp: 789–798, Springer- Verlag, 2007, IFSA 2007. doi:10.1007/978-3-540-72950-1_77

[7] D. Karaboga, B. Basturk Akay, Artificial Bee Colony Algorithm on Training Artificial Neural Networks, Signal Processing and Communications Applications, 2007. SIU 2007, IEEE 15th. 11–13 June 2007, Page(s):1–4, doi:10.1109/SIU.2007.4298679

[8] D. Karaboga, B. Basturk Akay, C. Ozturk, Artificial Bee Colony (ABC) Optimization Algorithm for Training Feed-Forward Neural Networks, LNCS: Modeling Decisions for Artificial Intelligence, Vol: 4617/2007, pp:318–319, Springer-Verlag, 2007, MDAI 2007. doi:10.1007/978-3-540-73729-2_30