ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

An Efficient Fuzzy Based Approach for Hepatitis Detection

Ancy Sudhakar[1] , Rajalekshmy K D[2], Prince Abraham[3]

Assistant Professor, Dept. of I.T., CEV, Kerala, India

Assistant Professor, Dept. of MCA., CEV, Kerala, India

Assistant Professor, Dept. of I.T., CEV, Kerala, India

ABSTRACT: Data mining means knowledge discovery in database. Its main aim is to extract information from a dataset and to transform it into an understandable form. Data mining deals with the discovery of hidden knowledge. It involves tasks such as classification, clustering, regression etc. Here we use the Ant colony Optimization algorithm for discovering the classification rules. Ant colony optimization algorithm is an algorithm for finding optimal path and it is based on the behavior of ants searching for food. Datasets of hepatitis are obtained from the benchmark repositories. Fuzzy based ant colony optimization algorithm is applied initially for classifying the attributes and constructing the rule base . Rule pruning method is used in this work to get the optimized rules. For detection of hepatitis we make use of fuzzy logic. By this work it is possible to predict the hepatitis in the earlier stage with maximum accuracy.

KEYWORDS: Rule Pruning, ACO Algorithm, Fuzzy logic , Rule Quality , Data mining

I. INTRODUCTION

Ant Colony Optimization mainly deals with the behavior of ants. It is based on heuristic optimization method inspired by biological systems. Most communication among ants is based on a chemical substance called Pheromone. These pheromone is used for marking and following shortest paths from nest to food source and vice versa. As more pheromone gets deposited the more optimized path is obtained. Each ant incrementally modifies or constructs a solution for the target problem. This is how the ants choose the optimized path. Thus the candidate solution for the problem depends on the amount of pheromone deposited.

Identification of classification rules is the target problem in Ant Colony optimization. The rule based classifier commonly use if-then rules for classification. The if part of the rule is called rule antecedent and then part called rule consequent. The rule is of the form:

IF<term 1 AND term2 AND…>THEN <class>

Each term is a triple <attribute,operator,value> . If the condition holds true for a given triple then the antecedent is satisfied. Fuzzy based Ant Colony Optimization method is used in this work. Fuzzy logic is an approach to computing based on degrees of truth rather than the usual true or false . Here if the rule satisfies the fuzzy value we then add the rule. A process of fuzzification, inference and defuzzification leads to the final decision of the problem.

In the existing system the quality of the rule is decreased as it deal with the whole database which includes both nominal and continuous data. As a result the accuracy is less in the existing system. The Ant colony algorithm here deal with only nominal attributes and for other attribute we do conversion in preprocessing phase. In the existing system some information's are lost during this conversion process.

In the proposed work we classify the nominal attribute using fuzzy based ant colony optimization algorithm. Then we classify the continuous attributes also. Rule pruning is applied to obtain the optimized rules. Pruning reduces the

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

complexity of the classifier. To improve the quality of the rule we use fuzzy rules. Pheromones' are updated whenever the rule is obtained.

II. RELATED WORK

In the work of Urmila M Diwekar and Berhane H Gebreslassie [1] they deal with an efficient ant colony optimization algorithm. Hammersley Sequence Sampling method is used introduced in this work. This new algorithm is useful for wide range of optimization problems. Threshold ant colony optimization algorithm is suggested by K Thangavel and P Jaganathan [2] swhose main goal is to provide comprehensible classification rules which have higher predictive accuracy and simpler rule list. In this method ACO-Miner algorithm is compared with C4.5 classification algorithm.In the work proposed by Rafael S. Parpinelli, Heitor S. Lopes and Alex A. Freitas [3] they compared CN2 a classification algorithm with Ant Miner algorithm. The results shows that the Ant Miner algorithm produces better results for four datasets where as CN2 produces better result for only one data set Bo Liu**,** Hussien A.Abbass and Bob Mckay proposed an Ant based rule discovery method which is identical to Ant Miner algorithm and its computational cost is less compared to Ant Miner algorithm. In the work done by S.Madhusudhanan, Marcus Karnan and K.Rajivgandhi [4] extracts the classified rules using fuzzy based ant miner algorithm(FACO). The FACO algorithm is used to classify the categorical attributes. In the work of P.V Sarath Chand , Dr.A Vinay Babu,Dr.A.Govardhan[5] they deal with metaheuristics. Metaheuristics solve instances of the problem that are believed to be hard in general. In the work of BO Liu1 Hussien A.Abbass Bob Mckay[6] they show that the heuristic value used in Ant miner algorithm is based on the concept of entropy.

III. PROPOSED ALGORITHM

A. *Ant Colony Algorithm:*

The Ant colony algorithm is based on the foraging behavior of some ant species. The foraging behavior of ants is their ability to find the shortest path between nest and food source. This idea of ant colony lead the researchers to find algorithm to solve many problems and it helped to find a least cost path between a source and destination.

Some ant species lay pheromone trails on the path they take and these pheromone trails act as stimuli since the ants are attracted to follow the paths that have relatively more trails. Ants uses this pheromone to communicate with each other and to find the shortest path between nest and food source. Fig 3.1 below shows the Ant colony system. Whenever an obstacle appears between their paths the ants choose the path where the pheromone concentration is more. More the pheromone deposit shorter the path they follow.

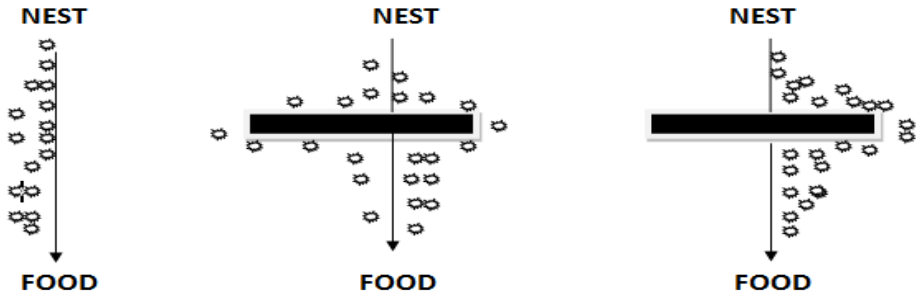

Fig 3.1 The Ant Colony System

The algorithm starts by selecting the training set that cover almost all possible cases of hepatitis. Next we initialize the rule list to an empty list. Repeat the following procedure until training set is less than or equal to user specified threshold, maximum uncovered value. Initially all terms of the rule which correspond to the attributes of the training set have the equal possibility of being chosen. This means that all trails are initialized with the same amount of pheromone. Construct the rule incrementally by adding one term at a time to the current partial rule which corresponds to the trail followed by the ants. The trail corresponds to the actual rule constructed that depends upon the particular case of hepatitis. Rule pruning is performed to find the optimized rules discovered which is performed by updating pheromone on the paths followed by ants and evaporating pheromones in other trails.

Select best rule based on the fuzzy logic since the main disadvantage of rule based classifier is sharp cutoffs for continuous attributes So we use fuzzy based approach and use graphical tools such as trapezoidal membership functions for converting the attribute values

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

to fuzzy truth values. More than one fuzzy rule may apply so we select the best rule based on the degree of membership of each applicable rule. The resulting fuzzy output is translated into crisp value.The quality of the rule is measured using sensitivity and specificity. Eliminate those training set that satisfy the best rule from the actual training set chosen initially.

B. *Fuzzy Rule:*

A fuzzy rule is based on if-then rules. That is if x is A then y is B where x and y are linguistic variables and A and B are linguistic values determined by fuzzy sets. Fuzzy rule is mostly based on fuzzy set theory and fuzzy logic. In this approach mainly three process are involved:

- Fuzzification
- Inference
- Defuzzification

Fuzzification- The problem is that in the real world some of our knowledge of facts is derived from the use of sensors such as weight measured in grams, heat measured in degree centigrade etc. This quantitative and precise information has to be mapped on to the linguistic variables. This process is called fuzzification.
Inference- Once mapped rules are checked to see which of the rules are fired and to what degree. In the fuzzy expert systems all rules may fire to a degree between 0 and 1.
Defuzzification- Finally we have to covert the fuzzy values outputted by the inference procedure into the crisp set that can be used in real world.
This fuzzy algorithm is actually an ordered sequence of instructions which may contain fuzzy assignments and conditional statements. We make use fuzzy in situations for example, in a medical dataset we cannot decide whether a person has headache or not. In this situation yes or no answers are not possible. In such a situation we make use of fuzzy.
In this work we can make use of this fuzzy logic to detect whether there is hepatitis or not. This work makes it easier to diagonize the disease. The data sets for hepatitis are taken from the benchmark repositories.

C. *Rule Pruning:*

Rule pruning is done to increase the accuracy of the rule. It is done immediately after ant constructs the rules. The rule pruning procedure improves the quality of the rule by removing irrelevant terms and repetitive terms. It is done until there is no term left to remove so that quality of the rule is improved. Rules may perform well on the training data but less well on subsequent data so rule pruning is required for better accuracy.

D. *Pheromone Updation:*

Pheromone is updated after each ant completes the construction of its rule. Here Pheromone is updated increases in the trail followed by ants and it decreases in the other trails.

We measure the rule quality using Sensitivity and Specificity. Sensitivity and specificity are used for medical informatics. Sensitivity measures the percentage of people with disease whose test is positive. Specificity measures the percentage of people who is not having disease with test negative.

Rule Quality= Sensitivity * Specificity
Sensitivity= TP/TP+FN
Specificity=TN/TN+FP
Rule Quality =Sensitivity* Precision
Precision=TP/TP+FP
Accuracy=(TP+TN)/(TP+FP+FN+TN)

	DISEASE	NO DISEASE
TEST (+)	TRUE POSITIVE(TP)	FALSE POSITIVE(FP)
TEST (-)	FALSE NEGATIVE(FN)	TRUE NEGATIVE(TN)

Fig.3.2 Sensitivity and Specificity Measuring Table

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

E. *The Proposed System Architecture:*

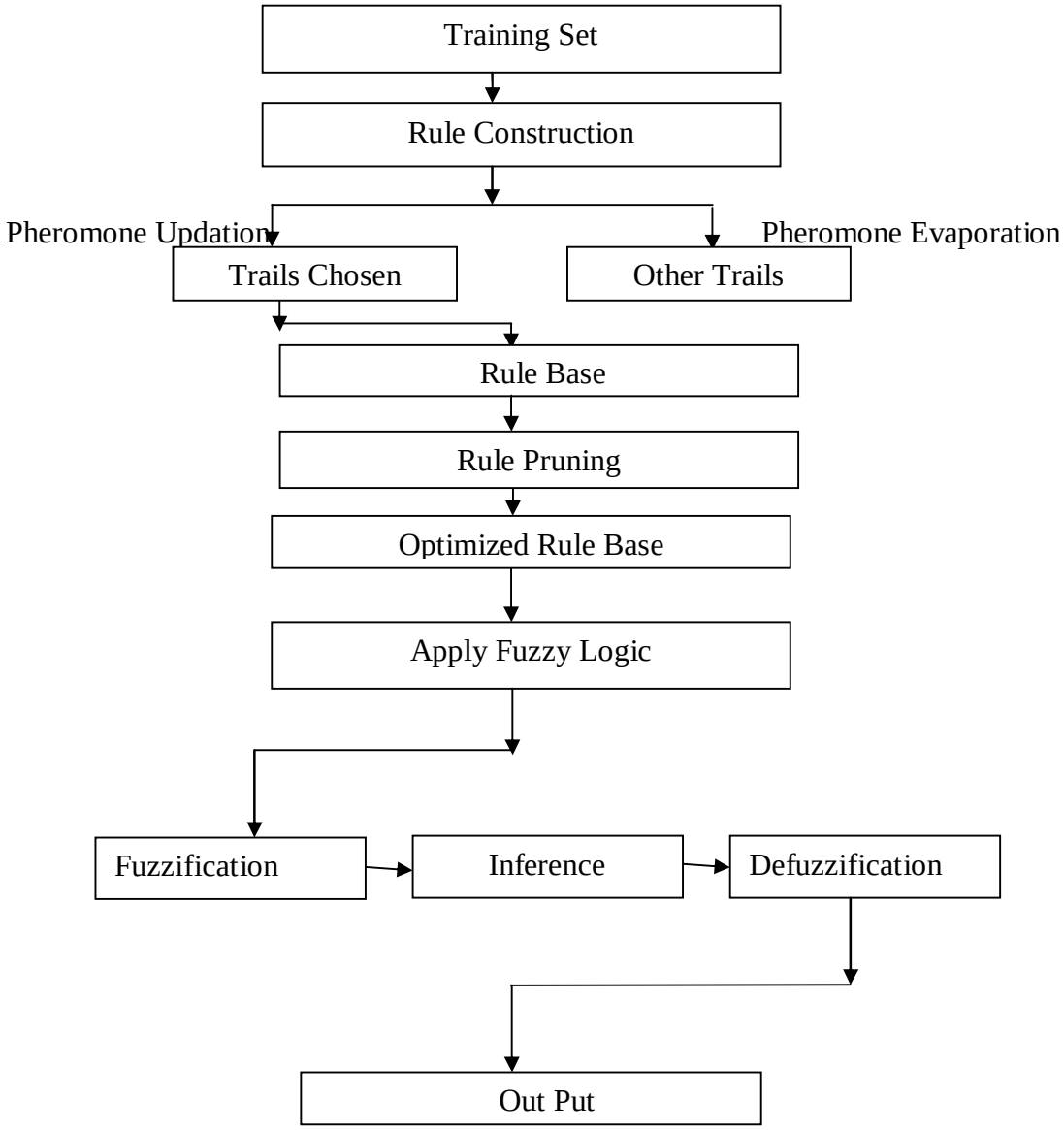

Fig. 3.3 The Proposed System Architecture

IV. RESULT

For experiment we make use of benchmark repository for hepatitis which contains 155 instances with 19 features such as age, sex, steroid, antivirus, fatigue, malaise,anorexia,big liver, liver film, palpable spleen, spiders,ascites etc. and two classes(LIVE or DIE). In this experiment we consider 123 diseased cases and 32 non diseased cases. We obtained an average accuracy of 92.99% , sensitivity 84.25% and specificity 58.9%.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

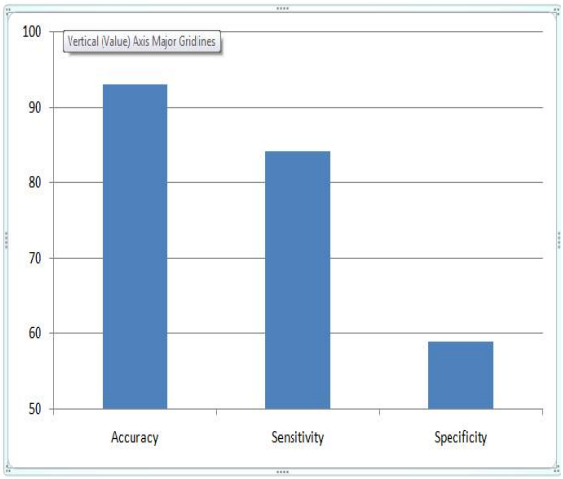

Fig. 3.4 Disease Prevalence graph

V. CONCLUSION AND FUTURE WORK

This paper "Detection of Hepatitis Using Ant Colony Optimization Based on Fuzzy" makes use of fuzzy logic and ACO algorithm for detecting hepatitis. The accuracy is measured using test cases. This works helps in easy diagnosis of hepatitis. It is also possible to apply this work different areas. The project finds its application in areas such as medical, financial, banking, education system etc. Through various data mining techniques it is possible to extract hidden knowledge. In future we try to extend our work for prediction of droughts in India.

REFERENCES

1. Urmila M Diwekar*and Berhane H Gebreslassie, "Efficient Ant Colony Optimization(EACO) Algorithm for Deterministic Optimization" , International Journal of Swarm Intelligence and Evolutionary Computation,Volume 5 • Issue 2 • 1000131, 2016
2. K. Thangavel P.Jaganathan, "Rule Mining Algorithm With A New Ant Colony Optimization" , International Conference on Computational Intelligence and Multimedia Applications ,2007.
3. P. Shivakumar and K Elakia , " Survey of Ant Colony Optimization" , International Journal of Advanced Research in Computer Science and Software Engineering, Volume 6, Issue 3, ISSN: 2277 128X, March 2016
4. S.Madhusudhanan, Marcus Karnan and K.Rajivgandhi, "Fuzzy Based Ant Miner Algorithm in Datamining for Hepatitis", International Conference on Signal Acquisition and Processing, IEEE Computer Society,2010.
5. P.V Sarath Chand , Dr.A Vinay Babu,Dr.A.Govardhan, T. F. Gonzalez, "Data Streams With Ant Miner Approximation Algorithms and Metaheuristics", CRC Press, 2007.
6. BO Liu1 Hussien A.Abbass Bob Mckay, "Density Based Heuristic For Rule Discovery With Ant-Miner",IEEE Computational Intelligence Bulletin, Vol.3 No.1, February 2004
7. Hisao Ishibuchi, Tomoharu Nakashima, and Tetsuya Kuroda, "Hybrid Fuzzy Genetics-Based Machine Learning Algorithm:Hybridization Of Michigan Approach And Pittsburgh Approach, Pattern Recognition", vol. 37, no. 6, pp. 1287–1298, 2005.
8. Parpinelli,R.S., Lopes,H.S., and Freitas,A.A, "Datamining with an ant colony optimization Algorithm",IEEE Transactions on Evolutionary Computing 6,321– 332,2002.

BIOGRAPHY

Ancy Sudhakar received her Masters Degree in Computer Science and Engineering from KMCT College of Engineering Calicut, Calicut University, Kerala, India. Now working as Assistant Professor, Information Technology Department, College of Engineering,Vadakara, Calicut Dist, Kerala, India. Her research areas include datamining and image processing

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

Rajalekshmy K D did her Masters Degree in Computer Application from JKK College Komarapalayam, Madras University, Kerala, India. Now working as Assistant Professor, Computer Applications Department in College of Engineering Vadakara, Calicut Dist, Kerala, India. Her research area include datamining.

Prince Abraham received his Masters Degree in Computer and Information Science from , Cochin University, Kerala, India. Now working as Assistant Professor, Information Technology Department, College of Engineering Vadakara, Calicut Dist,Kerala, India. His research areas include image processing and datamining .