International Journal of Advanced Technology and Engineering Exploration
ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
Volume-1 Issue-1 December-2014

An Efficient Data Mining and Ant Colony Optimization technique (DMACO) for Heart Disease Prediction

Animesh Dubey*, Rajendra Patel and Khyati Choure
Assistant Professor, Patel College of Science and Technology, Bhopal, India
animeshdubey123@gmail.com

Abstract

India is set to witness a spike in deaths due to heart diseases. Early stage detection may prevent the death due to the heart diseases. In this paper we provide an efficient approach which is based on Data Mining and Ant Colony Optimization technique (DMACO) for Heart Disease Prediction so that we can prevent it in the earlier stages. For this we first taken the concept of data mining to finding the support, generated support is used as a weight of the symptom which will be the initial pheromone value of the ant. Then we consider Pain in the chest, Discomfort radiating to the back, choking feeling (heartburn), Nausea, Extreme weakness and Irregular heartbeats as the factor of heart attack. According to risk level identified we find the max pheromone value), max pheromone value is the addition of weight and the risk level. After applying the DMACO algorithm we can observe the increasing detection rate. So by this approach we can increase the detection probability in the early stage which is not generally detected in the earlier stage.

Keywords

DMACO, Heart Attack Factors, ACO, Support Count

1. Introduction

The goal of data mining is to extract knowledge from huge amount of data. Data mining is an interdisciplinary field, whose fundamental lies on data analysis and pattern recognition. In this paper we use data mining to emphasize to discover knowledge that is not only accurate, but also comprehensible for the user [1], [2], [3]. According to Yang Jianxiong et al. [4] Distinctness is standard whenever discovered understanding mettle be second-hand for supporting a human decision. Tab round, if discovered acquaintanceship is groan eliminate for a owner, it spine war cry be postcard to scrutinize and validate the knowledge.

According to chen et al. [5] by using ant colony algorithm, we count item sets according to the pheromone concentration. Moreover, using the accumulating and volatilization characteristic of pheromones, we simulate the change of frequency of Item set. So in this paper we want to combine data mining techniques with ACO for better heart disease prediction.

The rest of this paper is arranged as follows: Section 2 introduces Literature Review; Section 3 shows the proposed approach; Section 4 shows the result analysis; Section 5 describes Conclusion.

2. Literature Review

In 2012, Peter, T.J. et al. [6] proposes use of pattern recognition and data mining techniques into risk prediction is proposed. It can be separated by using classification technique. There are some limitations in the traditional medical scoring systems are that there is a presence of intrinsic linear combinations of variables in the input set and hence they are not adept at modelling nonlinear complex interactions in medical domains. They work on the limitation which can implicitly detect complex nonlinear relationships between dependent and independent variables as well as the ability to detect all possible interactions between predictor variables. In 2012, Muhammed et al. [7] present and discuss the experiment that was executed with naïve bayes technique in order to build predictive model as an artificial diagnose for heart disease based on data set which contains set of parameters that were measured for individuals previously. Then they compare the results with other techniques according to using the same data that were

*Author for correspondence

given from UCI repository data. In 2012, Ruchita Gupta et al. [8] present an efficient range partitioning method for finding frequent pattern from huge database. This is based on separation of key for finding the local frequent pattern (LFP). After finding the partition frequent pattern from the subdivided local database, then they find the global frequent pattern from the local database and perform the pruning. Subset super set mining is also suggested in [9]. In 2012, Shashank Singh et al. [10] apply the apriori algorithm with transaction reduction on cancer symptoms. They consider five different types of cancer and according to the classification they generate the candidate sets and minimum support to find the spreading of cancer. They can find the symptoms by which the cancer is spreading more and also about the highest spreading cancer type. In 2012, Neha Purohit et al. [11] explores about the mining of data and finding essential information from huge amounts of data. Extracting or “mining” knowledge from large amounts of data is known as Data Mining. They suggest that the use of algorithms to extract the information and patterns is derived by the KDD process. In 2012, Nikhil Jain,Vishal Sharma et al. [12] find the near distance of rule set using Euclidean distance formula and generate two class higher class and lower class .the validate of class check by distance weight vector. Author suggests that distance weight vector maintain a threshold value of rule item sets, the whole process they used genetic algorithm for optimization of rule set. In 2013, Anshuman Singh Sadh et al. [13] survey several aspects of optimization techniques by which they can optimize the association rules. So the main motivation of their survey is to minimize the rule generation or optimize rule generation larger size of rules can be minimized.
In 2013, U.Chandrasekhar et al. [14] surveys various latest frequent pattern mining algorithms on data streams to understand various advantages and disadvantages, so they provides a way of using new insights in the direction of frequent pattern. The combination of optimization and data mining is also suggested in [15].

3. Proposed Methodology

The Ant Colony Optimization (ACO) calculation is a meta-heuristic that has a mix of circulated processing, autocatalysis, and valuable insatiability to discover an ideal arrangement. This calculation away at burrowing little creature's conduct in this present reality. Since its presentation, the ACO calculation has gotten much consideration and has been joined in numerous enhancement issues, to be specific the system directing, voyaging salesperson, quadratic task, and asset assignment issues [16].

The ACO calculation has been enlivened by the trials run by Goss et al. [17] utilizing a settlement of ants. They watched that genuine ants had the capacity select the briefest way between their home and nourishment asset, in the presence of exchange ways between the two. The inquiry is made conceivable by a backhanded correspondence known as stigmergy among the ants. While voyaging their direction, ants store a synthetic substance, called pheromone, on the ground. This conduct has an autocatalytic impact on account of the very truth that a ground dwelling insect picking a way will build the likelihood that the comparing way will be picked again by different ants later on. When they return back, the likelihood of picking the same way is higher (because of the increment of pheromone). New pheromone will be discharged on the picked way, which makes it more alluring for future ants. Right away, all ants will choose the most limited path.

This conduct was detailed as Ant System (AS) by Dorigo et al. [18].in ACO calculation, the advancement issue is figured as an issue G = (C; L), where C is the situated of parts of the issue, and L is the situated of conceivable associations or moves among the components of C. The arrangement is communicated as far as practical ways on the diagram G, concerning a set of given demands. The number of inhabitants in operators (ants) by and large tackles the issue under thought utilizing the diagram representation. In spite of the fact that every ground dwelling insect is fit for discovering a (presumably poor) arrangement, great quality arrangements can rise as an issue of aggregate association among ants. Pheromone trails encode a long haul memory about the entire ground dwelling insect inquiry process. Its quality relies on upon the issue representation and the streamlining target.

In this methodology we consider the information set of patient perception from a few days. In the event that the manifestation is available in the patient we put 1 if the indication is not located then we put 0. Taking into account the attention a plain database is made like table 1. At that point we discover the recurrence or the help estimation of every indication by applying affiliation principle mining. By this we set the weight of the thing set. At first pheromone will be introduce, so it will be upgraded each time or for our situation it will be redesigned for 10-20 cycles which is rely on upon the quantity of days the indications saw by the patient. For expanding the discovery proficiency we apply the ACO which enhances the identification exactness, so it will be discovered in the prior stages and legitimate treatment can be conceivable.

The procedure is as per the following [16]:

1. For each one continuous pattern discover the help tally:

$$support = \frac{(X \cup Y).count}{n}$$

2. For each one successive example discover the solution:
3. An ant let k has a memory M^k that it can use to store data on the way it emulated in this way. The put away data can be utilized to construct possible arrangements, assess arrangements and backtrack the way retrogressive.
4. An ant *k* can be assigned a start state s^k_s and more than one end conditions e^k.
5. Ants begin from a begin state and move to achievable neighbor states, fabricating the arrangement in an incremental way. The system stops when no less than one termination condition e^k for ant *k* is satisfied.
6. An ant *k* located in node *i* can move to node *j* chosen in a feasible neighborhood N^k_i through probabilistic decision rules. This can be formulated as follows:
   An ant *k* in state $sr =< s_{r-1}; i >$ can move to any node *j* in its feasible neighborhood N^k_i , defined as $N^k_i = \{j \mid (j \in Ni) \wedge (< sr, j > \in S)\}$ $sr \in S$, with *S* is a set of all states.
7. The qualities put away in a hub neighbourhood information structure Ai = [a_{ij}] called ant routing table obtained from pheromone trails and heuristic values,
   The ground dwelling ants own particular memory from past emphasis, and the issue obligations.
8. When moving from node *i* to neighbour node *j*, the ant can update the pheromone trails τ_{ij} on the edge (*i*, *j*).
9. Once it has manufactured an answer, a ground dwelling insect can backtrack the same way retrogressive, upgrade the pheromone trails and bite the dust.

The trail intensity is determined by the below formula:

$$T_{ij}\,(t + n) = p \,.\, T_{ij}(t) + \Delta T_{ij}$$

And the next move will be determine

$$\Delta T^k_{ij} = \begin{cases} \frac{Q}{L_k} & \text{if the kth ant uses edge}(i, j) \text{ in its tour} \\ & (\text{between time } t \text{ and } t+n) \\ 0 & \text{otherwise} \end{cases}$$

And then we concentrate on the group highest value, so that we increase the prediction rate.

4. Result Analysis

For better understanding we consider the following symptoms of a heart attack for our example. And provide a symbolic notation like A,B etc.

Pain in the chest (A)
Discomfort radiating to the back (B)
Choking feeling (heartburn) (C)
Sweating, nausea (D)
Extreme weakness (E)

Table 1: Sample database

Symptoms Days of Observation	**A**	**B**	**C**	**D**	**E**
Day1	1	0	1	0	0
Day2	1	1	1	1	1
Day3	0	0	0	0	0
Day4	0	0	1	0	0
Day5	0	1	1	0	0

Then we find the support count based on table 1 data which is shown in table 2.

Table 2: Support Count

Symptoms	**Support Count**
A	0.4
B	0.4
C	0.8
D	0.2
E	0.2

Based on the above data we plot the graph of detection as shown in figure 1.

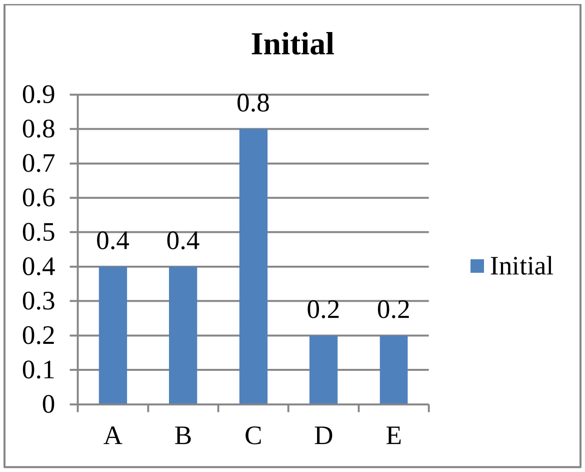

Figure 1: Initial Graph

If the support value is less than P_T then it is replaced by P_T value, otherwise remains unchanged.

From the next trail we also subtract the evaporation value, in our case we assume as 0.2 which is increases as 0.6, 0.8 and then it is fixed for 0.8, because we put the evaporation value in the range of [0.2-0.8]. If the value of P_T is less than 0 or greater than 1 then we consider the previous value only. After the first iteration the lower values from the P_{T1} is replaced by P_{T1} as shown in table 3.

Table 3: First Trail

Symptoms	Support Count
A	0.4
B	0.4
C	0.8
D	0.4
E	0.4

Based on the above data we plot the graph of detection as shown in figure 2.

Second Trail

P_{T2} $=P_{T1}$ + $\sum$ (support count (A+B+C+D+E+F))/ Number of Symptoms -0.2

P_{T2} $=0.4 + 0.48-0.2$
$P_{T2}$$=0.68$

After the second iteration the lower values from the P_{T2} is replaced by P_{T2} as shown in table 4.

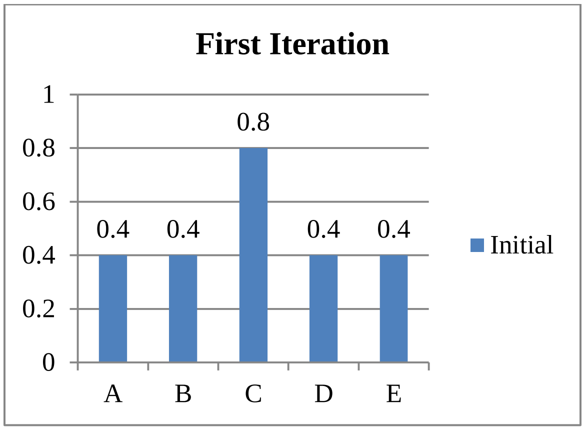

Figure 2: Graph after First Iteration

Table 4: Second Trail

Symptoms	Support Count
A	0.68
B	0.68
C	0.8
D	0.68
E	0.68

Based on the above data we plot the graph of detection as shown in figure 3.

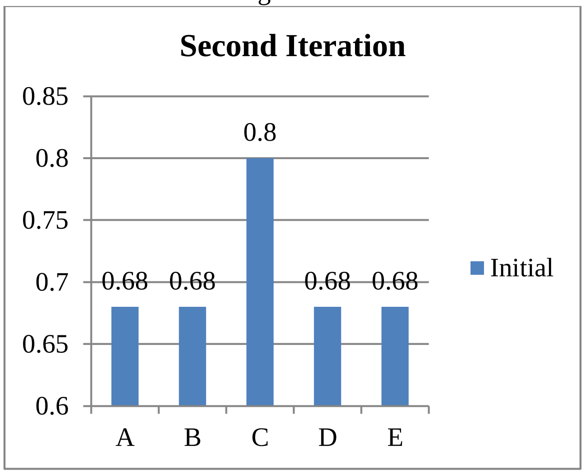

Figure 3: Graph after Second Iteration

P_{T3} $=P_{T2}$ + $\sum$ (support count (A+B+C+D+E+F))/ Number of Symptoms -0.6
P_{T3} $=0.68 + 0.704-0.6$

$P_{T3} = 0.784$

After the third iteration the lower values from the P_{T3} is replaced by P_{T3} as shown in table 5.

Table 5: Third Trail

Symptoms	Support Count
A	0.784
B	0.784
C	0.8
D	0.784
E	0.784

Based on the above data we plot the graph of detection as shown in figure 4.

As we observe in the examples in each subsequent iteration the detection rate is increases, so we can use ACO to improve the detection rate and it is helpful in proper heart attack detection.

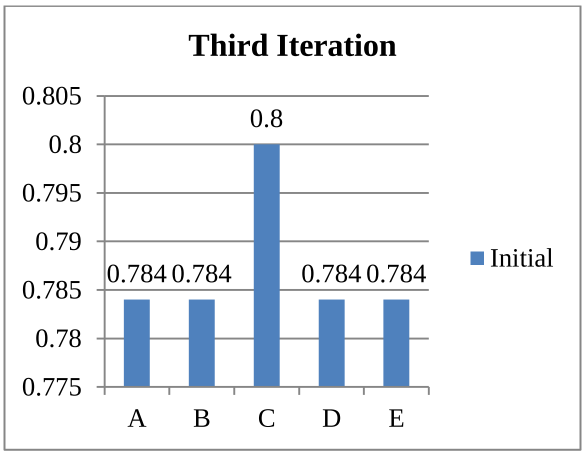

Figure 4: Graph after Third Iteration

5. Conclusion

In this paper we present a combination of data mining and ACO technique which improves the detection rate. By applying above two approaches we can increase the pheromone intensity and make the direction in a singular way.

References

[1] M. Dorigo, G. Di Caro, and L. M. Gambardella, "Ant algorithms for discrete optimization," Artif. Life, vol. 5, no. 2, pp. 137–172, 1999.

[2] U. M. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery: An overview," in Advances in Know Discovery & Data Mining, U. Fayyad, G. Piatetsky-Shapiro, P. Smyth,and R. Uthurusamy, Eds. Cambridge, MA: MIT Press, 1996, pp. 1–34.

[3] Junzo Watada, Keisuke Aoki, Masahiro Kawano, Muhammad Suzuri Hitam, Dual Scaling Approach to Data M Journal of Advanced Computational Intelligence Intelligent Informatics (JACIII), Vol. 10, No. 4, pp. 441-447, 2006.12.

[4] Yang Jianxiong and Junzo Watada, "Wise Mining Method through Ant Colony Optimization", Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics San Antonio, TX, USA - October 2009.

[5] Chen ShengBing, Wang XiaoFeng and Wang XiaoFang, "Mining Dynamical Frequent Itemsets Based on Ant Colony Algorithm", IEEE 2011.

[6] Peter, T.J.; Somasundaram, K., "An empirical study on prediction of heart disease using classification data mining techniques", IEEE 2012.

[7] Muhammed, Lamia AbedNoor, "Using data mining technique to diagnosis heart disease", IEEE 2012.

[8] Ms. Ruchita Gupta, Dr.C.S.Satsangi," An Efficient Range Partitioning Method for Finding Frequent Patterns from Huge Database", International Journal of Advanced Computer Research (IJACR), Volume 2 Number 2 June 2012.

[9] Dubey, A.K.; Dubey, A.K.; Namdev, M.; Shrivastava, S.S., "Cloud-user security based on RSA and MD5 algorithm for resource attestation and sharing in java environment," Software Engineering (CONSEG), 2012 CSI Sixth International Conference on, pp.1,8, 5-7 Sept. 2012.

[10] Shashank Singh, Manoj Yadav and Hitesh Gupta, "Finding the Chances and Prediction of Cancer through Apriori Algorithm with Transaction Reduction", International Journal of Advanced Computer Research (IJACR), Volume 2 Number 2 June 2012.

[11] Neha Purohit, Sapna Purohit, Ritesh Kumar Purohit," Data Mining, Applications and Knowledge Discovery", International Journal of Advanced Computer Research (IJACR), Volume-2 Number-4 Issue-6 December-2012.

[12] Nikhil Jain,Vishal Sharma,Mahesh Malviya, "Reduction of Negative and Positive Association Rule Mining and Maintain Superiority of Rule Using Modified Genetic Algorithm", International Journal of Advanced Computer

Research (IJACR),Volume-2 Number-4 Issue-6 December-2012.

[13] Anshuman Singh Sadh, Nitin Shukla, "Association Rules Optimization: A Survey", International Journal of Advanced Computer Research (IJACR), Volume-3 Number-1 Issue-9 March-2013.

[14] U.Chandrasekhar, Sandeep Kumar. K, Yakkala Uma Mahesh," A Survey of latest Algorithms for Frequent Item set Mining in Data Stream", International Journal of Advanced Computer Research (IJACR),Volume-3 Number-1 Issue-9 March-2013.

[15] Dubey, Ashutosh Kumar, Umesh Gupta, and Sonal Jain. "A Survey on Breast Cancer Scenario and Prediction Strategy." In Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014, pp. 367-375. Springer International Publishing, 2015.

[16] M. Dorigo and M. Maniezzo and A. Colorni. The Ant Systems: An Autocatalytic Optimizing Process. Revised 91-016, Dept. of Electronica, Milan Polytechnic, 1991.

[17] S. Goss, S. Aron, J. L. Deneubourg, and J. M. Pasteels. Self-organized Shorcuts in the Argentine Ant. Naturwissenschaften, 76:579–581, 1989.

[18] M. Dorigo and G. Di Caro. New Ideas in Optimisation. McGraw Hill, London, UK, 1999.