

Ant colony clustering and classification

Surya.K

P.G scholar, department of CSE,

Karpagam academy of higher education, Coimbatore-21

Abstract: - In this paper, to produce the system performance higher, a new algorithm is developed to combine clustering and classifying technique. On considering the objects, it may be similar or dissimilar. The process of organizing objects based on whether the objects are similar or dissimilar is done by CLUSTERING technique and CLASSIFICATION which helps to classify the data objects when the data is large. In this paper clustering and classification is done by class values and attributes. It adopts the improved ANT COLONY algorithm as the method of selection classification characteristics parameters. Totally it includes five modules to show the accuracy and effectiveness of the proposed system.

Keywords— clustering, classifying, Class values, Attributes, Ant colony algorithm.

I. INTRODUCTION

This paper researches the intrusion detection problem of the network defence, small sample training, and puts forward the algorithm of Support Vector Machine. Aimed at the important influence of CLASSIFICATION with ant colony function on classification performance, this paper adopts the improved Ant Colony Algorithm as the method of selection CLASSIFICATION characteristics parameters. This algorithm is significantly will produce higher results than the other algorithm in training and the detection speed, and have a high enhance of the detection rates of attacking sample.

There are many algorithm for new intrusion detection, to overcome the disadvantage the new approach is introduced for the better accuracy and understanding. The basic task is to collect the dataset. Elimination process takes place in preprocessing which tries to eliminate the duplicate and incomplete data followed

by training the dataset and finally the result is achieved which is significantly will produce the higher results than the algorithm in training.

All ant colony algorithms as Ant Colony Optimization (ACO) in general, which proposed an unique framework model. This algorithm has not only great robustness, positive feedback characteristic and also with parallel and distributed computing feature.

Problem statement:

The disadvantage of the existing system is long time convergence, easy stagnation and local optimization. It selects a point as a random (potentially poor) solution, and make small changes iteratively to the solution, each time improving it a little. It terminates when this algorithm cannot see any improvement. Stagnation divides the data into smaller parts which takes more iteration to get completed.

PROPOSED SYSTEM:

MODULE DESCRIPTION:

1. Data set

The first module is the process of uploading datasets. The proposed system uses car dataset. The dataset has a list of descriptions in the table. The dataset may contain 10000 records. This modules collects those data's and stores into the database for further process

2. Preprocessing and training:

Preprocessing is the process of elimination, which eliminates duplicate and incomplete data's from the dataset before processing. There are many techniques are in preprocessing mainly focused on elimination of duplicate and inconsistent data. Because there may be the availability of missing data, incomplete data, noise data and inconsistent data. This Preprocessing helps to remove all those things before preprocessing for better results.

3. Ant clustering:

The ant colony in the natural world has an intellectual character. The system implements the ant colony techniques for intrusion detection. This module describes the “Ant system” method based on such character of ants, which received great lab results. This module implements the ant clustering phase.

After the third module clustering, the neighbors of those marked objects are stored in the CLASSIFICATION training data file, which is used by the component CLASSIFICATION. All ant colony algorithms as Ant Colony Optimization (ACO) in general, which proposed a unique framework model. This algorithm has not only great robustness, positive feedback characteristic and also with parallel and distributed computing feature.

4. Classification

It establishes the enhanced model by classifying the large dataset. Ant conveying process is done with the Ant cluster algorithm. Here in this paper, the clustering algorithm is based on the training set with each data point labeled as positive or negative (class labels) and one should randomly select data points from each class and with the help of classifier it retains using the updated training set.

5. Reports and results

The final module provides the classification results and test bed approach to show the accuracy and effectiveness of the proposed system. The results will be generated as a graphical form.

ALGORITHMS:

Algorithm Steps

Step 1: Normalizing data

Read the dataset D. from D select each transaction Tid.

Consider z be the detection rate, initially 0;

while z < ZZ do

for r = 1, . . . ,N do

Step 2: phase relates to training

Read the training dataset T. from T selects each transaction Tid.

For each (T1,T2...Tn)

Find the SVM(T1)

Calculates mean M and variance V

While more points to add to training set do

Find support vectors among the selected

points;

Step 3: Ant clustering phase

Cluster© the T and store in cluster Ci.

Find similar data and add into the cluster Ci.

Step 4: Constructing classifiers

Step 5: Classifier modification

Clustering Algorithm

Input: A training set with each data point labeled as positive or negative (class labels).

Output: A classifier.

1 begin

2 Randomly select data points from each class.

3 Generate a classifier.

4 while more points to add to training set do

5 Find support vectors among the selected points;

6 Apply proposed clustering around the vectors;

7 Add the points in the clusters to the training set;

8 Retrain the classifier using the updated training set;

9 end

10 end

Training Algorithm

Input: A data set relates with training

Input: N – number of training iterations.

Input: ZZ – detection rate threshold.

Output: ACO Classifiers.

1 begin

2 Normalize the data;

3 Let z be the detection rate, initially 0;

4 while z < ZZ do

5 for r = 1, . . . ,N do

6 phase relates to training;

- 7 phase relates to ant clustering;
- 8 end
- 9 Construct classifiers;
- 10 Do testing to update z;
- 11 end
- 12 end

(A)home page

Dataset description
Acco implementation

(d)training phase:

Dataset description
Class values UNACC, ACC, GOOD,VGOOD
Attributes Buying: VH, H, M, L. Maint: VH, H, M, L. Doors: 2, 3, 4, 5MORE Persons: 2,4,MORE Lug_boots: S, M, B. Safety: L, M, H.

(b)dataset description

Training process started
Attribute analysis
Total transaction 1727
Total attribute 7
Training process completed

(d)classifying and cluster

Test your data using ACO

THRESHOLD = 5
VH-H-H-5-4-S>DETECTED CLASS IS UNACC
VH-H-H-5-4-M>DETECTED CLASS IS UNACC
VH-H-H-5-4-B>DETECTED CLASS IS UNACC
VH-H-H-5-MORE-S>DETECTED CLASS IS UNACC
VH-H-5-MORE-M>DETECTED CLASS IS UNACC
VH-H-H-5-MORE-B>DETECTED CLASS IS UNACC

(c)training dataset

class	buying	Maint	doors	person	Lug_boot	safety
UNACC	VH	VH	2	2	S	L
UNACC	VH	VH	2	2	S	M
UNACC	VH	VH	2	2	M	H
UNACC	VH	VH	2	4	M	L
UNACC	VH	VH	2	4	B	M
UNACC	VH	VH	2	4	B	H
.
.

Here number of clusters formed is 345 and (ACC-accurate;B-big: H-high: L-low; M-medium: S-small; UNACC-unacc; VH-veryhigh)

Conclusion:

The classification of clustered data is obtained by uploading dataset and by entering the threshold. As compared with the traditional method, the input design,

the code design, the test case design are simple and produce the system performance higher.

Future enhancement

On considering all application, each and every system have advantages and more disadvantages. There is the possible of attacks in the proposed work, so the further enhancements must be can be made to the application, so that at the system will be immediately blocked while attacks take place. In future all transaction will be processed in a secure manner and can find the intruders activity by getting all relevant details. In future there will be capture all actions of the intruder by using screen capturing mechanism

REFERENCES:

1. A. Colomi, M. Dorigo et V. Maniezzo, Distributed Optimization by Ant Colonies, actes de la première conférence européenne sur la vie artificielle, Paris, France, Elsevier Publishing, 134-142, 2010.
2. B. Pfahring, "Multi-agent search for open scheduling: adapting the Ant-Q formalism," Technical report TR-96-09, 2006
3. Dorigo et L.M. Gambardella, Ant Colony System : A Cooperative Learning Approach to the Traveling Salesman Problem, IEEE Transactions on Evolutionary Computation, volume 1, numéro 1, pages 53-66, 2007.
4. D. Martens, M. De Backer, R. Haesen, J. Vanthienen, Evolutionary Computation, volume 11, number 5, pages 651—665, 2007.
5. Gupta, D.K.; Arora, Y.; Singh, U.K.; Gupta, J.P., "Recursive Ant Colony Optimization for estimation of parameters of a function," Recent Advances in Information Technology (RAIT), 2012 1st International Conference on , vol., no., pp.448,454, 15-17 March 2012
6. Jiawei Han and Micheline Kamber, "DATA MINING concepts and techniques" 2011 edition 978-81-312-0535-8.
7. M. Dorigo, V. Maniezzo, et A. Colomi, Ant systemPart B , volume 26, numéro 1, pages 29-41, 2008
8. M. Dorigo, Optimization, Learning and Natural Algorithms, PhD thesis, Politecnico di Milano, Italie, 2012.
9. M. Zlochin, M. Birattari, N. Meuleau, et M. Dorigo, Model-based search for combinatorial optimization: A critical survey, Annals of Operations Research, vol. 131, pp. 373-395, 2004.
10. Roger S Pressman, '*Software Engineering*', 2000 Edition, Dreamtech Publications.
11. T. Stützle et H.H. Hoos, MAX MIN Ant System, Future Generation Computer Systems, volume 16, pages 889-914, 2000.
12. X Hu, J Zhang, and Y Li (2008). Orthogonal methods based ant colony search for solving continuous optimization problems. Journal of Computer Science and Technology, 23(1), pp.2-18.