Journal of Information Technology and Society
Volume: 01 Issue: 01 June 2021 Page No.1-7
ISSN: XXX-XXXX

An Efficient and Novel Approach for Neural Network Supervised Classification in AI Using SCCM Technique

[Semi Conquer Classification Method]

M.Jayakandan[1], Dr. A. Chandrabose[2]
[1]Research Scholar, [2]Research Advisor
[1,2]Department of Computer Science, Edayathangudy G.S. Pillay Arts and Science College, Nagapattinam

ABSTRACT - Supervised classification is a machine learning procedure that includes the gathering of targeted information which is a strategy for unsupervised learning and is a typical procedure for factual information investigation utilized in numerous fields. Supervised classification of information is a technique by which expansive arrangements of information are gathered into groups of little arrangements of comparative information. Supervised classification (ANN) is a technique for supervised classification which enables one bit of information to have a place with two or more supervised classification. It is a kind of partition supervised classification where more incredible data can be Classificationed simultaneously on basis of their size and its functionality as it grouping the data in a relational manner with both implicit and explicit data sets. The ANN classification method is obviously used classification model which classifies the data entirely however the size is in a common manner. In this paper, a set of data sets is implanted and the experimental supervised classification report is verified with the frequent parameters such as overlapping, data partitioning, high dimensional data and irrelevant data supervised classification. On comparing with existing supervised classification processes, this proposed approach shows the high efficiency than other supervised classification models with approximate effective results on both the implicit and the explicit data in the association rules.

KEYWORDS: Supervised classification, ANN, Classification, Semi conquer, C-Means

I. INTRODUCTION

A. Data Analysis

Data mining technology is gradually developed from the 1990s which automatically detects hidden and potentially valuable information from a large data repository and generalizes useful structure in order to help people to make proactive, knowledge-based decision. Nowadays people need to deal with more and more amount of data, so data mining receives people's more and more attention. Because of its effectiveness in dealing with massive data, it becomes internationally the most popular topic at present [1].

B. Supervised classification

Supervised classification analysis is grouping the data objects based only on information found in the description of objects and their relationships. The goal is to detect the objects in one group which are similar to each other (related) and different groups of objects which are different (not related). The greater similarity within the group, and the greater difference between the groups provides the better supervised classification. supervised classification is one of the main methods of data mining, which has been widely used in pattern recognition, trend analysis, similarity search and other areas [1].

C. ANN

ANN supervised classification algorithm (ANN) is the most well-known and popularly used supervised classification algorithm **[2], [3],[12]** which is based on the neural networkset theory **[12]**,**[4].** To run ANN, the Classification number must be set first. But in real applications, users may not have the information about the Classification number which leads to the Classification validity problem. The partition coefficient (PC) **[12]** ,**[5]**

and partition entropy (PE) **[12]** , **[6]** are two typical neural network supervised classification indices which consist only the membership values in the neural networkpartition. PC and PE are easy to compute because of its their simple form. There are several validity indices involving in the membership values have been proposed after PC and PE. These indices do not always get absolute result since they do not consider the geometric structure of formula. To avoid such issues, new indices involving the geometric structure are proposed. Xie-Beni's Classification validity index (XB) **[7],[12]** is the most representative among them. With further research, Classification validity indices consisting of compactness and separation criteria are put forward in the literature **[12], [8].** Generally speaking, the compactness criterion indicates the variation or scattering of the data within a Classification, and the separation criterion indicates the reparability between two different Classifications. To decrease the monotonous tendency for the Classification number, some indices are combined with the term of the Classification number in themselves **[12]** ,**[9], [10], [11].**

D. Classification

Classification is an information mining capacity which divides the things in a proper way to target the various categories or classes and the main objective of arrangement is to precisely anticipate the objective class for each case in the information. Classification is a method used either to mine models discussing with important data classes or to predict the future data. Classification is a two-step process and the first step is learning or training step where data is analyzed by a classification algorithm. Second is testing step where the data are used for classification and to calculate the accuracy of the classification **[14, 13].**

E. Association Rule Mining

Association rule mining is a technique which is implemented to discover the known designs, connections, affiliations, or causal structures from informational collections found in different sorts of databases such as social databases, value-based databases, and different types of data repository. Most of the machine learning algorithms work with numeric data sets and hence tend to be mathematical. However, association rule mining is suitable for non-numeric, unconditional data and requires just a little bit more than simple counting. This technique is suitable for both implicit and explicit of data in a set of relation manner.

II. LITERATURE SURVEY

Many Classification algorithms were proposed for generalizing known structure and focused on classifiers weakness and boosting its performance by combination of models C.Preisach et.al **[17],[20]** proposed a generic relational ensemble model which improves Classification accuracy of scientific publications by combining the probability distribution of several relational attributes and local attributes. Relational attributes probability is determined using graph representation and local graph using traditional text Classification. Heterogeneity becomes a major issue since models are combined in a specific format. Xiaoxin Yin et.al **[18][23]** proposed cross mine tree and cross mine rule. Both Classification approaches make use of Tuple ID propagation which helps in virtual join among relation rather than physical join. In Tuple ID selection, key attributes are used for spanning among relations. For non-target relation, all relations are joined together for computing foil gain. However both the methods are unable to handle database imbalances for complex application.

Guo et.al **[19],[23]** proposed multi relation Classification by multiple view creation without upgrading or flattering the original data sets. In this approach Multi Relational Classification (MRC) algorithm is proposed for Classification which in turn makes use of conventional data mining methods at different stages. Finally views are validated by correlation based viewed validation algorithm. Yet this approach makes use of training data set for different views for learning the target concept which may not be helpful for large complex relational database. J.M.Serrano et.al **[21],[23]** proposed query system architecture for retrieving information of olive crop information along with geographical data, crop management and soil attributes. Precise and imprecise data Classifications are determined by using neural networkrelational database. Priori and posteriori neural networkdata processing is used for storing data and querying process. A. Jiménez et.al **[22],[23]** proposed a tree mining approach for multi relational Classification. In this approach two different schemes are proposed for representing multi relational database as sets of trees namely Key-based tree representation and Object-based

tree representation. In Key-based tree representation, primary key attribute is taken as a root node and remaining attributes of relation are taken as child nodes. In Object-based tree representation intermediate nodes act as roots of sub trees.

III. METHODOLOGY

The data mining principle usually deals with various associations mining process that probably handles the both implicit and the explicit data, both of these data are always find by using the supervised classification functions. Thus, this proposed technique also deals the supervised classification technique in a different manner on the basis of both implicit and the explicit data without overlapping on a subsets and finds a subset automatically for the Classificationed data allocation.

A.PROPOSED TECHNIQUE

SCCM:

The supervised classification significance of ANN is superior to anything as usual hard c-means grouping. The supervised classification rate of ANN will turn out to be reasonable, particularly when the informational index is substantial. SCCM is another supervised classification result that dependent on the ANN, which isolates the example space into various classes as indicated by the membership level of tests to the supervised classification. The primary idea of SCANN is to think about the most extreme membership degree with a given control edge parameter in the calculation iterative process. In the event, the most intense involvement degree is more prominent than the limit at that points by incrementing them in the mean time for smoothing other member degree without any changes.

$$\sum_{k=0}^{n} \mathbf{m = 0, y = 1, 2 \ldots\ldots n} \_\_\_\_\_ (1)$$

//Initializing the value for supervised classification//

P>=0, x=1,2.....m;y=1,2,.....n_____(2)

$$0 \leq \sum_{k=0}^{n} \mathbf{m = 0, y = 1, 2 \ldots\ldots n} \_\_\_\_\_ (3)$$

k_i , x=1,2,.....m is the Classification with centre point m

$$\mathbf{k_i} = \frac{\sum_{k=0}^{n} \mathbf{mp[x+y]}}{\sum_{k=0}^{n} \mathbf{mp[x-y]}} \_\_\_\_\_ (4)$$

The Classification point k_i(x=1,2,..., y (x) x n = which is defined in each step of the functioning and the iteration process is defined as follows, Suppose that the extreme relationship of j x is $_{Pmn}$ a and 1 max ij $\leq k_i \leq$ =P , then the value after update is

When P= X_{ij},.....X_{in}

// the value taken automatically for supervised classification with its nearest relation member//

X_{ij}= α m,1-α, β.....n_____(5)

//where 0 ≤1 ≤ α is called suppression factor, 1−α is overthrow rate, and β is overthrow threshold//

Equation **(4)** demonstrates that group focus in the weighted normal value as everything equal, each example plays a certain job in pulling in Classification focus **[15].** μij means the quality of the absorption and the biggest estimation of μij is the highest possible absorption. It maintains a strategic space from the issue of surpasses on centre in HCM, yet it will make the combination speed faster. The most extreme participation degree increment of 1− α + Xij after restore is indicated by **(5)** , and the approval for the Classification focal point of class r is expanded. In the mean time, the other participation degree is lower; consequently the combination rate of grouping calculation is made strides. The most extreme participation degree esteem with the past limit value

provided, rather than any other membership degree, which is corrected in the iterative procedure of SCANN, and it will influence the effect of grouping to make it better. At the same time, the fundamental group focus is situated by the tests which have bigger participation degree, therefore to upgrade the estimation of bigger membership and smother participation can enhance the union rate of the calculation.

B. THE EFFECTIVENESS OF SUPERVISED CLASSIFICATION

By SCANN, we can find that the parameter u and v are partial by introduced supervised classification focus, which will have an effect on the exactness rate of the calculation. It is important to pass the judgment to supervised classification authority with the end goal to guarantee that using the Classification focus as the powerful preparing tests. The supervised classification which is more significant should make the class facility limit value, while at the same time it make intra-class partition record to a maximum level.

$$\mathbf{EC}= \frac{\sum_{i=0}^{u} \mathbf{p|xi-uj|}}{\sum_{j=0}^{v} \mathbf{p|xi-uj|}} \_\_\_\_\_ (6)$$

Where, numerator express tightness degree of supervised classification, therefore the denominator expresses detachment degree where the Classification is more fixed and free from one another if S is small.

C.IMPLICIT AND EXPLICIT DATA

Implicit Data will be data which is not provided purposefully however, accumulated from accessible information streams, either specifically or through examination of explicit data. Explicit Data will be data that is given deliberately, for instance through studies and employment recruitment frames. The SCANN technique is easily composed with both implicit and the explicit of data on the basis of supervised classification, the data are Classificationed with separated region with minimum density with high efficiency by using the association rule mining. Generally a Classification is defined as a maximal set of density oriented points and the proposed technique is also associated with this.

D.ANN ON SCCM

ANN is produced from optimal isolating hyper plane when the information is directly removable; Whenever the information isn't directly noticeable, one can change the information procedures of the space by means of a nonlinear mapping into a higher dimensional highlight space with the end goal to upgrade the acceptability of straight division.At the point when looked with expansive informational indexes, the preparation time will be long; in this manner, it is exceptionally important to decrease the number of preparing tests before preparing ANN. In the short-term, if the preparation set contains samples tests, it will decrease the grouping exactness of the conventional ANN. In this way, there will be favourable position in the event that we can lessen the quantity of preparing tests and expel the sample tests before preparing ANN.

IV. RESULT AND DISCUSSION

To make the supervised classification with more efficient, here we deals with the crime data sets and this shows separation and grouping of data after and before Classification. Initially ANN is applied and the data is classified by the ANN Classification method and our proposed system SCANN [Semi Conquer ANN] shows the effectiveness of supervised classification.

A.DATASET

A dataset is a gathering of information which is usually an informational collection compares to the substance of a private database table, or a private measurable information framework, where each segment of the table speaks to a specific variable and each line relates to a given individual from the informational collection being referred to. Informational collections which are large to the point when the regular information preparing applications are incomplete to manage them are known as large information. The basic function states some of the parameters with some average set and it has only the function of supervised classification high dimension data on comparing with other process which has the membership relation on providing the subset allocation when the

another function has the membership with high dimension data supervised classification. On comparing all those principles, the proposed technique has overcome all the difficulties in supervised classification both implicit and the explicit data with the rule mining. Generally the data sets are larger in size and can be imported for supervised classification and here, the data sets implies is up to 1GB in size, and it is imported by 100mb per supervised classification.

Table.1. Description of Dataset Attributes

S.NO	PRIMARY DESCRIPTION	CASE	SECONDARY DESCRIPTION	LOCATION DESCRIPTION
1.	THEFT	JB245397	$500 AND UNDER	RESIDENCE PORCH/HALLWAY
2.	ROBBERY	JB241116	ARMED:KNIFE/CUTTING INSTRUMENT	ALLEY
3.	WEAPONS VIOLATION	JB241792	FOUND SUSPECT NARCOTICS	POLICE FACILITY/VEH PARKING LOT
4.	NARCOTICS	JB242018	POSS: HEROIN(WHITE)	SIDEWALK

V. CONCLUSION

The proposed SCCM is behaved with the training sets for analysing its supervised classification efficiency where it has the prospects of association mining and the Classification centers of valid supervised classification are used for training an ANN. Classified data are performed with SCCM and it is superior than ANN on two-class classification. Moreover, as the training samples is increasing, the advantage of the algorithm becomes more obvious and it is more frequent in supervised classification all class of data with their relationship with both implicit and the explicit data In future, we will apply the proposed SCCM model with multi-class classification problems for mapping the Classified data in association mining.

REFERENCES

[1]. Chen Yanyun, Qiu Jianlin, Gu Xiang,Chen Jianping, Ji Dan, Chen Li “Advances In Research Of ANN supervised classification Algorithm” IEEE International Conference on Network Computing and Information Security 2017 P.P:28.
[2]. J. C. Bezdek, Pattern Recognition with neural networkObjective Function Algorithms, Plenum Press, New York, 1981
[3]. J. C. Dunn, “A neural networkrelative of the ISODATA process andits use in detecting compact, well-separated Classifications,” J. Cybern., vol. 3, pp. 32–57,1974.
[4]. L.A.. Zadeh, “neural networksets,” Inf. Control, vol. 8, pp. 338–353, 1965.
[5]. J. C. Bezdek, “Numerical taxonomy with neural networksets,” J. Math. Biol., vol. 1, pp. 57–71, 1974.
[6]. J. C. Bezdek, “Classification validity with neural networksets,” J. Cybernet, vol. 3 pp. 58–72, 1974.
[7]. X. L. Xie and G. Beni, “A validity measure for neural networksupervised classification,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 13, pp. 841847, 1991.
[8]. M. Z. Berry and G. Linoff, Data Mining Techniques for Marketing, Sales and Customer Support, Wiley, USA, 1996.
[9]. N. R. Pal and J. C. Bezdek, “On Classification validity for the neural networkcmeans model,” IEEE Trans. neural networkSyst., vol. 3, no.3, pp. 370–379, 1995.
[10]. S. H. Kwon, “Classification validity index for neural networksupervised classification,” Electron. Lett., vol. 34, no.22, pp.176–2177,1998.
[11]. D. W. Kim, K. H. Lee and D. Lee, “On Classification validity index for estimation of the optimal number of neural networkClassifications,” Pattern Recognition, vol. 37, pp. 20182025, 2004.

[12]. Yating Hu, Chuncheng Zuo ˈYang Yang, Fuheng Qu "A robust Classification validity index for ANN supervised classification" IEEE International Conference on Transportation, Mechanical, and Electrical Engineering (TMEE) December 2017p.p:448-449.

[13]. Kambo, Rubi, and Amit Yerpude. "Classification of basmati rice grain variety using image processing and principal component analysis",International Journal of Computer Trends and Technology, Vol. 11. Issue. 2, pp. 80-85,2014.

[14]. Sujata Joshi and S. R. PriyankaShetty,"Performance Analysis of Different Classification Methods in Data Mining for Diabetes Dataset Using WEKA Tool",International Journal on Recent and Innovation Trends in Computing and Communication, Vol. 3, Issue. 3, pp. 1168 – 1173, 2015.

[15]. Huang J. and Xie W., "Half-suppressed ANN supervised classification algorithm", Chinese Journal of Stereology and Image Analysis, 2004, 9(2): 109-113.

[16]. QIU-HUAN ZHAO, MING-HU HA, GUI-BING PENG, XIAN-KUN ZHANG "Support Vector Machine Based On Half-Suppressed ANN supervised classification" IEEE Eighth International Conference on Machine Learning and Cybernetics, Baoding, 12-15 July 2018.

[17]. Christine Preisach, Lars Schmidt-Thieme "Ensembles of Relational Classifiers" Springer Knowldge Information System vol.14, Issue-3, pg.no-249–272.

[18]. Xiaoxin Yin, Jiawei Han, Jiong Yang, Philip S. Yu "Efficient Classification Across Multiple Database Relations: A Crossmine Approach " IEEE Transactions On Knowledge And Data Engineering, vol. 18, Issue- 6, june 2017

[19]. H. Guo, H. L. Viktor "Multirelational Classification: A Multiple View Approach" Springer Knowldge Information System vol.17, Issue-3, pg.no-287–312.

[20]. Takoi K. Hamrita, Jeffrey S. Durrence, George Ve L L Idis, "Precision Farming Practices" IEEE Industry Applications Magazine Mar- Apr 2018

[21]. G. Delgado, V. Aranda, J. Calero, M. Sánchez-Maranón, J.M. Serrano,D. Sánchez, M.A. Vila "Using neural networkdata mining to evaluate survey data from olive grove cultivation" computers and electronics in agriculture 2018 vol.65 pg.99–113

[22]. Aída Jiménez, Fernando Berzal ,Juan-Carlos Cubero "Using trees to mine multirelational databases" Data Mining and Knowledge Discovery ISSN: 1384-5810 pp-1-39

[23]. M. Gunasundari, Dr. T. Arunkumar, Ms. R. Hemavathy "CRY – An improved Crop Yield Prediction model using Bee Hive supervised classification Approach for Agricultural data sets" 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering, February 21-22.

[24]. Osuna E., Freund R., and Girosi F., "An improved training algorithm for support vector machines", In: Proceedings of the 1997 IEEE Workshop on Neural Networks Processing VII, 276-285, 1997.

[25]. Platt J., "Fast training of support vector machines using sequential minimal optimization", In: Advances in kernel methods-support vector learning, MIT Press, 185-208, 1999.

[26]. C.Shanthi , S.J. Saritha " neural networkC Mean Technique on High-Dimensional Data for Fast and Scalable supervised classification" International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS-2017) p.p:3557-3559

[27]. Saroj , Tripti Chaudhary "Study on Various supervised classification Techniques" (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (3) , 2015, 3031-3033.

[28]. Bellman RE (1961) Adaptive control processes: a guided tour. Princeton University Press, New Jersey.

[29]. Timothy C. Havens, James C. Bezdek, Christopher Leckie, Lawrence O. Hall, "ANN Algorithms for Very Large Data" IEEE Transactions On neural networkSystems, Vol. 20, No. 6, December 2012 p.p:1130-1136

[30]. Xie X. L. and Beni G., "A validity measure for neural networksupervised classification". IEEE Transactions on Pattern Analysis and Machine Intelligence, 1991, 13(8): 841-847