International Journal of Engineering, Business and Management (IJEBM)
ISSN: 2456-7817
[Vol-7, Issue-4, Jul-Aug, 2023]
Issue DOI: https://dx.doi.org/10.22161/ijebm.7.4
Article DOI: https://dx.doi.org/10.22161/ijebm.7.4.11

Classifier Model using Artificial Neural Network

Inderjit Kaur[1], Dr. Pardeep Saini[2]

[1]Research Scholar, Sunrise University, Alwar, Rajasthan, India
[2]Professor, Sunrise University, Alwar, Rajasthan, India

Received: 12 Jul 2023; Received in revised form: 09 Aug 2023; Accepted: 16 Aug 2023; Available online: 25 Aug 2023

Abstract— *When it comes to AI and ML, precision in categorization is of the utmost importance. In this research, the use of supervised instance selection (SIS) to improve the performance of artificial neural networks (ANNs) in classification is investigated. The goal of SIS is to enhance the accuracy of future classification tasks by identifying and selecting a subset of examples from the original dataset. The purpose of this research is to provide light on how useful SIS is as a preprocessing tool for artificial neural network-based classification. The work aims to improve the input dataset to ANNs by using SIS, which may help with problems caused by noisy or redundant data. The ultimate goal is to improve ANNs' ability to identify data points properly across a wide range of application areas.*

Keywords— ***Artificial Neural Network, supervised instance selection, Data classification, machine learning.***

I. INTRODUCTION

The primary goal of any data classifier is to appropriately categorize patterns into one of many groups that may or may not be known. The field of data classification has attracted neural networks because to its impressive not-linear function approximation and adaptive learning capabilities. The first step in any data classification process is to create a model that stands in for the various data classes, and the second is to use a model that was specifically made for classification.

These fundamentals of artificial neural networks demonstrate the sufficiency of a Feed Forward Neural Network in tackling difficult data classification problems. The development of classification models using ANNs is similarly fraught with difficulties.

Training samples for the k-Nearest Neighbor Data Classification technique are stored uniformly across n dimensions. When an unknown sample is provided, the algorithm calculates the Euclidean distance between the sample and the unknown and then searches the pattern space for the k samples that are closest to the unknown. Classification schemes that use neighboring nations as examples retain all training samples and wait to create a classification until a new sample is classified. When comparing an unlabeled sample to a large pool of potential neighbors, they may rack up hefty computational costs.

II. LITERATURE REVIEW

Narender Kumar, (2020) In machine learning, you may go one of two ways: supervised or unsupervised. Supervised learning may be used to the classification approach. Among the many classification methods available, the Artificial Neural Network stands out as the most widely used. Neural networks are useful for classifying data and creating models, but their accuracy is debatable. The artificial neural network is optimized to provide more precise and timely results. The Bat Algorithm is a metaheuristic algorithm that may be used with ANN to create a hybrid system. Optimizing the neural network has several benefits, including better classification accuracy, better data interpretation, lower costs, less time spent, etc. In this research, we evaluate the ANN Back propagation model's results for medical diagnosis against those of our proposed ANN-Bat model. Results showed that the ANN-Bat approach was superior, cutting delivery times and improving precision.

Wanto.et.al Anjar (2017) The creation of artificial neural networks is a computing paradigm that borrows heavily from the biologically inspired structure of intelligent brains.

There are many uses for artificial neural networks in the computing world. One of them stores information used for making predictions. Since the back spread algorithm can learn from historical data and identify data patterns, artificial neural networks of the back spread kind are quite popular. It's possible to analyze and forecast future events based on this background reproduction pattern. The Human Development Index from 2011-2015 is the source of information for this analysis. North Sumatra statistics from the Central Bureau of Statistics. The research used the 3-8-1, 3-18-1, 3-28-1, 3-16-1, and 3-48-1 architectural models. With an epoch of 5480 iterations and 0.0006386600 with error 0.001 to 0.05, Model 3-48-1 in architectural design has the highest accuracy of the five models, at 100%. Therefore, when employed for data prediction, the 3-48-1 back propagation approach is adequate.

MARCIN BLACHNIK (2019) Preprocessing techniques such as selecting instances and characteristics may drastically decrease computational complexity and improve prediction accuracy. Despite the widespread academic interest in finale prediction models, only few authors have delved into set selection methods. To fill this need, this research looks at four sets specifically designed for instance selection: bagging, function bagging, adaboost, and extra noise. This is the first time that last one has been seen in print. The study relies on an empirical comparison of 43 datasets and 9 fundamental instance selection procedures. There are three different types of testing. In the first, the impact of ensembles on the compression relation is shown using a single dataset for evaluation. The second case is concerned with optimizing for predicted accuracy, whereas the third case involves balancing many criteria, including data compression. The gathered data demonstrates that, with the exception of unstable methods like CNN and IB3, instance selection ensembles improve upon the fundamental instance selection algorithms, although with a compression cost. In most cases, Bagging and AdaBoost are superior. Specifically, 1NN, kNN, and SVM are tested and compared in the studies. We also discover that the prediction accuracy of robust classifiers (kNN and SVMs) based on data filtered by installation (including ensembles) decreases when compared to the results obtained in the whole training set for the training of these classifiers.

Sonam Saxena.et.al (2019) In recent years, data mining has seen rapid growth and widespread use of associated technologies. Quick conclusions may be reached by using it to analyze past data. A formalized method of making decisions has the potential to improve data protection as well. An example data mining application is shown in the material on offer. The proposed use of data mining enhances data protection. As a result, we consider the problem of how to classify URLs. In this research, we propose using association rule-mining technology to resolve URL classification, however the supervised learning technique might also be useful. Phishing and legitimate websites' URLs may be analyzed using this technique. It is proposed to apply a classification strategy based on rules to this domain. This approach may be used to classify URL information based on calculated association criteria. The inspiration for this originates from the usage of apriori algorithms for the creation and categorization of phishing URLs. The computational and memory requirements of the apriori method for generating candidate sets are high. We use the FP-Tree method, which efficiently generates lightweight association rules. This method has potential use in the development of phishing toolbars. This approach is used to compare the results of the Phish tank dataset to those of other datasets. The results indicate that the suggested approach requires less mental effort and storage space. There will soon be a more efficient and less cumbersome approach for classifying potential phishing URLs.

Jonathan Schmidt (2019) Among the many fascinating new techniques in materials science, machine learning stands out. It has been shown that basic and applied research may benefit considerably from this suite of statistical techniques. Recent years have seen a proliferation of research into using machine learning to semiconductors. We review and discuss the most recent studies on the topic. We introduce the fundamentals of machine learning, including algorithms, descriptors, and databases for materials science. We continue to detail other machine learning-based strategies for locating stable materials and predicting their crystal structure. We give studies on several strategies for replacing fundamental principles with machine learning, as well as many quantitative linkages between structures and characteristics. We investigate the potential of active learning and surgical optimization to improve rational design and associated processes. Two perennial issues with machine learning models are their lack of interpretability and physical understanding. As a result, we discuss the significance of interpretability in materials science and the different facets of this concept. Finally, we provide solutions to a variety of computational materials science problems and suggest directions for further study.

III. ARCHITECTURE OF FEED FORWARD NEURAL NETWORK

An artificial neural network is a paradigm for processing data that takes cues from the brain. It's made up of a network of neurons all working together to find a solution to a certain problem. The architecture of a three-layer feed forward neural network (FFNN) is shown in Figure 1.

All of the designs in this subclass of neural networks have one thing in common: they all use unidirectional connections between neurons in successive layers. That is to say, information may go in just one direction (the "forward direction") via a given set of branches and links. The weights of the connected branches may be adjusted according to a user-defined learning policy. Neuronal connections to other architectural layers are not made possible by feedback networks. The neuron's response is generated by feeding the linear combiner's output (the neuron's activity level) into a non-linear active function f (.).

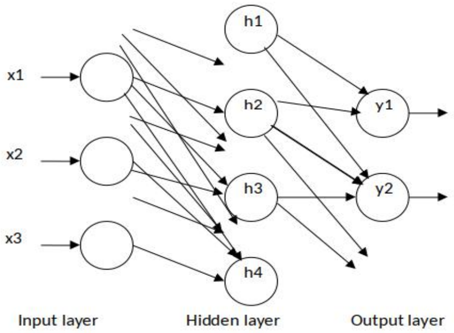

Fig.1: Architecture of Feed Forward Neural Network

The network's neuronal activity typically falls between -1 and 1, however the range [0, 1] is useful in certain contexts. There are really three distinct layers in Figure 1; The signal for the "second" layer neurons is input in the "third" layer (or the output layer), and no computations are done in the "first" layer. The network's responsiveness is measured by what comes out of the last layer, the output layer.

Non-linear mapping between inputs and outputs is possible in this network. While there may be many theoretically possible hidden levels in architecture, in practice just one or two are often used. To approximate non-linear mapping, all that's needed is a multi-layer perceptron with a single hidden layer and enough neurons. The identification of a large enough number of neurons to achieve the required approximation accuracy is notoriously challenging in practice. Therefore, the trial-and-error method is used to determine the density of the hidden layer.

IV. LEARNING IN NEURAL NETWORK

The ability to adjust to new conditions is the primary source of their resilience and strength. Throughout the process of readjustment, they construct mental models using information about their surrounding environment. These mental representations are written down as various "structured" vectors of importance. Learning algorithms describe a process that is architecture-dependent and entails the encoding of data input into weights to generate these internal models. Strengthening and weakening connections is how learning occurs.

Postsynaptic channels in biological learning systems are affected by the efficiency of the synapse, both in terms of the amount of neurotransmitters produced by a synaptic terminal and the physical shape of the axon-dender junction. In artificial systems, learning alters the model's synaptic weights.

Data is the primary engine of most learning. Input-output batteries representing data from a (perhaps unknown) probability distribution are possible. In this scenario, the output pattern may reveal the system's reaction to a given input pattern, and the learning task would then be to approximatively determine the unknown function. It's also possible that learning will be difficult since the data contains patterns that naturally cluster into several unknown classes.

For the purpose of training and testing neural networks, several different learning algorithms are at your disposal. In this research, a backpropagation-based learning algorithm is developed for use in training and evaluating the neural feed network. Details of the back propagation algorithm are laid down below.

V. BACK PROPAGATION ALGORITHM

The Back Propagation learning method is a step-down strategy for minimizing the mean square error between the observed and desired output of a multi-layer perceptron. When training a network using back propagation, a non-linear relationship is created between the input and output values. To account for the nonlinear relationship between the input and output pairs, the network may adjust its weights using the rear propagation strategy.

The method for background propagation includes:

Step 1. Weigh and offset initialization

Weights and node offsets are first set to arbitrary small values.

Step 2. Present vector input and output desired

Put forth the input x as a continuous vector and specify the output you seek. d. All members of the vector output are 0 unless they belong to the current input class.

Step 3. Compute current outputs

Get the vector of output values right now, then apply the sigmoid nonlinearity to them.

$$f(net_i) = \frac{1}{1 - e^{-net}}$$

Step 4 Adapt weights Adjust weights by $w_{ij}(t+1) = w_{ij}(t) + n_j^{\&} x_i$

Where the output of the node I, η is the sensitivity of node j, and the learning rate constant. If node j is a destination node, then $\delta_j = f(net_j)(d_j - y_j)$

Where $f(net_j)$ netj's estimated activation function is the target value for node j's output, whereas yj represents the actual value. Sensitivity is defined as where j is the index of the node and if it is an internal node.

$$\delta_j = f^{'}(net_j) \sum_k \delta_k w_{jk} \qquad (3.4)$$

where k is the accumulated weight of all nodes above layer j. Using the LMS training criteria function and the chain derivation procedure, we can derive the update equations.

Step 5. Repeat by going to step 2

If the shift in the exercise criterion is smaller than a set threshold, the workout might be considered complete. When the training error in one validation set is small enough, the cross-validation approach ceases.

After being trained, networks with fixed weights may be able to provide an output for a given input. Once the network has been trained, it may be used as a classifier model in any engineering context.

VI. MUTUAL INFORMATION-BASED FEATURE SELECTION

Concept Of Mutual Information

Entropy is a measure of the average amount of uncertainty around a random experiment. Let Y be a discrete random variable with potential values yi, I = 1, 2,... NY, and let Prob(Y=yi) = Pi be its probability distribution function to characterize a random experiment. Then, the formula gives a definition of the entropy of the random experiment,

$$H(Y) = -\sum_{i=j}^{N_y} P_i log p_i$$

The initial entropy of a random experiment may be decreased if we know more information X about it. If you know X, then the conditional entropy of a random experiment is

$$H(Y|X) = \sum_{j-1}^{N_y} P_j \left(\sum_{j-1}^{N_y} p(y_{i¿} x_j) \log P \, (y_{i¿} x_j) \right)$$

Where P_j is the probability distribution function of X with possible values x_j, j=1,2,…,N_x and $P(y_i/x_j)$ is the likelihood that yi will occur if xj does. The conditional entropy is always less than or equal to the original entropy. For any two sets of information Y and X, the mutual information I(Y; X) is the amount by which the entropy (uncertainty) is reduced:

$$I\,(Y;\,X) = H(Y) - H(Y/X)$$

As a result, the mutual information level lowers the typical degree of uncertainty regarding the experiment's random outcome Y. Mutual information is the symmetrical metric. In other words, the amount of knowledge gained about Y after seeing X is the same as the amount of knowledge gained about X after seeing Y. X is the raw data, and Y is the final class label for this function selection problem.

Computation Of Mutual Information

To compute mutual information, we must use the best information at our disposal, which is the histogram of data, to represent the probability distribution of variables that does not exist in reality. Here are the steps required to derive the inverse data from the training data histogram:

Step 1: Sort the output patterns from most numerous to least many, and then divide the sorted patterns into equality in Ny classes.

Step 2: If you don't know anything about the input variable, you can figure out the initial entropy of the output Y.

Step 3: Separate X1 into Nx equal subsets based on descending pattern similarity.

Step 4: Determine Y's entropy if and only if we know the value of X1.

Step 5: Find out what knowledge about X1 Y has that X1 does not have.

Step 6: To account for the remaining variables, repeat Steps 3–5.

VII. CONCLUSION

A classifier with strong generalizability might be constructed using a neural network with optimal topology. It is possible that the cutting process will reveal the optimal structure of neural networks. Act swiftly to find a solution by starting with a large network and gradually shrinking it to a smaller network with the goal of increasing generality. When pre-processing and/or pruning improve the classifier's performance, the data used to train it may be as basic as a set of rules for making a classification. The classification rules may be extracted with the aid of the rule extraction technique from the cut network, which is easier to comprehend as a condensed trained network. The thesis focuses on essential principles that facilitate the efficient use of neural networks in the creation of the classifier. It has led to advancements in discretization methods, pattern recognition, and neural network design. This discrete algorithm's findings show that the proposed discrete system requires less discrete time and produces more accurate classifications with a less number of intervals.

REFERENCES

[1] Anjar Wanto.et.al 2017," Analysis of Artificial Neural Network Accuracy Using Backpropagation Algorithm in Predicting Process (Forecasting)," International Journal Of Information System & Technology Vol. 1, No. 1, (2017), pp. 34-42

[2] Narender Kumar,2020," Classification using Artificial Neural Network Optimized with Bat Algorithm," International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-3, January 2020

[3] Marcin Blachnik 2019," Ensembles Of Instance Selection Methods: A Comparative Study," Int. J. Appl. Math. Comput. Sci., 2019, Vol. 29, No. 1, 151–168 DOI: 10.2478/amcs-2019-0012

[4] Sonam Saxena.et.al 2019," A Data Mining approach to Deal with Phishing URL Classification Problem," International Journal of Computer Applications (0975 – 8887) Volume 178 – No. 41, August 2019

[5] Jonathan Schmidt 2019," Recent advances and applications of machine learning in solid-state materials science," npj Computational Materials volume 5, Article number: 83 (2019)

[6] R. Vijaya Kumar Reddy.et.al 2018," A Review on Classification Techniques in Machine Learning,"InternationalJonroual of Advance Research in Sciences and Engineering Volumn no.7

[7] Anu Sharma.et.al. 2017," Literature Review and Challenges of Data Mining Techniques for Social Network Analysis," Advances in Computational Sciences and Technology ISSN 0973-6107 Volume 10, Number 5 (2017) pp. 1337-1354

[8] statistical learning theory; optimisation theory; financial econometrics; support vector machine; SVM; kernel methods. DOI: 10.1504/IJBIDM.2019.10019195

[9] Dharmender Kumar 2017," Classification Using ANN: A Review," International Journal of Computational Intelligence Research ISSN 0973-1873 Volume 13, Number 7 (2017), pp. 1811-1820

[10] Tameru Hailesilassie 2016," Rule Extraction Algorithm for Deep Neural

[11] Networks: A Review," (IJCSIS) International Journal of Computer Science and Information Security, Vol. 14, No. 7, July 2016

[12] María Pérez-Ortiz.et.al. 2016," A Review of Classification Problems and Algorithms in Renewable Energy Applications," Energies 2016, 9, 607; doi:10.3390/en9080607 www.mdpi.com/journal/energies

[13] Alvaro Osornio-Vargas.et.al. 2017," A systematic review of data mining and machine learning for air pollution epidemiology," BMC Public Health volume 17, Article number: 907 (2017)

[14] Abhishek, K., M. P. Singh, S. Ghosh and A. Anand (2012). Weather forecasting model using artificial neural network, Procedia Technology, Vol. 04, pp. 311-318.

[15] Ali-Khashashneh, E. A. and Q. A. Al-Radaideh (2013). Evaluation of discernibility matrix based reduct computation techniques, 5 th International Conference on Computer Science and Information Technology - IEEE, Amman, pp. 76-81, Jordan.