An Empirical Comparison of Various Classification Techniques

R.Vinitha\(^1\) & Dr. N. Sujatha\(^2\)

\(^1\)PG & Research Department of Computer Science, Raja Doraisingham Govt. Arts College, Sivagangai.

\(^2\)Assistant Professor, PG & Research Department of Computer Science, Raja Doraisingham Govt. Arts College, Sivagangai.

**Abstract:** Classification is a data mining process it denotes assigning the data into different classes relation to specific constraints and classification models are predicting categorical class labels and prediction models predict continuous valued functions. Classification plays the major role in data analysis. A classification task begins with a data set in which the class assignments are known. It denotes of predicting a certain outcome based on a given input. To predict the outcomes, the algorithm processes a training set containing a set of attributes and the respective outcome, usually called prediction attribute. Classification techniques for object detection were surveyed. There are many successful prediction techniques like Support vector machine SVM, GLCM, Neural Network, Back propagation neural network and decision tree.

**Keywords:** Neural Network, Support vector machine, Back propagation neural network.

1. INTRODUCTION

Data Mining is a part of knowledge discovery process. It is one of a clever technique that can be applied to extract useful patterns. Images are considered as one of the most important roles of communication in the field of computer vision. There is a use for understanding and extracting patterns. For classification of any data, the first step is data analysis, which can be used to extract models describing important classes or predict future data. Classification techniques in data mining are capable of processing a large amount of data. It can predict categorical class labels and classifies data based on the training set and class labels and can be used for classifying newly available data. Classification and prediction are forms of data analysis that can be needed to extract models describing the important data classes or to predict the future data trends. The classification predicts categorical (discrete, unordered) labels, prediction model, and continuous assignment function. Few of the most famous classification methodologies including decision tree induction, max-margin classifier (SVM), back propagation neural network classification, were discussed in this survey.

2. CLASSIFICATION TECHNIQUES

- Support Vector Machine (SVM)
- Gray Level Co-Occurrence Matrix (GLCM)
- Neural Network
- Back Propagation Neural Network
- Decision Tree

2.1 SUPPORT VECTOR MACHINE

Support vector machine are a set of uniform supervised learning methods used for classification. The goal is to learn few function that describes the relationship between observation and their labels [3]. A support vector machine constructs a hyper plane or stipulated of hyper planes in a high or infinite dimensional space, which can be used for classification, regression or other tasks [3]. Support vector machines (SVM) was developed for building an optimal binary (2-class) classifier but thereafter the technique was extended to regression and clustering problems. The working principle of SVM is to ascertain a hyper plane (linear/non-linear) which maximizes the margin.
Algorithm:
candidateSV = { closest pair from opposite classes }
while there are violating points
do
Find a violator
candidateSV = candidateSV \ violator
if any αp < 0 due to addition of c to S
then
candidateSV = candidateSV \ p
repeat till all such points are pruned
end if
end while

Advantage:
- It lure flexibility in the choice of the form of the threshold.
- Involve a non-linear transformation.
- It provides a health generalization capability.
- The issue of over fitting is eliminated.
- Discount in computational complexity.
- Few of manage decision rule complexity and Error frequency.

Disadvantage:
- Result transparency is flat.
- The track is time-consuming.
- Structure of algorithm is laborious to understand
- Determination of optimal parameters is not easy when there is nonlinearly independent training data

2.2 GRAY LEVEL CO-OCCURENCY MATRIX (GLCM):
A GLCM is recycled to extract texture feature from the images. The volume of GLCM is equal to a number of gray levels in the image [2]. This method is used to assessment image properties by considering the relationship between two neighboring pixels where the first pixel is called the reference and the second is called the neighbor pixel. The GLCM enclose information about how often a pixel with value i occurs one or other horizontally, vertically, or diagonally to adjacent pixels with the value j [4]. Where i & j are the gray level values (tone) in an image [4]. GLCM directions of investigation are Horizontal (0˚ or 180˚), Vertical (90˚ or 270˚), Right Diagonal (45˚ or 225˚), Left diagonal (135˚ or 315˚).

Genetic Algorithm to find best GLCM features:
a. Isolate image to a number of blocks each with the same size and give a label to each block.
b. Each chromosome is used to produce a sort of block Matrix.
c. Texture Features eradication using GLCM Matrix.
d. Calculate vigor function based on GLCM features.

Algorithm:
The basic GLCM algorithm is as follow:
1. Count all couple of pixels in which the first pixel has a value i, and its identical pair displaced from the first pixel by d has a value of j.
2. This count is filled in the ith row and jth column of the matrix Pd[i,j]
3. Note that Pd[i,j] is not symmetric, Because the number of pairs of pixels having gray levels[i,j]does not necessarily balanced I the number of pixel pairs having gray levels [j,i].
4. The elements of Pd[i,j]can be normalized by the break down each entry by the total number of pixel pairs.
5. Normalized GLCM N[i,j], defined by:

\[ N[i,j] = \frac{P[i,j]}{\sum\sum P[i,j]} \]

2.3 NEURAL NETWORK
A neural network includes a collection of processing elements that are highly interconnected and transform a set of inputs to a set of coveting outputs. The result is complete by the characteristics of the elements and the weights accomplish with the interconnections between them. By modifying the connections between the nodes, the network can adapt to the coveting outputs. Neural networks have been used in both anomaly detection and exploitation detection. For anomaly detection, neural networks abide modeled to learn the typical characteristics of system users and analyze significant variations from the user’s established behavior as an anomaly. In misuse detection, the neural network would receive data from the network stream and analyze the information for instances of exploitation [5].

Comparison among three back propagation algorithms used in intrusion detection. These three algorithms were:
a. The basic On-Line BackProp algorithm,
b. The Batch BackProp algorithm and
c. The Resilient BackProp algorithm.

Algorithm:
function Perceptron-Learner(problem) returns a set of (learned) weights
inputs: examples : training examples
targets : label values
rate: learning rate
local variables: weights : a node
weights ← Initialize-Weights(weights)
while not termination condition satisfied do
for i ← 1 to number of examples do
for j ← 0 to number of features do
weights[j] ← weights[j] + rate * (Target-Sum[i] - Weighted-Sum[i]) * examples[i][j]
end
if termination condition is satisfied then return weights
end

Advantage:
• Fashion to be parallelized.
• Robust on boisterous training data.
• NN advantages are that they can acclimate to new scenarios, they are fault tolerant and can deal with noisy data.

Disadvantage:
• Detriment and industries where they are being used.
• Time to train NN is probably analyze as biggest disadvantage.
• Large complication of the network structure.

2.4 BACK PROPAGATION NEURAL NETWORK:
The Back Propagation algorithm benefit supervised learning, which means that we provide the algorithm with examples of the inputs and outputs we want the network to compute and then the error is calculated. The main concept of the back propagation algorithm is to reduce the error until the neural networks learn the training data. One of the most important neural network algorithms is back propagation algorithm; this algorithm is stopped when the value of the error function has become sufficiently small. Neural network model could be created to help with classifying the new data.

Back propagation uses these error values to calculate the gradient of the loss function with respect to the weights in the network. In the second phase, this gradient is filling to the optimization method, which in turn uses it to update the weights, in an attempt to minimize the loss function.

Algorithm:

1. \( \frac{\partial E}{\partial w_{ij}} = \delta_j \cdot x_i \) for all weights and biases
2. \( \delta_j^o = c_j \cdot (1 - c_j) \) for output layer nodes using softmax
3. \( \varphi_j^i = (1 - \varphi_j^i) \cdot (1 + \varphi_j^i) \) for hidden layer nodes using tanh
4. \( \varphi_j^i = \varphi_j^i \cdot (1 - \varphi_j^i) \) for hidden layer nodes using logistic sigmoid
5. \( c_j = (c_j - \epsilon_j) \) for hidden and output layer nodes
6. \( \delta_j = \epsilon_j \cdot \varphi_j \) if j is an output node
7. \( \delta_j = \left( \sum \delta_j \cdot w_j \right) \cdot \varphi_j \) if j is a hidden node
8. \( \Delta W_{ij} = \alpha \cdot \frac{\partial E}{\partial w_{ij}} \) delta for all weights and biases
9. \( w_{ij}' = w_{ij} + \Delta W_{ij} \) update for all weights and biases

Advantage:
• Using high accuracy neural networks are able to approximate complex nonlinear mappings.
• The noise tolerance in a neural network is very flexible with respect to incomplete, missing and noisy data.
• Neural networks can be updated with present data, making them useful for dynamic environments, because it is ease of maintenance.

Disadvantage:
• There are no general methods to determine the optimal number to solving any problem.
• It is difficult to select a training data set which fully describes the problem to be used.

2.5 DECISION TREE:
A decision tree is a tree-related graph of decisions. Each branch represents the decisions to be built graphically. It is a non-parametric managed approach. It partition input into homogeneous classes. Decision tree classification approach is very useful in classification problems. [8] It is a flow chart denotes to a tree structure. Trees are constructed of a top-down recurrent divide and
conquer manner. The basic algorithm for decision tree introduction is a greedy algorithm that constructs decision trees in a top-down recursive divide-and-conquer manner.

Algorithm: ID3 (Examples, Target_Attribute, Attributes)

Create a root node for the tree

If every example is positive, Return the single-node tree Root, with label = +.

If every example is negative, Return the single-node tree Root, with label = -.

If number of predicting attribute is empty, then Return the solitary node tree Root, With label = most common value of the target attribute in the examples.

Otherwise Begin

A ← The Attribute that most excellent classified examples.

Decision Tree attribute for Root = A.

For each possible value, $v_i$, of A,

Add a new tree branch under Root, corresponding to the test $A = v_i$.

Let Examples($v_i$) be the subset of examples that contain the value $v_i$ for A

If Examples($v_i$) is empty Then below this new branch add a leaf node with label = most frequent target value in the examples

Else below this new branch add the subtree ID3 (Examples($v_i$), Target_Attribute, Attributes – {A})

End

Return Root

Advantage:
- Package generate understandable rules
- Implementation classification without much computation
- Package handle continuous and categorical variables
- Provide a fair indication of which fields are most important for prediction or classification.

Disadvantage:
- Not applicable for prediction of the continuous attribute.
- Perform ill with many class and small data.
  - Computationally costly to train. At each node, each candidate splitting field must be sorted before its best split can be found.
  - The sequence of fields is used and a search must be fashioned for optimal combining weights.
  - Pruning algorithms can also be costly since many candidate subtrees must be formed and compared.
- Do not treat strong non-rectangular regions

3. CONCLUSION:
This paper spotlight on various classification techniques (statistical and machine learning based) needs in data mining. Data mining may be used in a wide area that integrates techniques from assorted fields including machine learning, Network incursion detection, spam filtering, artificial intelligence, statistics and pattern recognition for analysis of broad volumes of data. Classification methods are typically able in modeling communications. These classification algorithms bottle implemented on different types of data sets. Hence these classification techniques show how a data canister determined and grouped when a new set of data is available. Every technique has got its own feature and limitations as given in the paper. Based on the Conditions, corresponding performance and feature each one as needed can be selected.
<table>
<thead>
<tr>
<th>Classification Techniques</th>
<th>Descriptions</th>
<th>Features</th>
<th>Limitations</th>
</tr>
</thead>
</table>
| Decision Tree             | A decision tree is a support tool that uses a tree-like graph. | • Decision trees are simple to understand and interpret.  
• It can be combined with other decision techniques. | • Low-Performance  
• Poor Resolution on Data With Complex Relationships Among the Variables |
| GLCM (Gray level co-occurrence matrix) | A statistical method of examining texture that considers the spatial relationship of pixels is the gray level co-occurrence matrix. | • A GLCM matrices are computed.  
• GLCM dissimilarity and correlation. | • Texture Features extraction using GLCM Matrix. Calculate fitness function based on GLCM features |
| Support Vector Machine    | A support vector Machine is a discriminative classifier formally defined by a separating hyper-plane. | • High accuracy.  
• Work well even if data is not linearly separable in the base feature space | • Speed and size requirements are both in training and testing is more.  
• High complexity and extensive memory requirements for classification in many cases. |
| Backpropagation algorithm | The backward propagation of errors is a common method of training artificial neural networks and used in conjunction with an optimization method such as gradient descent. | Intends to be significantly faster for training recurrent neural networks than general-purpose optimization techniques such as evolutionary optimization. | • Gradient descent with backpropagation is not guaranteed to find the global minimum of the error function, but only a local minimum.  
• It does not require normalization of input vectors; however, normalization could improve performance |
| Neural network algorithm  | A neural network is a system of hardware and software patterned after the operation of neurons in the human brain. | • It is easy to use, with few parameters to adjust.  
• A neural network learns and reprogramming is not needed. | • Requires high processing time if neural network is large.  
• Difficult to know how many neurons and layers are necessary. |

4.REFERENCES


