



IMAGE CLASSIFICATION OF AGRICULTURAL DATA USING SUPERVISED LEARNING TECHNIQUES:-A SURVEY

Nisha S Sukhwani

Department of Computer Science and Engineering
Nirma University,
Ahmedabad, Gujarat, India

Abstract: Nowadays with the increasing progress in the field of machine learning and image processing, image classification plays a vital role and provides various advantages like to classify the different varieties of wheat/rice seeds or any other agricultural seeds/grains. The literature basically includes the algorithm for feature extraction and dimensionality reduction algorithms in order to reduce error. Accuracy totally depends on a number of the ratio of samples which are divided for training and testing phase. This survey includes an artificial neural network with back-propagation having multiple hidden layer and support vector machine, these supervised learning models are taken into consideration in order to classify the types, by which through the accuracy could be known. Various issues and challenges are highlighted in this survey. It is evident from the literature that back-propagation with multiple hidden layers can reduce more error in comparison to single layer hidden layer.

Keywords: image processing, segmentation, supervised learning, computer vision, classification, machine learning, neural networks.

1. INTRODUCTION

Electric On digital images for performing image processing that uses the computer algorithms it is called to be Digital Image processing. Digital image processing has been taken into consideration. More noise and signal problems can be avoided by using various algorithms. Digital image processing is the main viable innovation for Classification, Feature extraction, Pattern recognition, Projection, Multiscale signal analysis.

According to the research and experiments, reduction of dimension can be done using feature extraction algorithms like PCA-SIFT [9]. From an initial set of measured data, features are extracted that are quite informative and non redundant by using generalization steps. Lots of memory and calculation control are required for examining most of the part. Preparation of tests are needed for grouping the calculation overfit and new specimens are summed up inadequately.

Techniques in order to develop blends of factors for overcoming the issues, feature extraction termed to be used and used for portraying important information with satisfactory precision.

Various researchers have done research using different supervised models for classification. Datasets of seeds taken in which some samples taken for training and testing phase. Accuracy totally depends on the no. of the ratio of samples which are divided for training and testing phase and which model to be selected that also totally depends on the dataset. Problem occurs by doing image resizing like noise distortion, blurring of image and signal problems.

Section 2 describes the related work on image classification using supervised learning techniques. Supervised learning techniques like ANN and SVM are described in Section 3. Section 4 gives the brief information about image processing

steps to overcome issues which troubles the researcher nowadays and feature extraction algorithms are described. Section 5 contains the conclusion of this survey paper.

2. RELATED WORK

Literature survey have done by keeping in mind the open issues facing researchers nowadays. Classification is done using various types of techniques. Image classification has been done by taking out various datasets of rice seed varieties [11] or any other seeds of crops like almonds[7], maize and wheat.

H. Saad and A. Hussain [5] proposed model for checking the ripeness of papayas which are mature, over-mature or immature using the artificial neural network as well as threshold rule but accuracy shows more while using neural networks rather than threshold rule. Neural network toolbox of some MATLAB version to be used in order to classify it.

Various researchers have done research using different supervised models for classification. Datasets of seeds taken in which some samples were taken for training and testing phase. Accuracy totally depends on the no. of the ratio of samples which are divided for training and testing phase. Which model to be selected that also totally depends on dataset.

i.Kavdir [3] proposed approach for discrimination of Helianthus, weed, and soil have done victimization by artificial neural network (ANN) with back-propagation (BPNNs) containing multiple hidden layers provides improved recognition compared to at least one hidden stratified BPNN topologies [3].

Crop Classification has additionally been done victimization mathematical logic [4] and neural network. Classification for the maturity of papayas is done victimization ANN and

threshold rule. Use by ANN is more practical than threshold rule even if time interval have taken a lot of by ANN. W. Tan, L. Sun, D. Zhang, D. Ye, and W. Che [2] proposed approach in order to improve results of Support vector machines that used for characterizing wheat grains in several quality classes done by close to infrared spectrographic analysis and support vector machine. Results obtained from SVM were higher from NIR Technology.

D. Halac, E. Sokic, and E. Turajlic [7] researchers have proposed amazing model for almond classification in which they have used support vector machine supervised model instead of the neural network. SVMs gives more accuracy in order to classify parted, broken almond or another hazelnut/almonds.

S. Gupta and S. G. Mazumdar [6] done great research on image edge detection as one of the steps in digital image processing in which Sobel operator has been used by which focused object can be detected. They have proposed sobel edge detection algorithm is useful for detecting edges.

M. R. Golzarian and R. A. Frick [8] have proposed approach for dimensionality reduction called Principal component analysis in order to classify images of wheat, ryegrass and brome grass. Image processing includes conversion of true image to gray scale image. Gray scale image is then converted into binary image. After that, color segmented image is to be further used for feature extraction process.

3. SURVEY ON SUPERVISED LEARNING TECHNIQUES

The task of inferring operation in the machine learning from labeled training knowledge is called Supervised learning. A group of training examples is to be included in training knowledge. Examples consisting of a vector that is input object and supervisory/higher signals that are called desired output object that whole components come in supervised learning. For performing mapping of new examples, production of an inferred output that gets from analyzing the training input, it is to be ruled out by supervised learning. An optimum state of affairs can afford the rule to properly confirm the category labels for unseen instances. this needs the learning| the coaching the educational rule to generalize from the training knowledge to unseen things in a very "reasonable" approach.

A. Artificial neural network

Set of connected units are predicated by ANN are called as artificial neurons (basically included in the animal brain as analogous it to biological neurons). ANN model is shown in figure 1.

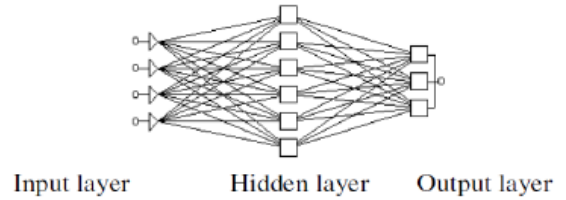


Figure 1: ANN architecture

Transmission of proof to different neuron by every association or synapse between neuron cells.[10] Signal downstream neurons which are connected to it after processing signal by receiving neuron that is post-synaptic neuron. Calculation of output of each neuron is done by the nonlinear function of the sum of its output where a synapse is a real number.

Variations of neurons and synapses occur as learning rule moves forward, by which strength of signal increases/decreases while sending downstream. [1] Aggregate signal is lower or above then only it contains threshold value and that sent by the downstream signal.

Artificial neural network having the flexibility to find out and model non linear and complicated relationships, that is basically necessary. It will infer unseen relationships on unseen information moreover, therefore creating the model generalize and predict on unseen information.

B. Support vector machine approach

Supervised Learning models includes learning algorithms which are useful in classification and regression analysis by which we get by analyzing data. A model built up by SVM training algorithm which is useful in assigning new examples to anyone classification or regression category. By assigning, it just makes a non probabilistic binary linear classifier. SVM model is shown in Figure 2.

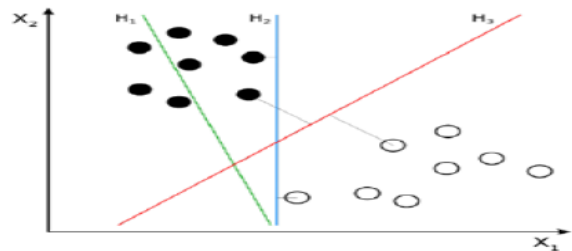


Figure 2: SVM architecture

Kernel trick is used by SVM through which nonlinear classification can be performed efficiently by the implicit method of mapping in which inputs are to be mapped into high dimensional feature spaces.

Set of hyperplanes is constructed by support vector machine which are to be useful in classification like classification in various varieties of wheat grains and also in regression. Mainly hyper plane is used in order to achieve good separation

and the hyper plane should have a maximum distance to the selected training data point and that selected training data point is called as functional margin [14,15]. So we can say that, as the margin is larger, that is directly proportional which lower the generalization error of that particular classifier which is selected.

4. DIGITAL IMAGE PROCESSING

Image Processing is the thought to be basic part as though the picture resizing and feature extraction. If it is done in the inappropriate process, the order will end up noticeably mistaken. Two fundamental operations is constrained in image processing in order to get information for an ANN classifier. Accelerating learning procedure is the first operation and edge slicing is second in order to increase the level of pixels in images which are contrasted with pixels of soil.

Image Resizing

Resizing a unique picture in image processing is the initial phase into a certain size which is perfect for programming which we are using. New image with the higher or lower no. of pixels was to be generated, when scaling or resizing of the digital image[13].

Edge cutting

Conversion of the advanced image into numerous portions called a set of pixels or superpixels. The fundamental objective of edge cutting to focus on the important object rather focusing on a less demanding object. For finding out articles and limits like lines and curves of images and many more properties, image segmentation is used. Pixels having the same name can share some of the attributes if image segmentation allots mark to every pixel in a particular image up to that extent.

Algorithm for Image segmentation

- Step 1: Sample image (Input) had been taken.
- Step 2: For noise reduction, median filtering technique has been applied to the image.
- Step 3: Find edges using Sobel edge detection algorithm.
- Step 4: Edges of the image detected (Output image).

The feature extraction algorithm has been described in next section which includes overall steps for performing feature extraction method.

Feature extraction algorithm

Principal component analysis Scale-invariant feature transformation (PCA-SIFT) feature extraction algorithm is proposed by Y. Ke and R. Sukthankar [9] for extracting local image descriptors and for extracting features. Research done in order to detect the edges of image for, edge detection algorithms are useful in smoothing of image so the connected componenets can be decreased.

Sobel edge detection

Edges to be found out by using the Sobel method [6] thereby uses a derivative approximation. when the gradient of a

The advantage of SVMs includes square measure useful in text and machine readable text categorization as their application will considerably scale back the requirement [12]. By using SVMs, one can perform classification of digital images. Higher search accuracy can be gained by using SVMs rather than using ancient schemes which use 3 to 4 rounds of connection feedback.

particular image found the maximum, then points of edges would get in output. Images having 33 dimensions have been taken into consideration for edge detection operations, so a pair of horizontal and vertical gradient matrices have been used.

Sobel edge detection algorithm steps

- Stage 1:** Input color image had been taken.
- Stage 2:** Smoothing of the image has been done in order to decrease the no. of connected components of the image.
- Stage 3:** Apply cover Gx, Gy to the information picture.
- Stage 4:** Apply Sobel edge recognition calculation and the angle.
- Stage 5:** Masks control of Gx, Gy independently on the input picture.
- Stage 6:** Results consolidated to locate the total greatness of the angle.
- Stage 7:** The total size is the yield edges

Principal component analysis Scale invariant feature transform (PCA-SIFT)

Scale-invariant feature transform (SIFT) algorithm is used for describing local features and for the detection process, it mainly includes which types of pixels we needed that can be highlighted. Uniform scaling, brightness change and invariant disturbances all these factors are invariant to uniform scaling. PCA-SIFT [9] considered to a variant of SIFT. For feature extraction, descriptor algorithm has been used called PCASIFT.

Image inclinations towards x and the y-direction located inside support region, these vector is contained in PCA-SIFT. 39*39 areas have been inspected called inclination area, in this way the vector is of measurement 3042. The measurement is decreased to 36 with PCA. Principal Component Analysis (PCA) [8] is a standard procedure for dimensionality decrease and has been connected to an expansive class of PC vision issues, counting feature determination, object recognition and face recognition.

PCA-SIFT algorithm

- Step1: Take the initial image for processing.
- Step2: Scale-Space Extrema Detection method to be done.
- Step3: Perform Key-point localization.
- Step4: Note accurate keypoint localization.
- Step5: Do orientation assignment.
- Step6: Construct keypoint descriptor.

Following points describes the whole process of the algorithm:

Scale-Space Extrema Detection : Values of sigma and no of

Octave can be modified. The first image can be generated in the first octave that could be obtained by interpolating the original one. After obtaining an original image, DoG (difference of Gaussians) pyramid would be created.

Keypoint Localisation : In order to find the extreme point, use the DoG map for the examination of each pixel.

Accurate Keypoint localization Eliminate the point with low contrast or poorly localized on an edge. If the edge is with low contrast or location of edge to be seen poorly, then elimination would have been done. On the image edge, it would also eliminate the points on it. After elimination, perform the computation of eigenspace of each extreme point.

Orientation Assignment (Main orientations assignment) : Perform searching on a certain scale. The point which too is taken into consideration, make sure to be present in the searchable area. After that, construct

weight matrix. Region pixel magnitude and region orientation would be calculated. Construct histogram on the basis of that. After construction of histogram, maximum histogram bar and higher than 80 percent maximum to be found out. Unsearchable points would be deleted and the addition of minor orientation points has to be done.

Keypoint Descriptor : The area to be divided into 4*4. After dividing, location rotation would be prepared and then after distorting, coordinate it.

Next section describes the related work of the project in which many researchers have done research in this field of computer vision using supervised learning models.

5. OPEN ISSUES AND CHALLENGES

Dealing an image classification with supervised learning is not an easy task. Due to some issues, researchers face some problems to perform the dedicated task. Following are some of the issues and challenges which needs to be addressed by researchers:

- Back-propagation with multiple hidden layers gives the highest accuracy but efficiency of model gets reduced.
- Model selection issues are facing researchers nowadays.
- Important challenge is to do classification among various combination of seeds.
- Some supervised learning model faces efficiency issues while some of giving inaccurate classification results.
- How much the efficient is classification algorithm and how many epochs to be taken to gain highest accuracy rate.

6. CONCLUSIONS

In this survey paper, research on this field had done in order to summarize the current scenario of the whole topic of the problem which researchers are facing issues so this survey would be helpful for the researcher and could bring

improvement in accuracy of classification result. A survey was done in order to solve issues coming in computer vision using machine learning algorithms. Keeping in mind the various issues, this paper provides the detailed survey which model to be selected and which algorithm is beneficial to that provides the highest accuracy that totally depends on the dataset.

Artificial neural network with back-propagation of multiple hidden layers gives the highest accuracy in comparison to threshold rule and support vector machine as well as from single layer hidden layer of backpropagation, this is possible only if there are a large number of datasets. While support vector machines (SVM) gives the best accuracy if datasets are small. Principal component analysis model is used for dimensionality reduction of images that overcome the issue of blur-ring of images. To obtain the highest accuracy of classification that will be a challenging task.

REFERENCES

- [1] Li, Z., An, Q. and Ji, C., 2008, October. Classification of weed species using artificial neural networks based on color leaf texture feature. In International Conference on Computer and Computing Technologies in Agriculture (pp. 1217-1225). Springer, Boston, MA.
- [2] Tan, W., Sun, L., Zhang, D., Ye, D. and Che, W., 2016, October. Classification of wheat grains in different quality categories by near infrared spectroscopy and support vector machine. In Cloud Computing and Internet of Things (CCIOT), 2016 2nd International Conference on (pp. 124-128). IEEE.
- [3] Kavdr, ., 2004. Discrimination of sunflower, weed and soil by artificial neural networks. Computers and Electronics in Agriculture, 44(2), pp.153-160.
- [4] Murmu, S. and Biswas, S., 2015. Application of fuzzy logic and neural network in crop classification: A review. Aquatic Procedia, 4, pp.1203-1210.
- [5] Saad, H. and Hussain, A., 2006, June. Classification for the ripeness of papayas using artificial neural network (ANN) and threshold rule. In Research and Development, 2006. SCORED 2006. 4th Student Conference on (pp. 132-136). IEEE.
- [6] Gupta, S. and Mazumdar, S.G., 2013. Sobel edge detection algorithm. International journal of computer science and management Research, 2(2), pp.1578-1583.
- [7] Halac, D., Sokic, E. and Turajlic, E., 2017, October. Almonds classification using supervised learning methods. In 2017 XXVI International Conference on Information, Communication and Automation Technologies (ICAT) (pp. 1-6). IEEE.
- [8] Golzarian, M.R. and Frick, R.A., 2011. Classification of images of wheat, ryegrass and brome grass species at early growth stages using principal component analysis. Plant Methods, 7(1), p.28.
- [9] Ke, Y. and Sukthankar, R., 2004, June. PCA-SIFT: A more distinctive representation for local image descriptors. In Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on (Vol. 2, pp. II-II). IEEE.
- [10] Najib, M.S., Taib, M.N., Ali, N.A.M., Arip, M.N.M. and Jalil, A.M., 2011, June. Classification of Agarwood grades using ANN. In Electrical, Control and Computer Engineering (INECCE), 2011 International Conference

- on (pp. 367-372). IEEE.
- [11] Pazoki, A.R., Farokhi, F. and Pazoki, Z., 2014. Classification of rice grain varieties using two Artificial Neural Networks (MLP and Neuro-Fuzzy). *The Journal of Animal Plant Sciences*, 24(1), pp.336-343.
- [12] Ahmed, F., Al-Mamun, H.A., Bari, A.H., Hossain, E. and Kwan, P., 2012. Classification of crops and weeds from digital images: A support vector machine approach. *Crop Protection*, 40, pp.98-104.
- [13] da Fontoura Costa, L. and Cesar Jr, R.M., 2009. Shape classification and analysis: theory and practice. CRC Press, Inc..
- [14] SUN, X.R., LIU, C.L., WU, J.Z., DONG, X.L. and HAN, M.L., 2011. Application of SVM method in starch classification. *Science and Technology of Food Industry*, 11, p.115.
- [15] Tellaeche, A., Pajares, G., Burgos-Artizzu, X.P. and Ribeiro, A., 2011. A computer vision approach for weeds identification through Support Vector Machines. *Applied Soft Computing*, 11(1), pp.908-915.