Study: Artificial Neural Networks classification for semantic labeling

Srinivasan Rajappa
University at Buffalo, State University of New York
New York, USA
srajappa@buffalo.edu

Abstract

Providing semantic labels to the segments in images is seen as an effort to make computer vision more in line with how humans perceive the real world. The latest research work by Google [1] in computer vision that shows how a system is able to provide auto-caption to the images. The principle of such systems is that perform feature extraction and then assign labels which may be supervised by a human or a further intelligent system. These labels along with the set of features help the system to make prediction on the new images that have these specific features. This process of semantic labeling involves feature extraction and myraid extrapolations. There exists techniques wherein a system can be trained to identify labels based on a given set of features. The training can lead the system to come up with correct labeling with minor error rates of classification. One of the classification technique is Artificial Neural Networks, which gives better results with a relatively short running time. In this paper Artificial Neural Networks is used for performing classification. The performance metric that is used here is classification error rate, which is the percentage of misclassified entities to the total number of entities. Artificial Neural Networks (ANN) classifies with a higher accuracy in comparison to other algorithms used for classifying

1. Introduction

Consider the image in figure 2. When a human sees this image the labels are identified viz. sky, building, tree, grass, road etc. The reason for this immediate recollection is the fact that there has been enough conditioning in the human brain, which allows them to identify the image. The same situation will not necessarily apply to an infant as there has not been sufficient conditioning that enables it identify the labels. Similarly, a system needs to be trained (conditioned) with all the features and the resultant labels over a plethora of images. The process of training imparts knowledge to the system and as a result it is able to perform predictions on the images that it observed in the future. This technique in computer vision consists of two steps. The first, feature extraction and the second part is using the feature set and the corresponding label values and ensure the machine is trained.

This paper demonstrates how such a learning is imparted and what results are obtained when *Artificial Neural Networks* is used for classification. Neural networks for classification provide a small error rates[1] and they are used in wide variety of applications. Such a learning where neural networks is used is often called deep learning as it has mulitple levels of non-linear transformations over a data. This is typically abstracted and helps to replicate biological neural networks. This classification is used for variety of applications apart from computer vision. The Google self-driving car is an example of such application where deep neural networks observes the environment and classifies obstacles to pedestrians to vehicles. This enables it to take corrective steps just like a human driver will take.

Thus, this is a promising area of research which can provide us with applications where pattern recognition is effective, fast and is humane.

2. Related Work

There are four key research areas which have helped and motivate the research work performed here.

Problem of Labeling Time is essential components when it comes to solving problems with asymptotic input of images. Markov Random fields optimization over a neighboring pixels helps to find labels with local features in a superpixel. The work of Joseph Tighe and Svetlana Lazebnik [2] showed that feature representation can be done without training over a large set of input images. This reasearch establishes an effective way to label images using a set of image descriptors.The other inclusive method is to detect objects in scale of images using bag of keypoints. The work by Cristopher R. Dance *et al*. [3] help to identify

[1] Google's research on providing auto-caption to images http://googleresearch.blogspot.com/2014/11/a-picture-is-worth-thousand-coherent.html

Figure 1. **Hayes Hall, UB** This image comprises of various components. The components that are involved here include sky, building, grass, tree etc.

objects which is robust and scalable. Thus, such research work helps to extract features and objects at ease for moving into the training phase.

Classification techniques SIFT features are Histogram based features that require classifiers like SVM. Due to large dimensions of the histograms the SVM is the right chocie of classifier[4]. As training data is more innate to the sphere of machine learning thus techniques like decision forests come in purview. Such techniques are better in scalability but they have their own disadvantages. This is used to concur on classification while training. The work of Jamie Shotton *et al*. [5] helps us to understand the nuances in applying such techniques.

3. Approach

In this section the approach to how classification is performed using Artificial Neural Networks is explained. For the experimental purpose dataset is used that comprises of 572 training images and 143 test images. The images comprise of various scenarios viz. a farm, an urban location etc. This introduces the disparity in the features that will be acquired from the image set. In this section the approach to how classification is performed using Artificial Neural Networks is explained. For the experimental purpose dataset is used that comprises of 572 training images and 143 test images. The images comprise of various scenarios viz. a farm, an urban location etc. This introduces the disparity in the features that will be acquired from the image set.
The strategy incorporated here is to compute the features of the image set. Since, each image will have different types of objects, scenery etc. it becomes important to modularize the image into various sets that will pertain to a group which contains the likes. Next the characteristic feature of each set is computed for all the image set in the dataset. In addition to the given data we also have labels of each set in each image which will be our ground truth. This process is also called classification as each label corresponds to a class. Afterwards, we will apply Advanced Neural Networks (ANN) on the dataset of these features and labels to train the system. The state information of the trained system is recorded and it is used on testing data that will return the error rate.
Thus, the entire process has three steps which is mentioned here in the following:

1. Image set creation.
2. Feature identification.
3. Training the system using ANN technique and Testing

The Image set creation can be understood as creating special pockets within an image where pixels share similar characteristics. One of the examples which is performed is superpixel creation. Here each superpixel consists of like pixels that help to create a segmented image. The basic rules applied while creating such pixels is to observe the boundaries and edges when there is a diffusion process that happens within a marked pixel. These superpixels can be generated over an image using Simple Linear Iterative Clustering (SLIC).
Feature Identification can be computed for each superpixel in each image. The superpixel features may include information about the color space it possess. Information about the shape of the superpixel can be used. Feature space can also contain the texture information about that superpixel. It may also contain the information about the edge information at that particular superpixel. This can be calculated by applying LM filter to the superpixel, extracting SIFT features for the superpixel etc.
Training the system using ANN will yield some state information and variables. These variables are weights that can be used in conjugate with an arbitrary data and it should yield an output that represents the class/label to which the superpixel should belong.

3.1. Artificial Neural Network

Neural Network is inspired from how a human brain works. Neural Network Model is based out of how nuerons behave. The first study that helped us to understand the potential, the experiments of Hubel and Wiesel[2] exhibits the same. Here receptors and sensors were connected to the hind side of the brain of a cat. The cat was shown a flouroscent bars in various orientations about a 2D-axis. The results that was obtained were maximum in only a specific orientation of the bar. It concluded that the nerves were behaving like a special filters that were only responding to

[2]Hubel and Wiesel- The neural basis of visual perception

a particular orientation of light. This concluded the fact that each neuron has the capability to use the inputs and relay the information to the other neurons and so on. This brought up the conclusion that the model was effective because the error-rates are neglible as biological animals tend to provide an accurate perception using the neurons. Moreover, the communication and switching time is slower in comparison to the present day computers but still face recognition/ speech recognition are all performed exceptionally well by our brain. This gives a motivation to try the Neural networks in modern computer systems that will provide low cost cum high computing systems.

Today ANN is being used to perform regression, classification and the complexity and efficacy can be increased using the hidden nodes.

Here it is noteworthy to mention the types of Neural networks. The two types are Feed forward ANN and the Recursive ANN. The feed forward neural network is faster and less complex to train but has limitations in its use. The Recursive Neural Network on the other hand can perform a lot more complex tasks but is very complex to train due to the recursive nature. The recursion here means that as opposed to other Neural Networks, we can branch from one layer to a previous layer and train the weights again. This allows us to modify the weights more easily but requires complex training algorithms. There are various types of neurons in neural networks like the Binary Threshold, Linear, Sigmoidal Neural Networks.

Despite a promising face to Artificial neural networks and techniques, it is difficult to implement a large scale neural network like the one in our brain. As an immense amount of hidden layer units are required to accomplish such a task. The mathematical expression of the same is represented below:

$$y_k(x,w) = f(\sum_{j=1}^{M} w_{kj}^{(2)} h(\sum_{i=1}^{D} w_{ji}^{(1)} x_i + w_{j0}^{(1)}) + w_{k0}^{(2)})$$

Here The hidden neural network is represented as a suprscript in the above functions. (1) represents hidden layer 1. In this layer the function applied *h(.)* is a sigmoid function.

The function *f(.)* is a softmax function which is represented by the following formula. Where a_j are activations.

$$a_j = \sum_{i=1}^{D} w_{ji}^{(1)} x_i + w_{j0}^{(1)}$$

Cross entropy error function is calculated and it is used to perfrom gradient descend. After calculation and converging upon a correct minima Error *Backpropogation* is performed. This ensures a global optimal minima is reached for the myriad cost function which we have. The same is performed using the following function.

$$\delta_k = y_k - t_k$$

$$\delta_j = z_j(1 - z_j) \sum_{k=1}^{K} w_{kj} \delta_k$$

This will find the hidden units and we use these above values to perform the gradient descent. The derivative with respect to weights in the respective units will help us converge to a relavent minima. *Note:* The amount of hidden nodes in a nueral is taken to be 2/3rd the number of features. The more the number of hidden nodes the better they will yield results.

4. Algorithm

For the given dataset the algorithm applied to accomplish the classification training and generate the error rate of the misclassifiction is here below:

1. Create Super-pixels.
2. Create Features for each superpixel over all the images in the dataset. Add the information to vector pertaining to each superpixel.
   (a) *Find and add the Orientation feature information.*
   (b) *Find and add the color component information.*
   (c) *Find and add texture information*
3. Train the data set on a Neural Network using the data information for each superpixel.[6]
   (a) *Initialize weights with random values.*
   (b) *Calculate activation units and values of hidden nodes.*
   (c) *Calculate soft max function.*
   (d) *Calculate the cross entropy.*
   (e) *Calculate derivatives with respect to weights to get error forms.*
   (f) *Perform the gradient descent till the last iteration.*
4. Testing with a given input yields erro rates.
   (a) *Use the weight vectors and calculate the prediction values.*
   (b) *Calculate Error rate with the given ground truth.*

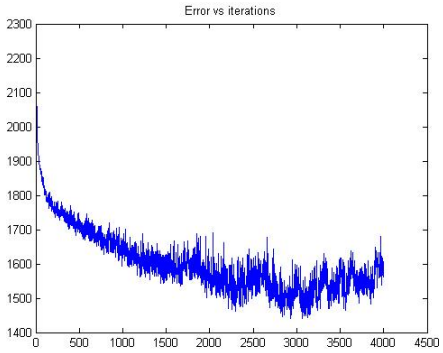

Figure 2. **Error vs Iteration** This graph shows the variation of cross entropy with varying step size.

Table 1. Error Rate with different fold

1	5.040213e+01
2	5.040614e+01
3	5.039966e+01
4	5.039403e+01
5	5.038793e+01

5. Analysis

On running the algorithm. The results mentioned in Table 1 were obtained. 5 folds of random data were used and the number of iterations for gradient descent were 4000. This yields an error rate of 50 %. Considering the results that we obtained for an SVM classifier (10 %) this implementation (ANN) has significantly high accuracy rate. The results thus obtained raise questions to why the error rate is poor. The reason may be linked to the number of feature vectors provided in comparison to the total number of data. Since, a higher order of features helps us find better classification.

6. Conclusion

The Neural network offers machine learning capabilities which help to improve efficiency and improve the running time of classification algorithms. Due to implementation constrainsts where the ANN is limited by the features and the amount of data. This yields performance below the expected standards of how Neural networks should function.

References

[1] R. Diaz, L. Gil, C. Serrano *et al*. *Comparison of three algorithms in the classification of table olives by means of computer vision*. Journal of Food engineering, Vol 61 Issue 1: 101-107, 2003.

[2] Joseph Tighe, Sletvana Lazebnik. *SuperParsing: Scalable Nonparametric Image Parsing with Superpixels* Computer Vision ECCV 2010, Volume 6315, pp 352-365, 2010

[3] Gabriella Csurka, Christopher R. Dance, Lixin Fan, Jutta Willamowski and Cdric Bray `Visual Categorization with Bags of Keypoints` Xerox Research Centre Europe 6, chemin de Maupertuis 38240 Meylan, France

[4] Chapelle, O. `Support vector machines for histogram-based image classification` Neural Networks, IEEE Transactions on (Volume:10 , Issue: 5), pp 1055 - 1064, 1999

[5] Antonio Criminisi,Jamie Shotton,Ender Konukoglu `Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning` Journal Foundations and Trends in Computer Graphics and Vision archivev Volume 7 Issue 23, pp 81-227 , February 2012

[6] Algorithm for Neural Network `Provides information to how to apply neural network program.` http://msdn.microsoft.com/en-us/magazine/jj190808.aspx