International Journal of Applied Research 2020; 6(5): 358-360

ISSN Print: 2394-7500
ISSN Online: 2394-5869
Impact Factor: 5.2
IJAR 2020; 6(5): 358-360
www.allresearchjournal.com
Received: 20-03-2020
Accepted: 24-04-2020

G Sai Sirisha
Department of Computer Science and Engineering, Lendi Institute of Engineering and Technology, Jonnada Village, Denkada Mandal, Vizianagaram district, Andhra Pradesh, India

I Mounika
Department of Computer Science and Engineering, Lendi Institute of Engineering and Technology, Jonnada Village, Denkada Mandal, Vizianagaram district, Andhra Pradesh, India

K Mukesh
Department of Computer Science and Engineering, Lendi Institute of Engineering and Technology, Jonnada Village, Denkada Mandal, Vizianagaram district, Andhra Pradesh, India

B Dharani
Department of Computer Science and Engineering, Lendi Institute of Engineering and Technology, Jonnada Village, Denkada Mandal, Vizianagaram district, Andhra Pradesh, India

Corresponding Author:
G Sai Sirisha
Department of Computer Science and Engineering, Lendi Institute of Engineering and Technology, Jonnada Village, Denkada Mandal, Vizianagaram district, Andhra Pradesh, India

Image classification using machine learning

G Sai Sirisha, I Mounika, K Mukesh and B Dharani

Abstract
Flower Species Recognition is a very hard and challenging mission to identify different types of flowers as they are very similar. The idea of automated flower recognition is bewildering as the flowers are not rigid objects and their images can be affected by many external influences .The existing system is based on classifying flowers into different categories which has very less accuracy. Now the proposed system use machine learning algorithm to fully automate and increase the accuracy of flower classification. Machine learning model will be used to extract image features and predicts the class and scientific name of the flower. The identification of flower name from an input image is based on RGB histogram data. The researchers found that the proposed system is able to classify flower images with an accuracy of 80.67%. The training set consists of 17 different types of flower species. In the proposed model we use feature extraction algorithm to extract features.

Keywords: Classification, RGB Histogram, Feature extraction

1. Introduction
Machine learning is a subject that studies how to use computers to simulate human learning activities, and to study self-improvement methods of computers that to obtain new knowledge and new skills, identify existing knowledge, and continuously improve the performance and achievement. Compared with human learning, machine learning learns faster, the accumulation of knowledge is more and facilitate the results of learning spread easier [1].

Machine learning implementations are classified into two major categories, depending on the nature of the learning “signal” or “response” available to a learning system which are as follows.In Supervised learning text classification is the automated assigning of text documents to pre-defined classes. The common approach to building a text classifier is to manually label some set of documents to pre-defined categories or classes, and then use a learning algorithm to produce a classifier. This classifier can then assign classes to future documents based on the words they contain [2].

In unsupervised learning the machine simply receives inputs x1, x2, but obtains neither supervised target outputs, nor rewards from its environment. It may seem somewhat mysterious to imagine what the machine could possibly learn given that it doesn’t get any feedback from its environment. However, it is possible to develop of formal framework for unsupervised learning based on the notion that the machine’s goal is to build representations of the input that can be used for decision making, predicting future inputs, efficiently communicating the inputs to another machine, etc. Two very simple classic examples of unsupervised learning are clustering and dimensionality reduction [3].

Semi-Supervised Learning is halfway between supervised and unsupervised learning and so, the data set is divided into labeled and unlabeled data sets.SSL is particularly devoted to application domains in which unlabeled data are plentiful, such as images processing, information retrieval, and bioinformatics [4].

Image processing includes those methods that start with an image (an array of pixels, each with a brightness or "grey scale" value or perhaps with color information) and end with an image. Usually, the resulting image is of similar size (number of pixels and number of grey levels). The brightness of the resulting pixels will, in most cases, have been modified using rules that taken into account the original value of the pixel and its neighbors, or perhaps its position or spatial relationship to many other pixels [5].

Color is one of the most important features of images. Color features are defined subject to a particular color space or model. A number of color spaces have been used in literature, such as RGB, HSV. Once the color space is specified, color feature can be extracted from images or regions. A number of important color features have been proposed in the literatures, including color histogram, color moments etc. Texture is a very useful characterization for a wide range of image. In general, color is usually a pixel property while texture can only be measured from a group of pixels. A large number of techniques have been proposed to extract texture features. Based on the domain from which the texture feature is extracted, they can be broadly classified into spatial texture feature extraction methods and spectral texture feature extraction methods.Shape is known as an important cue for human beings to identify and recognize the real-world objects. Shape feature extraction techniques can be broadly classified into two groups viz., contour based and region based methods. The former calculates shape features only from the boundary of the shape, while the latter method extracts features from the entire region as we know, images are represented in rows and columns we have the following syntax in which images are represented

$$f(x,y) = \begin{bmatrix} f(0,0) & f(0,1) & f(0,2) & \dots & f(0,N-1) \\ f(1,0) & f(1,1) & f(1,2) & \dots & f(1,N-1) \\ . & . & . & & . \\ . & . & . & & . \\ . & . & . & & . \\ f(M-1,0) & f(M-1,1) & f(M-1,2) & \dots & f(M-1,N-1) \end{bmatrix}$$

Every element of this matrix is called image element , picture element or pixel. [6]

The arrangement of paper is as follows. In Section 2, we present the recent literature on image classification. Section 3, describes the procedure and model for the identification of class of a flower.Section 4 presents the results obtained by using machine learning technique.Section 5 presents the conclusion with the extension of the future work.

2. Current Approaches of Classifying Images

Image classification is done by using different models, some of the models and techniques are review as the recent literature.

Kumar, Sandeep, Zeeshan Khan, and Anurag Jain [7] had review of machine learning approach for image classification. Various researchers apply different approach for image classification such as segmentation, clustering and some machine learning approach for the classification of image. Content of image such as color, texture and shape and size plays an important role in semantic image classification. but the proper selection of feature are challenging task of classification, so various authors apply some machine learning approach for image classification such as decision tree, RBF network, Markova model and support vector machine.

Zhou, Xi, *et al.* [8] introduces a framework for image classification using local feature descriptors. The pipeline first produces a nonlinear feature transformation on descriptors, then aggregates the results together to form image-level representations, and finally applies a classification model. For all the three steps we suggest novel solutions which make our approach appealing in theory, more scalable in computation and transparent in classification.

Wang, Fei, and Min-Yen Kan [9] have implemented a system which uses semi-supervised machine learning to create its classifier. It does this by first using class-specific keywords to build a corpus of associated images via an image search engine. Textual features are extracted from the filename, comments and URLs of the images and content-based image retrieval features are also extracted. These features are strung together as a single feature vector and fed to a machine learner to learn a model.

3. The proposed method for predicting species

The details of the proposed approach are presented in this section.

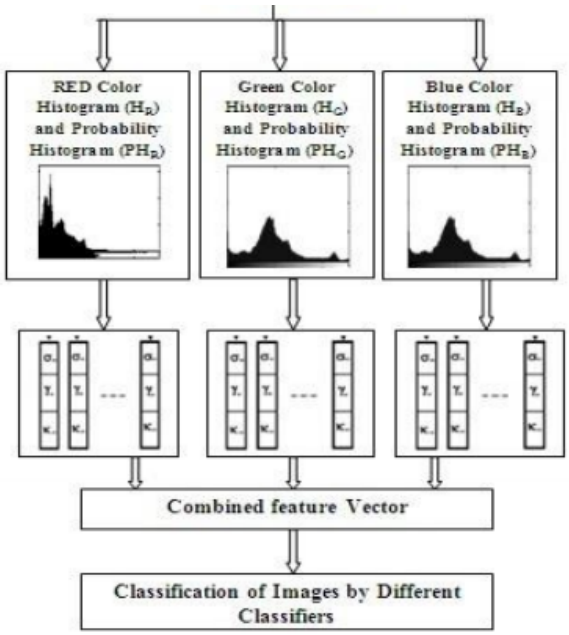
Initially an image is taken in a test folder. The global features are extracted from an image and all those global features are combined to form a single feature and store in stack using open cv library to get vector with a label and the label is compared with the training set of images present along with the labels with images and if it is found the label is printed on the image.

To obtain the features we are using a feature extraction algorithm and we are converting the colour image to gray scale image by using mahotas library. All the data and labels are stored in an hdf5 file format this format is used because it can store huge amount of data compared to .csv format.

Algorithm:

The RandomForestClassifier algorithm is a machine learning technique that is increasingly being used for image classification.Random forests is an ensemble model which means that it uses the results from many different models to calculate a response.

1. Start
2. Import required libraries and files.
3. Get the training path of the image folders.
4. Color, shape and texture extraction from images in each image folder.
5. Merge the feature extraction (Global extraction).
6. Normalize and convert the Global Extraction(to simple array using numpy)
7. Store the converted array into h5py file.
8. Train the different algorithm used using training data.
   Extract the three features of the image
   Concatenate the three features of the image into a single feature
9. Predict the label of the flower using RandomForestClassifier model.
10. End.



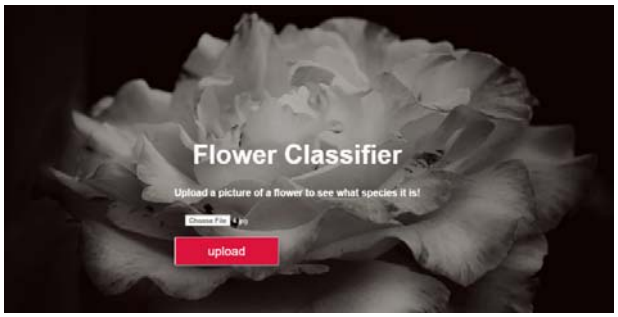
4. Results

The experimental results of the proposed approach are presented in the below section.

Use the flask application to display a front end web page for taking an input image from the user.

Steps for taking an input image are:

1. Set flask_ env=development
2. Set flask_app=predict.py
3. Flask run



Extract the global features such as color, shape, texture from an image. Taking an individual features will display the incorrect result because two images are of same color. For example daffodil is of same color like sunflower etc.

So we extract individual global features and concatenate into a single feature vector to identify the prediction.

Three methods for extracting features are:

1. fd_histogram
2. fd_haralick
3. fd_hu_moments

We combine all these features into a single one and the output of the single feature is a vector and that is compared with all the training labels along with images present in the training set.

Here we are taking different flower classifiers like support vector machine, logistic regression, decision tree and randomforestclassifier etc. Among them Random forest Classifier will give the highest accuracy which is of 80.37%.So we take random forest model to predict the image label. If the image found it displays the correct label of an image along with its scientific name.

Prediction

Flower Name: Cowslip

Scientific Name: Primula Veris

5. Conclusion

The overall flower species recognition problem is divided into three parts. Firstly, features from the images in the training dataset are extracted using Feature Extraction and indexed into a HDF5 file format. Secondly, the network is trained using various machine learning classifiers such as Linear Discriminant Analysis, Gaussian Naïve Bayes, K Nearest Neighbor, Logistic Regression, Decision Trees, Random Forests. Finally, random test images are given to the network for label prediction to evaluate the accuracy of the system. It is observed that the system correctly identifies flower species with a accuracy of 82.32% using Logistic Regression as the machine learning classifier on FLOWERS17 dataset. The FLOWERS17 dataset is split into 1360 training images. The accuracy obtained by training different machine learning classifiers on the machine learning algorithm it extracts features from the training images. Different global feature descriptors are applied on the training dataset and evaluated using random forests classifier which consistently outperformed all the other classifiers.

In future work, we will like to extend our system for more number of images per class and we can identify by combining local features for increasing accuracy and also display full information of the given flower.

6. References

1. Wang, Hua, Cuiqin Ma, Lijuan Zhou. "A brief review of machine learning and its application." 2009 international conference on information engineering and computer science. IEEE, 2009.
2. Liu, Bing, *et al.* "Partially supervised classification of text documents." ICML. 2002, 2.
3. Ghahramani, Zoubin. "Unsupervised learning." Summer School on Machine Learning. Springer, Berlin, Heidelberg, 2003.
4. Chapelle, Olivier, Bernhard Scholkopf, Alexander Zien. "Semi-supervised learning (chapelle, o. *et al.*, eds.; 2006) [book reviews]." IEEE Transactions on Neural Networks, 2009; 20.3:542-542.
5. Russ John C. "Image processing." Computer-assisted microscopy. Springer, Boston, MA, 1990, 33-69.
6. Ping Tian, Dong. "A review on image feature extraction and representation techniques." International Journal of Multimedia and Ubiquitous Engineering, 2013; 8.4:385-396.
7. Kumar Sandeep, Zeeshan Khan, Anurag Jain. "A review of content based image classification using machine learning approach." International Journal of Advanced Computer Research, 2012, 2.3:55.
8. Zhou Xi, *et al.* "Image classification using super-vector coding of local image descriptors." European conference on computer vision. Springer, Berlin, Heidelberg, 2010.
9. Wang Fei, Min-Yen Kan. "NPIC: Hierarchical synthetic image classification using image search and generic features." International Conference on Image and Video Retrieval. Springer, Berlin, Heidelberg, 2006.