International Journal of Latest Trends in Engineering and Technology
Vol.(10)Issue(2), pp.347-352
DOI: http://dx.doi.org/10.21172/1.102.56
e-ISSN:2278-621X

BIRDS SPECIES CLASSIFICATION

Potluri Rohit Sai Chowdary[1], L. Hemanth Reddy[2], K. Sugamya[3]

Abstract- Birds- a body of endothermic vertebrates and classified by wings, beaks, belly patterns, tails, forehead, etc., Birds are found globally ranging from 2in to 9ft in size. There are about ten thousand living species of birds and are ranked as the world's most numerically successful class of tetra-pods. It's almost impossible even for an ornithologist to classify each and every bird by looking at its features. We are making use of Caltech-UCSD Birds-200-2011 (CUB- 200-2011) for training and testing. We have used different algorithms for classifying different bird species into different classes and we have achieved the accuracy of 44.44% using LDA as compared to 43.7% using SVM, 33.58% using random forests, 20.4% using KNN, 19.22% using Naive Bayes.
Keywords – KNN, SVM, LDA.

1. INTRODUCTION

In our daily life we see different kinds of birds such as parrots, sparrows, etc... It is really a hectic task to figure out the specie of the bird on the first sight. Even professional photographers also get confused while classifying the birds based on their species. Classification of birds by their species is valuable for biological research and environmental monitoring applications, especially in detecting and locating them. The simple way for the researches to determine the ram of human activities on birds is identify them and maintain a count of them in the specific location.

Our main aim is to determine an algorithm that can accurately classify the bird species based on their divergent features i.e. by considering a collection of different attributes that describes the bird characteristics. Some of the characteristics for classifying birds are their color, beaks, wings, eye size, tail, etc,. Here we build a model for classifying different birds based on the above features. in our dataset we have 312 different attributes that differentiates a bird from others. We have used different algorithms for classifying such as KNN, SVM, Naive Bayes, LDA. Our model gives the accuracy for these algorithms by classifying them based on training and testing. An accuracy plot is drawn for every algorithm and at the end, the algorithm that classifies the birds with greater accuracy is determined. A final plot is shown which separates different algorithms form each other based on the accuracy.

2. CLASSIFICATION USING MACHINE LEARNING

We have used several Machine Learning classification algorithms such as:

1. Naive Bayes
2. KNN
3. SVM
4. LDA
5. Random Forest.

2.1 Classification using Naive Bayes–

Every pair of features being classified is independent of each other. Naive Bayes is a conditional probability model. Naive Bayes classifier is based on Bayes Theorem. Bayes Theorem finds the probability of an event occurring given the probability of another event that has already occurred.

Equation of Bayes theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Figure 1. Probability function for Naive Bayes

– Here, we try to find the probability of an event A, given the event B is true. Event B is termed as evidence

[1] Student, Department of Information Technology, Chaitanya Bharathi Institute of Technology, Hyderabad, Telangana, India
[2] Student, Department of Information Technology, Chaitanya Bharathi Institute of Technology, Hyderabad, Telangana, India
[3] Associate Professor, Department of Information Technology, Chaitanya Bharathi Institute of Technology, Hyderabad, Telangana, India

– P(A) is the probability of the event before the evidence is seen.
– P(A/B) is posterior probability of B, i.e. probability of event after evidence is seen.

2.2 Naive Assumption–
Now we apply the Naïve assumption on Bayes theorem to obtain the Naive Bayes theorem, which is, independence among the features. We now split the evidence into independent parts. If events are independent among each other, we get the result as

$$P(y|x_1,...,x_n) = \frac{P(x_1|y)P(x_2|y)...P(x_n|y)P(y)}{P(x_1)P(x_2)...P(x_n)}$$

Figure 2. Naive assumption for n events which can be expressed as

$$P(y|x_1,...,x_n) = \frac{P(y)\prod_{i=1}^{n} P(x_i|y)}{P(x_1)P(x_2)...P(x_n)}$$

Figure 3. Simplified function for Naive assumption

where P(y) is called class probability and P(Xi/y) is called conditional probability.
Variations or assumptions made to the distribution of P (xi / y) gives us the possibility to create different Naive Bayes classifiers. Now we apply these formulae manually on our bird species dataset.

2.3 Gaussian Naive Bayes classifier–
Here, the continuous values associated with each feature are assumed to be distributed according to gaussian distribution. Gaussian distribution is also called normal distribution. The shape of normal distribution curve is bell-shaped, and which is symmetric about the mean of the feature values.

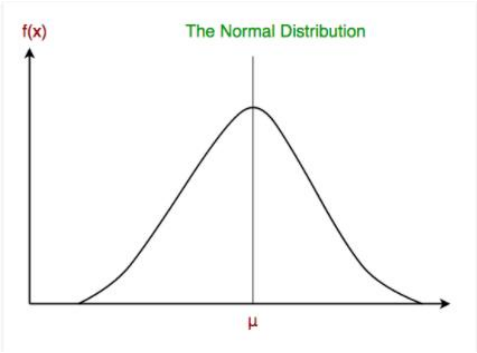

Figure 4. Normal distribution curve for Gaussian Naive Bayes classifier
Conditional probability is given by

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} exp\left(-\frac{(x_i-\mu_y)^2}{2\sigma_y^2}\right)$$

Figure 5. Conditional probability function for Gaussian Naive Bayes classifier

2.4 K-Nearest Neighbour (K-NN) –
Nearest-neighbor classifiers are built on training by metaphor, i.e. a testing tuple is compared with a given training tuple that are identical to it. The training tuples are described by n attributes. Every tuple is represented by a point in an n-dimensional space. Thus, every training tuple is stored in an n-dimensional pattern space. When an unknown tuple is given, the K-NN classifier searches for the pattern space for K training tuples that are near to the unknown tuple. These k training tuples are the k "nearest neighbors" of the unknown tuple.
When a large training sets are given K-Nearest Neighbor is labor intensive. It is widely used in the field of pattern recognition. In addition to this it also has many applications in intrusion detection and data mining.
Let n be the training data samples and p be the unknown point.
– we store n training samples in an array a []. Each element of an array represents a training tuple (x, y),
– Now calculate Euclidean distance for every training data sample. Euclidean distance id d (a [], p).
Euclidean distance is calculated by:

$$dist(\boldsymbol{X}_1, \boldsymbol{X}_2) = \sqrt{\sum_{i=1}^{n} (x_{1i} - x_{2i})^2}.$$

Figure 6. Euclidean distance for calculating nearest tuple value

– The smallest distance of k is put into S.
– Return the majority labels among S.
We are given some training data, it classifies the classes into attributes of different groups.
Consider the below figure which consists of two different classes

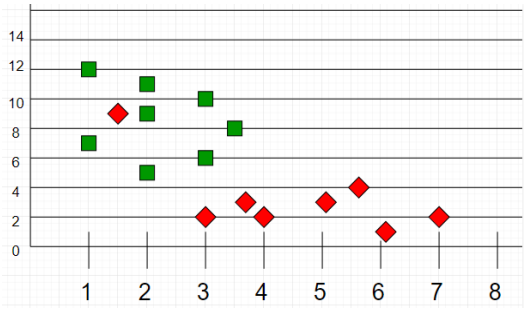

Figure 7. Represents two different classes

Now we are given with another set of classes i.e. testing data and the points are allocated by the given testing data. The unclassified tuples are uncolored.

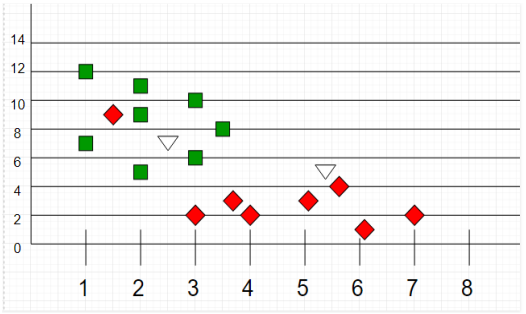

Figure 8. Represents unclassified tuples

Nearest-neighbor classifiers are very slow during the classification of testing tuples. Let is assume D as the training database of |D| tuples and k = 1, then number of comparisons required to classify the given testing tuple are O(|D|). The number of comparisons can be reduced to O(log(|D|)) by sorting the tuples into search trees. The
running time can be reduced to O (1) by parallel implementation, independent of |D|.
Other techniques to speed up classification time include the use of partial distances and editing the storing tuples. In editing method, the training tuples that are proven useless are removed. editing method is also called as pruning or condensing. In partial distances the distance between the n-tuples is calculated, if the value exceeds a threshold, the computation is stopped, and it is returned to the next step.

2.5 SVM–
Support Vector Machine(SVM) is a Machine Learning algorithm that can be used for both regression and classification. Support Vector Machines mainly consists of decision planes which define decision boundaries. A decision plane separates different objects of different classes.

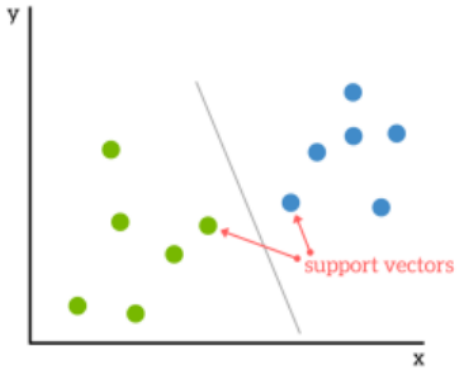

Figure 9. Represents support vectors.

The data points which are closer to the decision plane are called support vectors. A decision plane linearly separates the data. It also classifies the set of data and divides it into different classes. If the data points are far away from the plane, then they are considered as correctly classified. So, we must take care that the data points are as far as away from the decision plane. The decision plane decides the class of the data points.

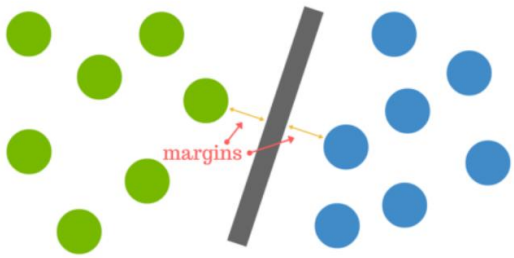

Figure 10. Represents a decision plane.

The main is to pick the decision plane with the greatest margin between the decision plane and the point within the training set, such that there is more chance for the data to be classified accurately. Margin is nothing but the distance between the decision plane and nearest data point. The illustration of margin plane is shown in the above figure.

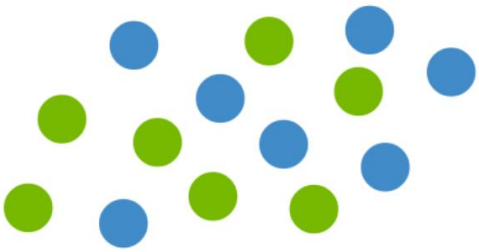

Figure 11. Represents Tuples randomly arranged without decision plane

If there is no decision plane the dataset will be jumbled representing a non-linearly separable dataset. Thus, SVM considers decision planes and decision boundaries in order to separate the dataset linearly.

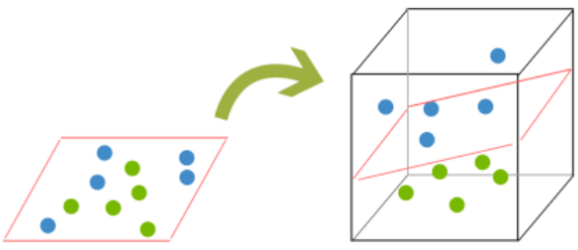

Figure 12. Represents 3-D view of tuples and decision planes in space

To classify the dataset a 3-D view is necessary, which is very simple to explain. Imagine that our data points are randomly arranged on the sheets and these sheets are raised all at a sudden moving the balls into the air. When the data points are in the air we use the sheet to separate them. The lifting of the balls means the mapping of the data into a higher dimension. The process of lifting the data points in known as kernelling.

2.6 LDA–

LDA–Latent Dirichlet allocation. Suppose you are given with the following sentences:

• I consume water and food
• Birds do fly
• Birds drink water

The main function of LDA is that it automatically detects topics which are of these type from the document.

When we give these kind of sentences, LDA classifies the different colored words into different topics. For example, when given the above sentences LDA may classify the cyan words under the topic of C, that are labeled as Consume. In the same way red words are also classified under a different topic B, that are labeled as Birds. In this way LDA labels every word and categorizes them into different topics.

The advantages of defining words under different topics are

• We can assume the content expansion of every
sentence by counting the words

Sentence 1: 100% Topic C.
Sentence 2: 100% Topic B.
Sentence 3: 50% Topic C and 50% Topic B.

• Each word can be described by determining the proportions of each word from the given sentences. If we consider the above example, C is comprised of following proportions- 30% water, 30% food, and 40% consume. and B is comprised of following proportions- 50% birds and 50% fly.

2.7 Random Forests–

Random forests are one of the ensemble methods. Assume that all the classifiers in the ensemble is a decision tree classifier so that the group can be called a 'forest'. At each node, random selection of attributes or labels is done to generate individual decision trees and to find the split. While classifying the tuples into appropriate classes, each tree votes and the majority class is returned.

Decision tree can be constructed in 2 different ways:

While constructing the decision tree classifier, if the selection of labels at each node is done randomly to determine the split, then the trees formed this way with random input selection is called Forest-RI.

Another type of Random Forest, is Forest-RC which determines the split at each node by choosing attributes that are formed by linearly combining the existing labels.

As long as the number of trees is sufficiently large, the generalization error tends to a minimum. So, overfitting isn't a problem. The accuracy of a random forest depends on the strength of the individual classifiers and the magnitude of dependence among them.

Random forests are built by aggregating N decision trees

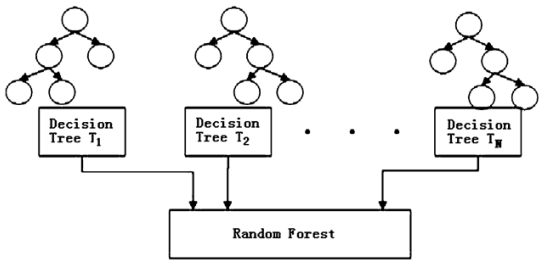

Figure 13. An image showing how a Random Forest is constructed

3. RESULTS AND DISCUSSION

We trained and tested the models built using different algorithms on the complete dataset to begin with. Later,
we randomly divided the complete dataset into training and testing data so that the sampling is done well to cover all the classes. From the complete dataset 70% of the data set was used to train the model and 30% was used as test data. The accuracy is calculated by identifying the number of correctly identified labels out of the 12,000 (approximate) tuples in the dataset. The following table shows the results that were observed after we implemented the algorithms as mentioned above.

Method	Training Accuracy	Testing Accuracy
Naïve Bayes	33.07	19.22
KNN	45.43	31.18
Random Forests	99.39	33.58
LDA	63.56	45.44
SVM	50.67	43.91

Figure 14. Comparison of testing and training accuracy

The following graph shows how different algorithms that we implemented performed in correctly classifying the bird species.

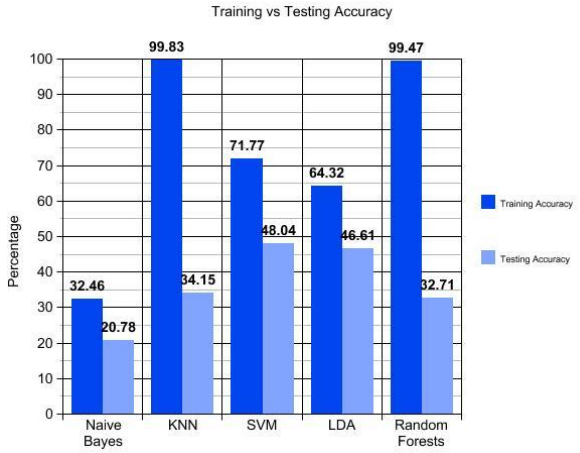

Figure 15. A graph for testing and training accuracy

We observed low accuracies with basic implementations of Naïve Bayes, KNN and Random Forests. We then observed improved accuracies by using library implementations of SVM and LDA. The accuracies observed from SVM and LDA were fairly similar and close to 50%. We believe such an accuracy for a 200 classes classification problem is decent.

4. CONCLUSION

Bird species classification is a tedious task. Machine learning algorithms can be used to build models that can effectively identify the bird species based on its features. In this paper, we initially used the Naïve Bayes approach to classification using Bayes Theorem, Naive Assumption, and Naive Bayes Classifier. Then we built a model using the K-Nearest Neighbor approach which represents every tuple in a N-dimensional space and classifies them based on the Euclidean distance. Then we built a model using the Support Vector Machine approach that tries to find the maximum marginal hyperplane using support vectors and margins to classify followed by the Latent Dirichlet Allocation approach that does the classification by assuming that every attribute is a different topic. Finally, we built a model which used the Random Forest Algorithm to conclude our experiment. The different accuracies obtained from different approaches were plotted on the graph. The work can be further enhanced by optimizing the techniques by implementing various feature selection and feature reduction techniques.

5. REFERENCES

[1] Steve Branson et al. “Bird Species Categorization Using Pose Normalized Deep Convolu- tional Nets”. In: CoRR abs/1406.2952 (2014).
[2] F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: Journal of Machine Learning Research 12 (2011), pp. 2825–2830.
[3] Wah et al. The Caltech-UCSD Birds-200-2011 Dataset. Tech. rep. CNS-TR-2011-001. California Institute of Technology, 2011.
[4] H.Kopka and P.W. Daly, A Guide to LATEX, 3rd ed Harlow, England: Addison-Wesley, 1999.
[5] J. Han, M. Kamber and J. Pei, Data mining, 3rd ed. Amsterdam: Elsevier/Morgan Kaufmann, 2012.
[6] Create A Graph", Nces.ed.gov, 2018. [Online]. Available: http://nces.ed.gov/nceskids/createAgraph/.
[7] "Classifying data using Support Vector Machines(SVMs)", GeeksforGeeks, 2018. [Online]. Available: https://www.geeksforgeeks.org/classifying-data-using-support-vector-machinessvms-in-python/.
[8] "Naive Bayes Classifiers", GeeksforGeeks, 2018. [Online]. Available: https://www.geeksforgeeks.org/naive-bayes-classifiers/.