2019 2nd International Conference on Intelligent Autonomous Systems (ICoIAS)

A Novel Solution to the Curse of Dimensionality in Using KNNs for Image Classification

Aditya D Bhat
Department of
Computer Science and Engineering
PES University
Bangalore, India
adityadb24@gmail.com

Harshith R Acharya
Department of
Computer Science and Engineering
PES University
Bangalore, India
harshithracharya@gmail.com

Srikanth H.R.
Department of
Computer Science and Engineering
PES University
Bangalore, India
srikanthhr@pes.edu

***Abstract*— The k-Nearest Neighbors (KNN) is one of the simplest and widely used algorithms in Machine Learning applications such as Image Classification. Being based on the Euclidean distance the algorithm is quite simple and effective in most cases. However, it suffers from the problem of "The Curse of Dimensionality" as the Euclidean distance becomes meaningless when the dimension of data becomes significantly high. In this paper we present a novel solution to this problem by making use of the Convolutional Neural Network (CNN) which can extract the most important features automatically from the images. These features extracted by the CNN are of reduced dimensions and can effectively be used by the KNN to recognize the images. The results and comparisons show that the proposed method is also seen to reduce the time taken for testing while retaining high accuracy. The proposed technique achieved an accuracy of 96.92% on MNIST, 85.09% on Fashion MNIST and 95.17% on the A-Z Alphabets databases respectively.**

***Index Terms*—Curse of Dimensionality, k-Nearest Neighbor (KNN), Convolutional Neural Network (CNN), Image Classification**

I. Introduction

Image Classification and Character Recognition have been widely researched areas in Computer Vision having numerous applications in various fields. Present day research [1] is focused towards developing accurate, efficient and robust classification systems using Machine Learning techniques such as K-Nearest Neighbors (KNN) [2], Support Vector Machines (SVM) [3], etc.

The developments in Deep Learning has achieved groundbreaking results in Image Classification with the emergence of Convolutional Neural Networks (CNN). CNNs are capable of extracting automatically the important features from images and have proven to be quite successful in Image Classification tasks. CNNs can be effectively used as feature extractors to

The KNN [7] [8] is a simple and lazy instance based algorithm which is based on the Euclidean distance, and is widely used for classification. Lazy meaning that the training involves only storing of the features and class labels from the training images. Classification is simply done taking a majority vote from the k nearest neighbours of an instance.

Although the KNN has been widely and effectively used in various applications with good results, it suffers from the problem of Curse of Dimensionality and irrelevant features as the dimension of the feature list increases to a high number. This arrives due to fact that the algorithm makes use of the Euclidean distance which becomes irrelevant when the dimension of features becomes substantially high [9] [10].

In this paper we propose a novel solution to this problem in using KNNs for Image Classification by using a CNN, which extracts the most important features automatically from the image. These features extracted by the CNN are of reduced dimensions and can effectively be used by the KNN to recognize the images. By doing so, we have overcome the curse that was cast by the dimensionality of the data. The hybrid model obtained from using the CNN in conjunction with the KNN is also seen to be highly accurate while also being more efficient than the stand-alone model making it more reliable as depicted from the results obtained. To test the performance of the proposed technique and compare it with that of the stand-alone models, three benchmark databases namely: MNIST, Fashion MNIST and the A-Z handwritten alphabets databases were used.

The rest of this paper is organized as follows: Section II describes the working of the k-Nearest Neighbour algorithm, Convolutional Neural Networks (CNNs) and the proposed CNN-KNN hybrid model. Section III highlights the databases used in this paper to test the models. Section IV shows