Chapter 48
Machine-Learning-Based Image Feature Selection

Vivek K. Verma
Manipal University Jaipur, India

Tarun Jain
Manipal University Jaipur, India

ABSTRACT

This is the age of big data where aggregating information is simple and keeping it economical. Tragically, as the measure of machine intelligible data builds, the capacity to comprehend and make utilization of it doesn't keep pace with its development. In content-based image retrieval (CBIR) applications, every database needs its comparing parameter setting for feature extraction. CBIR is the application of computer vision techniques to the image retrieval problem that is the problem of searching for digital images in large databases. In any case, the vast majority of the CBIR frameworks perform ordering by an arrangement of settled and pre-particular parameters. All the major machine-learning-based search algorithms have discussed in this chapter for better understanding related with the image retrieval accuracy. The efficiency of FS using machine learning compared with some other search algorithms and observed for the improvement of the CBIR system.

INTRODUCTION

In the last few decades, the dimensionality of the data associated with machine learning and information mining errands has hugely expanded. Information with high degree and dimensionality has becomes major challenge with classical learning techniques (Belarbi, Mahmoudi, & Belalem, 2017). With the expansive number of features, a learning model can be non-fitted, and it results poor accuracy. To address the information mining challenges and issue of the FS, a wide range of research reference are available. Image feature descriptors selection is the best way to eliminate noisy and redundant set from feature to be used in the classification. Major purpose of any FS method is to choose a subset of features that reduce the redundancy and maximize the significance to the target set. FS is used for refining learning

DOI: 10.4018/978-1-6684-6291-1.ch048

performance, reducing computation, and shrinking required space. To discriminate one class object from another for any pattern recognition algorithm features takes a major role. This chapter focuses on CBIR systems in which image features are extracted, and classified accordingly with one aspect of the relevancy of features with desired outcome. As a consequence FS plays a significant role CBIR and the improved selection process usually results in greater retrieval precision. Image classification is a generally contemplated issue in the analysis of images and also with, computer vision. Most of the classification frameworks can be separated into two major stages, include extraction of each image feature by a high-dimensional element vector. Next, these vectors are selected by help of classifier based on various search algorithms. Among feature extraction and the feature classification, an additional step can take place that is feature selection. This step, is about selection of a subset of features to improve the accuracy of information retrieval. Then again, include selection techniques have presently increased significant role to lessen semantic gap. Machine learning gives instruments by which vast amounts of information can be naturally examined. Feature Selection (FS) by distinguishing the most remarkable elements for learning, concentrates a learning computation on those parts of the information most helpful for investigation and future expectation. In such manner, this chapter is dedicated to show an idea of ways to deal with decrease the semantic gap using machine learning between low level visual components and irregular state semantics, through concurrent feature adjustment and highlight FS. To solve Feature Selection problem, a type of heuristic search algorithm can be used. Machine learning based search method is attractive intelligence optimization technique and many of the powerful method that has motivated and discussed in this chapter. Image feature subset selection is the way toward distinguishing and expelling however much immaterial and excess data as could reasonably be expected. This decreases the dimensionality of the information and may enable learning computations to work quicker and all the more adequately. Now and again, exactness on future characterization can be enhanced; in others, the outcome is a more reduced, effortlessly deciphered portrayal of the objective idea. On the basis of evaluation environments, FS procedures are divided into three major models as filter model, wrapper model and embedded model.

Importance of FS

Any machine learning based classification works on the simplest rule if set of features are waste (noisy data) it gives only waste as outcome. In case when size of feature set is large this becomes necessary for the relevant outcome. In most of the cases it is not necessary to include all the feature set for creating efficient algorithm. Only few set of features should be use for efficient and optimal result of the algorithms. Sometimes less is more accurate but accuracy depends on the optimal number of features i.e. not more or less but accurate. An optimal number of feature set generally reduces the time train as well as the performance time.

- **Faster Training:** One of the toughest difficulties in machine learning is getting the right data in the optimal size. An efficient CBIR system need machine learning algorithm and it commonly needs good size of training set for more accurate result but it upturns the complexity as well. If FS methods utilizes while constructing the training data set it improves the time complexity of algorithm.
- **Reduces Complexity:** Generally features selection procedures effort to discover the finest set of features that can isolate the classes but there is no open concern for problematic or informal samples and what should be used as training data. In enhancing, the algorithm picks out the fea-