Abstract

Conventionally, pattern recognition problems involve both samples and features that get collected over time or that gets generated from distributed sources. The system starts to falter when the number of features reaches a certain threshold and exhibits the curse of dimensionality. Traditionally dimensionality reduction (DR) is performed to prevent the curse of dimensionality when all features are available or when the system starts to degrade in its performance. But in the current digital age systems, the enormous and continuous influx of both samples and features mandates performing DR at regular intervals to keep up with the system performance. The massive amount of feature space and sample space that gets accumulated instantaneously allows little chance to extract the knowledge effectively that can be used promptly and hence mandates performing the DR at regular intervals of time. In real time scenarios, for any domain, decisions have to be made as and when the data is made available to realize the best outcome and to mitigate the risks. The various ways in which the features flow or get generated can be different depending on the domain of the dataspace. Due to its ever changing environment, extraction of knowledge can get more challenging. To overcome
this problem of big data, an incremental dimensionality reduction (IDR) approach to extract, carry forward, build and accumulate the knowledge without recalling the previous data is explored in this case study. Both Feature subsetting and Feature transformation methods are employed for the purpose of illustrating the incremental reduction of attributes. The hyperspectral image generated from an AVIRIS sensor provides a versatile environment required to demonstrate the in depth study of an IDR approach. This case study attempts to showcase a novel approach of maximizing the knowledge while minimizing the information loss through the use of IDR techniques in a multifaceted environment with hyperspectral data.

References

1. https://purr.purdue.edu/publications/1947/1

Index Terms

Computer Science  Information Sciences

Keywords

Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), dimensionality reduction (DR), Incremental dimensionality reduction (IDR), Principal Component Analysis (PCA), Prims like approach, Kruskals like approach