Abstract

Dimensionality Reduction process is a means to overcome curse of dimensionality in general. When all features are available together, it is a way to extract knowledge from a population in a big feature space. On the contrary, dimensionality reduction is intriguing when update to feature space is streaming and the question arises whether one could reduce the feature space as and when the features become available instead of waiting for all the features to arrive. This could not only enable the creation of knowledge that can incrementally align with the incremental access to feature space, but would also facilitate decision making locally at every incremental stage. While facilitating the local decision making parameters, it would eventually generate the most optimal reduced feature space. Moving in this direction, the possibility of implementing feature subsetting in an incremental framework is explored. The incremental streaming could be due to the temporal arrival of features or due to collection of features arriving from distributed sources. In this paper, the adoption of incremental dimensionality reduction model is also explored to observe the complexity reduction of working with a big feature space. The speciality of the proposed incremental framework is that the dimensionality reduction is performed to obtain a cumulative reduced feature space at every stage without having to look back at the earlier features.
Incremental Feature Subsetting useful for Big Feature Space Problems

References

- C. Hsu, Y. Huang, "Incremental Clustering of mixed databased on distance hierarchy, 2008.
- Jieping Ye, "IDR/QR: an incremental dimension reduction algorithm via QR."
Incremental Feature Subsetting useful for Big Feature Space Problems

decomposition, Knowledge and Data Engineering, IEEE Transactions, Vol17, Issue09, 2005
- Sheng Uei Guan et al., Incremental Learning with Respect to New Incoming Input Features, Neural Processing Letters 14, 2001
- Courses. cs.tamu.edu/rgutier/cs790_w02/l5.pdf – United States
- http://dss.princeton.edu/training
- Agrawal, H. Mannila, R. Srikant, H. Toivonen, and A. Verkamo. Fast discovery of
- Flip Korn et al., Quantifiable datamining using PCA, VLDB Journal: Very Large Data Bases, 1997
- Anil K. Jain's talk: Clustering Big Data, University of Notre Dame, Nov. 29, 2012
- http://www.coffeechoiceguide.co.uk/coffee-regions.htm
- http://www.fusion2014.org/tutorials/t14-mtpa

Index Terms
Computer Science
Information Sciences

Keywords
Big Feature Space Incremental Dimensionality Reduction Feature Selection
Optimal Feature Subset

Local Knowledge