

# Evolutionary Approach to Dimensionality Reduction

**Amit Saxena**

*Guru Ghasida University, Bilaspur, India*

**Megha Kothari**

*St. Peter's University, Chennai, India*

**Navneet Pandey**

*Indian Institute of Technology, Delhi, India*

## INTRODUCTION

Excess of data due to different voluminous storage and online devices has become a bottleneck to seek meaningful information therein and we are information wise rich but knowledge wise poor. One of the major problems in extracting knowledge from large databases is the size of dimension i.e. number of features, of databases. More often than not, it is observed that some features do not affect the performance of a classifier. There could be features that are derogatory in nature and degrade the performance of classifiers used subsequently for dimensionality reduction (DR). Thus one can have redundant features, bad features and highly correlated features. Removing such features not only improves the performance of the system but also makes the learning task much simpler. Data mining as a multidisciplinary joint effort from databases, machine learning, and statistics, is championing in turning mountains of data into nuggets (Mitra, Murthy, & Pal, 2002)

## Feature Analysis

DR is achieved through feature analysis which includes feature selection (FS) and feature extraction (FE). The term FS refers to selecting the best subset of the input feature set whereas creating new features based on transformation or combination of the original feature set is called FE. FS and FE can be achieved using supervised and unsupervised approaches. In a supervised approach, class label of each data pattern is given and the process of selection will use this knowledge for determining the accuracy of classification whereas in

unsupervised FS, class level is not given and process will apply natural clustering of the data sets.

## BACKGROUND

### Feature Selection (FS)

The main task of FS is to select the most discriminatory features from original feature set to lower the dimension of pattern space in terms of internal information of feature samples. Ho (Ho, 1998) combined and constructed multiple classifiers using randomly selected features which can achieve better performance in classification than using the complete set of features. The only way to guarantee the selection of an optimal feature vector is an exhaustive search of all possible subset of features (Zhang, Verma, & Kumar, 2005).

### Feature Selection Methods

In the FS procedures, four basic stages are distinguished:

1. **Generation procedure:** In this stage a possible subset of features to represent the problem is determined. This procedure is carried according to one of the standard methods used for this purpose.
2. **Evaluation function:** In this stage the subset of features selected in the previous stage is evaluated according to some fitness function.
3. **Stopping criterion:** It is verified if the evaluation of the selected subset satisfies the stopping criterion defined for the searching procedure.

4. **Validation procedure:** It will be used to test the quality of the selected subset of features.

FS methods can be divided into two categories: the wrapper method and the filter one. In the wrapper methods, the classification accuracy is employed to evaluate feature subsets, whereas in the filter methods, various measurements may be used as FS criteria. The wrapper methods may perform better, but huge computational effort is required (Chow, & Huang, 2005). Thus, it is difficult to deal with large feature sets, such as the gene (feature) set of a cDNA data. A hybrid method suggested by Liu (Liu, & Yu 2005) attempts to take advantage of both the methods by exploiting their different evaluation criteria in different search stages. The FS criteria in the filter methods fall into two categories: the classifier-parameter-based criterion and the classifier-free one.

An unsupervised algorithm (Mitra, Murthy, & Pal, 2002) uses feature dependency/similarity for redundancy reduction. The method involves partitioning of the original feature set into some distinct subsets or clusters so that the features within a cluster are highly similar while those in different clusters are dissimilar. A single feature from each such cluster is then selected to constitute the resulting reduced subset.

Use of soft computing methods like GA, fuzzy logic and Neural Networks for FS and Feature ranking is suggested in (Pal, 1999). FS method can be categorized (Muni, Pal & Das, 2006) into five groups based on the evaluation function, distance, information, dependence, consistency (all filter types) and classifier error rate (wrapper type). It is stated in (Setiono, & Liu, 1997) that the process of FS works opposite to ID3 (Quinlan, 1993). Instead of selecting one attribute at a time, it starts with taking whole set of attributes and removes irrelevant attribute one by one using a three layer feed forward neural network. In (Basak, Pal, 2000) neuro-fuzzy approach was used for unsupervised FS and comparison is done with other supervised approaches. Recently developed Best Incremental Ranked Subset (BIRS) based algorithm (Roberto Ruiza, José C. Riquelmea, Jesús S. Aguilar-Ruizb, 2006) presents a fast search through the attribute space and any classifier can be embedded into it as evaluator. BIRS chooses a small subset of genes from the original set (0.0018% on average) with similar predictive performance to others. For very high dimensional datasets, wrapper-based methods might be computationally unfeasible,

so BIRS turns out a fast technique that provides good performance in prediction accuracy.

## Feature Extraction (FE)

In a supervised approach, FE is performed by a technique called discriminant analysis (Steve De Backer, 2002). Another supervised criterion for non-Gaussian class-conditional densities is based on the Patrick-Fisher distance using Parzen density estimates. The best known unsupervised feature extractor is the principal component analysis (PCA) or Karhunen-Lo`eve expansion, that computes the  $d$  largest eigenvectors from  $\mathbf{D} \times \mathbf{D}$  covariance matrix of the  $n$ ,  $\mathbf{D}$ -dimensional patterns. Since PCA uses the most expressive features (eigenvectors with largest eigenvalues), it effectively approximates the data by a linear subspace using the mean squared error criterion.

## MAIN FOCUS

### Evolutionary Approach for DR

Most of the researchers prefer to apply evolutionary approach to select features. There are evolutionary programming (EP) approaches like particle swarm optimization (PSO), ant colony optimization (ACO), genetic algorithms (GA); however GA is the most widely used approach whereas ACO and PSO are the emerging area in DR. Before describing GA approaches for DR, we outline PSO and ACO.

### Particle Swarm Optimization (PSO)

Particle swarm optimizers are population-based optimization algorithms modeled after the simulation of social behavior of bird flocks (Kennedy, Eberhart, 2001). A swarm of individuals, called particles, fly through the search space. Each particle represents a candidate solution to the optimization problem. The position of a particle is influenced by the best position visited by itself (i.e. its own experience) and the position of the best particle in its neighborhood (i.e. the experience of neighboring particles called  $g_{best}/l_{best}$ ) (Engelbrecht, 2002).

5 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: [www.igi-global.com/chapter/evolutionary-approach-dimensionality-reduction/10913](http://www.igi-global.com/chapter/evolutionary-approach-dimensionality-reduction/10913)

## Related Content

---

### Clustering of Time Series Data

Anne Denton (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 258-263).  
[www.irma-international.org/chapter/clustering-time-series-data/10830](http://www.irma-international.org/chapter/clustering-time-series-data/10830)

### Text Mining Methods for Hierarchical Document Indexing

Han-Joon Kim (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 1957-1965).  
[www.irma-international.org/chapter/text-mining-methods-hierarchical-document/11087](http://www.irma-international.org/chapter/text-mining-methods-hierarchical-document/11087)

### A Novel Approach on Negative Association Rules

Ioannis N. Kouris (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 1425-1430).  
[www.irma-international.org/chapter/novel-approach-negative-association-rules/11008](http://www.irma-international.org/chapter/novel-approach-negative-association-rules/11008)

### Storage Systems for Data Warehousing

Alexander Thomasian (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 1859-1864).  
[www.irma-international.org/chapter/storage-systems-data-warehousing/11072](http://www.irma-international.org/chapter/storage-systems-data-warehousing/11072)

### Data Mining in the Telecommunications Industry

Gary Weiss (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 486-491).  
[www.irma-international.org/chapter/data-mining-telecommunications-industry/10864](http://www.irma-international.org/chapter/data-mining-telecommunications-industry/10864)