A Review on Dimensionality Reduction Techniques

Priyanka Jindal
Department of CSE, GJU S&T,
Hisar, Haryana-125001

Dharmender Kumar
Department of CSE, GJU S&T,
Hisar, Haryana-125001

ABSTRACT
Progress in digital data acquisition and storage technology has resulted in exponential growth in high dimensional data. Removing redundant and irrelevant features from this high-dimensional data helps in improving mining performance and comprehensibility and increasing learning accuracy. Feature selection and feature extraction techniques as a preprocessing step are used for reducing data dimensionality. This paper analyses some existing popular feature selection and feature extraction techniques and addresses benefits and challenges of these algorithms which would be beneficial for beginners.

Keywords
Feature Selection, Feature Extraction, Principal Component Analysis (PCA), Filter methods, Wrapper Methods.

1. INTRODUCTION
Nowadays, the growth of high-throughput technologies has resulted in exponential growth in data with respect to both dimensionality and sample size. Efficient and effective management of this data becomes challenging. It become a major problem for users that how to find useful or needed information from this vast amount of data quickly and easily. Information retrieval technology can help people to partially solve this problem.

High dimensional data is the data which contains hundreds or thousands of features. Many of these features in such dataset contains useful information for understanding the data but it also contains a lot of irrelevant or redundant features which reduce the performance, quality of data and computational efficiency [1]. DOROTHEA [2] is such a dataset used for drug discovery consist of about 1 lakh features and 2000 instances. To avoid these problems, a pre-processing step called ‘dimensionality reduction’ is used to remove these irrelevant and redundant features.

Dimensionality reduction is preprocessing in high dimensional data analysis, visualization and modeling. It has become the focus of much research in areas of applications including text processing of internet documents, gene expression array analysis, combinatorial chemistry and especially medical field [3]. Removing irrelevant features from data not only allows search algorithm to run faster and accurate but also results in high accuracy and increased computational efficiency.

A motive of this paper is to present a review on various dimensionality reduction approaches. The rest of the paper is organized as follows: Section II presents a brief review on dimensionality reduction approaches existing in literature: feature selection and feature extraction and various algorithms available for each method. Finally conclusion part is given to motivate the researchers.

2. REVIEW ON DIMENSIONALITY REDUCTION TECHNIQUES
High dimensional data is problematic due to high computational cost and memory usage [4]. There are two types of dimensionality reduction techniques in literature in data mining: 1) Feature Selection and 2) Feature Extraction / Transformation.

Feature selection [5] is the process of identifying and removing irrelevant and redundant information from data. It helps us to find good feature subsets related to target concepts. It describes several tools and techniques for reducing dimensionality of data.

Feature extraction [6] creates new feature sets that are more significant from original features by applying some transformations. The new feature sets have comparatively low dimensional subspace preserving most of the relevant information.

Feature extraction and selection methods can be used in isolated or in combination with the aim to improve performance such as estimated accuracy, visualization and comprehensibility of learned knowledge. Among the two methods of dimensionality reduction, feature extraction is more general.

2.1 Feature Selection
Feature selection [3], [7] selects a subset of original features, without any loss of useful information. It is a heuristic search process of selecting descriptors that are most effective in characterizing a specified domain. It addresses the particular task of finding a subset of given features that are useful to solve the domain problem.

The aim of feature selection is to select a subset of features as small as possible. Before applying data mining tasks, this steps needs to be applied as a pre-processing step. Feature selection offers various advantages [8] as:

- Reducing storage requirement.
- Avoiding over fitting.
- Facilitating data visualization.
- Speeding up execution of mining algorithms.
- Reducing training times.

Feature selection process consists of four steps: feature subset generation, subset evaluation, stopping criterion and result validation. It uses various search methods like complete, sequential and random search to generate small feature subset. The goodness of the generated subset is evaluated using an evaluation criterion. If the newly generated subset is better than the previous subset, it replaces the previous subset with the best feature subset. The above two processes are repeated until the stopping criterion is reached. The final best feature subset is then validated by using different tests [9]. The best feature subset is the one with least number of dimensions that most contribute to learning accuracy. The whole process of feature selection is summarized in figure 2 given below.
Feature selection methods are classified into three classes:
- Filter methods,
- Wrapper methods and
- Embedded/hybrid methods.

Relying on characteristics of data, Filter methods [10] evaluate features without utilizing any classification algorithm. Filter models use statistical properties of variables to filter out poorly informative variables. These models can be either ‘Univariate’ or ‘multivariate’. In the Univariate [11] scheme each feature is ranked independently of feature space while the multivariate scheme evaluates features in a batch way. Filter models are easily scalable to very high dimensional datasets, computationally simple and fast, the drawbacks of filter models is that they totally ignores the effect of selected feature subset on performance of induction algorithm [12].

Wrapper methods [13] utilize a predetermined learning algorithm to evaluate the quality of selected features and offer a simple and powerful way to address the problem of feature selection. The accuracy measured by this algorithm is very high as this method considers the interaction between feature subset searches and model selection. But this method is computationally expensive and lacks generality. This method is more demanding than filter methods.

Hybrid models [14] are combination of both filter models and wrapper models. They include the features of 1) filter models: they are less computationally intensive. 2) Wrapper models: they include the interaction with model construction. A block diagram for hybrid model is presented in fig 3.

The features of these three methods are summarized in table 1 given below.

Table 1: Comparison of Feature Selection Methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Features</th>
<th>Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filter</td>
<td>Suitable for very large features.</td>
<td>- Focus</td>
</tr>
<tr>
<td></td>
<td>- Computationally simple and fast.</td>
<td>- Relief</td>
</tr>
<tr>
<td></td>
<td>- Posses generality.</td>
<td>- Correlation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Coefficient</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Bw-ratio</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- mRmR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Interact</td>
</tr>
<tr>
<td>Wrapper</td>
<td>High accuracy.</td>
<td>Forward selection</td>
</tr>
<tr>
<td></td>
<td>- Lacks generality.</td>
<td>- Backward</td>
</tr>
<tr>
<td></td>
<td>- Considers interaction between feature</td>
<td>elimination</td>
</tr>
<tr>
<td></td>
<td>subset searches and model</td>
<td>- Evolutionary</td>
</tr>
</tbody>
</table>
Some of the algorithms for feature selection techniques are mRmR, Relief, Correlation Coefficient, CMIM, BW-ratio, evolutionary local selection algorithm, two-phase feature selection algorithm, interact etc. The following table provides a comparative analysis of various feature selection algorithms.

### Table 2: Various Feature Selection Algorithms

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Type</th>
<th>Approach used</th>
<th>Features</th>
<th>Drawbacks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interact</td>
<td>Filter</td>
<td>Uses symmetric uncertainty and backward elimination approach</td>
<td>-Improves accuracy</td>
<td>-mining performance decreases as dimensionality increases.</td>
</tr>
<tr>
<td>Relief</td>
<td>Filter</td>
<td>Relevance evaluation</td>
<td>-scalable to dataset with high dimensionality</td>
<td>-cannot eliminate redundant features.</td>
</tr>
<tr>
<td>Fast clustering based feature selection [17]</td>
<td>Filter</td>
<td>Graph theoretic clustering method used for clustering</td>
<td>-Highly reduced dimensionality.</td>
<td>-works well for microarray data only.</td>
</tr>
<tr>
<td>Correlation based Feature Selection[18]</td>
<td>Filter</td>
<td>Uses symmetric uncertainty</td>
<td>-handles both irrelevant and redundant features.</td>
<td>-works on smaller datasets.</td>
</tr>
<tr>
<td>Condition</td>
<td>Filter</td>
<td>Mutual information</td>
<td>-better performance</td>
<td>-Sensitive to noise</td>
</tr>
<tr>
<td>Two phase feature selection[22]</td>
<td>Hybrid</td>
<td>(filter) Artificial Neural Network Weight Analysis used to remove irrelevant features.  (Wrapper) Genetic Algorithm used to remove redundant features.</td>
<td>-handles both irrelevant and redundant feature.</td>
<td>- Improves accuracy.</td>
</tr>
</tbody>
</table>
Feature selection method also has some drawbacks:

- Search for subset of relevant features introduces an additional layer of complexity in modeling task.
- Consume more time for learning.

### 2.2 Feature Extraction/Transformation

Feature extraction [24] is a method to obtain new features from available features by applying some transformations to reduce complexity and to give a simple representation of data representing each variable in feature space as a linear combination of input variables. Feature extraction is more general method than feature selection methods. Various feature extraction methods are Principal component analysis (PCA), Non-Linear principal component analysis, independent component analysis etc.

The most popular and widely used feature extraction approach is Principal Component Analysis (PCA) [25] introduced by Karl. Principal component analysis (PCA) consists into an orthogonal transformation to convert samples belonging to correlated variables into samples of linearly uncorrelated variables. It can project the data from the original space into a lower dimensional space in an unsupervised manner. The main reason for the use of PCA concerns the fact that PCA is a simple nonparametric method used to extract the most relevant information from a set of redundant or noisy data.

Principal components (PC) are new variables with two properties: 1) each PC is a linear combination of the original variables; 2) the PC’s are uncorrelated to each other and also the redundant information is removed [12]. Main application areas of PCA include data compression, image analysis, visualization, pattern recognition, regression, and time series prediction.

PCA suffers from some limitations:

- It assumes that the relationships between variables are linear.
- Interpretation of PCA is only meaningful if all of the variables are assumed to be scaled at the numeric level. It lacks a probabilistic model structure which is important in many contexts such as mixture modeling and Bayesian decision.

### 3. CONCLUSION

In this paper, various widely used dimensionality reduction techniques are investigated with the purpose of how effectively these techniques can be used to achieve high performance of learning algorithms that ultimately improves the predictive accuracy of various classifiers. Among all the existing dimensionality reduction algorithms, some algorithms involve in removal of either irrelevant or redundant features while others involve in removal of both types of features. These algorithms develop small feature subset consisting of either same types of features as in the original feature set or derive some new features from original features depending on the need. Our study concludes that due to availability of wide variety of variables, data distributions and objectives, still there is no feature selection algorithm that is universally better than others after a lot of research. So, there is substantial need to count on criteria for feature selection which should involve in selection of best feature subset in certain situations.

### 4. REFERENCES


