Journal of Engineering and Applied Sciences 12 (Special Issue 1): 5735-5739, 2017
ISSN: 1816-949X

Feature Selection for Optimization Algorithms: Literature Survey

[1]R. Mythily and [2]W. Aisha Banu
[1]Department of Information Technology,
[2]Department of Computer Science and Engineering, B.S. Abdur Rahman Crescent University,
600048 Chennai, Tamil Nadu, India

Abstract: Feature selection and feature extraction is a preprocessing step for the effective analysis of high dimensional data. They widely used nowadays to improve the performance such as estimated accuracy, low computational performance and reducing storage. Optimization algorithms give better results but the process involved to find the optimal features are expensive. A lot of data are generating from various sources like the internet, weather forecast, medical data, etc. with huge in dimensions along with noise. The main objective is to extract useful information by reducing the dimensions of the data and to increase the efficiency and accuracy of the result. Feature selection is one of the methods used in dimensionality reduction the main aim is to select the small subset of best features from the original feature set without affecting the originality of data. Analyzing the suitable feature selection method for high dimensional data is very essential and hence it is required to survey on the various feature selection methods to improve the performance of the machine learning task. This study covers the concept of feature selection and its various methods and algorithms. We conclude this research with future research directions.

Key words: Feature selection, classification, clustering, optimization, supervise, performance

INTRODUCTION

Currently exabytes of data are growing day by day and the production of the data continues to enlarge. Feature selection has become the focus of much research in areas of application for which datasets with tens or hundreds of thousands of variables are available. Because the volume, variety, veracity and velocity are continuously growing, machine learning techniques are important to extract useful information from the huge amount of noisy data. Different machine learning techniques are involved in which feature selection is one of the methods that chooses a relevant subset of features from among the irrelevant and redundant features. Preprocessing is the data mining task its necessary step to clean the data. Mining algorithms are classified into classification and clustering.

Feature selection is a process of removing the irrelevant and redundant features from a dataset to improve the performance of the machine learning algorithms in terms of accuracy and time to build the model. The process of feature selection is classified into two categories namely feature subset selection and feature ranking methods based on how the features are combined for evaluation. The feature subset selection approach generates the possible number of combinations of the feature subsets using any one of the searching strategies such as a greedy forward selection, greedy backward elimination, etc. to evaluate the individual feature subset with a feature selection metric such as correlation, consistency, etc.

Classification of data comes under supervised learning algorithm in which the classes are already predefined before examining the data. Clustering comes under unsupervised learning algorithm in which classes are not predefined. Classification process comprises of two parts:

Training: In this, we select a certain part of the data and obtain the best features by looking at the predefined classes.

Classification: In this, we process with new data and assign to the class depends on the features we have learned during training set.

MATERIALS AND METHODS

Feature selection aims to reduce the data by selecting optimal features while reducing the features its aims to

Corresponding Author: R. Mythily, Department of Information Technology, B.S. Abdur Rahman Crescent University, 600048 Chennai, Tamil Nadu, India

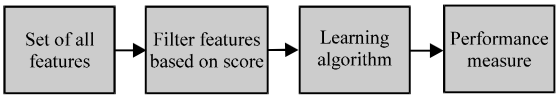

Fig. 1: Filter method

improve the classification algorithm accuracy. There are many potential benefits of variable and feature selection: facilitating data visualization and data understanding, reducing the measurement and storage requirements, reducing training and utilization times, defying the curse of dimensionality to improve prediction performance. Some methods put more emphasis on one aspect than another and this is another point of distinction between this special issue and previous research. The rest of the study deals with feature selection techniques and different optimization algorithm. The process of feature selection encompasses four steps:

- Generate the possible subset
- Evaluate the newly obtained subset
- Stopping condition
- Validation procedure-checking the performance of the model

Feature selection algorithms are classified into three types: filter, wrapper and embedded method.

Filter method: Filter method applies some measuring techniques to assign a score for each feature. Then the features are ranked or selected based on their score value. A unique characteristic of a filter method is that relevance score is calculated based only on the single feature without considering the values of neighbor features. Filter methods are based on some measures like dependency, distance and consistency measures (Fig. 1).

Examples of a ranking algorithm which are used for this process are information gain, Fisher score, Pearson's correlation, Chi-square, mutual information, correlation coefficient, etc. The advantage of this method is fast in progress and scalable with high dimensional data. The disadvantage is every feature is considered separately without dependency with other feature so it reduces the classification accuracy.

Wrapper method: In wrapper method, instead of ranking every individual feature, it forms a subset and trains the learning model using the subset. Based on the inference of the previous result, we can add or remove certain features from the subset. This model computationally more expensive than the previous model. Filter methods might fail to improve the classification accuracy because

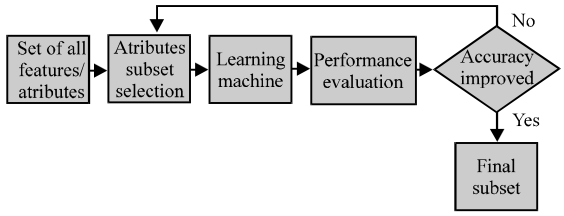

Fig. 2: Wrapper method

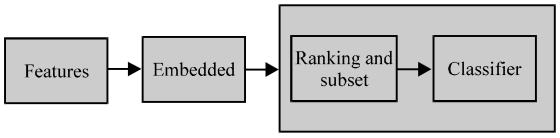

Fig. 3: Embedded method

Table 1: Comparison of the methods

Methods	Computation time	Complexity
Filter	Time efficient	Low
Wrapper	Slow (cross-validation procedure on every iteration)	High
Embedded	Quite like wrapper model	High

of the selected features but wrapper method can always provide the best subset of features. Wrapper Model uses a search algorithm to search for a possible set of features.

Examples of searching algorithms are forward selection, backward selection, Genetic algorithm, Particle Swarm Optimization, Ant Colony Optimization and Fish Swarm Optimization (Fig. 2).

Embedded method: Embedded Model is the combination of Filter and Wrapper Model. It is implemented by their own built-in feature selection methods and it does not separate learning model from the feature selection. It selects features based on the training. Examples of embedded methods include Least Angle Regression (LARS), decision trees and feature elimination. Table 1 shows the comparison of these three methods (Fig. 3).

RESULTS AND DISCUSSION

Optimization algorithms for feature selection

Ant colony optimization methods: Ant colony methods work based on the behavior of real ants searching for a food in a short way.

Sarac and Ozel (2014) worked on automated web page classification using Ant Colony Optimization to select best features better than information gain and Chi-square feature selection methods. Here, they considered each term as a different feature instead of considering a bag of words alone and they worked on with large scale feature

space. Experiments carried out in WebKB and conference dataset and result proved that tagged terms as feature extraction give better classification performance with respect to bag of terms.

Kabir *et al.* (2012) worked on ACO using neural networks. Here, they considered hybrid technique combination of both filter and Wrapper Model to select the features. They designed a new set of rules and heuristic measurements and guided the ants in the proper direction. Experimental results proved that new rules and measurements improved the performance compared to the previous feature selection algorithms.

Imani *et al.* (2012) worked on with ACO algorithm and Genetic algorithms using support vector machine as a classifier for Persian font's database classification and lead a better accuracy compared to other methods.

Meena *et al.* (2011) worked on to categories longer documents into related categories using ACO algorithm and Naive Baiyes classifier. Here, the heuristic value for each word is calculated. In this study, they have chosen Wrapper Model and the execution time was taken was high in order to reduce that algorithm are parallelized, so they implemented the ACO algorithm in Hadoop framework using map-reduce programming. Experiments carried out with the Newsgroup 20 dataset and Reuters-21578 benchmark. They observed that accuracy of the classifier is improved and time also reduced because of parallelization and without losing the quality.

Aghdam *et al.* (2009) focus on ant colony algorithm along with Bayesian classification. In this study, they compared the proposed ACO algorithm with PSO on selecting best features with the postsynaptic dataset. Evaluation of the classifier proves that proposed method is higher in accuracy and performance.

Welikala *et al.* presented (Xue *et al.*, 2013) certain feature selection research, the Genetic algorithm is adopted to generate the feature subsets for evaluation and the supervised machine learning algorithm is used to evaluate the generated subsets of feature selection using Genetic algorithm with Support Vector Machine (SVM) for mining the medical dataset.

Tu *et al.* (2006) presented a hybrid feature selection method by combining filter and wrapper method for text classification. In this method, information gain measure is used for ranking the noteworthy features and the Genetic algorithm is used as the searching strategy with support vector machine.

In general, processing the high-dimensional data is a difficult task with the wrapper method therefore Bermejo *et al.* developed a hybrid feature selection method known as filter-wrapper approach. Kannan and

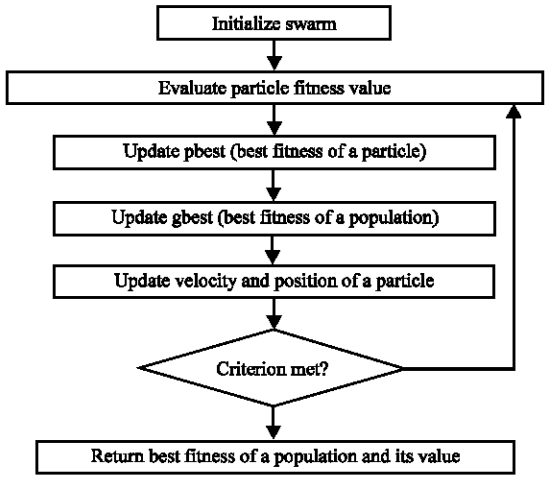

Fig. 4: PSO based feature selection

Faez presented a hybrid feature selection framework. In this approach, Ant Colony Optimization (ACO) based Local Search (LS) is used with the symmetric uncertainty measure to rank the features.

Particle swarm optimization: PSO is meta-heuristic approach it simulates the behavior of birds and it was proposed by Kennedy and Eberhart in the year 1995. PSO finds for optimal solution by continuously changing the position and velocity based on the experience of the flying particle and it group towards gbest and pbest in following iterations. Best fitness value of population represents gbest and best fitness value of particle represents pbest (Fig. 4).

Xue *et al.* (2013) focus on particle swarm optimization with multi-objective approach and they designed an algorithm with the goal of selecting smaller features and to reduce the error rate. First they cover about commonly used PSO technique with single objective that only helps to reduce the error rate and they enhanced to 2SFS (Two-Stage Feature Selection process) technique with multi-objective approach and in the first stage the fitness function is used to minimize the error and second level takes output of the first level as input and focused on minimizing the features.

Tu *et al.* (2006) proposed a feature selection method using PSO and classified using SVM and they used five classification problems. Experimental results proved that accuracy improved than other feature selection methods.

Brahim and Limam (2016) present a novel feature selection algorithm from trained by neural network using a genetic algorithm which is meant for inding the optimal relevant features which maximize the output function of trained artificial neural network. Singh propose a Fast

Correlation Based Filter (FCBF) method which can identify relevance and redundancy among relevant features using the concept of predominant correlation. Therefore, the new feature selection methods are constantly increasing to tackle the specific problem with different strategies. Many approximation methods have been proposed for feature selection based on computational intelligence methods. Due to their acceptable performances, they have attracted a lot of attention. Since, most of these methods use the learning algorithm in their processes, they are classified as varieties of the wrapper approach. Swarm intelligence is a computational intelligence based approach which is made up of a population of artificial agents and inspired by the social behavior of animals in the real world. Each agent performs a simple task while the colony's cooperative work will solve a hard problem. Swarm intelligence based methods can be used to find an approximate solution to a hard-combinatorial optimization problem. One of the most popular algorithms in the research field of swarm intelligence is Ant Colony Optimization (ACO). ACO uses knowledge from previous iterations to achieve better solutions. Moreover, it can be easily implemented; thus, it is widely used in the feature selection area.

In recent years, some ant colony optimization based methods for feature selecting are reported. Ant colony optimization is used to select features for predicting post-synaptic activity in proteins. An ant colony optimization based algorithm to reduce features size in automatic speaker verification is also proposed. The research by Das *et al*. (2012) proposed an ant colony optimization based feature selecting algorithm for text categorization. An improved ant colony optimization algorithm for feature selection in face recognition. Their algorithm can select the optimal feature subset in terms of shortest feature length and the best performance of classifier proposed an algorithm for feature selection based on two cooperative ant colonies which minimizes two objectives: the number of features and the classification error. Two pheromone matrices and two different heuristics are used for these objectives.

In most ACO-based feature selection methods, the nodes in the graph are used to represent the features. Ants traverse on the graph to look for a path containing part of the nodes which indicate the features selected. For n features, most ACO-based feature selection methods use a complete graph with O(n2) edges. This means that O(n2) pheromone and heuristic information will be computed and stored.

This literature review studies of existing feature selection methods have been done an experimental study of eight filter methods is used in 33 datasets and for the text classification problem, 12 feature selection methods are compared. The capability of the survival algorithm and tuned approach are evaluated by Ferreira *et al*. Table 2 shows the different algorithms and procedures that are available for feature selection. Feature selection algorithms can also be divided into two categories according to the way they process and evaluate features: feature ranking methods and subset selection methods. Feature ranking methods rank the features by a metric and eliminate all features that do not achieve an adequate score. Subset selection methods search the set of possible features for the optimal subset. Another efficient approach to feature selection is using Non-negative Matrix Factorization (NMF). Many feature selection algorithms involve heuristic or random search strategies in order to reduce the computing time.

Table 2: Selection strategies

Methods	Exponential	Sequential	Random
Filter			
Distance	B&B, BFF, BOBRO, OBLIVIIN	Relif S, SFS, Segen's, SBS	-
Information	MDLM, CARDIE	DTM, Koller, FG, FCBF, BSE	Dash's, SBUD
Dependency	Bobrowski's	CFS, RRESET, POE+ACC, DVMM	Mitra's
Consistency	FOCUS, ABB, MIFESI, Schlimmer's	Set Cover, WINNOW	LIV, QBB, LVF
Wrapper			
Predictive accuracy or cluster goodness	BS, AMB&B, FSLC FSBC, CARDIE, OBLIVIIN	SBS-SLASH, WSFG, WSBG, BDS, PQSS, RC, SS, Queiros, BSE, K2-AS, RACE, SBS-W, SBS-SLASH	SA, RGSS LVW, RMHC-PF, GA, RVE
Embedded			
Filter+wrapper	BBHFS, Xing's	Dash-Liu's	-

CONCLUSION

In this research, the literature comprises many definitions of feature selection process and relevance for optimal feature subsets. The general procedure of feature selection is described with subset generation, evaluation of subsets and stopping criteria. Three general approaches of feature selection methods, namely filter, wrapper and embedded methods are described in detail and their pseudo code is also presented. Image feature selection is an important task which can significantly affect the performance of image classification and recognition.

The feature selection approaches are categorized into filter, wrapper and embedded methods. The review shows that feature selection as a crucial step in data mining and machine learning applications. Feature selection is to

identify a related subset of features in the datasets and remove any other feature as irrelevant and redundant information content. Data mining algorithms can be activated faster with improved accuracy by using feature selection methods. The categorization and characteristics of feature selection are reviewed and the interesting facts regarding the advantages and disadvantages of feature selection methods to handle the distinctive characteristics of the real-world applications are enumerated.

REFERENCES

Aghdam, M.H., J. Tanha and M. Basri, 2009. Combination of ant colony optimization and Bayesian classification for feature selection in a bioinformatics dataset. J. Comput. Sci. Syst. Biol., 2: 186-199.

Brahim, A.B. and M. Limam, 2016. A hybrid feature selection method based on instance learning and cooperative subset search. Pattern Recognit. Lett., 69: 28-34.

Das, N., R. Sarkar, S. Basu, M. Kundu, M. Nasipuri and D.K. Basu, 2012. A genetic algorithm based region sampling for selection of local features in handwritten digit recognition application. Applied Soft Comput., 12: 1592-1606.

Imani, M.B., T. Pourhabibi, M.R. Keyvanpour and R. Azmi, 2012. A new feature selection method based on ant colony and genetic algorithm on Persian font recognition. Intl. J. Mach. Learn. Comput., 2: 278-282.

Kabir, M.M., M. Shahjahan and K. Murase, 2012. A new hybrid ant colony optimization algorithm for feature selection. Expert Syst. Appl., 39: 3747-3763.

Meena, M.J., K.R. Chandran, A. Karthik and A.V. Samuel, 2011. A parallel ACO algorithm to select terms to categorise longer documents. Intl. J. Comput. Sci. Eng., 6: 238-248.

Sarac, E. and S.A. Ozel, 2014. An ant colony optimization based feature selection for web page classification. Sci. World J., 2014: 1-16.

Tu, C.J., L.Y. Chuang, J.Y. Chang and C.H. Yang, 2006. Feature selection using PSO-SVM. IAENG Int. J. Comput. Sci., 33: 138-143.

Xue, B., M. Zhang and N. Browne, 2013. Particle swarm optimization for feature selection in classification: A multi-objective approach. IEEE. Trans. Cybern., 43: 1656-1671.