M.Ananthi *et al*, International Journal of Computer Science and Mobile Applications,
Vol.5 Issue. 9, September- 2017, pg. 23-31 **ISSN: 2321-8363**
Impact Factor: 4.123

SURVEY ON CLASSIFICATION OF FEATURE SELECTION STRATEGIES

M.Ananthi
M.Phil. Scholar
Avc College of Autonomous
Mayiladuthurai

K.Palanivel
Asst. Prof
Avc College of Autonomous
Mayiladuthurai

ABSTRACT: Feature Selection is the important component in the preprocessing of data mining classification problem. In this paper we can examine the different feature selection methodologies in terms of need, method adopted for feature selection. The three types of method are mainly available, such as forward selection strategies, backward elimination and optimized selection. In this article, we aim to provide a basic introduction to feature selection including basic concepts, classifications of existing systems, recent development, and applications.

Keywords: Feature selection, Feature selection methods, classification, preprocessing.

INTRODUCTION

Data mining can be used in context of many things such as pattern recognition, statistics, and database management and computer science. It can be used to discover relation between large sets of data. Feature selection is an important task in effective data mining. In this paper we discuss three different strategies for the reduction or the selection of the feature of the given dataset. They are backward elimination, forward selection and optimized selection. Feature selection, as a dimensionality reduction technique, aims to choosing a small subset of the relevant features from the original ones by removing irrelevant, redundant or noisy features. Feature selection usually leads to better learning performance, i.e., higher learning accuracy, lower computational cost, and better.

Feature selection methods can be broadly classified into supervised, unsupervised, and semi-supervised methods. In terms of different selection strategies, feature selection can be categorized as filter, wrapper and embedded models [1]. In machine learning and statistics Feature selection also known as variable selection, attribute selection is the process of selecting a subset of relevant features**.** A feature selection can be seen as the combination of search

techniques for proposing new feature subsets. Along with an evaluation measure which score the different feature subsets.

Importance of feature selection

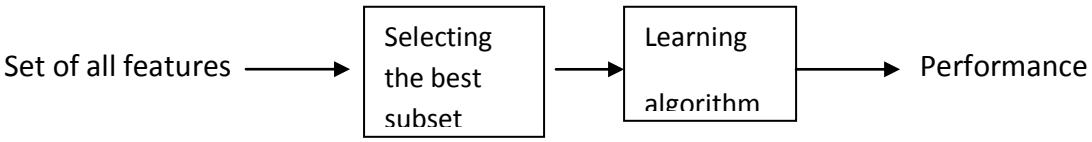
When the numbers of features are very large .you need not use every feature at your disposal for creating an algorithm. You can assist your algorithm by use only those features that is really important.

Main reason to use feature selection are:

- It enables the machine learning algorithm to train faster
- It reduce the complexity of a model and makes it easier to interpret
- It reduce the over fitting
- It improve the accuracy of a model if the right subset is chosen

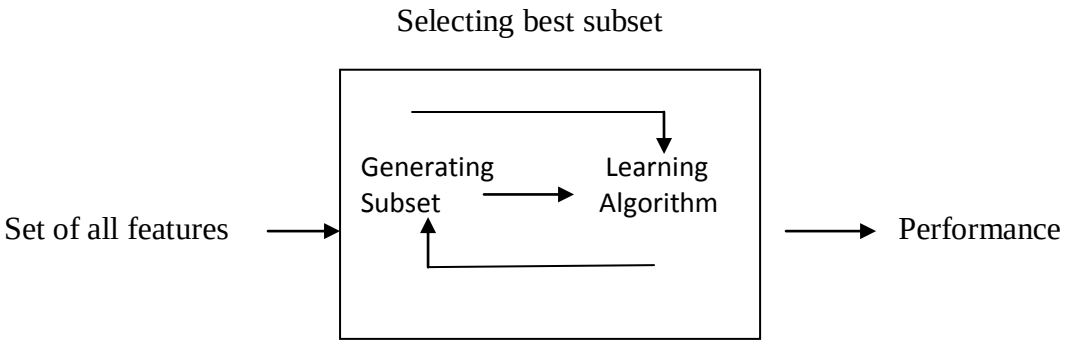
FILTER METHOD

Filter methods are generally used as a preprocessing step. The selection of feature is independent of any machine learning algorithms. Instead features are selected on the basis of their scores in various statistical tests for their correlation with the outcome variable. For filter models, features are selected based on the characteristics of the data without utilizing learning algorithms. This approach is very efficient. However, it doesn't consider the bias and heuristics of the learning algorithms. Thus, it may miss features that are relevant for the target learning algorithm. A filter algorithm usually consists of two steps [2]. In the first step, features are ranked based on certain criterion. In the second step, features with the highest rankings are chosen. A lot of ranking criteria, which measures different characteristics of the features, are proposed [3]. the ability to effectively separate samples from different classes by considering between class variance and within class variance, the dependence between feature and the class label, the correlation between feature-class and feature, the ability to preserve the manifold structure, mutual information between the features, and so on.



WRAPPER METHOD

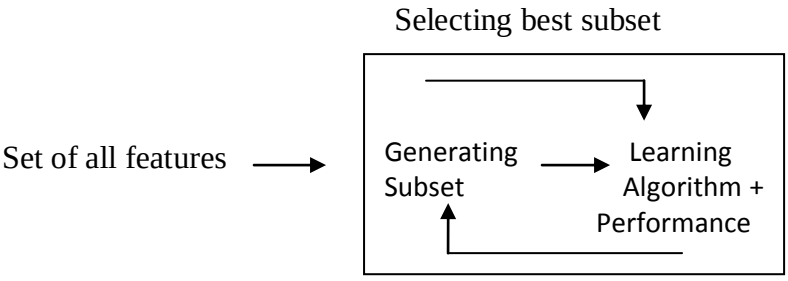
In wrapper method, we try to use a subset of features and train a model using them. Based on the inference that we draw from previous model. We decide to add or remove features from your subset. Wrapper models use a specific learning algorithm to evaluate the quality of the selected features. Given a predefined learning algorithm, The feature search component will produce a set of features based on certain search strategies. The feature evaluation component will then use the predefined learning algorithm to evaluate the performance, which will be returned to the feature search component for the next iteration of feature subset selection. The feature set with the best performance will be chosen as the final set.



EMBEDDED METHOD

Embedded methods combine the Qualities of filter and wrapped methods .The embedded methods catch all group of techniques which perform feature selection as part of the model construction process. Filter models are computationally efficient, but totally ignore the biases of the learning algorithm. Compared with filter models, wrapper models obtain better predictive accuracy estimates, since they take into account the biases of the learning algorithms. However, wrapper models are very computationally expensive.

Embedded models are a tradeoff between the two models by embedding the feature selection into the model construction. Thus, embedded models take advantage of both filter models and wrapper models, they are far less computationally intensive than wrapper methods, since they don't need to run the learning models many times to evaluate the features, and they include the interaction with the learning model. The biggest difference between wrapper models and embedded models is that wrapper models first train learning models using the Candidate features and then perform feature selection by evaluating features using the learning model, while embedded models select features during the process of model construction to perform feature selection without further.



RELATED WORK

D.Donoho.et.al., Presented about Filter methods select features based on a performance measure regardless of the employed data modeling algorithm. Only after the features are found, the modeling algorithm can use them. Filter method can rank individual feature or evaluate entire feature subset. We can roughly classify the developed measure for feature filtering into: Information, distance, consistency, similarity, and statistical measures.

W.Daelemans et.al., Describe about pearson's correlation statistics or pearson's correlation coefficient is also known as statistical models the value .For any two variables ,it return a value that indicates the strength of correlation.pearson's correlation coefficient is computed by taking the co variance of two variables and dividing by the product of their standard deviations

Jantawan et al., provide that the fisher score (also called the Fisher method, or Fisher combined probability score) is sometimes termed the information score, because it represents the amount of information that one variable provides about some unknown parameter on which it depends .The score is computer by measuring the variance between

expected value of the information and the observed value. When variance is minimized information is maximized .Since the expectation of the score is zero the fisher Information is also the variance of the score

Ghayur Naqvi et al., Presented about Wrapper methods consider the selection of a set of features as a search problem, where different combination are prepared, evaluated and compared to other combinations. A predictive model us used to evaluate a combination of feature and assign a score based on model accuracy. The search process may be methodical such as a best search , it may stochastic such as a random hill-climbing algorithm, or it may use heuristics, like forward and backward passes to add and remove features.

Ron Kohavi et al., provide that in Wrapper method , a search strategy iteratively adds or remove features from the data to search a best possible features subset that maximizes the accuracy. A search approach decides the order in which the variable subsets are evaluated like best-first , exhaustive search ,stimulated annealing ,Genetic algorithm, branch and bound ,etc.,Search strategies are crucial in designing feature selection algorithm, especially for wrappers, where optimal subset is to be identified by searching through the subset space.

S.Ma et al., Presented about embedded methods learn which features best contribute to the accuracy of the model while the model is being created . It facilates interaction between the classifier and features selection methods . The most common type of embedded feature selection method are regularization method. Regularization method are also called penalization method that introduce additional constraints into the optimization of a predictive algorithm that bias the model toward lower complexity .

Embedded methods the learning part and the features selection part cannot be separated the structure of the class of function under consideration plays a crucial role. Common embedded methods include various type of decision tree algorithm like CART, C4.5, and Random forest

Method adopted for feature selection in various application domain.

Paper Title: Text Document Classification using Ant Colony Optimization and Genetic Algorithm
Author Name: Monica Bali, Deipali Gore

Year: December 2015

The amount of information in digital form available with us is increasing rapidly day-to-day. The available information would be useful if they are able to access the

relevant information efficiently. The main problem is to improve the efficiency and accuracy of text classification [4]. To improve the efficiency of information they need to search, sort, index, store, and analyze the available information with the help of specific tools. In line with this one can read the texts and categorize them manually when amount of information is less, but what can be done when information is in huge amount e.g. in terms of hundreds, and thousands of texts? To answer this they require a tool which uses a supervised learning task that assigns the predefined category or class to new text documents. This initiates the need of some kind of automated application which works on the text categorization. There are several algorithms used for text categorization.

Paper Title: Ant colony optimization-based selected features for Text-independent speaker verification

Author Name: Hunny Pahuja, Jitender Chhabra, Ajay Khokhar

Year: June 2012

With the growing trend toward remote security verification procedures for telephone banking, biometric security measures and similar applications, automatic speaker verification (ASV) has received a lot of attention in recent years. The complexity of ASV system and its verification time depends on the number of feature vectors, their dimensionality, the complexity of the speaker models and the number of speakers. At present there are several methods for feature selection in ASV systems. To improve performance of ASV system they present another method that is based on ant colony optimization (ACO) algorithm [5]. After feature reduction phase, feature vectors are applied to a Gaussian mixture model universal back-ground model (GMM-UBM) which is a text-independent speaker verification model. The results of experiments indicate that with the optimized feature set, the performance of the ASV system is improved. Moreover, the speed of verification is significantly increased since by use of ACO, number of features is reduced over 80% which consequently decrease the complexity of our ASV system.

Paper Title: A novel feature selection method based on normalized mutual information

Author Name: La The Vinh, Sungyoung Lee, Young-Tack Park

Year: August 2011

In this thesis, a novel feature selection method based on the normalization of the well-known mutual information measurement is presented. Our method is derived from an

existing approach, the max-relevance and min redundancy (mRMR) approach. They, however, propose to normalize the mutual information used in the method so that the domination of the relevance or of the redundancy can be eliminated [6]. They borrow some commonly used recognition models including Support Vector Machine (SVM),k-Nearest-Neighbor (kNN), and Linear Discriminate Analysis(LDA) to compare our algorithm with the original(mRMR) and a recently improved version of the mRMR, the Normalized Mutual Information Feature Selection (NMIFS)algorithm. To avoid data-specific statements, they conduct our classification experiments using various datasets from the UCI machine learning repository. The results confirm that our feature selection method is more robust than the others with regard to classification accuracy.

Paper Title: Effective Wrapper-Filter hybridization through GRASP Schemata

Author Name: M.A. Esseghir

Year: November 2010

Of all of the challenges which face the selection of relevant features for predictive data mining or pattern recognition modeling, the adaptation of computational intelligence techniques to feature selection problem requirements is one of the primary impediments [6]. A new improved meta heuristic based on Greedy Randomized Adaptive Search Procedure (GRASP) is proposed for the problem of Feature Selection [7]. Our devised optimization approach provides an effective scheme for wrapper-filter hybridization through the adaptation of GRASP components. The thesis investigates the GRASP component design as well as its adaptation to the feature selection problem. Carried out experiments showed Empirical effectiveness of the devised approach.

Paper Title: A novel ACO–GA hybrid algorithm for feature selection in protein function prediction

Author Name: Shahla Nemati, Mohammad Ehsan Basiri, Nasser Ghasem-Aghaee

Year: April 2009

Protein function prediction is an important problem in functional genomics. Typically, protein sequences are represented by feature vectors. A major problem of protein datasets that increase the complexity of classification models is their large number of features. Feature selection (FS) techniques are used to deal with this high dimensional space of features [8]. In this thesis, they propose a novel feature selection algorithm that combines genetic algorithms (GA) and ant colony optimization (ACO) for faster and better search capability.

The hybrid algorithm makes use of advantages of both ACO and GA methods. Proposed algorithm is easily implemented and because of use of a simple classifier in that, its computational complexity is very low. The performance of proposed algorithm is compared to the performance of two prominent population-based algorithms, ACO and genetic algorithms

Feature selection Algorithm

Now we can see the discussion about the various feature selection algorithms. The choice of feature selection methods are problem dependent. When the training sample size is small, filter are expected to perform better for their ability to provide more stable estimation.

Neural Network

An artificial Neural Network (ANN) is an information processing paradiagram that is inspired by the way biological nervous system, such as the brain, process information. An ANN is configured for a specific problem such as pattern recognication or data classification, through a learning process. Generally Neural Network consists of layers of interconnected Nodes where each node producing a non linear function of its input and input to a node may come from other nodes. Some nodes produced output values.

ID3 Algorithm

The ID3 algorithm is decision tree algorithm. It is used to generate Decision tree from a dataset.ID3 is the precursor to the C4.5 algorithm and it's typically used in the machine learning process.ID3 does not guarantee an optimal solution. It uses a greedy approach by selecting the best attribute to split the dataset on each iteration.ID3 can over fit the training dataset to avoid over fitting, Smaller Decision trees should be preferred over large ones.

Every element in subset belonging to the same class (+ or -) then the node is turned into a leaf and labeled with the class of the examples. If no more attribute to be selected but the examples still do not belonging to the same class(some are + and some are -) then the node is turned into a leaf and labeled with the most common class of the examples in that subset.

C4.5 Algorithm

C4.5 algorithm is an algorithm used to produce a decision tree which is an expansion of prior ID3 calculation. It enhances the ID3algorithm by managing both continuous and discrete properties, missing value and pruning trees after construction.C4.5 creates the decision tree from a set of training data same way as ID3 algorithm.

CONCLUSION

In this paper, we have presented a survey of feature selection for classification techniques. A lot of research has been done in this field. Still the work is going on to improve the accuracy of feature selection. However the different methods of feature selection methods are very effective and useful for new researchers.

REFERENCES

[1]. Jennifer G Dy and Carla E Brodley. Feature selection for unsupervised learning,The Journal of Machine Learning Research, Vol. 5:845–889, 2004
[2]. Girish Chandrashekar, Ferat Sahin, "A survey on feature selection methods". Computers and Electrical Engineering.year 2004
[3]. Yvan Saeys, Inak Inza, Pedro Larranaga, (2007). "A review of Feature Selection techniques in bioinformatics". Bioinformatics, Oxford University press.
[4]. Monica Bali, Deipali Gore,Text Document Classification using AntColony Optimization and Genetic Algorithm,Vol. 3, Issue 12, December 2015.
[5]. Hunny Pahuja, Jitender Chhabra, Ajay Khokhar,Ant colony optimization-based selected features for Text-independent speaker verification,Vol. 2, Issue 3, May-Jun 2012, pp.1466-1476.
[6]. La The Vinh, Sungyoung Lee, Young-Tack Park,A novel feature selection method based on normalized mutual information,Appl Intell (2012) 37:100–120.
[7]. M.A. Esseghir,Effective Wrapper-Filter hybridization through GRASP Schemata,November 2010.
[8]. Shahla Nemati, Mohammad Ehsan Basiri, Nasser Ghasem-Aghaee,A novel ACO–GA hybrid algorithm for feature selection in protein function prediction,April 2009.
[9]. Dr.J.Arunadevi, M.Josephine Nithya, Comparison of Feature Selection Strategies for Classification using Rapid Miner, Vol. 4, Issue 7, July 2016
[10]. Mary Walowe Mwadulo, A review on feature selection methods for feature selection volume 5 issue 6, 395-402, 2016

