An evaluation of filter and wrapper methods for feature selection in classification

[1]Mitushi Modi, [2]Swati Patel

[1]M.E.Scholar, [2]Assistant Professor
[1]Computer Department,
[1]L.D.College of Engineering, Ahmadabad, India
[1]mmitushi.it@gmail.com , [2] swati.ldce@gmail.com

***Abstract*— Data mining is a form of knowledge discovery required for solving problems in a specific domain. Classification is a technique used for discovering class labels of unknown data. Different methods for classification exists like bayesian, decision trees, rule based, neural networks etc. Before applying any mining technique, irrelevant and redundant features needs to be removed. Filtering is done using different feature selection techniques like wrapper, filter, and hybrid. The central idea of feature selection is to select a subset of input variables by eliminating features with little or no predictive information. Its direct benefits included building simpler and more comprehensible models, improving performance, and helping organize, clean, and understand data. This paper presents different feature selection methods and their accuracy and performance which show the better technique for improving classification accuracy.**

***Key words*—Classification, data mining, feature selection technique, weka in classification, wrapper method, wrapper method**

I. INTRODUCTION

In real world, many irrelevant features may be present in data that has to be mined. Those irrelevant features have to be removed. Previous researches have shown that many algorithms don't perform well with many features. [1] Feature Selection is an important pre-processing step to choose subset from the original large amount of attributes. The main objective of feature selection is to eliminate irrelevant features which have no predictive information. Feature selection is a productive field of research area in machine learning, pattern recognization and data mining. In fact feature selection techniques have been broadly used in a variety of applications, such as genomic analysis, information retrieval, and text categorization. The advances of this process are improving learning accuracy, speeding up a data mining algorithms and better model comprehensibility.

II. FEATURE SELECTION TECHNIQUES

2.1 Filter Method

In filter method, a lot of statistic and information theoretical methods have been used on feature estimation. Feature selection is performed as a pre-processing step on training data and it relies on the common characteristics of the training dataset to choose important features without including any learning algorithms. Many well-known algorithms of filter methods are CBF (Correlation Based Feature Selection), FCBF (Fast Correlation Based Feature Selection), FOCUS. The advantages of filter methods are its structure is simple. Filter techniques are easily scale to high dimensional dataset and computationally fast. Filter method is independent of the mining algorithm so feature selections has to be performed once and based on it many algorithms can be calculated. The disadvantages of filter method are they cannot interact with the classifier. Also in this features are considered individually which may tend to worst classification performance when compared to other feature selection techniques.

2.2 Wrapper Method

Wrappers can find feature subsets with high accuracy because the features match well with the learning algorithms. Wrappers are feedback methods which incorporate with the machine learning algorithm in feature selection process. Wrapper methods search through the space of feature subsets and calculate the approximated accuracy of a single learning algorithm for each feature that can be added to or removed from the feature subset. [2] The feature space can be searched with different strategies such as forward selection and backward elimination. Forward selection approach begins with a small subset and adds additional features to the subset if they improve the performance of the learned hypothesis. Backward elimination begins with nearly all the original features and eliminates features as long as there is no reduction in the performance of the learning hypothesis. [3] The advantage of wrapper approach is interaction between feature subset search and model selection. The drawback of wrapper method involve computational expensive. Also wrapper approach tends to be much slower than the filter approach.

2.3Hybrid Method

This method is a combination of filter approach and wrapper approach. It takes the advantages of both methods. Filters can provide a quick parameter for wrappers, such as a reduced search space or a shorter search path, which help scale wrappers to

larger size problems. Commonly hybrid method uses the independent measure to decide the best subsets and then uses the learning algorithm to select the best subsets.

III. THE COMPARATIVE STUDY

3.1Tool Description

WEKA is extensively used tool for machine learning and data mining that was originally developed at the University of Waikato in New Zealand. It contains a large collection of modern machine learning and data mining algorithms written in Java. WEKA contains tools for regression, classification, clustering, association rules, visualization, and data pre-processing. WEKA has become very popular with the academic and industrial researchers, and is also widely used for educational purposes. [4]

3.2 Data Sketch

A number of data sets are selected for running the check; we have downloaded ILPD data from the UCI repository. Table 1 shows the description of data for testing purposes, the dataset is described by the data type being used, the types of attributes; whether they are categorical, real, or integer, the number of instances stored within the data set, the number of attributes that describe dataset. This data was selected because it has different characteristics and have addressed different areas, such as the number of instances which range greater than 1000. Also, the number of attributes; which range has 10 features and 583 instances.

Table 1 Data Sketch

Data Set Name and Type	Attribute Type	Instances	Attributes
ILPD	Integer, Real	583	10

IV. EXPERIMENTAL EVALUATIONS

For the experimental evaluation we have used WEKA toolkit to compute the feature selection subsets (filter and wrapper approach) and evaluate these feature sets.

4.1 Feature Subset

For ILPD data set we determined different feature subset For the FS subsets (Filter and wrapper), differ only in the amount of columns (attributes), for the subsets they differ in the number of columns and in their interpretation.

4.2 Selection of Attributes in Wrapper Approach and Filter Method

In ILPD dataset we calculated 2 feature subset methods. In that from the original relation R we calculated new relation R1. For feature subset (wrapper), in both dataset i.e. original R and new relation R1, change only the numbers of features (attributes or dimensions).Generally pre-defined threshold value has to be given but for this paper, neither a maximum nor a minimum number of features is pre-defined. The most select number of attributes is automatically considered within the wrapper subset evaluator. For retrieving the wrapper subset we have used WEKA's subset evaluator as attribute evaluator with the searching method greedy stepwise. As filtered approach WEKA has select Ranker algorithm by default and this ranking is independent of a specific learning algorithm and contains – before selecting a subset – all attributes.

V. RESULT

5.1 Classification on original dataset

We use ILPD data from the UCI repository for classification. Table 2 shows accuracy of classification without using filter or wrapper method on original data set.

Table 2: Classification Accuracy with Original Dataset

	Result
Correctly Classified Instances	324
Incorrectly Classified Instances	259
Mean absolute error	0.4428
Root mean squared error	0.6564
Relative absolute error	108.62%
Root relative squared error	145.45%
Accuracy	55.57%

5.2 Filter Attribute Evaluation

For attribute selection, in filter attribute evaluation, it has ranked to all features in range of [0, 1]. Table 3 shows rank of all attributes of ILPD dataset.

Table 3: Data of Filter Approach

Rank	Weightage	Attributes
1	0.102	total Bilirubin
2	0.085	direct Bilirubin
3	0.066	A/G ratio

4	0.064	total proteins
5	0.059	Albumin
6	0.027	Alkphos
7	0.022	Age
8	0.021	SGOT
9	0.004	Gender
10	0.000	SGPT

Filter approach gives rank to all attributes and minimum ranked attribute can be eliminated. We remove last two attribute from dataset, through which we are able to increase accuracy of classification. Table 4 shows accuracy of classification after removing attributes from dataset.

Table 4: Classification accuracy using Filter approach

	Result
Correctly Classified Instances	325
Incorrectly Classified Instances	258
Mean absolute error	0.4434
Root mean squared error	0.6571
Relative absolute error	108.76 %
Root relative squared error	145.61 %
Accuracy	55.75 %

5.3 Filter Subset Evaluation

For feature selection, in filter Subset evaluation, it already wrapped most selected features automatically. In ILPD dataset, out of 11 attributes wrapper selects most 6 relevant features for the evaluation. Table 5 shows the selected attributes of wrapper approach.

Table 5: Selected Features after Wrapper Approach

Rank	Attribute
1	Age
3	total Bilirubin
4	direct Bilirubin
5	total proteins
6	Albumin
7	A/G ratio

Wrapper approach applies on all attributes using algorithm and selected features are extracted from dataset. Table 6 shows accuracy of classification of wrapper approach.

Table 6: Classification accuracy using Wrapper approach

	Result
Correctly Classified Instances	325
Incorrectly Classified Instances	258
Mean absolute error	0.4434
Root mean squared error	0.6571
Relative absolute error	108.76 %
Root relative squared error	145.61 %
Accuracy	55.75 %

Figure 1 shows accuracy comparison between filter and wrapper approach. Both approach perform significantly much better than original dataset.

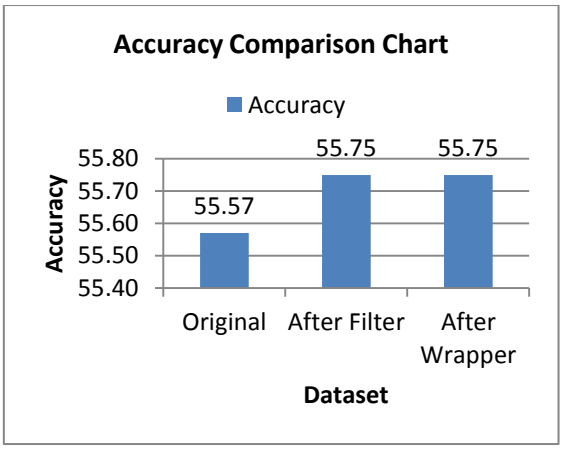

Fig 1: Accuracy Comparison Chart

VI. CONCLUSION AND FUTURE WORK

In conclusion, both approaches have advantages as well disadvantages. Thought the fact is by using feature selection method attributes in a relation has been reduced and the space of dataset is also being less thought reduction of dimensions. So for a large dataset by using filter or wrapper approach good features can be achieved. This paper shows that, accuracy is increasing, while we used filter and wrapper approach on original dataset. Extensions of the algorithm presented here to multi relational problems are an important direction for future work.

REFERENCES

[1] Sunita Beniwal*, Jitender Arora "Classification and Feature Selection Techniques in Data Mining", International Journal of Engineering Research & Technology (IJERT), Vol. 1 Issue 6, August – 2012

[2] Andreas G.K. Janecek, Wilfried N. Gansterer "On the Relationship between Feature Selection and Classification Accuracy", JMLR: Workshop and Conference

[3] LAWRENCE B. HOLDER, ANTHONY G. PIPE, BRIAN CARSE "Current And Future Trends In Feature Selection And Extraction For Classification Problems"

[4] Mitushi Modi,V.B. Vaghela "Comparison of different classification techniques using weka for car data and ILPD dataset", NCIET-2014,SRPEC

[5] Huan Liu, Hiroshi Motoda, Rudy Setiono "Feature Selection: An Ever Evolving Frontier in Data Mining", JMLR: Workshop and Conference Proceedings

[6] Dhafer Lahbib, Marc Boull´e,Dominique Laurent "Informative Variables Selection for Multi-Relational Supervised Learning "

[7] M. Dash, H. Liu "Feature Selection for Classification", Intelligent Data Analysis 1 (1997)

[8] Amir Navot "On the Role of Feature Selection in Machine Learning",Submitted to the Senate of the Hebrew University December 2006