Feature Selection for Classification of Kidney-Renal Failure

Phyu Phyu Htun, Moe Sanda Htun
Computer University, Loikaw
phyuphyuhtun24@gmail.com, moesdhtun@gmail.com

Abstract

Several recent machines learning publication demonstrates the utility of using feature selection algorithm in supervised learning tasks. Among these, sequential feature selection algorithms are receiving attention .In the feature subset selection problem , a learning algorithm is faced with problem of selecting a relevant subset of feature upon which to focus its attention to achieve the highest predictive accuracy with the learning algorithm on this domain , a feature subset selection method should consider how the algorithm and the training data interact with wrapper method .This paper is described the use of feature selection techniques that uses sequential forward selection to improve the performance of classifier and compute the performance of Naive Bayesian with complete feature set and selected feature set.

Keywords: Feature Selection, Sequential Forward Selection, Naive Bayesian Classification.

1. Introduction

Health plays an essential role in human lives. Nowadays, kidney or renal failure is one kinds of threading cases in our human health. Kidney failure can cause two kinds, Chronic or Acute renal failure. Chronic kidney failure (CKF) also called chronic renal failure (CRF) is a progressive loss of renal function over a period of months or years .Acute renal failure is an immediate loss of renal function. This proposed system is intended for classifying based on those two parts of kidney failure [7].

Feature selection is the process of identifying and reducing as much of the irrelevant and redundant features as possible. In Supervised machine learning, feature selection is used as a preprocessing step. It is also select a subset of original features so that the feature space is optimally reduced according to a certain evaluation criterion [1].

In classification problems, the issue of high dimensionality, feature is often considered important. To lower feature dimensionality, feature selection methods are often employed. To select a set of features that will span a feature space is as good as possible for the classification task. Therefore, feature selection becomes very necessary for selection of features tasks for classification when facing high dimensional features.

In the case of wrapper approach, the relevance feature selection measure is directly defined from the induction algorithm. The induction algorithm is considered as a black box. The induction algorithm is run on the dataset usually partitioned into the training and testing set with different set of features removed the data or features. The resulting classifier is then calculated on an independent test set that was not used during the search.

2. Motivation

Feature selection is the process of selecting a feature subset from the training examples and ignoring feature not in this set during induction and classification, is an effective way to improve the performance and decrease the training time of supervised learning algorithm. Feature selection typically improves classifier performance when training set is small without significant degrades performance on large training set. The problem of feature selection can be defined as finding relevant features among the original attributes to define the data in order to minimize the error probability or some other reasonable selection criteria.

3. Feature Selection

The goal of feature subset selection is to find a minimum set of feature or attributes subset. It reduces the number of feature appearing in the discovered patterns.

Ideally, feature selection methods search through the subsets of features, and try to find the best features among the competing 2^N candidate subsets according to some evaluation function. This procedure is based on the heuristic methods attempt to reduce computational complexity by compromising performance. This method needs a stopping criterion to prevent an exhaustive search of subsets.

There are four basic steps in a typical feature selection method:

1. A generation procedure to generate the next candidate subset;

2. An evaluation function to evaluate the subset under examination;
3. A stopping criterion to decide when to stop; and
4. A validation procedure to check whether the subset is valid.
5. The generation procedure is a search procedure; it generates subsets of features for evaluation.

The generation procedure can start (a) with no feature, (b) with all features, (c) with a random subset of features. In this system, the generation procedure starts with no feature [1].

3.1 Sequential Forward Selection

Sequential forward selection (SFS) is the popular method in the heuristic generation function. SFS starts from the empty set, and in each iteration generates new subsets by adding a feature selected by some evaluation function .This simple setting is (i) always assume independence of features(do not consider the previously selected features), (ii) never assume independence (consider the previously selected features) .Bayesian classifier error rate as the evaluation function was used.

3.2 Wrapper approach

The type of feature selection algorithm falls into three different approaches as; filter approaches, wrapper approaches, and embedded approaches. A filter is defined as a feature selection that it is not suitable for huge amount of dataset. Wrapper approach is suitable for large amount of datasets.

In wrapper approach, the feature subset selection algorithm exists as a wrapper around the induction algorithm itself as part of the function evaluation feature subsets. The induction algorithm is considered as a black box [5].

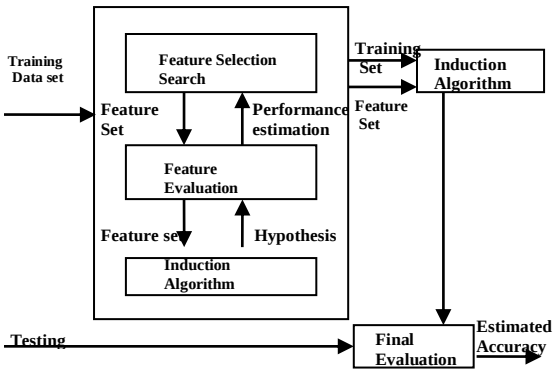

Figure1. Wrapper approach to feature subset selection.

4. Experimental Methodology

The experiments descried in this paper compute the performance of supervised machine learning algorithms with Sequential forward Selection of feature selection approach on the renal –kidney data set .The data that used in this system has fourteen features and therefore is more suitable for high dimensional data. In addition to feature selection algorithms, Naive Bayesian classifier is used to evaluate the accuracy on selected features on the original data set as well as newly obtained data set containing only the selected features from selection algorithm and record overall accuracy by k-fold cross validation.

4.1 Naive Bayesian classifier

The Naive Bayesian classifier is also known as the simple Bayesian classifier. In the Naive Bayesian Classifier is one of the simplest probabilistic classifier.

It works as follows:

1. Each data sample is represented by an **n**-dimensional feature vector, $\mathbf{X(x_1,x_2,.....,x_n)}$, depicting **n** measurement on the sample from **n** attributes respectively,$\mathbf{A_1,A_2,.....,A_n}$.

2. Suppose that there are m classes, $C_1,C_2,.....$ and C_m. The Naive Bayesian classifier assigns an unknown sample X to the class C, if and only if

$$\mathbf{P(C_i \mid X) > P(Cj \mid X) \text{ for } 1 \le j \le m, j \ne i.}$$

Thus we maximize $\mathbf{P(C_i \mid X)}$. The class $\mathbf{C_i}$ for which $\mathbf{P(C_i \mid X)}$ is maximized is called the maximum posterior hypothesis. By Bayes Theorem,

$$\mathbf{P(C_i \mid X) = P(X \mid C_i)\, P(C_i) \mid P(X)}$$

3. As **P (X)** is constant for all classes, only $\mathbf{P(C_i \mid X)\, P(C_i)}$ needs to be maximized. If the class prior probabilities are not known, then it is commonly assumed that the class are equally likely, that is, $\mathbf{P(C_1) = P(C_2) = ... P(C_m)}$, and we would therefore maximize $\mathbf{P(X \mid C_i)}$. Otherwise, we maximize $\mathbf{P(X \mid C_i)\, P(C_i)}$. Note that $\mathbf{P(C_i) = s_i/s}$, where '$\mathbf{s_i}$'is the number of training samples of class $\mathbf{C_i}$, and's' is the total number of training samples.

4. Given data sets with many attributes, it would be extremely computationally expensive to compute $\mathbf{P(C_i \mid X)}$. In order to reduce computation in evaluating $\mathbf{P(C_i \mid X)}$, the naive assumption of class conditional independent is made. This presumes that the values of the attributes are conditionally independent of one another, given the class label of

the sample, that is, there are no dependence relationships among the attributes.

Thus

$$P(C_i \mid X) = \prod_{k=1}^{n} P(x_k \mid C_i)$$

The probability $P(x_1 \mid C_i)$, $P(x_2 \mid C_i)$, $P(x_n, C_i)$ can be estimate from the training samples.

5. In order to classify an unknown sample **X**, $P(X \mid C_i)\, P(C_i)$ is evaluated for each class C_i, sample **X** is then assigned to the class C_i if and only if

$$P(X \mid C_i)\, P(C_i) > P(X \mid C_j)\, P(C_j)$$

For $1 \leq j \leq m, j \neq i$

In other words, it is assigned to the class C_i for which $P(X \mid C_i)\, P(C_i)$ is the maximum.

4.2 k-fold cross validation

Estimating classifier accuracy is important in that it allows one to evaluate how accurately a given classifier will label future data that is data on which the classifier has not been trained. In k-fold cross validation, the initial data are randomly partitioned into k- mutually exclusive subsets or "folds", $S_1, S_2, \ldots S_k$, each of approximately equal size. Training and testing is performed k-times. In experiment i, the subset Si is reserved as the test set, and the remaining subsets are used to train the classifier. That is, the classifier of the first experiment is trained on the subsets $S_2, \ldots, S_k$ and tested on S1; the classifier of the second experiment is trained on subsets $S_1, S_3, \ldots, S_k$ and tested on S_2; and so on. The accuracy estimate is the overall number of correct classification from the k experiments, divided by the total number of samples in the initial data.

In the k-fold cross-validation, the data was randomly partition into k mutually exclusive subsets or folds of approximately equal size. A learning algorithm was trained and tested k times; each time it is tested on one of the k – folds and trained using the remaining k-1 folds. The cross- validation estimate of accuracy was the overall number of correct classifications from the k experiment, divided by the number of examples in the initial data.

4.3 Performance Evaluation

In addition to accuracy, classifiers can be compared with respect to their speed, robustness that is accuracy on noisy data, scalability, and interpretability. Sensitivity, specificity, and precision are useful alternatives to the accuracy measure. These measures are defined as

$$\text{sensitivity} = \frac{t\text{-}chro}{chro}$$

$$\text{specificity} = \frac{t\text{-}acu}{acu}$$

$$\text{precision} = \frac{t\text{-}chro}{(t\text{-}chro + f\text{-}chro)}$$

$$\text{accuracy} = \text{sensitivity}\frac{chro}{(chro + acu)} + \text{specificity}\frac{acu}{(chro + acu)}$$

t-chro means true chronic in the renal data classifying by classifier and t-acu also means true acute class in the dataset. chro and acu can be defined as the class chronic and class acute in the renal training datasets.

5. System Overview

This system contains five stages. In the first stage, the data are partitioned into training and testing data using k-fold cross validation strategy over the whole data set. The second stage consists of a feature selection process by using the sequential forward selection with wrapper approach as a search engine to find the optimal subset of features. In the third stage, Naive Bayesian is used for classification the instances and evaluated the selected feature set. The selected feature sets are finally evaluated using the testing data set in the fourth stage. Finally, compute the performance of classifier with selected feature and complete feature set.

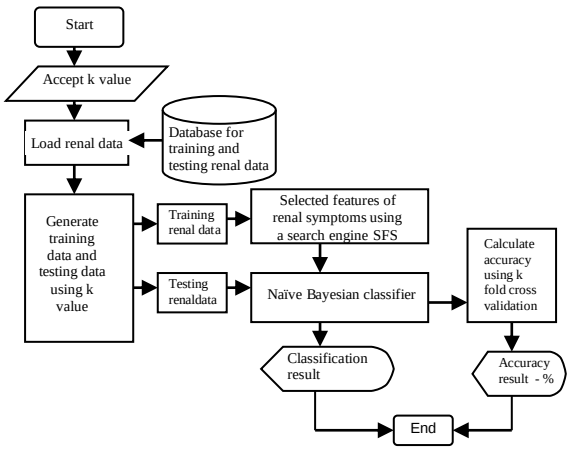

Figure 2. System Flow diagram.

6. Experimental Result

The sequential forward selection starts empty set of features and the evaluation function was a single 5- fold cross validation. The same 5 way split was done for all subsets. At each step, sequential forward selection chooses to select an unselected feature with the highest estimated accuracy. The experiments described in this system compute the performance of classifier with selected feature set and complete feature set on the real-world dataset. The data used for this experiment are renal –kidney failure data. The data set contains totally fourteen features and two classes. The selected features are evaluated by

using Naive Bayesian Classifier to estimate the renal data is whether chronic or acute renal failure.

6.1. Data Preprocessing

This system use the attribute mean to fill the missing value for all samples belonging to the same class as the given instance. Then attribute value is normalized by using the Min-max normalization. This system used the preprocessed renal data for feature selection processes. The renal data have fourteen features. They are Age, Sex, Diabetes, ChronicWeight Loss (CWL), Hypertension, ChronicNSAIDUsed, Blood Pressure, Hemoglobin, Calcium, Phosphate, Potassium, Bone pain, SerumCreatinine, RcentTX . This sample data have two class :chronic renal failure of acute renal failure .

Table1: Interface for Preprocessing

No	Features Name	Description
1	Age	Age in years
2	Sex	Male=1 female=0
3	Diabetes	Presence=1 Absence=0
4	Chronic Weight Loss(CWL)	One week=1 One month=0.3 Three month=0.6 One year=0.9
5	Hypertension	Presence=1 Absence=0
6	Chronic NSAID used	None=0 One month=0.3 Three month=0.6 One year=0.9
7	Blood pressure	90,140,240
8	Hemoglobin	<11mmol/dl=1 >11mmol/dl=0
9	Calcium	30,40,60mmol/dl
10	Phosphate	50,60,80 mmol/dl
11	Potassium	30,60,70 mmol/dl
12	Bone pain	Presence=1 Absence=0
13	SerumCreatinine	250,280,300,500
14	RecentTX	None=0 Toxins=1 Severe infection=2 Injury to kidney=3 Shock=4
15	Class	Chronic=1 Acute=0

6.2. Feature Selection for Renal Kidney Failure data

This system use sequential forward selection with wrapper approach to reduce the number of feature and maximize the performance of Naive Bayesian Classifier. In this proposed system, 5-fold cross validation was used and the learning algorithm was trained and tested 5 times.

Table 2: Selected feature of dataset

Dataset	#of feature in kidney-renal data set	Naive Bayesian Classifier with (Sequential forward selection) no. selected feature
Renal-kidney failure data	14	10

Table 3: Show Naive Bayesian Classification on renal-kidney data

Dataset	Total number of feature in renal – kidney data set	Naive Bayesian (Sequential forward selection) no. selected feature	Naive Bayesian with Complete Feature set
Renal-kidney failure data	14	10	14

6.3 Performance Evaluation

The accuracy rate of Naive Bayesian, which is calculated, based on forward selection is better than the complete feature subset.

Table 4: Accuracy of Naive Bayesian Classifier

Data set	Naive Bayesian Classifier's accuracy on the selected feature	Naive Bayesian Classifier's accuracy on the complete feature

Renal – kidney failure data	98%	86%

7. Conclusion

The feature selection problem in supervised learning, which involves identifying the relevant or useful feature in a dataset and giving only that subset to the learning algorithm. The wrapper approach requires a search engine and an evaluation function .For the search engine this system used Sequential Forward Selection and for evaluation function as use cross validation as accuracy estimation technique. Evaluate the Renal failure is chronic of acute by using Naive Bayesian Classifier.

Feature selection methods are more suitable for large-dimension applications than other. This method can be used in supervised learning mode. Naive Bayesian is efficient but suffer the attribute independence assumption. This system is limited dependence Bayesian Classifier. This system gives the classifier accuracy and result by instance of Renal –kidney failure dataset.

This system can be extended by using Sequential Backward Selection and other search algorithm such as Hill-climbing .This system also extend Naive Bayesian classifier to work on the other datasets.

REFERENCES

[1] A. Tsymbal, S.Puuronen, D.Patterson,"**Feature Selection for ensembles of Simple Bayesian Classifiers"**.

[2] G. Bonte "**Structural feature selection for wrapper methods"**
ULB Machine Learning Group
Universit´e Libre de Bruxelles
email: gbonte@ulb.ac.be

[3] H. Liu and R. Setiono,'' **A Probabilistic Approach to Feature Selection- a Filter Solution**", in Processings of International Conference on Machine Learning, pages 319-327,1996.

[4] M.Dash & H.Liu **"Feature Selection for Classifications "** National University of Singapore, 1997.

[5] R. Kohavi , George H. John , " **Wrappers for feature subset selection**"
Received September 1995; revised May 1996

[6] S. Shah, Andrew Kusiak, and Bradley Dixon
"**Data Mining in Predicting Survival of Kidney Dialysis Patients -Invariant object approach"**

[7] **Chronic Renal Failure**

http://www.nlm.nih.gov/medlineplus/ency/article/000471.htm.