Feature Selection Effects on Classification Algorithms

Laconic description of Machine Learning Algorithms

Sonal Singh , Shyam S Choudhary,S. Bhavishya
Computer Science
Veltech Technical University
Chennai, India

Abstract— This Feature Engineering is cardinal to the application of Machine Learning. It enables the use of turf knowledge of data to form features which makes Machine Learning Algorithms to be efficient and better. With the advent of use of huge databases, feature selection is in great demand and much work has been done in the same. In this paper we have compared the effect of Classification algorithms and their efficiency before and after applying the technique of feature selection. The vitality and vulnerability of K Nearest Neighbor, Naive Bayes and Support Vector Machines (classification algorithms) is altercated. Moreover, guidelines to perform feature selection on categorical dataset is also explained laconically.

Keywords— Feature Engineering; Machine Learning; Classification; Algorithms; Nearest Neighbour; Naïve Bayes; Support Vectors.

I.   INTRODUCTION

Feature Selection is a knack of selecting relevant features from huge data bases for its usage in model construction. A dataset may contain many features which are either irrelevant or redundant and whose elimination does not cause any loss of information in the data set. This method enables to select features that are most relevant and important in forming decision over algorithms. Feature Selection enhances generalization by reducing the overfitting in order to avoid the curse of dimensionality. It simplifies the model making it easier for researchers to interpret it by reducing the training time of the data. It also improves the classifiers by removing the redundant features. So our main goal by feature selection is to reduce the number of features used and analyze it effects on various classification algorithms. There are three basic methods to perform Feature Selection .Filter Method, Wrapper Method and Embedded Method.

A.   FILTER METHOD

Filter method is a technique for feature selection where we select the best subset from all the available features randomly and apply our learning algorithm. The obtained performance is then evaluated. These steps are repeated continuously until there is no change in the resultant value. The selection of features for the analysis is independent of any algorithms .Some of the basic example of this method include CHI SQUARE TEST, INFORMATION GAIN and CORRELATION COEFFICIENT RANKS, Anova.

B.   WRAPPER METHOD

Wrapper methods wrap up the subset with the algorithm in a block which is experimented first and then go for the performance analysis .In this method, we generally train a model and draw inference by testing them. Based on the results, we decide either to add a feature or remove feature from the given subset.It is implemented in three ways. FORWARD SELECTION METHOD – In this method we start without any feature in our model. With every iterations, we add the required feature which suits our model the best. BACKWARD SELECTION METHOD – in this method, start with all the features in our model and eliminate the least required feature at every iteration. RECURSIVE FEATURE ELIMINATION – In this method, we create a model with feature and at each iteration we segregate the best and worst performing features

C.   EMBEDDED METHOD

It is a combination of qualities of both wrapper and filter method. This method does not require any special technique as the algorithms used for implementing embedded methods have built in features of feature selection to perform dimensionality reduction. Some of the most known example of this methods are LASSO and RIDGE. Lasso performs L1 regularization while L2 regularization is performed by Ridge.
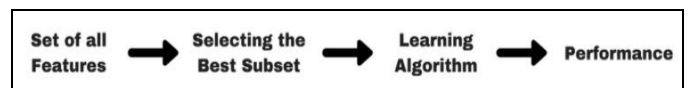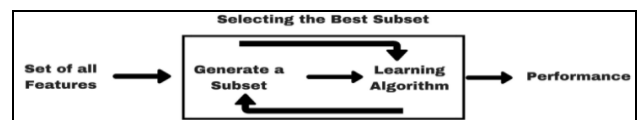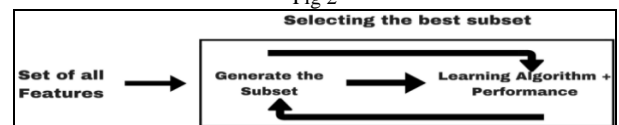

Fig 1

Fig 2

Fig 3

II.   MACHINE LEARNING ALGORITHMS

Machine learning is germaneness of Artificial Intelligence which provides the ability to a system to learn and improve itself. These are the algorithms that are not called explicitly but they learn by accessing the data and training themselves. It can

be supervised where we train and label the instances, unsupervised where labelling is not entertained or reinforcement where it gives actions and learn from their errors. These algorithm can be of classification, prediction, description or regression types. It gives the novel users the power to analyze massive data and work hoyole. However in this paper will study the classification algorithms and how they behave after undergoing Feature selection.

A. K NEAREST ALGORITHM

K NEAREST ALGORITHM is a supervised learning technique which works well with both classification and regression problems. It is a non-parametric, instance based lazy learning algorithm.. The unlabeled data points are actually segregated based on the nearest neighbor by calculating the distance between them into well-defined groups. The unlabeled points, get labelled with the nearest instance .To nail down the similarity between the training data set and the new input, a distance measure is used. This measure could be of any sort like EUCLEDIAN distance, HAMMING distance, MANHATTAN distance, TANIMOTO distance, JACCARD distance, MAHALANOBIS distance and etc. K is a user defined constant which is also an unlabeled vector. Large values of K, reduces the effect of noise on the classification, simultaneously making the boundaries of classes less distinct. To avoid tied votes, it is advisable to select k as an odd number when number of classes are even and k as an even number when number of classes are odd. In general, we can break ties even by incrementing the value of K by 1.

B. NAÏVE BAYES ALGORITHM

Naïve Bayes is a supervised, non-parametric, machine learning algorithm used for classification purposes. It is based on the Bayes probability theorem. To reduce the complexity of Bayes classifier, naïve Bayes algorithm was introduced. As the name suggest, it assumes that all the properties of a certain features belong to a group and are independent of the occurrence of the other feature. Hence, the name naïve contributes to the fact that all the properties individually accord to the probability of one occurrence. And the term Bayes was coined after a scientist Thomas Bayes, a famous mathematician and statistic scholar. The conditional probability formula based on theorem is:

Posterior = ((Likelihood) (Posterior Prior Probability)) / Evidence Prior Probability

There are basically three variations to find the probability. The primordial is Gaussian, which is used for methodical and conventional problems, the next in order is Multinomial which works well when there are multiple occurrences of data and the triennial be Bernoulli, which is used when we have binary valued data sets. The probability of response variable is predicted by dealing with the probability distributions of the variables in the data set. It is mainly used for text classification, spam filtration, sentimental analysis, recommendation system and etc.

C. SUPPORT VECTOR MACHINES

Support Vector Machines also a Supervised Machine learning algorithm, which follows a discriminative approach for classification and regression. The intrinsic use of SVM is for classification problems. The subset of training data points is used as support vectors for Decision function. Classification is carried on by constructing HYPER PLANES In multidimensional space to distinguish different class labels. This is brought to pass by plotting data items in n dimensional space, where each feature holds the value of a coordinate. Then a HYPER PLANE is endowed to classify the data items into different classes. It converts categorical variable into dummy binary variables of 0 and 1. It uses many parameters like Kernel Trick which converts non separable problems into a separable ones, other being Regularization and Gamma. SVM for classification is of two types, C-SVM, Nu- SVM To Identify Correct Hyperplane, we select the plane which segregates the two classes the best or plane which gives higher margin for data points is selected as robustness is increased. The outliers are ignored by the inbuilt feature of SVM. . It is one of the best approaches to segregate the classes and thus acts as a frontier.

III. ANALYSIS OF EXPERIMENT

As per above described methods, we conducted our experiment over Loan Data set and received a confusion matrix as a outcome. The algorithms were intercessed based on some evaluation metrics which is described further.

A. EVALUATION METRICES

The algorithms were trained and tested both with and without feature selection using Embedded Method. The intercessor was based on accuracy, true positive rate and false positive rates, where Accuracy is the ration total correct predictions in relation to all the predictions. TPR is the relationship of true instances with total classified instances and FPR is the proportion of all wrong predictions done.

$$AC = \frac{TN+TP}{TP+TN+FP+FN}$$

$$TPR = \frac{TP}{TP+FN}$$

$$FPR = \frac{FP}{FP+TN}$$

B. With Feature Selection

Loan Data The data set after undergoing feature selection give insignificant different result. By the data we observe that efficiency of every classifier has improved to an extent. The table underneath sums it up.

Classifier	TPR	FPR	Accuracy
KNN	0.84	0.02	84
NB	0.90	0.21	90.38
SVM	0.88	0.09	88.97

C. Without Feature Selection

The data set has been vitally used to train three machine learning classifiers which are K Nearest Neighbor, Naïve Bayes and Support Vector Machines. The obtained analyzed results are described in the table below.

Classifier	TPR	FPR	Accuracy
KNN	0.65	0.14	65.02
NB	0.74	0.41	74.41
SVM	0.79	0.12	79.12

D. OBSERVATIONS

The following observations were studied to bring about conclusion on effects of Feature Selection on classification algorithms.
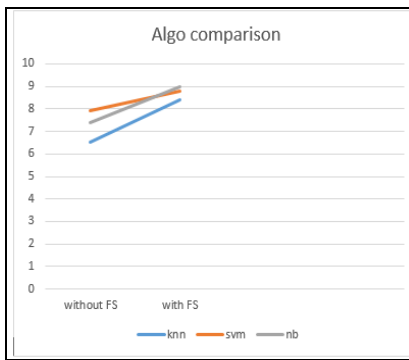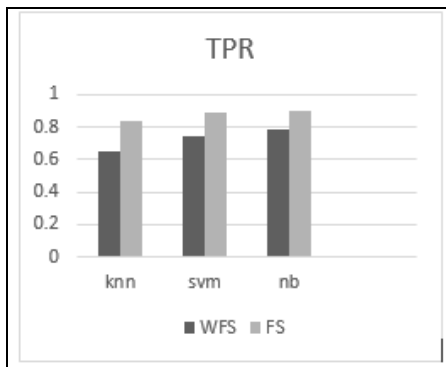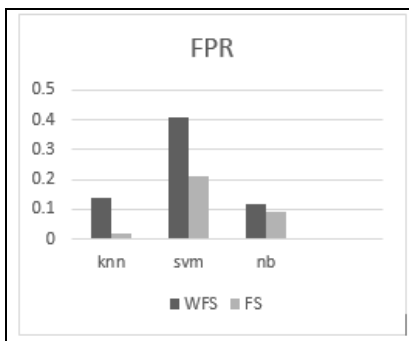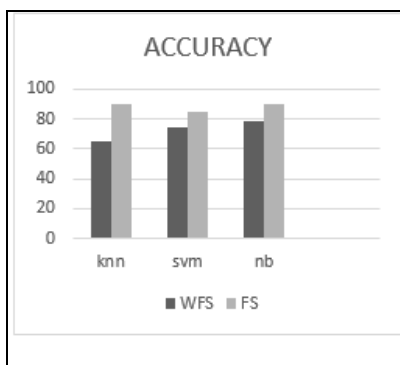

Fig 4

Fig 5

Fig 6s

Fig 7

The graphs above depicts the overall behavior of machine learning algorithms used here as classifiers. Fig 4. Represents the performance of the classification algorithm. It shows

resemblance to the improvement in classification efficiency. Fig 5-7. Explains about the effects of Feature Selection on prediction of instances for classification algorithms. We can clearly observe that true positive rate has increased immensely for all the three algorithms and simultaneously rate of false predictions have decreased. The overall accuracy increased for the algorithms as the selected were the best ones.

IV. RESULT

The Machine learning classification algorithm performed veraciously. The K nearest algorithm is robust to noisy data but at the same time its is difficult to choose attributes for distance based learning which makes the computation cost very high. Naive Bayes makes strong assumptions about the independency of the instances. However it has high various and works well with text classifications. It even yields qualitative instances especially in Loan data set. Time computation is visibly very low as compared to Knn. SVM can handle significantly large data set well by showing similar attributes to that of linear regression. At the same time it works well with non linear boundaries based on kernel use. For the Loan Data Set, we can have perception that Naïve Bayes shows excellent properties for classification. Its basic property of Text Classification have supported well for our Dataset.

FUTURE WORKS

Several scholars from University of Toronto and even IBM is conducting various researches to increase the efficiency of classification algorithms to a minimum of ninety plus. We can expect a trend of Machine learning classification algorithms which are not bias, taking visibly less computation time and shows immense improvement in the performance after undergoing Feature Selection.

REFERENCES

[1] Dipali Bhonsale and Roshni Ade,"Feature selectiom based on classification using naïve bayes,j48 and support vector machines," International Journal of Computer Applications, vol. 99, August 2014.

[2] C.sandhiya,P.anandhkumar,D.preetha,T.Rajendran,"Feature subset selection for irrelevant data removal using decision tree algorithms",Fifth international conference on Advance Computing, April2012.

[3] Desai devanshi and r.Rajamenakshi,"Large scale image feature extraction on medical image analysis,"International Journal of Advanced Research and Science, vol.3.

[4] Noelia Sanchez ,Amparo Alonso and Maria Tombilla,"Filter methods for feature selection, a comparative study," 2002.

[5] C.Arunkumar and S.Ramakrishnan "Ahybrid approach to feature selection using correlation coefficient and fuzzy rough quick reduct algorithm applied to microarray of data," 1998.

[6] Zuan khan and Vijeta kalkal,"Classification algorithms and their efficiency," in press.

[7] Vineeta Andrews,"laconic overiew of algorithms," unpublished.

[8] M .marques,K.NIcoleeti and Jinny P,"Using algorithms on eductional data,"2006.

[9] Sakshi yadav, Jaya tiwari and Romesh Pilai,"Using feature selection on cancer data set," in press.

[10] M.Dash and H liu,"Feature selectionn for classification,"Elsevier,1997.

[11] V.Radha, "Feature Selection at its best," in press.

Essra Mahesci and Selma Zayich,"A comparative study on feature selection accuracy,"Elsevier,2012.

IJERTV7IS020109

www.ijert.org

185

(This work is licensed under a Creative Commons Attribution 4.0 International License.)