International Journal of Soft Computing and Engineering (IJSCE)
ISSN: 2231-2307, Volume-4, Issue-ICCIN-2K14, March 2014

Inductive-Analytical Learning based Stepwise Support Vector Machine (SVM) Model

Anshu Bharadwaj, Sonajharia Minz

Abstract- Support Vector Machines tend to perform better when dealing with multi-dimensions and continuous features. They work well with the data with high dimension also. This paper introduces the support vector machine (SVM) approach to the classification task in a step-wise manner to address mainly the high dimensional datasets. The task here is modeled as a supervised learning problem using SVM classifier in multiple steps. This approach has enabled to combine the inductive and analytical learning. Integrated learning systems, (i.e., systems that combine empirical and explanation-based learning or inductive and analytical learning) have the potential of overcoming the weakness of either method applied individually. This approach has been employed in proposing a step-wise SVM model in this paper.

Keywords: Support Vector Machines, Inductive Learning, Analytical Learning, Attribute Relevance Analysis, Information Gain.

I. INTRODUCTION

Integrated learning systems, (i.e., systems that combine empirical and explanation-based learning or inductive and analytical learning) have the potential of overcoming the weakness of either method applied individually. This approach has been employed in proposing a step-wise SVM model in this paper using Inductive and Analytical learning methods. Support vector machine (SVM) has gained much attention, since their inception [4], [2]. It presents a powerful new generation learning algorithm based on recent advances in statistical learning theory. It delivers state-of-the-art performance in real-world applications and presents a useful algorithm for classification. A Support Vector Machine (SVM) maps input (real-valued) feature vectors into a higher dimensional feature space through some nonlinear mapping. SVMs are powerful tools for providing solutions to classification, regression and density estimation problems. These are developed on the principle of structural risk minimization [8]. Computing the hyper plane to separate the data points i.e. training a SVM leads to quadratic optimization problem [8], [5]. Speed of the SVMs is one of its main advantages. SVMs can learn a larger set of patterns and be able to scale better, because the classification complexity does not depend on the dimensionality of the feature space. SVMs also have the ability to update the training patterns dynamically whenever there is a new pattern during classification.

Inductive methods seek general hypotheses that fit the observed training data, whereas analytical learning methods seek general hypotheses that fit prior knowledge while covering the observed data. Inductive learning methods depend on the amount of training data available. Pure analytical learning methods offer the advantage of generalizing more accurately from less data by using prior knowledge to guide learning.

Manuscript received March 2014

Anshu Bharadwaj, Indian Agricultural Statistics Research Institute New Delhi, India.

Sonajharia Minz, Jawaharlal Nehru University New Delhi, India.

However, they can be misled when given incorrect or insufficient prior knowledge. Purely inductive methods offer the advantage that they require no explicit prior knowledge and learn regularities based solely on the training data. However, they can fail when given insufficient training data, and can be misled by the implicit inductive bias they must adopt in order to generalize beyond the observed data. These two learning paradigms are based on fundamentally different justifications for learned hypotheses and offer complementary advantages and disadvantages. Combining them offers the possibility of more powerful learning methods. This approach has enabled to combine the inductive and analytical learning in this paper. In general, the all classification algorithms consider the entire set of attributes as input for modeling the classifier. However, the proposed approach considers a partition of a set of attributes and accepts as input a block of the partition in each of the steps in this step-wise SVM classification model. The subset of attributes considered in step *i* include the predicted decision attribute of step *(i-1)*.

II. PRELIMINARIES

A. Inductive Learning

Inductive learning methods are those that generalize from observed training examples by identifying features that empirically distinguish positive from negative training examples. Decision trees, neural network learning, inductive logic programming and genetic algorithms are all examples of inductive learning methods that work in this manner. The limitations of these inductive learning methods are that, they do not perform well when there is insufficient data available. It is observed that through theoretical analysis that when learning inductively from a given number of training examples, there are fundamental bounds on accuracy that can be achieved.

In inductive learning, the learner is given a hypothesis space H from which it must select an output hypothesis, and a set of training examples D = $\{\langle x_1, f(x_1)\rangle,\ldots,\langle x_n, f(x_n)\rangle\}$ where $f(x_i)$ is the target value for the instance x_i. The desired output of the learner is a hypothesis *h* from *H* that is consistent with these training examples.

B. Analytical Learning

Analytical learning uses prior knowledge and deductive reasoning to augment the information provided by the training examples, so that is not subject to these same bounds. In analytical learning, the input to the learner includes the same hypothesis *H* and training examples *D* as for inductive learning. In addition, the learner is provided an additional input: A *domain theory B* consisting of background knowledge that can be used to explain observed training examples. The desired output of the learner is a hypothesis *h* from *H* that is

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

consistent with both the training examples *D* and the domain theory *B*.

C. Data Reduction

Data reduction techniques have been helpful in analyzing reduced representation of the dataset without compromising the quality of the knowledge extracted. The concept of data reduction is commonly understood as either reducing the volume or reducing the dimensions (number of attributes). There are a number of methods that have facilitated in analyzing a reduced volume or dimension of data and yet yield useful knowledge. Certain partition based methods work on partition of data tuples. However, partitioning the data table with respect to set of attributes is considered in this paper.

Attribute Relevance Analysis

The general idea behind attribute Relevance analysis is to compute some measure that is used to quantify the relevance of an attribute with respect to a given class or concept. It is not possible to be definitely sure that all attributes are informative for the given target. In real-world data, the representation of data often uses too many attributes, but only a few of them may be related to the target concept. Attribute selection and ordering procedures help us to solve this kind of problem. They order the set of input attributes according to their information content, and then select a relatively small subset of the most informative attribute. This procedure is also called attribute evaluation method (attribute relevance) as the attributes are evaluated and ordered in terms of their relevance with respect to the information content. There are a number of methods available for attribute relevance analysis. Some of them are enumerated as Information Gain, Gain Ratio, Gini Index, Principal Component Analysis, Factor Analysis.

Information Gain

The method of attribute relevance analysis used in the proposed model is Information Gain. In general terms, the expected information gain is the change in information entropy from a prior state to a state that takes some information as given:

$$IG(Ex,a) = H(Ex) - H(Ex \mid a) \tag{1.1}$$

Let *Attr* be the set of all attributes and E_x the set of all training examples, *value(x,a)* with defines the value of a specific example *x* for attribute , *H* specifies the entropy. The information gain for an attribute is defined as follows:

$$IG(E_x,a) = H(E_x) - \sum_{v \in values(a)} \frac{\left|\{x \in E_x \mid value(x,a) = v\}\right|}{|E_x|} \bullet H\left(\{x \in E_x \mid value(x,a) = a\}\right) \tag{1.2}$$

This is the heuristic that was originally used by [7].

III. PROPOSED WORK

In order to address the issues of high dimensional multi-variate data for classification dimension reduction is has been one of the essential preprocessing steps before application of a classification algorithm. The complexity of decision tree or classification-rule based classifiers may be a concern for high dimensional data, learning a sizable weight vector for SVM based classifier also poses challenge. Therefore, in order to reduce the computational effort of learning a large weight vector at one step for an SVM based classifier, working on a subset of the attributes describing the data tuples at a time in each step has been considered.

Step-wise Support Vector Machine (S-SVM)

In order to address the issue of high dimensionality in multi-variate data, dimension reduction becomes an essential pre-processing step before modeling classifier by applying a suitable algorithm. Besides complexity of decision tree or classification-rule based classifiers, large number of attributes directly impact the learning speed. Learning a weight vector corresponding to the set of attributes for an SVM-based classifier also poses significant challenge. Therefore, in order to reduce the computational effort of learning a large weight vector at one step to model, a SVM based classifier, using a subset of the attributes describing the data subtables at each step has been considered for a step-wise approach to design an SVM based model.

Algorithm Design

The set of disjoint subsets of attributes in particular partition of the attribute set may be obtained in various ways. However, attribute relevance based criterion may yield a partition of the set of all conditional attribute of a dataset with some blocks containing attributes with very low or negligible relevance and some with higher relevance. The SVM-based algorithm developed in this paper considers to analyse data-subtable corresponding a block of the partition of set of conditional attributes at a time in a step-wise manner. The algorithm for S-SVM is given below:

Algorithm

Step1. A Dataset X= {x: x } of labelled instances N and conditional Attribute A

Step2. Apply SVM and obtain the classification accuracy C_{svm} and Attribute Relevance Analysis on entire set of Attribute A.

Step3. Arrange the attributes in order of their ranks

Step4. Vertically Partition the dataset X, according to ascending order of their ranks. These n attributes have been partitioned into two subtables A_1 and A_2 $A = A_1 \cup A_2$.

Step5. Take one partition A_1 of A with the decision attribute Y and apply SVM classifier, obtain the classification accuracy C_{svm} (A_1) and predicted value attribute.

Step6. Compare $C_{svm}(A_1)$ with the threshold value of the classification accuracy C_{svm}

Step7. If $C_{svm}(A_1) \geq C_{svm}$, then end, else take the next subset of A, i.e. A_2

Step8. Add the predicted value attribute obtained from step 5 and add it to A_2 as conditional attribute

Step9. Apply SVM on set (A_2+Predicted Value Attribute) with the decision attribute Y and obtain classification accuracy $C_{svm}(A_2)$ and the new predicted value attribute.

Step10. Repeat Step 7 with $C_{svm}(A_2)$

The overall computational complexity of the algorithm for the proposed model has been worked out to be O*(k.n.m),* where *k* is the number of partitions, *n* is the number of training instances and *m* is the number of attributes .

IV. EXPERIMENTS USING PROPOSED MODEL

For the proposed method of classification namely step-wise SVM classification, the attribute evaluation of the attributes has been done using WEKA and Phase II and Phase III of the method has been carried out in StatSoft Software

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

STATISTICA Data Miner. The experiments were carried out on the following seven labeled datasets obtained from the UC Irvine ML repository [WWW] datasets. All the datasets are of varied sizes and the number of attributes in each datasets is also different. The data is numeric in all the datasets except for mushroom dataset which is categorical in nature. To check the feasibility of the approach the data sets considered are relatively medium in size and have neither very large nor very small number of attributes so that the effectiveness of the proposed method can be tested.

Using the step-wise SVM model, it has been clearly observed that vertical partitioning of the dataset into sub-tables or two subsets of the same dataset is possible, simplifies the classification by considering block of attribute set of half (50%) of the attributes. The classifier is modeled using both the subsets of the original dataset using SVM technique without any data loss in terms of information all the attributes are used to model the classifier in two steps. The ultimate motive of using step-wise SVM method for classification of the datasets instead of SVM has been evaluated and it is studied whether the step-wise procedure helps in improving the performance of learning and classification accuracy.

The accuracy of an SVM model is largely dependent on the selection of the kernel parameters such as *C* and Gamma. A grid search tries values of each parameter across the specified search range using geometric steps. To avoid over fitting, cross-validation is used to estimate the fitting provided by each parameter value set tried during the grid or search process. Here, the selection of parameters for SVM classifier has also been given importance as they affect the classification accuracy. [4] have said "In the support-vector networks algorithm one can control the trade-off between complexity of decision rule and frequency of error by changing the parameter *C*,...". As per [3] "...the coefficient *C* affects the trade-off between complexity and proportion of non-separable samples and must be selected by the user". For the step-wise SVM model, the value of *C* has been chosen through the grid search method. The value of *C* has been chosen from the range of 1 to 80. The kernel used here for training SVM is Radial Basis Function (RBF).

The experiment is conducted through 10-fold cross validation and the evaluation of the performance of the classification model is done using Confusion Matrix. The two classifiers are also evaluated using the statistical non-parametric test Wilcoxon Signed-Ranks test.

A. Data Description

The proposed method has been applied on the following seven datasets: Australian Credit Card, Credit Approval, Chess, Breast Cancer, Ionosphere, Image Segmentation, German Credit Card. All the datasets used in this study have been taken from UCI Machine Learning repository. A detailed description of the data sets is presented in Table 1.

TABLE 1. DESCRIPTION OF THE DATASETS

S.No.	Dataset	No. of Attributes	No. of Instances
1.	Australian Credit Card	15	690
2.	Credit Approval	16	690
3.	Chess	37	3195
4.	Breast Cancer	10	699
5.	Ionosphere	34	351
6.	Image Segmentation	20	2310
7.	German Credit Card	20	1000

B. Experimental Setup

The experiment was conducted with 10 runs each for each dataset. Each run means, to classify the data at split of different seed value. Seed values used for splits are 1500, 1000, 900, 800, 750, 600, 500, 350, 200, 100. The seed values were randomly selected. The RBF Kernel has been used for SVM and the parameters *C* have been selected using grid search. Experimental design is given in table 2.

V. RESULTS AND ANALYSIS

The results obtained from the application of step-wise SVM model on the selected data sets are presented in the Table 2.

TABLE 2. CLASSIFICATION ACCURACY OF SVM AND STEP WISE SVM

Dataset	**SVM**	**Step-wise SVM**
Australian Credit	85.49	**86.65**
Chess	93.698	**97.12**
Credit Approval	86.26	**86.65**
Breast Cancer	95.80	**96.56**
Ionosphere	96.01	**97.57**
Image Segmentation	94.97	**95.99**
German Credit Card	**81.30**	80.8

The proposed method has the best advantages of both the inductive learning and analytical learning. In the second phase of the method, when the first dataset is classified using SVM classifier, the predicted values of the classes for the dataset are stored and used as a new variable in second dataset. These predicted values have the information about the classes. This information about the classes is then used as the domain knowledge as in analytical learning and the second dataset is classified using SVM and this domain knowledge when used as a new variable, gives better classification accuracy and also reduces the number of support vectors.

As the preprocessing step of data mining, in this paper, attribute selection method is used but it is used as the combination of attribute selection and ordering, more precisely, attribute evaluation. Attribute relevance analysis is a process that chooses a subset of M features from the original set of N features (M<N), so that the feature space is optimally reduced according to a certain criterion (Blum and Langley, 1997).

VI. EVALUATION

For carrying out the statistical analysis of the two classifiers, Wilcoxon signed-ranks test has been used. The Wilcoxon signed-ranks test [9] is named after Frank Wilcoxon (1892–1965) and is a non-parametric alternative to the paired t-test, which ranks the differences in performances of two classifiers for each data set, ignoring the signs, and compares the ranks for the positive and the negative differences. The test was popularized by Siegel [6]. The accuracy of the two classifiers has been used to calculate the Wilcoxon test statistic and then obtained value of Wilcoxon test statistic is compared with the critical value Wilcoxon test statistic at $\alpha = 0.05$ level of significance. The obtained value of Wilcoxon test statistic for the two classifiers i.e. the proposed model and SVM is **-0.16903**, and the critical value of Wilcoxon test statistic at

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

N=7 and α =0.05 is 2, since the obtained value is less than the critical value, it has been concluded that the difference between the two classifiers are significantly different i.e. the difference between their performance is unlikely to occur by chance.

VII. CONCLUSION

There have been several SVM models used to classify data, but the SVM models have been used only as inductive learning methods. When SVM classifier has been combined with the advantages of analytical learning methods also, they are seen to perform better. In the proposed method, the model not only combines the benefit of both inductive and analytical learning methods but also takes care of the fact that no information is lost in terms of attribute selection. Here the attribute selection is also been combined with attribute ordering resulting in attribute evaluation. This has given an advantage that all the attributes have been used in the classifier and thus no information is lost. Since the model is not very complex to be used, it can be used for small as well as large data sets. One advantage of the proposed method is that the datasets that have large number of attributes and there is no other way to classify them but to have the attribute subset to classify it, there this model can be used efficiently as it doesn't carry out the classification on the entire set of attributes in one go, but classify the data in two phases using all half of the attributes at a time. This eases the work of the classifier and a dataset with large number of attributes is also classified efficiently.

REFERENCES

[1] Blum, A. L. and Langley, Pat.: Selection of relevant features and examples in machine learning. Artificial Intelligence. Vol. 97, Issues 1–2, pp. 245–271. (1997).

[2] Burges, C. J. C. and Scholkopf, B.: Improving the accuracy and speed of support vector learning machines. In: M. Mozer, M. Jordan, and T. Petsche, (eds.). Advances in Neural Information Processing Systems, Vol. 9, pp. 375–381. Cambridge, MA. MIT Press. (1997).

[3] Cherkassky, V., F. Mulier.: Learning From Data: Concepts, Theory and Methods. John Wiley & Sons, New York, NY (1998)

[4] Cortes, C. and Vapnik, V.: Support-vector networks. Machine Learning. Vol.20 (3), pp. 273-297. (1995).

[5] Joachims, T.: Text categorization with support vector machines: Learning with many relevant features. In European Conference on Machine Learning (ECML). (1998).

[6] Siegel, S.: Nonparametric Statistics for the Behavioral Sciences. McGraw-Hill. New York, (1956).

[7] Quinlan, J.R.: Induction of decision trees. Machine Learning, Vol. 1, pp. 81-106. Kluwer Academic Publishers. Boston (1986).

[8] Vapnik, 1995

[9] Wilcoxon, F.: Individual comparisons by ranking methods. Biometrics Bulletin, Vol.1, pp. 80-83. (1945).

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.