Performance analysis of support vector machine based classifiers

Zulfiqar Ali 1,2,*, Syed Khuram Shahzad 3, Waseem Shahzad 2

1Department of Computer Syde and Information Technology, University of Lahore, Lahore, Pakistan
2Department of Computer Science, National University of Computer and Emerging Science, Islamabad, Pakistan
3Department of Computer Science and Information Technology, The Superior College, Lahore, Pakistan

A R T I C L E  I N F O

Article history:
Received 28 March 2018
Received in revised form
10 July 2018
Accepted 12 July 2018

Keywords:
Classification
Support vector machine
KEEL
SVM kernel

A B S T R A C T

Classification is a challenging problem in the various fields of knowledge i.e., Pattern Recognition, Data Mining, Knowledge Discovery from Database etc. There is various classification methods are proposed in the contemporary literature. The choice of an appropriate classifier to achieve the optimal performance on a specific problem needs more empirical studies. There are various algorithmic paradigms like, Associative Classification; Decision Trees based classification, Statistical Classification and Support Vector Machines etc. which are exploited for the classification purposes. This paper investigates the performance of Support Vector Machine (SVM) based classifiers namely SMO-C, C-SVM-C, and NU-SVM-C. SVM is a very successful classification approach for the binary classification as well as non-binary classification problems. This study, performance comparative analysis of SVM based classification approach on public data sets; exploit the implementation of the corresponding classifiers in the KEEL. The SVM-C approach wins one time, draw 5 times and lost 6 times with respective other approaches. The NU-SVM-C win one time, draw 4 times and lost 7 times while SMO-C wins 5 times, draw 3 times and loss 4 times. It is shown that the performance of SMO-C is promising with respect to other SVM based classifiers.

© 2018 The Authors. Published by IASE. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

1. Introduction

Classification is a method used to build predictive models to separate and classify new data points (Elder IV, 1996; Michie et al., 1994). Classification is also known as supervised learning. Classification is a challenging problem in the field of Pattern Recognition (Tou and Gonzalez, 1974), Data Mining (Zaki et al., 1999), Knowledge Discovery from Databases (Kwasnik, 1999) and Image Processing (Van Heel et al., 1996) and Remote Sensing (Mills, 2011). There are various classification paradigms proposed in the contemporary literature. Following are the few examples of classification algorithmic paradigms i.e. Associative Classification (AS), Decision Trees based classification, Statistical Classification and Support Vector Machines etc.

Associative Classification (AS) (Ma and Liu, 1998) is a hybrid approach by combining the classification rule mining and association rule mining. The association rule mining is unsupervised learning means there are no class labels available at the time of generation of association rules. The purpose of the technique of association rules mining is to find association and relationship among items present in the transactional database. A consequent part (right-hand side) of association rule can include more than one attribute. The associative classification is supervised learning which means class label is known and provided at the time of generation of association rules. The goal of associative classification technique is to develop a classifier which can predict the class of data object which comes from testing data. Only the class attribute is on the right-hand side of the rule which is basically called consequent. In associative classification rule generation, the problem of over fitting is significant. Following sections are describing the summary of selective Associative Classification techniques for the purpose of performance analysis regarding this study. Examples of Associative Classification approaches are like CBA (Ma and Liu, 1998), CMAR-C (Li et al., 2001), FARC-HD-C (Alcala-Fdez et al., 2011a), ACO-AC (Shahzad and Baig, 2011), AntMiner, cAnt-Mine (Otero et al., 2008) and ACO-Miner etc. in (Jin et al., 2006).

The decision tree based classification approaches are also successfully applied in various fields. For the
determination of the class of a given instance, a decision procedure is represented by the decision tree (Moret, 1982). In a decision tree, each node of the tree specifies either a class name or a particular test. The decision tree based algorithms work like a divide and conquers strategy for object classification (Stone, 1984). There are various decision tree-based classification approaches available in the literature like ID4 (Schlimmer and Fisher, 1986), Quinlan’s ID3 (Quinlan, 1986), ID5R (Utgoff, 1989), SLIQ (Mehta et al., 1996), AdaBoost.NC (Wang et al., 2010).

The paradigm of statistical classification approaches possesses the explicit underlying probability model. The statistical classifiers provide a probability of being in each class rather than simply a classification (Wang et al., 2010). There are various statistical based classification approaches proposed in the literature like NB (Domingos and Pazzani, 1997; Maron, 1961), LDA-C (Fisher, 1936; Friedman, 1989; McLachlan, 2004) and Particle Swarm Optimization - Linear Discriminant Analysis (PSOLDA-C) (Lin and Chen, 2009).

Section 2 provides information of the SVM based classifiers under the focus of this study. Section 3 explains prominent Kernels for the SVMs. Section 4 explains the experimental system exploited for this study. Section 5 provides the results and discussion and final Section concludes the work.

2. Support vector machine-based classifiers

This section provides the description of the SVM based classification approaches focuses on this empirical study.

2.1. C_SVM-C

Corinna Cortes Vladimir Vapnik proposed a new classification approach based on the artificial neural networks so-called named as Support Vector Network in Cortes and Vapnik (1995). The support vector network implementation in KEEL is denoted by C-SVM-C. In this study, we use abbreviation C-SVM-C for the support vector networks that is a letter known as Support Vector Machine (SVM) in literature. The working procedure of Support Vector Network is as that it maps the input vectors into some high dimensional feature space Z via some non-linear mapping chosen a priori. The support vector network constructs a linear decision surface space possessing special properties. These properties provide the capabilities to the high generalization of the network. The C-SVM-C exploits Radial Bases Function (RBF) kernel. The general example of support vector network for a separable problem in a 2-dimensional space is given in Fig. 1. The support vectors, marked with grey squares, define the margin of largest separation between the two classes. The objective of SVM is to maximize the separation margin of two classes.

2.2. NU_SVM-C

Scholkopf et al. (2000) proposed a new version of support vector machine based learning algorithm for classification in Schölkopf et al. (2000). The abbreviation used for this support vector algorithm in KEEL implementation is NU_SVM-C. Authors incorporated a quantity $v$ in the basic SVM learner that lets one control the number of support vectors and errors. This addition of new parameter results in improvement in the SVM on the benchmark data sets. NU_SVM-C uses the radial based function kernel. In this empirical study, we exploited default parameters of the NU_SVM-C algorithm given in Table 1.

2.3. SMO-C2

John C. Platt proposed a new version of support vector machine learning algorithms named Sequential Minimal Optimization (SMO) in Zeng et al. (2008). The SMO algorithm is comparatively simple, easy in implementation, better in scaling and faster than another state of the art approached exploiting projected conjugate gradient (PCG) (Benzi et al., 1996). The SMO uses an analytic quadratic programming (QP). The SMO approach solves the SVM QP problem without storage for the extra matrix. There no requirement of iterative numerical routine invoking for each sub-problem in SMO. The SMO approach performs well on sparse data sets, with either binary or non-binary input data. This comparative study uses Polynomial Kernel implementation in SMO algorithm. The parameters exploited in this study for SMO are given in Table 1.

3. Support vector machine kernels

This section provides the mathematical description of prominent kernels exploited by the support vector machines.

3.1. Linear kernel

Eq. 1 is an example of Linear Kernel (Shimodaira et al., 2002). The linear kernel is the simplest kernel function for support vector machines. The linear
kernel is the inner product $<x, y>$ and the addition of an optional constant $c$. Where in Eq. 1, alpha ($\alpha$) shows the slope, $c$ constant term, and $T$.

$$k(x, y) = x^T y + c$$  \hspace{1cm} (1)

**3.2. Polynomial kernel**

The Polynomial kernel relation is represented in Eq. 2 (Fan et al., 1995). The Polynomial kernel is a non-stationary kernel. The more appropriate application of Polynomial kernel is for the domain of problems where all the training data is normalized. In Eq. 2, $\alpha$ shows the slope, $c$ constant term, and $d$ for the degree of the polynomial.

$$k(x, y) = (\alpha x^T y + c)^d$$  \hspace{1cm} (2)

**3.3. Gaussian kernel**

Eq. 3 shows the Gaussian kernel relation (Babaud et al., 1986). The Gaussian kernel is an example of radial basis function kernel. In Eq. 3, sigma ($\sigma$) is an adjustable parameter of the kernel. The sigma parameter plays a major role in the performance of the kernel.

$$k(x, y) = \exp \left( -\frac{|x-y|^2}{2\sigma^2} \right)$$  \hspace{1cm} (3)

**3.4. Exponential kernel**

Eq. 4 shows the mathematical relation of Exponential kernel (Choi and Williams, 1989). The exponential kernel is also a member of radial basis function kernel family. The Exponential kernel is similar to Gaussian kernel except for the square of the norm. Sigma ($\sigma$) is an adjustable parameter of the kernel and plays a major role in the performance of the Exponential kernel given in Eq. 4.

$$k(x, y) = \exp \left( -\frac{|x-y|^2}{2\sigma^2} \right)$$  \hspace{1cm} (4)

**3.5. Laplacian kernel**

The Laplace kernel is given in Eq. 5. The Laplace Kernel is also a member of the family of radial basis function kernels (Netsch and Peitgen, 1999). The Laplace kernel is equivalent to the exponential kernel except for being less sensitive to changes in the $\sigma$. The $\sigma$ value significantly influences the performance of the Laplacian Kernel.

$$k(x, y) = \exp \left( -\frac{|x-y|}{\sigma} \right)$$  \hspace{1cm} (5)

**4. Experimental set-up**

The experimental set-up used for this empirical study is given in this section. The description of datasets used for the experiment, experiment graph, parameters for the SVM based classifiers under focus and experimental results discussion is provided in the subsections.

**4.1. Data sets description**

The description of datasets used for the comparative performance analysis of the selective Associative Classifiers under this study is given in Table 1. The number of attributes (#Attributes), a number of instances in the database (#Examples) and a number of classes (#Classes) are shown in Table 1. The missing values (Missing_V) in the dataset are representing by “Yes” (missing values present)/ “No” (missing values not present). The missing values of the datasets are imputed with the KMean-MV module implemented in KEEL. The datasets are discretized with the Ameva-D module included in KEEL as the associative classifiers accept the discretized form of datasets. We use the 10-fold cross-validation model for the datasets provided in KEEL. Table 1 summarizes the main characteristics of the 12 datasets which are given at Knowledge Extraction based on Evolutionary Learning (KEEL)-dataset repository (Alcala-Fdez et al., 2011b).

**4.2. KEEL**

The Knowledge Extraction based on Evolutionary Learning (KEEL) is an open source software tool to assess Evolutionary Algorithms for data mining problems including regression, classification, clustering, and pattern mining and so on. This tool provides a simple GUI based on a data flow to design experiments with different datasets. KEEL provides a good collection of computational intelligence algorithms which can be used by the researchers in order to assess the behavior of the algorithms. Moreover, it may also be used to compare newly proposed techniques with the state-of-the-art approaches to their corresponding areas (Alcala-Fdez et al., 2011b).
results of the classifier and a statistical module for the analysis of the results produced by the algorithms used in the experiment.

Table 1: Data sets considered for the experimental study

| Data-sets | #Attributes | #Examples | #Classes | Missing V |
|-----------|-------------|-----------|----------|-----------|
| Pupa      | 6           | 345       | 2        | No        |
| Cleveland | 13          | 297       | 5        | Yes       |
| Ecoli     | 7           | 336       | 8        | No        |
| Glass     | 9           | 214       | 7        | No        |
| Haberman  | 3           | 306       | 2        | No        |
| Iris      | 4           | 150       | 3        | No        |
| Monks     | 6           | 432       | 2        | No        |
| Newthyroid| 5           | 215       | 3        | No        |
| Pima      | 8           | 768       | 2        | No        |
| Vehicle   | 18          | 846       | 4        | No        |
| Wine      | 13          | 178       | 3        | No        |
| Wisconsin | 9           | 683       | 2        | Yes       |

Fig. 2: The experiment graph

4.4. Parameters of the methods

The parameters of the SVM based classification methods (SVM-C, NU_SVM-C, and SMO-C) under the focus of this comparative study are shown in Table 2. The parameters of the methods are selected according to the recommendation of the corresponding authors of each proposal which are the default parameters settings included in the KEEL software tool. In Table 2 'N/A' indicated that the corresponding parameter does not apply to the corresponding method.

Table 2: Parameters of the methods for experiment

| Parameter Description                  | SVM-C | NU_SVM-C | SMO-C |
|---------------------------------------|-------|----------|-------|
| KERNEL Type                           | RBF   | RBF      | PloyKernel |
| C                                     | 100   | 1000     | 1     |
| epc                                   | 0.001 | 0.001    | 1.00E-12 |
| degree                                | 1     | 10       | N/A   |
| gamma                                 | 0.01  | 0.01     | N/A   |
| coeff                                 | 0     | 0        | N/A   |
| nu                                    | 0.1   | 0.1      | N/A   |
| p                                      | 1     | 1        | N/A   |
| shrinking                             | 1     | 1        | N/A   |
| toleranceParameter                    | N/A   | N/A      | 0.001 |
| RBFKernel.gamma                       | N/A   | N/A      | 0.01  |
| Normalized-PolyKernel.exponent        | N/A   | N/A      | 1     |
| Normalized-PolyKernel.useLowerOrder   | N/A   | N/A      | FALSE |
| PukKernel.omega                       | N/A   | N/A      | 1     |
| PukKernel.sigma                       | N/A   | N/A      | 1     |
| StringKernel.lambda                   | N/A   | N/A      | 0.5   |
| StringKernel.maxSubsequenceLength     | N/A   | N/A      | 3     |
| StringKernel.maxSubsequenceLength     | N/A   | N/A      | 9     |
| StringKernel.normalize                | N/A   | N/A      | FALSE |
| StringKernel.pruning                  | N/A   | N/A      | None  |
| FitLogisticModels                     | N/A   | N/A      | FALSE |
| ConvertNominalAttributesToBinary      | N/A   | N/A      | TRUE  |
| PreprocessType                        | N/A   | N/A      | Normalize |
5. Experimental results

Table 3 provides the comparative performance analysis of support vector machine based classification approaches namely SVM-C, NU_SVM-C, and SMO-C. The description of the stated approaches is given in Section 3. The critical observation of the results given in Table 3 reveals the performance of the SMO-C is better as compared to the other competitive approaches. The performance on monk’s data set is 100% for all classifiers while the minimum performance of SVM-C, NU_SVM-C, and SMO-C is on glass dataset (43.48%, 17.39%, and 47.83 %) respectively. The average performance in terms of accuracy of the classifiers on selected 12 datasets is also given in Table 3. The average performances of SVM-C, NU_SVM-C, and SMO-C are 77.635, 66.99% and 77.96% respectively.

Table 3: Comparative performance analysis of SVM based classifiers in terms of accuracy

| Dataset        | SVM-C | NU_SVM-C | SMO-C |
|----------------|-------|----------|-------|
| bupa           | 66.04 | 44.27    | 68.30 |
| cleveland      | 50.00 | 53.33    | 63.33 |
| ecoli          | 76.47 | 70.59    | 73.53 |
| glass          | 43.48 | 17.39    | 47.83 |
| haberman       | 74.19 | 35.48    | 67.74 |
| iris           | 93.33 | 86.67    | 93.33 |
| monks          | 100.00| 100.00   | 100.00|
| new-thyroid    | 95.45 | 90.91    | 95.45 |
| pima           | 67.53 | 48.05    | 68.83 |
| vehicle        | 76.47 | 67.06    | 74.12 |
| wine           | 94.44 | 94.44    | 88.89 |
| wisconsin      | 94.20 | 95.65    | 94.20 |
| Average        | 77.63 | 66.99    | 77.96 |
| Min            | 43.48 | 17.39    | 47.83 |
| Max            | 100.00| 100.00   | 100.00|

Table 4 describes the performance of SVM-C, NU_SVM-C, and SMO-C in terms of Win/Draw/Loss. The SVM-C approach wins one time, draw 5 times and lost 6 times with respective other approaches.

Table 4: Analysis in terms of win/draw/loss

|       | SVM-C | NU_SVM-C | SMO-C |
|-------|-------|----------|-------|
| Win   | 1     | 1        | 5     |
| Draw  | 5     | 4        | 3     |
| Loss  | 6     | 7        | 4     |

Fig. 3 describes the results in the graph for more insight of the performance of the SVM based classification methods. The graph presents the performance of the SVM-C, NU_SVM-C, and SMO-C on 12 public datasets described in Table 1 and the performance of methods in terms of Average, Min and Max values by considering accuracy. Fig. 3 shows that the performance of all SVM based approaches is lower on glass dataset with respect to other datasets. By considering the minimum (Min) and average values, the performance of NU_SVM-C is lower than other competitive methods.

6. Conclusion

In this paper, we perform a comparative performance analysis of classifiers based on Support Vector Machine. The selective SVM based approaches namely SMO-C, C-SVM-C, and NU-SVM-C. SVM is focused on this study on public datasets. The results of the study reveal that the performance of SMO-C is promising with respect to other SVM based classifiers. The SVM-C approach wins one time, draw 5 times and lost 6 times with respective other approaches in terms of accuracy. The NU_SVM-C win one time, draw 4 times and lost 7 times while SMO-C wins 5 times, draw 3 times and loss 4 times. Table 4 shows that the performance of SMO-C is promising with respect to other SVM based classifiers.

Acknowledgment

This research work is sponsored by the Higher Education Commission of Pakistan in the form of Scholarship for the PhD program.

References

Alcala-Fdez J, Alcala R, and Herrera F (2011a). A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning. IEEE Transactions on Fuzzy Systems, 19(5): 857-872.
