STRATEGIC MANAGEMENT, Vol. 20 (2015), No. 4, pp. 046-053

UDC 004.8
007.52

Received: May 20, 2015
Accepted: September 15, 2015

Estimating Performances of Learned Knowledge for the RBF Network as an Artificial Intelligence Method

Jasmina Dj. Novaković
Belgrade Business School, Higher Education Institution for Applied Science, Belgrade, Serbia

Abstract
The main aim of this paper is to experimentally verify the impact of filter methods on the classification accuracy of the radial basis function (RBF) network. The goal of this research is also to present and compare different algorithmic approaches for constructing and evaluating systems that learn from experience in order to make decisions and predictions, and minimize the expected number or proportion of mistakes. Fifteen real data sets and three artificial data sets have been used to compare the results of classification accuracy with the RBF network. We can conclude that it is possible to improve the system performance of inductive learning rules in different problems, using the filter methods for reducing the dimensionality of data.

Keywords
Artificial intelligence, classification accuracy,feature selection, filter method, neural network, RBF.

Introduction

Machine learning is a field of artificial intelligence that deals with the construction of adaptive computing systems that are able to improve their performances by using information from experience. Machine learning is a discipline that studies the generalization, construction and analysis of algorithms that have the ability to generalize. However, as much as the applications of machine learning are diverse, there are repetitive tasks. Therefore, it is possible to talk about the types of learning tasks that often occur. One of the most common tasks of learning that occurs in practice is classification. Classification is an important recognition of object types, for example, whether a particular tissue represents a malignant tissue or not.

There is a wide range of classification algorithms available to us, each with their own strengths and weaknesses. There is, however, no learning algorithm which works best with all problems of supervised learning. Machine learning involves a large number of algorithms such as: artificial neural networks, genetic algorithms, probabilistic models, rule induction, decision trees, statistical and pattern recognition methods, k-nearest neighbors, Naïve Bayes classifiers and a discriminatory analysis.

In this paper, the radial basis function (RBF) network is used. The RBF network offers a number of advantages, including requiring less formal statistical training, the ability to implicitly detect a complex nonlinear relationship between dependent and independent variables, the ability to detect all possible interactions between predictor variables and the availability of multiple algorithms for training. The main objective of this paper is to show that it is possible to improve the performance of the system for inductive learning rules with the RBF network for classification problems, using the filter methods and data dimensionality reduction techniques.

Various aspects of the feature selection have been studied. A search is the key topic in the study of a feature selection (Doak, 1992), such as search starting points, search directions, and search strategies. Another important aspect is how to measure the goodness of a feature subset

(Doak, 1992). Algorithms for a feature selection may be divided into filters (Fayyad & Irani, 1992; Liu & Setiono, 1996), wrappers and embedded approaches (Das, 2001). Filters methods evaluate the quality of selected features, independently from the classification algorithm; wrapper methods require the application of a classifier to evaluate this quality, whereas embedded methods perform a feature selection during the learning of optimal parameters. According to class information availability in data, there are supervised feature selection approaches as well as unsupervised feature selection approaches.

The main aim of this paper is to experimentally verify the impact of filter methods on the classification accuracy of the RBF network. For this purpose, the paper is structured in the following way. In the first part of the paper, a model of the RBF network is presented; in the second part, a description of data sets is given. The third part of the paper describes the methodology of experimental research. In the fourth part, we will try to solve a problem by using the RBF network as a supervised learning algorithm. To achieve greater forecasting accuracy and make more appropriate decision, the filter method for reducing the dimensionality of data is used. Also, in the fourth section, the results of an experimental study that have been collected during the survey are presented. In the last part of the paper, the obtained results are discussed and directions for further research are given.

1. The representation of the RBF network model

The classification of neural networks has proved to be a very good one only for more serious classification problems, where it is difficult or impossible to use the classical technique. Besides, neural networks are well-suited to work in conditions of noise in data. From the point of view of a layered mode of the organization of neurons in a network, the network can be classified into a single-and a multi-layered one. The first layer is called the input layer, the last is the output layer; all other layers are called hidden layers. As a rule, each layer receives inputs from the previous layer and sends their outputs to the next layer.

From a structural point of view, depending on the model used to build neural networks, neural networks can be divided into static and dynamic ones. In this paper, a static neural network is used. The main characteristic of static neural networks is that neurons are organized beforehand, so that neurons are connected in a way with no form of feedback. These networks cannot contain dynamic members, making them structurally stable. Since there are no dynamic members, the static response of the neural network depends only on the current state and the input values of the network parameters. Static neural networks are commonly used in the identification process, process management, and signal processing and pattern recognition. The most common types of static neural networks are the MLP and the RBF neural networks. The pseudo-code for RBF training (Basir, 2015) is shown in Figure1.

```
trainRBF (in, out, width, MaxError, data) {
hidden = 0;
net = initRBFNetwork (in, out, hidden); // init network nodes
do {
     //find the data vector that produces the largest error
     i = findMaxNetworkError (data, net); / / i = indexof vector
     / / add neuron to the RBF layer at same point as the above
     data vector
     addRBFNeuron (net, width, data (i)); / / data (i) = center
     point
     / / findthe overallnetworkerror
     NetError = trainOutputWeights (net, data);
   } while (NetError>MaxError);
}
```

Figure 1 The pseudo-code for RBF training
Source: Basir, 2015

2. The description of the data sets

Fifteen real data sets and three artificial data sets have been used for tests, taken from the University of California, Irvine (UCI) repository of machine learning databases (Frank, Asuncion, 2010). We used these data sets to compare the results of classification accuracy with the RBF network. These data sets are: breast cancer (bc),credit approval (ca), Statlog German credit data (cg), cardiography (ct), hepatitis (he), liver (li), lung cancer (lc), mammographic mass (mm), monk problems (monk1 (m1), monk2 (m2), monk3 (m3)), mushrooms (mu), Parkinson (pa), Pima Indians diabetes (pi), image segmentation (se), soybean (so), Stat log heart (sh) and congressional voting records (vo). Table 1accounts for the details for the benchmark data sets that have been used from the UCI repository of machine learning databases.

There are 18 data sets, out of which the 15 data sets are real, which means that they are obtained by collecting data from real, existing systems. The other three remaining data sets *m1*, *m2* and *m3* are artificial data sets, which means that the data have not been collected from the real sys-

tem, but rather created by the researchers for research purposes. To obtain the reference data, real and artificial data sets have been used to prove the stated hypothesis.

Five data sets have more than 20 attributes, *lc*– 56, *so*–35, *pa* and *ct*–23 and *mu*–22 attributes. The following data sets have the smallest number of attributes: *ma*–5, *li*, *m1*, *m2* and *m3*– 6 attributes. We conclude that the observed data sets are data sets with a very large number of attributes, as well as those sets that have a small number of attributes, which is good from the standpoint of research. The observed data sets are balanced because there are data sets containing only categorical or only numerical attributes, as well as data sets containing both categorical and numerical attributes.

Table 1 Display data sets. "CV" means cross-validation

Data set	Attributes			Number of classes	Training size	Test size	Reference accuracy
	Total	Categorical	Numerical				
bc	9	9	0	2	286	CV	70.30
ca	15	9	6	2	690	CV	55.50
cg	20	13	7	2	1000	CV	50.10
ct	23	0	23	3	2126	CV	95.00
he	19	13	6	2	155	CV	78.10
li	6	0	6	2	345	CV	58.10
lc	56	0	56	3	32	CV	26.80
ma	5	0	5	2	961	CV	84.00
m1	6	6	0	2	124	308	50.00
m2	6	6	0	2	169	263	67.13
m3	6	6	0	2	122	310	52.78
mu	22	22	0	2	8124	CV	51.80
pa	23	0	23	2	195	CV	76.00
pi	8	0	8	2	768	CV	65.10
se	19	0	19	7	2310	CV	14.30
so	35	35	0	19	683	CV	13.47
sh	13	3	10	2	270	CV	55.00
vo	16	16	0	2	435	CV	61.40

Source: Author

Among the analyzed data sets, only two data sets have a larger number of classes, *se*– 7 classes, and *so*–19 classes. The reason for this is the fact that in the majority of the problems of the classification, the existing instances are sorted into two, possibly three classes, and rarely into a larger number of classes.

Table 1 shows that the number of the instances provided for training varies from a small number of the collected instances, which is the case with *lc*–only 32 instances, to the events that have a much bigger number of the instances, for example *mu*–8124 instances for training. In all real datasets, the10-fold cross-validation (CV) is used. The researchers have created the artificial data sets *m1*, *m2* and *m3*, separating these data into two groups: those that will serve to train and those that will serve for testing, whereby a small number of data are used for training (on average, around 25%). The last column of the table shows the reference accuracy for the real and the artificial data sets.

3. The methodology of the experimental research

The experiment was performed by using the WEKA (Waikato Environment for Knowledge Analysis) tools for data preparation and research, developed at the University of Waikato in New Zealand. When searching for the model that best approximates the target function, it is necessary to provide measures of quality models and learning. In our experimental studies, different measures can be used, depending on the problem; classification accuracy, as a measure of the quality of the model, has been used.

To obtain a more reliable evaluation of learned knowledge, cross-validation has been used, where there is a full data set split into *n* approximately equal subsets. In doing so, in each iteration, there is an n-1 training subset; after the training, the quality of learned knowledge is assessed in the last remaining subset. The procedure described above is repeated for all other subsets extracted as the final quality score, obtained by taking the average score for each of the subsets. In our experimental study, the taken value of *n* is 10. Cross-validation has been used in our experimental study, because the procedure leads to a stable quality evaluation. The advantage of this method is that each of the *n* steps of cross validation using a large amount of data in their training and all available instances at one time have been used for the purpose of testing.

In the experimental research, filter methods have been used in order to reduce the dimensionality of the data. In our experiment, a solution with the number of attributes that will be used further in the study has been selected, which gives the highest classification accuracy. The results provide accuracy obtained as an average often

repetitions, each time with 10-fold cross-validation.

In our experimental research, the Paired t-test has been used, where the level of significance has been set to a value of 0.05.The Paired *t*-testis used if there is a simultaneous determination of classification accuracy for different data sets via two methods in order to determine whether the value obtained by different methods differs significantly. The Paired *t*-test is used to test the significance of the mean differences pairs *d*. according to the following equation:

$$t = \frac{\bar{d}\sqrt{N}}{s_d}$$

where s_d is a standard deviation of the obtained differences. If the calculated value of the parameter t is greater than the tabular one (a critical value), the null hypothesis is rejected and d is said to be significantly different from zero, or that the difference in the pairs is statistically significant.

The tables presented below, the ones accounting for classification accuracy, and the others accounting for the time needed for training data, use the signs "+" and "-" to indicate either statistically better (+) or worse (-) results. The basic level of significance of the classifier is specified in the value of 0.05. In the tables of classification accuracy, the sign "+" indicates a significantly higher value for classification accuracy, whereas the sign "-" indicates a significantly lower value for classification accuracy.

The tables that contain the data of the time needed for training data, the "+" sign indicates a significantly lower value for the required time, which means that it is a statistically better result, whereas the "-" sign indicates a significantly higher value for the required time, which means that it is a statistically worse outcome. Considering that the time required for training data can be changed, if different methods for reducing the dimensionality of the data are applied, it is good that during the experiment a smaller value for the required training time can be obtained, because then our algorithm works faster, which is especially important if we have a problem in real time.

In our experimental research, whenever two or more algorithms are compared, the table of classification accuracy and the table for the time needed for training data are given. The comparison is such that the second algorithm is an algorithm in which the pre-selection attributes are performed, and the first algorithm is the standard algorithm without the pre-selection of attributes. When the results for the time required for training data are shown, they are expressed in units of CPU seconds. The experiment was performed on the AMD Phenom (tm) 9650 Quard-Core Pocessor 2.31 GHz with 4GB of RAM.

Filtering methods work independently from the classification algorithms. The attribute value is estimated by analyzing the general characteristics of the data from the learning set. These methods use different techniques of the attributes selection, because there are several ways of the heuristic evaluation of attributes. Filtering methods are divided into two main groups, depending on whether the subsets of attributes or individual attributes are heuristically evaluated.

In this paper, the following filtering methods for ranking attributes, statistically and entropy-based, showing good performance in various domains are used: Information Gain (IG), Gain Ratio (GR), Symmetrical Uncertainty (SU), Relief-F (RF), One-R (OR) and Chi-Squared (CS). All these methods rank attributes for each data set. Considering that the method of ranking attributes ranks all the attributes in the order of their importance for the classification problem, these methods do not perform an automatic reduction in the number of the attributes. In order to realize a reduction in the number of attributes, there are two possibilities: (1) to use a threshold or (2) to use an appropriate number of attributes for each data set and each of the filtering methods. In this experimental study, the second possibility is used.

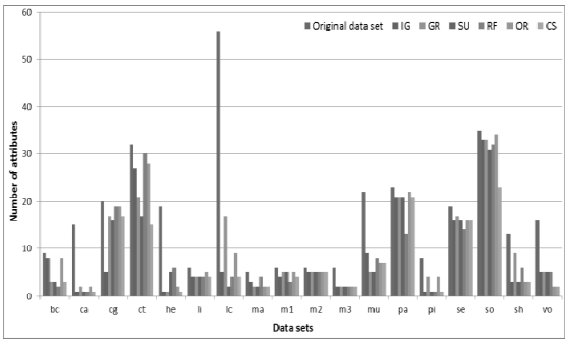

Figure 2 The number of the attributes in the original data set and the optimal number of the attributes obtained by the filter methods
Source: Author

The number of the attributes in the original data set and the optimal number of the attributes obtained by the filtering methods are shown in Figure 2. In the ten data sets out of the 18 observed ones, exactly one-half or more than one-half of the methods reduce the number of the

attributes to one-half. These data sets are *bc*, *ca*, *he*, *lc*, *ma*, *m3*, *mu*, *pi*, *sh* and *vo*. The greatest benefit of reducing the dimensionality of the data set belongs to *lc*, with 56 attributes; the filtering methods have selected a small number of attributes, even less than one-sixth, for each method, except the GR method. For the data set *ca* observe that all filtering methods have proved to be the most significant two attributes for the given classification problem, and that the other attributes do not affect the achievement of the greater reliability of the classification. For the *he* dataset, which originally has 19 attributes, all the filtering methods show that there are six most important attributes for the studied phenomenon.

For the artificial data set *m3*, all the filtering methods show that there are only two significant attributes for the given classification problem. The filtering methods for the *pi* data set show that there are four attributes important for the classification problem, and in the case of the *vo* data set, five attributes are important.

In the following experimental research, for the optimal number of the selected attributes for each data set and each filtering method, the classification accuracy of the RBF network is checked. The following text presents the results obtained. It should be noted that different scales are shown in the figures for the absolute classification accuracy in order to better determine the differences existing among the results.

4. The results of the experimental research

The classification accuracy of the RBF algorithm can be observed in Table 2. In the three data sets (*ca*, *m1* and *se*),there are the results for at least one of the filtering methods that are statistically better than the base classifier. In one data set (*ca*), all the filtering methods were statistically better than the base classifiers.

Figures 3, 4 and 5 show the absolute difference in the classification accuracy of the RBF algorithm on the original data set and the RBF algorithm with different filtering methods. The applied method of filtering IG on almost one-half of the data sets (8 sets) shows the same results as or better results than the RBF algorithm on the original data set, whereas in one data set, the result was statistically better. In two-thirds of the data sets (12 sets), the GR method shows the same results as or better results than the RBF algorithm on the original data set, whereas on one data set, the result was statistically better.

Table 2 The classification accuracy of the RBF algorithm for the original and the reduced data sets

Data set	RBF	RBF_ IG	RBF_ GR	RBF_ SU	RBF_ RF	RBF_ OR	RBF_ CS
bc	71.41	71.34	74.32	74.46	71.00	71.20	73.62
ca	79.55	85.51 +	85.43 +	85.51 +	85.51 +	85.10 +	85.51 +
cg	73.58	74.12	73.33	73.64	73.54	73.16	73.54
ct	97.93	98.35	98.41	97.65	98.13	96.90	96.27 -
he	85.29	81.31	83.45	83.05	80.49	82.69	81.25
li	65.06	58.16 -	58.16 -	58.16 -	57.33 -	60.96	58.16 -
lc	76.00	73.58	79.75	79.00	76.75	72.92	74.92
ma	77.31	77.66	79.67	79.24	77.07	77.51	79.16
m1	75.36	76.70	77.76	77.76	90.01 +	75.37	76.70
m2	67.82	63.53	63.54	63.53	64.77	64.77	63.53
m3	96.54	96.39	96.39	96.39	96.39	96.39	96.39
mu	98.61	98.06	98.99	98.99	98.43	98.55	98.55
pa	81.22	80.92	80.92	80.92	83.39	81.98	80.67
pi	74.04	73.84	76.28	73.84	73.84	75.32	73.84
se	87.31	87.84	87.56	87.84	88.88 +	87.84	87.84
so	90.79	91.11	91.20	91.59	91.29	90.57	91.42
sh	83.11	78.44	83.44	78.15	81.56	78.44	78.52
vo	93.73	94.60	94.60	94.60	94.92	95.63	95.63

Source: Author

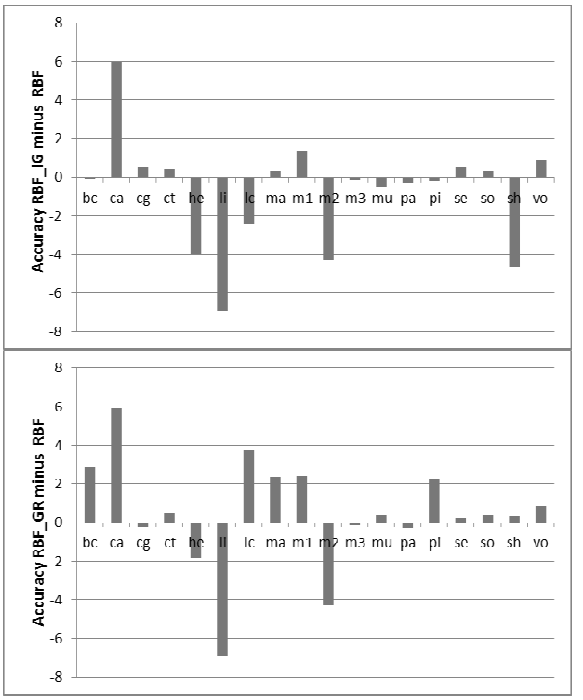

Figure 3 The absolute classification accuracy RBF_IG minus RBF and RBF_GR minus RBF
Source: Author

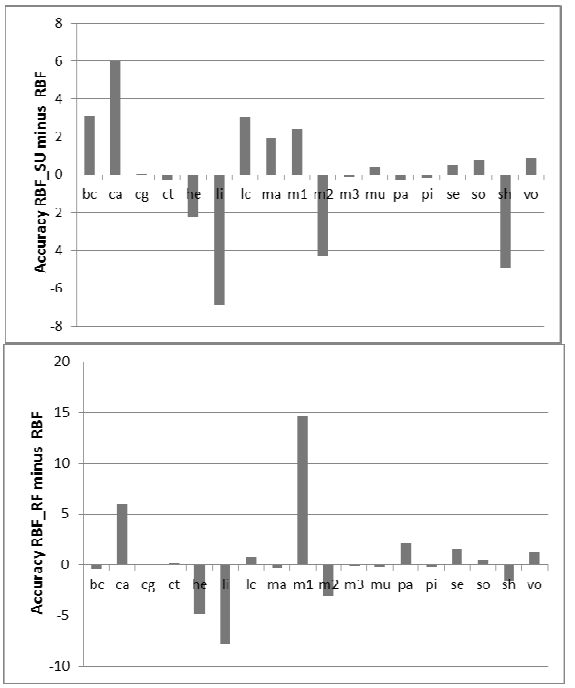

Figure 4 The absolute classification accuracy RBF_SU minus RBF and RBF_RF minus RBF
Source: Author

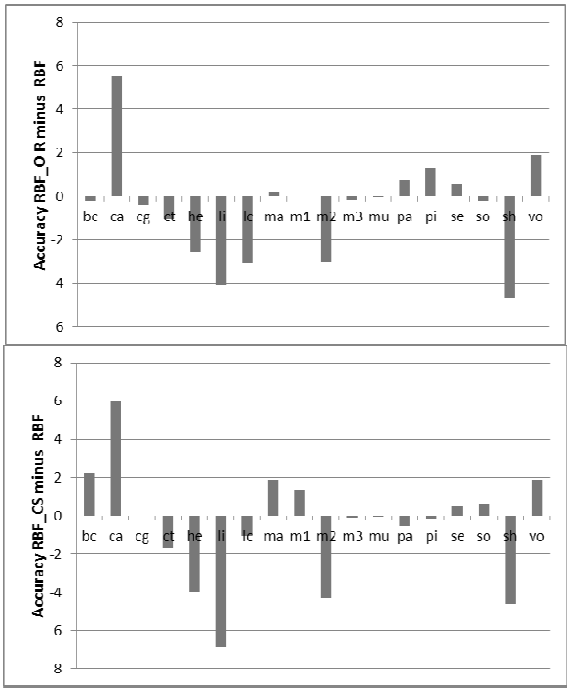

Figure 5 The absolute classification accuracy RBF_OR minus RBF and RBF_CS minus RBF
Source: Author

In more than one-half of the data sets (10 sets),the applied method of filtering SU shows the same results as or better results than the RBF algorithm on the original data set, whereas in one data set, the result was statistically better. In less than one-half of the data sets (8 sets),the RF filtering method shows the same results as or better results than the RBF algorithm on the original data set, whereas in the three data sets, the results were statistically better.

In slightly less than one-half of the data sets (7 sets),the OR filtering method shows the same or better results than the RBF algorithm on the original data set, and in one data set, the result was also statistically better. The applied method of filtering CS in less than one-half of the data sets (7 sets) shows the same results as or better results than the RBF algorithm on the original data set, and in one data set, the result was statistically better. In three cases, the RBF classifier together with the RF filtering method has led to statistically better results in the observed data sets, compared with the other filtering methods (none or one case).

Table 3 The standard deviation for the classification accuracy of RBF with the original and the reduced data sets

Data set	RBF	RBF_IG	RBF_GR	RBF_SU	RBF_RF	RBF_OR	RBF_CS
bc	7.88	7.66	6.41	6.16	6.81	8.34	6.28
ca	4.07	3.96	4.03	3.96	3.96	4.14	3.96
cg	4.30	3.46	3.92	3.74	4.31	4.25	4.03
ct	1.02	1.02	0.94	1.21	1.00	2.19	1.29
he	8.29	7.38	8.54	7.90	9.44	8.25	7.51
li	8.80	8.10	8.10	8.10	7.66	9.62	8.10
lc	22.91	22.91	21.10	20.70	23.31	22.17	22.52
ma	3.31	3.67	4.14	4.51	3.83	4.35	4.50
m1	5.92	9.44	7.97	7.97	8.10	7.97	9.44
m2	6.24	4.19	4.21	4.19	4.79	4.79	4.19
m3	2.19	2.20	2.20	2.20	2.20	2.20	2.20
mu	0.58	6.77	4.93	4.93	0.90	4.86	4.86
pa	7.37	7.49	7.42	7.53	7.39	7.24	7.35
pi	4.91	4.65	5.18	4.65	4.65	5.31	4.65
se	2.15	1.91	2.15	1.89	1.76	1.91	1.90
so	2.92	2.69	3.09	2.93	2.90	2.92	3.24
sh	6.50	7.28	6.44	7.25	7.29	7.13	7.28
vo	3.87	3.25	3.28	3.30	3.10	2.76	2.76

Source: Author

The standard deviation for the classification accuracy of RBF with the original and the reduced data sets by applying the filter methods are shown in Table 3. The Table accounts for the fact that standard deviations generally do not differ much between the standard algorithm and the algorithms that use a reduced data set. The least devia-

tion in the standard deviation is demonstrated by the RF method, so that for particular data sets, the standard deviation is less, but in some cases the standard deviation is greater.

The time required for training data with the RBF algorithm that uses the original and the reduced data sets is accounted for in Table 4. The time required for training data with the RBF classifier for all original data sets is below 1.00 seconds, except for the two sets of data *se* and *so*, where the required time is significantly longer. The time required for training data with some of the methods of filtering is longer, while for some it is less than for the original data sets. In all data sets, at least one method of filtering provides the same or better results for the time required for training compared to the original data sets.

Table 4 The time required for training data (in seconds) with the RBF algorithm with the original and the reduced data sets

Data set	RBF	RBF_ IG	RBF_ GR	RBF_S U	RBF_ RF	RBF_ OR	RBF_ CS
bc	0.01	0.01	0.01	0.00	0.02 -	0.02	0.01
ca	0.03	0.01 +	0.01 +	0.01 +	0.20 -	0.04 -	0.01 +
cg	0.05	0.02 +	0.04	0.04	0.53 -	0.10 -	0.04
ct	0.39	0.44	0.34	0.34	4.57 -	0.55 -	0.34
he	0.01	0.00	0.00	0.00	0.01	0.02 -	0.00
li	0.01	0.01	0.01	0.01	0.03 -	0.02	0.01
lc	0.00	0.00	0.00	0.00	0.00	0.02 -	0.00
ma	0.02	0.02	0.02	0.02	0.19 -	0.03 -	0.02
m1	0.01	0.01	0.01	0.01	0.06 -	0.02	0.01
m2	0.01	0.01	0.01	0.01	0.06 -	0.02	0.01
m3	0.01	0.01	0.01	0.01	0.06 -	0.02	0.01
mu	0.49	0.33 +	0.34 +	0.33 +	30.38 -	1.08 -	0.33 +
pa	0.02	0.02	0.02	0.02	0.04 -	0.04 -	0.02
pi	0.03	0.01 +	0.02 +	0.01 +	0.17 -	0.04 -	0.01 +
se	4.04	4.09	3.77	4.20	7.41 -	3.96	4.28
so	248.83	214.28	238.50	242.58	266.20	249.31	248.23
sh	0.01	0.01	0.01	0.01	0.04 -	0.02	0.01
vo	0.01	0.01	0.01	0.01	0.07 -	0.03 -	0.01

Source: Author

The applied method of filtering IG in two data sets shows worse results for the time required for training data; and for four data sets, the results are statistically better. The applied method of filtering GR in none data sets shows worse results for the time required for training data; and for three data sets, the results are statistically better.

The applied method of filtering SU in only one data set shows worse results for the time required for training data; in three data sets, the results are statistically better. The RF filtering method in all data sets shows the same or worse results for the time required for training data, and in almost all datasets, the results are statistically worse. The RF filtering method in all data sets shows the same results as or worse results than the RBF algorithm on the original data set for the time required for training data, and in almost all data sets, the results are statistically worse.

The OR filtering method in almost all data sets shows worse results than the RBF algorithm on the original data set, and in most cases, these results are statistically worse. The applied method of filtering CS in only one data set shows worse results than the RBF algorithm on the original dataset; in three data sets, the results are statistically better. Using the RBF classifier, the IG filtering method in four cases has led to statistically better results for the time required for training on the observed data sets, which is a better outcome compared with the other filter methods.

The standard deviation for the time required for training data with the RBF algorithm for the original and the reduced data sets with the filter methods is shown in Table 5. The Table accounts for the fact that the standard deviations generally do not differ a lot between the standard algorithm and the algorithms that use a selection of attributes, except for the data set *so*, where, with the help of some methods, this value is significantly higher or significantly lower than in the original data set. The maximum deviation of the standard deviation shows the method SU for the *so* data set.

Table 5 The standard deviation for the time required for training data (in seconds) with the RBF algorithm for the original and the reduced data sets

Data set	RBF	RBF_ IG	RBF_ GR	RBF_ SU	RBF_ RF	RBF_ OR	RBF_ CS
bc	0.01	0.01	0.01	0.01	0.01	0.01	0.01
ca	0.01	0.01	0.01	0.01	0.01	0.01	0.01
cg	0.01	0.01	0.01	0.01	0.02	0.01	0.01
ct	0.09	0.11	0.06	0.06	0.10	0.07	0.05
he	0.01	0.01	0.01	0.01	0.01	0.01	0.01
li	0.01	0.01	0.01	0.01	0.01	0.01	0.01
lc	0.01	0.01	0.01	0.01	0.01	0.01	0.00
ma	0.01	0.01	0.01	0.01	0.01	0.01	0.01
m1	0.01	0.01	0.01	0.01	0.01	0.01	0.01
m2	0.01	0.01	0.01	0.01	0.01	0.01	0.01
m3	0.01	0.01	0.01	0.01	0.01	0.01	0.01
mu	0.06	0.06	0.08	0.08	0.13	0.04	0.04
pa	0.01	0.01	0.01	0.01	0.01	0.01	0.01
pi	0.01	0.01	0.01	0.01	0.01	0.01	0.01
se	1.30	1.19	1.00	1.27	1.61	1.21	1.30
so	129.82	104.56	131.39	159.87	142.67	122.82	120.93
sh	0.01	0.01	0.01	0.01	0.01	0.01	0.01
vo	0.01	0.01	0.01	0.01	0.01	0.01	0.01

Source: Author

Discussion of the results and future research

According to the obtained results, a conclusion can be drawn that it is possible to improve the system performance of inductive learning rules in different problems, using the filter methods for reducing the dimensionality of data. To prove the hypothesis, the filter methods for reducing the dimensionality of data have been implemented and empirically tested. The experimental results reveal that the methods effectively applied contribute to the detection and elimination of irrelevant, redundant data and noise in data. In many cases, the filter methods select relevant attributes and contribute to greater classification accuracy. In the experimental study, the following has been demonstrated:

- the application of the previous selection of attributes by using the filtering methods with the RBF algorithm for classification leads to a reduction in the negative effects of the high dimensionality of data,
- the previous selection of an attribute by the method of filtering in some cases leads to a significant reduction in time to build a model,
- applying the method of filtering in the system for inductive learning, it is possible in some cases to significantly improve the accuracy of the existing learning methods.

In a further research, it would be interesting to apply other techniques to solve the problem of the dimensionality reduction in data, such as wrapper methods and the extraction of attributes, and analyze and compare effects of their implementation. These techniques could also improve the performance of classification learning algorithms. SM

References

Basir, O. (2015). *Neural networks.* Retrieved June 22, 2015, from University of Waterloo, UW Centre for Pattern Analysis & Machine Intelligence: http://pami.uwaterloo.ca/~basir/ECE457/week12.pdf

Das, S. (2001). Filters, wrappers and a boosting-based hybrid for feature selection. *Proceedings of the Eighteenth International Conference on Machine Learning* (pp. 74-81). San Francisco: Morgan Kaufmann Publishers.

Doak, J. (1992). *An evaluation of feature selection methods and their application to computer security.* Davis: University of California, Department of Computer Science.

Fayyad, U. M., & Irani, K. B. (1992). The attribute selection problem in decision tree generation. *Proceedings of the Ninth National Conference on Artificial Intelligence* (pp. 104-110). Cambridge MA: AAAI Press/The MIT Press.

Frank, A., & Asuncion, A. (2010). *UCI Machine learning repository.* Retrieved June 22, 2015, from University of California, School of Information and Computer Science: http://archive.ics.uci.edu/ml

Liu, H., & Setiono, R. (1996). A probabilistic approach to feature selection - a filter solution. *Proceedings of International Conference on Machine Learning (ICML-96)* (pp. 319-327). San Francisco: Morgan Kaufmann Publishers.

✉ **Correspondence**

Jasmina Dj. Novaković

Higher Education Institution for Applied Science,
Kraljice Marije 73, 11000, Belgrade, Serbia

E-mail: jnovakovic@sbb.rs