ESANN'2006 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2006, d-side publi., ISBN 2-930307-06-4.

Rotation-based Ensembles of RBF Networks

Juan J. Rodríguez[1], Jesús M. Maudes[1] Carlos J. Alonso[2] *

1- Escuela Politécnica Superior - Lenguajes y Sistemas Informáticos
Universidad de Burgos - Spain

2- Facultad de Informática - Departamento de Informática
Grupo de Sistemas Inteligentes - Universidad de Valladolid - Spain

Abstract. Ensemble methods allow to improve the accuracy of classification methods. This work considers the application of one of these methods, named Rotation-based, when the classifiers to combine are RBF Networks. This ensemble method, for each member of the ensemble, transforms the data set using a pseudo-random rotation of the axis. Then the classifier is constructed using this rotation data. The results of the ensembles obtained with this method are compared with the results using other ensemble methods (including Bagging and Boosting), over 34 data sets.

1 Introduction

Using ensemble methods [1] is one of the most natural methods for trying to improve the accuracy of a neural network classifier. This work considers an ensemble method, named Rotation-based, for the combination of RBF Networks.

One of the most used approaches for constructing an ensemble of classifiers is transforming the data set for each member of the ensemble. *Bagging* [2] transforms the data set by resampling the original data set. An instance can be selected several times, so some instances will not be present in the new data set. The *Random Subspaces* [3] method obtains a new data set deleting some attributes. *Boosting* [4] is a family of methods. The most prominent member is AdaBoost. In this case the data set is modified depending on the classification errors of the previously generated base classifier. The wrong classified examples are assigned a greater weight, hence the next classifier will give more importance to those examples. Comparatives among ensemble generation methods are presented in [5, 6, 7].

These method share the idea that it is necessary to modify the data set in a way that some information is lost (i.e., instances, attributes). None of the modifications would be considered if it was desired to obtain a unique classifier. They are used because ensemble methods need diverse base classifiers.

Rotation-based ensembles [8] transform the data set, but in a way that all the present information is preserved (although it is transformed). The idea is to group the attributes, and for each group to apply an axis rotation. Hence, all the available information (instance and attributes) in the data set is still available. When the method was presented [8] it was compared with other

*This work has been partially funded by Spanish Ministry of Education and Culture, through grant DPI2005-08498, and Junta Castilla y León VA088A05.

ensemble methods using decision trees. This work evaluates its performance when using RBF Networks.

The rest of the paper is organised as follows. Section 2 describes the method. The experimental validation of the method is presented in section 3. Finally, section 4 concludes.

2 Rotation-based Ensembles

2.1 Method Description

This method was presented in [8]. For each member of the ensemble a pseudo-random transformation of the data set is done. The transformation of the data set is done in the following steps:

- The input variables are randomly grouped.
- For each group of input variables:
  - Consider a data set formed by this input variables and all the examples.
  - Eliminate from the data set all the examples from a proper subset of the classes.
  - Eliminate from the data set a subset of the examples.
  - Apply PCA (Principal Component Analysis) with the remaining data set.
  - Consider the components of PCA as a new set of variables. None of the components is discarded.
- Transfrom all the training data set using as new variables the components selected by PCA for each group.

This transformation produces a rotation of the axis. The transformed data set has as many examples as the original data set. All the information that was in the original data set remains in the transformed data set, because none of the components is discarded.

The elimination of classes and examples of the data set is done because PCA is a deterministic method, and it is not difficult (specially for big ensembles) that some members of the ensemble had the same grouping of variables. Hence, an additional source of diversity is needed.

Rotation-based ensembles are not useful for any kind of classifiers because it needs a method that is not robust to axis rotations. In [8] decision trees were used, because they are very sensitive to rotations.

Data Set	Classes	Examples	Nominal	Numeric	Data Set	Classes	Examples	Nominal	Numeric
anneal	6	898	32	6	letter	26	20000	0	16
audiology	24	226	69	0	lymphography	4	148	15	3
autos	7	205	10	16	mushroom	2	8124	22	0
balance-scale	3	625	0	4	pendigits	10	10992	0	16
breast-cancer	2	286	10	0	pima-diabetes	2	768	0	8
cleveland-14-heart	5	307	7	6	primary-tumor	22	239	17	0
credit-rating	2	690	9	6	segment	7	2310	0	19
german-credit	2	1000	13	7	sonar	2	208	0	60
glass	7	214	0	9	soybean	19	683	35	0
heart-statlog	2	270	0	13	splice	3	3190	60	0
hepatitis	2	155	13	6	vehicle	4	846	0	18
horse-colic	2	368	16	7	vote	2	435	16	0
hungarian-14-heart	5	294	7	6	vowel-context	11	990	2	10
hypothyroid	4	3772	22	7	vowel-nocontext	11	990	0	10
ionosphere	2	351	0	34	waveform	3	5000	0	40
iris	3	150	0	4	wisconsin-breast	2	699	0	9
labor	2	57	8	8	zoo	7	101	16	2

Table 1: Characteristics of the data sets.

2.2 Application to RBF Networks

RBF Networks can be constructed using a lot of different methods. A classical approach consists of two steps. First, the centers and radii are selected. Then, the output-layer weights are calculated.

In this work the implementation of the RBF Networks available in WEKA [9] is used. For the first step, the centers are selected using the k-means clustering method. For each cluster and input variable, the radius is calculated as the deviation of the variable for the data in the cluster. For the second step, in classification problems, Logistic Regression is used.

This method can be used with Rotation-based ensembles because it is sensitive to rotations. The radii are calculated for each input variable. If the data is transformed using an axis rotation, the radii are calculated from different variables.

3 Experimental Validation

3.1 Data Sets

The data sets considered in this study appear in table 1. All of them are from the UCI Repository [10]. For the data sets "splice" and "zoo" one attribute was eliminated, because they were instance identifiers. For the data set "vowel" another attribute was eliminated. It indicated if the example should be used for training and testing. Moreover, two versions of the data set "vowel" were considered. This data set includes some context information that is discarded in some of the references that use this data set.

3.2 Settings

The experiments were done using WEKA [9], because it includes the implementation of ensemble methods and RBF Networks.

For the RBF Networks the default parameters were used, with the only exception of the number of centers. The default value is 2 centers for each class, in this work 5 were used.

The number of classifiers in the ensembles was 10. The methods considered were:

1. A single RBF Network.
2. A random ensemble. It is formed by several networks that were obtained with different random seeds. The RBF construction method is not deterministic, because it uses the k-means method for selecting the centers. Hence, if different random seeds are used, different networks are obtained from the same data set and the same set of parameters.
3. Bagging [2].
4. Boosting [4]. The original boosting method was for binary problems. There are several variants for the multiclass case, we consider AdaBoost.M1 [11].
5. Rotation-based. It has a parameter, the size of the groups of input variables. It was set arbitrarily to 3.

For comparing the methods the *corrected resampled t-test statistic* from [12] was used. One of the settings recommended in this reference was used: each method was constructed and evaluated 15 times, using a random partition of the data with 90% for training and 10% for testing. The same partitions were used for all the methods.

3.3 Results

Table 2 shows the classification accuracy of the different methods for each data set. Rotation-based ensembles were compared with the rest of the methods using the considered statistic. Rotation-based ensembles has more significant improvements that degradations for all the other methods. The best of the remaining methods is boosting. Rotation-based is better than boosting 4 times and worse 3 times.

Table 3 shows the number of data sets where the method of the column has a better result than the method of the row. For instance, Rotation-based is better than boosting for 26 data sets and worse for 8. According to this table, clearly the best ensemble-method is Rotation-based.

Table 4 is similar to the the previous table. The difference is that it only considers the cases where the difference is significant according to the test. This table also indicates that Rotation-based is the best method.

4 Conclusions and Future Work

This work has presented the application of an ensemble method, Rotation-based, using RBF Networks. For this type of networks, the experimental validation

Data Set	Rotations	Single	Random	Bagging	Boosting
anneal	96.81±1.73	96.89± 1.69	98.00± 1.13	97.93± 1.45	98.00± 1.40
audiology	78.63±6.06	70.15± 6.80 •	75.75± 6.89	76.49± 6.99	72.30± 8.71
autos	74.97±9.38	68.81±10.10	74.13± 8.94	74.96± 8.98	75.88± 6.76
balance-scale	89.45±2.93	86.35± 4.19 •	87.73± 3.69	87.09± 2.95 •	86.04± 4.84 •
breast-cancer	70.36±5.04	70.33± 7.92	71.08± 6.57	72.01± 7.43	67.07± 4.79
cleveland-14-heart	82.15±5.84	81.93± 5.00	82.83± 5.17	83.72± 6.12	81.75± 5.98
credit-rating	84.84±3.62	81.35± 4.52	82.14± 4.13	82.62± 3.37	81.74± 3.50
german-credit	74.00±3.23	71.80± 3.76	75.40± 4.03	75.33± 3.58	72.47± 4.67
glass	72.49±7.14	63.06±10.10 •	71.50± 9.09	70.03±10.87	64.65± 8.68
heart-statlog	81.98±7.90	82.72± 6.05	84.44± 7.03	84.20± 7.98	78.02± 8.45
hepatitis	84.13±7.72	82.85± 8.06	83.71± 5.95	84.18± 4.47	82.85± 7.25
horse-colic	80.99±6.96	77.54± 8.28	79.75± 8.20	80.66± 6.68	77.60± 6.49
hungarian-14-heart	78.26±5.23	78.94± 9.35	78.93± 6.93	79.57± 7.38	76.40± 9.25
hypothyroid	93.55±0.63	94.93± 0.87 ∘	95.44± 0.90 ∘	95.62± 0.91 ∘	95.69± 1.00 ∘
ionosphere	94.50±3.96	93.15± 4.30	94.28± 4.08	93.91± 3.70	93.55± 3.93
iris	97.78±4.11	96.00± 4.91	96.89± 4.95	96.89± 4.95	95.56± 5.44
labor	95.56±7.63	87.33±11.97	93.11±10.80	94.44±10.29	87.33±11.97
letter	91.13±1.01	85.35± 0.84 •	88.73± 0.92 •	89.27± 0.76 •	91.28± 0.66
lymphography	85.89±8.92	82.48± 9.64	83.01± 8.28	84.35± 7.75	82.45± 9.90
mushroom	99.30±0.26	99.58± 0.26 ∘	99.82± 0.19 ∘	99.86± 0.15 ∘	99.98± 0.06 ∘
pendigits	98.07±0.40	97.13± 0.65 •	97.51± 0.65 •	97.63± 0.60 •	98.84± 0.45 ∘
pima-diabetes	76.81±3.87	73.50± 4.53 •	74.81± 4.58	74.73± 5.17	73.15± 3.46 •
primary-tumor	40.52±7.70	31.75± 6.11 •	37.25± 7.44	38.56± 7.57	36.58± 6.29
segment	94.63±1.86	92.47± 1.80 •	93.51± 1.42 •	93.65± 1.71	94.98± 1.31
sonar	84.57±9.78	83.55± 8.46	85.18± 8.64	86.20± 9.12	82.28±10.46
soybean	94.40±2.15	86.97± 4.82 •	89.23± 3.10 •	90.01± 3.22 •	92.13± 2.63
splice	95.88±1.07	95.51± 1.18	96.49± 0.70	96.57± 0.92	95.63± 0.91
vehicle	76.66±3.78	71.38± 5.21 •	73.28± 5.29	73.83± 4.99	73.66± 4.51
vote	96.02±2.51	96.03± 2.34	95.88± 2.75	95.26± 2.50	94.19± 4.21
vowel-context	95.76±2.10	90.10± 2.76 •	95.89± 1.96	95.96± 2.06	95.08± 2.61
vowel-nocontext	96.77±2.48	89.76± 3.17 •	95.08± 1.98	95.69± 2.14	94.68± 2.07 •
waveform	87.14±1.19	86.64± 1.49	86.86± 1.41	86.85± 1.30	83.84± 1.26 •
wisconsin-breast	95.44±2.86	94.49± 1.77	94.68± 2.67	95.16± 2.45	94.30± 2.96
zoo	91.12±9.75	93.32± 6.52	94.50± 7.20	95.27± 5.30	93.32± 6.52

∘, • statistically significant improvement or degradation

Table 2: Classification accuracy.

	Single	Random	Bagging	Boosting	Rotations
Single	-	32	33	20	27
Random	2	-	24	9	22
Bagging	1	9	-	8	21
Boosting	12	25	26	-	26
Rotations	7	12	13	8	-

Table 3: Summary of results. Number of data sets where the method of the column has a better than the method of the row.

	Single	Random	Bagging	Boosting	Rotations
Single	-	9	10	8	12
Random	0	-	1	4	4
Bagging	0	0	-	3	4
Boosting	1	3	2	-	4
Rotations	2	2	2	3	-

Table 4: Summary of results. Number of data sets where the method of the column is significantly better than the method of the row.

shows that the ensembles obtained with the proposed method are better than the ensembles obtained with other methods.

Rotation-based ensembles are compatible with other ensemble methods, because it only defines a random transformation of the data set. It would be possible to do Bagging and Boosting using this transformation of the data set for each member of the ensemble.

This work has only considered RBF Networks for classification problems. Hence, it would be interesting to test different ensemble methods, including Rotation-based, for regression problems.

References

[1] L. I. Kuncheva. *Combining Pattern Classifiers: Methods and Algorithms.* Wiley-Interscience, 2004.

[2] L. Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996.

[3] T. K. Ho. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8):832–844, 1998.

[4] R. E. Schapire. The boosting approach to machine learning: An overview. In *MSRI Workshop on Nonlinear Estimation and Classification*, 2002. `http://www.cs.princeton.edu/~schapire/papers/msri.ps.gz`.

[5] E. Bauer and R. Kohavi. An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Machine Learning*, 36(1–2):105–139, 1999.

[6] T. G. Dietterich. Ensemble methods in machine learning. In *Multiple Classifier Systems 2000*, pages 1–15, 2000.

[7] T. K. Ho. A data complexity analysis of comparative advantages of decision forest constructors. *Pattern Analysis and Applications*, 5:102–112, 2002.

[8] J. J. Rodríguez and C. J. Alonso. Rotation-based ensembles. In *Current Topics in Artificial Intelligence: 10th Conference of the Spanish Association for Artificial Intelligence*, volume 3040 of *Lecture Notes in Artificial Intelligence*, pages 498–506. Springer, 2004.

[9] I. H. Witten and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques.* Morgan Kaufmann, 2nd edition, 2005. `http://www.cs.waikato.ac.nz/ml/weka`.

[10] C. L. Blake and C. J. Merz. UCI repository of machine learning databases, 1998. `http://www.ics.uci.edu/~mlearn/MLRepository.html`.

[11] Y. Freund and R. Schapire. Experiments with a new boosting algorithm. In *13th International Conference om Machine Learning (ICML-96)*, pages 148–156, 1996.

[12] C. Nadeau and Y. Bengio. Inference for the generalization error. *Machine Learning*, 52(239–281), 2003.