Using Quick Decision Tree Algorithm to Find Better RBF Networks*

Hyontai Sug

Division of Computer and Information Engineering, Dongseo University,
Busan, 617-716, Korea
hyontai@yahoo.com

Abstract. It is known that generated knowledge models for data mining tasks are dependent upon supplied data sets, so supplying good data sets for target data mining algorithms is important for the success of data mining. Therefore, in order to find better RBF networks of k-means clustering efficiently, we refer to the number of errors that are from decision trees, and use the information to improve training data sets for RBF networks and we also refer to terminal nodes to initialize the k value. Experiments with real world data sets showed good results.

Keywords: decision tree, radial basis function network, classification, large database.

1 Introduction

In the field of data mining tasks there are two challenges that may hinder success for the tasks. The first challenge is the fact that there can be a lot of data that can cause computational complexity problem, and the second challenge is the fact that data may not be complete for target data mining models so that the trained knowledge model might act poorly for future unseen cases. There are many data mining algorithms to deal with the problem [1], and decision trees and artificial neural networks are some of representative algorithms for the problem. In order to cope with the first problem decision trees can be used, since they are especially good for handling large data sets because of relatively shorter training time. On the other hand, for incomplete or imperfect data problem artificial neural networks can be used, since they are known to be good for the problem with increased computational complexity.

For tasks of data mining artificial neural networks like MLPs and radial basis function (RBF) networks are mostly used because of their good performance in many applications [2, 3, 4]. We are especially interested in RBF networks, because the neural networks have been applied successfully for classification tasks of data mining [5]. RBF networks make approximation based on the training data, and Gaussian functions are used mostly as the radial basis function. In order to train RBF networks first we should find appropriate centre and radius of radial basis function. For this

* This work was supported by Dongseo University, "Dongseo Frontier Project" Research Fund of 2010.

N.T. Nguyen, C.-G. Kim, and A. Janiak (Eds.): ACIIDS 2011, LNAI 6591, pp. 207–217, 2011.

task, we may use some unsupervised learning algorithms like k-means clustering. K-means clustering is one of the mostly used algorithms for clustering [6]. For k-means clustering an appropriate number of clusters has to be given for initialization. For this initialization we usually use domain knowledge to set the number of clusters. So, the task of setting the number of clusters is arbitrary in nature, so is true for the task of initializing the number of clusters of RBF networks. But the RBF networks have different performance depending on the number of clusters and training data sets, so we want to find better RBF networks exploiting decision trees that can be trained more quickly than RBF networks.

In section 2, we provide related work to the research, and in sections 3 we present the procedure. Experiments were run to see the effect of the method in section 4. Finally section 5 presents some conclusions and future work.

2 Related Work

Because it is easy for us to understand the structure of decision trees unless they are not very large, decision trees have been used often as a knowledge model for the task of data mining. Research efforts have been devoted to build better decision trees. Among them C4.5 is some representative decision tree algorithm, because it has been referred often in literature and freely available [7]. C4.5 uses entropy-based measure and generates decision trees in relatively quick time, but the generated tree size is relatively big.

When available data set size is relatively small, artificial neural networks are regarded as good data mining tools [8]. Among many artificial neural networks MLPs and RBF networks are some representative neural networks that have been often referred in literature [9]. A good point of neural networks is robustness to irrelevant features as well as erroneous data. RBF networks are one of the most popular feed-forward networks [10]. Even though RBF networks have three layers including the input layer, hidden layer, and output layer, they differ from MLPs, because the hidden units of RBF networks are constructed based on some clustering algorithms mostly.

There were some efforts to use decision trees to build better RBF networks. Kubat [11] tried to utilize the information of terminal nodes of C4.5 in building RBF networks. The terminal nodes were used as center points for clustering for RBF network. He showed that the RBF networks have better accuracy than decision trees of C4.5 in some data sets. Schwenker et al. also showed that decision trees can be used to initialize three kinds of RBF networks deterministically [12]. But, because the task of generating an optimal decision tree is NP-complete problem, the data space divided by decision tree is one of many possible ways to divide data space, so it is not easy to mention that the RBF networks are optimal.

Because training task of data mining models like neural networks is induction, the behavior of trained data mining models is dependent on the training data set. So, we can infer that the trained knowledge model will be dependent on sample size as well as the composition of data in the samples. Fukunaga and Hayes [13] discussed the effect of sample size for parameter estimates in a family of functions for classifiers. SMOTE method [14] used synthetic data generation method for minor classes, and showed that it is effective for decision trees. In [15] the authors showed that class imbalance has different effect in neural networks for medical domain data. In previous work [16] experiments with smaller sample sizes from original data sets were tried, and it showed good results. So in this paper we want to expand the work with modified method in biased sampling to cope with class imbalance for larger data sets.

3 The Method

Most target data sets for data mining have some skewed distribution in class values, and this fact can be checked easily by inspecting the terminal nodes of decision trees. Moreover, we may also use the information of the number of terminal nodes in the trees as the initial number of clusters for k-means clustering of RBF networks. If the generated tree is very large, the task of interpreting the structure of generated tree is difficult. So, we want to use the information of the number of terminal nodes of decision trees only to find better RBF networks. The method first builds a decision tree using some fast decision tree generation algorithms like C4.5. Then, we inspect the number of misclassified objects for each class. Then we choose classes that should be sampled more for more balanced training set of samples with respect to class value distribution in the samples.

We use the number of terminal nodes in the decision tress to determine the initial number of clusters for the RBF network. But the initial value for the number of clusters in RBF network might not be the best value for the given data set. So we first try to decrease the number of clusters from the initial value in arithmetical progression, then we also increase the number of clusters from the initial value. But increasing or decreasing the number of clusters sequentially and generating corresponding RBF networks may take a lot of computing time without much improvement in accuracy, so we increment or decrement the number as some multiple of the initial number of clusters. If the accuracy values of RBF networks do not increase within given criteria, the search stops. The following is a brief description of the procedure of the method.

```
procedure (Output)
  /* X, K, D,σ: parameters */
 1. Generate a decision tree;
  2. Inspect the terminal nodes of the decision tree to
      determine further sampling for inferior classes
      and count the number of terminal nodes;
 3. Do sampling of X % more for inferior classes
 4. Initialize the number of clusters of RBFN as C where
     C is the largest number that is less than the number
     of terminal nodes and the multiple of the number of
     classes;
  5. Generate a RBFN /* initial_accuracy = the accuracy
     of the network */
  6. loop_better_accuracy := initial_accuracy;
    global_better_accuracy := initial_accuracy;
```

```
/* check decreasingly */
7. Repeat K times
7.1 Generate a RBFN after decreasing the number
   of clusters by D;
7.2 If the accuracy of RBFN > loop_better_accuracy Then
      loop_better_accuracy := the accuracy of RBFN;
   End if;
8. End repeat
9. If loop_better_accuracy > global_better_accuracy
  Then
     K := K - σ;
     global_better_accuracy:=loop_better_accuracy;
     Go to 7;
  End If;
/* check increasingly */
10. loop_better_accuracy := global_better_accuracy;
11. Repeat K times
11.1 Generate a RBFN after increasing the number of
      clusters by D;
11.2 If the accuracy of RBFN > loop_better_accuracy
   Then
      loop_better_accuracy := the accuracy of RBFN;
   End if;
12. End repeat
13. If loop_better_accuracy > global_better_accuracy
Then
     K := K - σ;
     global_better_accuracy:=loop_better_accuracy;
     Go to 11;
  End If;
End.
```

In the above procedure there are four parameters to be defined, X, K, D, and σ. X represents additional percentage to do more sampling. K represents the number of repeats in generating RBF networks while we increase or decrease the number of clusters by D. Depending on the number of terminal nodes of the decision tree, we set the value of D and K appropriately. σ is for the adjustment of K value for the next round of the loop. In the following experiment X is set to 20%, K is set to five, and σ is set to two, and D is set depending on how many classes we have and how many terminal nodes exist in the generated decision tree. One may give smaller value of D, if he wants more through search. Increasing or decreasing the number of clusters will be stopped, when the accuracies of the generated RBF networks are not improved further. We decrease the number of clusters from starting point first, because training time of RBF networks for smaller number of clusters takes less time.

4 Experimentation

Experiments were run using data sets in UCI machine learning repository [17] called ‘adult’ [18] and 'statlog(Landsat satellite)' [19] to see the effect of the method. The number of instances in adult data set is 48,842, and the number of instances in statlog data set is 6,435. The data sets were selected, because they are relatively large, and adult data set may represent business domain and stalog data set may represent scientific domain. The total number of attributes is 14 and 36, and there are two classes and six classes for adult and statlog data set respectively. There are six continuous attributes for adult data set, and all attributes are continuous attributes for stalog data set. We used RBF network using K-means clustering [20] to train for various number of clusters. Because most applications of RBF network use relatively small-sized data sets, we did sampling of relatively small sizes for the experiment to simulate the situation. For adult data set sample size of 960 and 1,920 are used, and for stalog data set sample size of 480 and 960 are used. All the remaining data are used for testing. For each sample size seven random sample data sets were drawn. Used decision tree algorithm is C4.5 and default pruning parameter of 25% is used.

Before we sample in the above mentioned sample sizes, we sampled the sample size of 800 and 1,600 for adult data set, and the sample size of 400 and 800 for statlog data set to determine which class objects should be sampled more. Table 1 and table 2 shows average misclassification ratio of fourteen random sample sets for each class when we generate decision trees of C4.5. All sample data are used to generate the trees.

Table 1. Average misclassification ratio for each class of adult data set samples

Class	Misclassification ratio
>50K	35.5%
≤50K	4.7%

Table 2. Average misclassification ratio for each class of statlog data set samples

Class	Average misclassification ratio
1	1.2%
2	1.9%
3	1.6%
4	14.1%
5	8.1%
6	3.2%

So, 20% more objects were sampled from the object pool of ‘>50K’ class for adult data set, and 10% more objects were sampled for each class 4 and 5 for statlog data set. The following table 3 to 10 show accuracies RBF networks depending on the number of clusters for adult and statlog data set. In the tables, the first row contains results of decision tree C4.5, and in the second row ‘#cls’ means the number of clusters, and ‘Acc.’ means accuracy in percentage. ‘*’ in the body of the table

Table 3. Results of adult data set with sample size of 960

Sample 1: Accuracy: 81.9761 # terminal nodes: 17		Sample 2: Accuracy: 81.8153 # terminal nodes: 72		Sample 3: Accuracy: 84.0897 # terminal nodes: 40		Sample 4: Accuracy: 80.4662 # terminal nodes: 89	
#cls	Acc.	#cls	Acc.	#cls	Acc.	#cls	Acc.
		32	82.2038			24	82.1435
		40	82.3583	8	82.6904	32	81.9699
4	**83.4506**	48	82.2122	16	82.9933	40	81.7903
8	82.4168	56	82.0576	24	**83.0016**	48	**82.4085**
12	81.8989	64	82.2205	32	82.5296	56	81.8341
*16	82.3542	*72	82.4962	*40	82.8262	64	81.5209
20	82.0555	80	**82.776**	48	82.3918	72	81.8049
24	82.6211	88	81.7986	56	82.4001	80	81.6921
28	81.5605	96	81.1115	64	82.3792	*88	81.3392
32	82.1474	104	80.2098	72	81.8884	96	81.1303
36		112	81.2577	80	82.0242	104	81.3266
		120	80.9444			112	81.9761
		128	80.5393			120	82.0346
		136				128	81.9072

Table 4. Results of adult data set with sample size of 960 (cont.)

Sample 5: Accuracy: 83.3693 # terminal nodes: 30		Sample 6: Accuracy: 79.8793 # terminal nodes: 51		Sample 7: Accuracy: 82.7677 # terminal nodes: 67	
#cls	Acc.	#cls	Acc.	#cls	Acc.
		10	82.6319	26	**82.9431**
		18	**83.2209**	34	82.586
6	83.6678	26	83.1019	42	81.5041
14	**83.7033**	34	83.0413	50	81.4958
22	83.6762	42	83.1541	58	81.7151
*30	83.5091	*50	82.9619	*66	80.863
38	83.2564	58	82.4001	74	81.0656
46	83.342	66	82.2644	82	80.4181
54	83.3274	74	81.6169	90	80.0735
62	83.035	82	82.1913	98	80.0798
70		90	81.5376	106	80.0338

indicates the initial number of clusters. The number is based on the number of terminal nodes in the corresponding decision tree of C4.5. The best accuracy value in the experiment for each sample set is represented in bold numbers. Table 3 and 4 show the result of experiment for adult data set with sample size of 960.

If we give attention to sample 3 in table 3, the accuracy of RBF network is worse than the accuracy of decision tree of C4.5. But RBF networks for other sample sizes are better than C4.5, so we can say that better RBF networks can be found because of the repeated trials with different number of clusters. Table 5 and 6 show the result of experiment for adult data set with sample size of 1,920.

Table 5. Results of adult data set with sample size of 1920

Sample 1: Accuracy: 82.2915 # terminal nodes: 68		Sample 2: Accuracy: 82.6453 # terminal nodes: 94		Sample 3: Accuracy: 81.317 # terminal nodes: 102		Sample 4: Accuracy: 82.5387 # terminal nodes: 98	
#cls	Acc.	#cls	Acc.	#cls	Acc.	#cls	Acc.
4	**83.8388**	2	**83.6214**	6	**84.4717**	2	**83.615**
12	83.7791	10	83.4275	18	82.7565	14	82.8478
20	83.4253	22	83.1781	30	83.1087	26	82.8648
28	82.8158	34	82.9607	42	82.925	38	82.7476
36	82.6794	46	83.2378	54	82.4186	50	82.6389
44	81.959	58	82.6645	66	83.4303	62	82.9714
52	81.9526	70	83.3783	78	83.0475	74	82.7732
60	81.9974	82	82.7028	90	82.4962	86	82.5281
*68	81.9313	*94	82.7881	*102	82.7412	*98	82.5941
76	82.2894	106	82.7966	114	82.2511	110	82.6219
84	82.1849	118	82.0933	126	82.5727	122	82.7476
92	82.0123	130	82.2809	138	83.124	134	82.4577
100	81.5307	142	82.1892	150	83.0781	146	82.4663
104	81.1747	154	82.6943	162	82.7259	158	82.1849

Table 6. Results of adult data set with sample size of 1920 (cont.)

Sample 5: Accuracy: 82.2915 # terminal nodes: 68		Sample 6: Accuracy: 82.6453 # terminal nodes: 94		Sample 7: Accuracy: 81.317 # terminal nodes: 102	
#cls	Acc.	#cls	Acc.	#cls	Acc.
6	**83.875**				
14	83.3337	4	**83.777**	28	**82.7854**
22	82.9884	16	82.8542	44	82.6952
30	83.5042	28	82.2318	60	82.4247
38	82.9948	40	82.6751	76	82.8094
46	83.1419	52	81.633	92	82.4307
54	83.3998	64	82.7646	108	81.8898
62	82.9479	76	81.9761	124	82.016
70	82.786	88	81.9079	140	81.3007
*78	83.2783	*100	81.7203	*156	80.8019
86	83.4253	112	81.3666	114	81.1625
94	83.1078	124	82.0699	126	80.8499
102	82.445	136	81.9036	138	80.5854
108	83.0588	148	82.04	150	80.4532
116	82.9224	160	81.9548	162	80.8059

If we look at table 5 and table 6, we can notice that the best accuracies have been found in smaller number of clusters than the number of terminal nodes of C4.5. From the results of the both sample sizes for adult data set, we can find the fact that more accurate RBF networks could be found with relatively small number of clusters.

The following table 7 to table 10 show results of experiment for statlog data set. Table 7 and 8 show the result of experiment for statlog data set with sample size of 480.

Table 7. Results of statlog data set with sample size of 480

Sample 1: Accuracy: 78.623 # terminal nodes: 35		Sample 2: Accuracy: 80.1175 # terminal nodes: 31		Sample 3: Accuracy: 80.084 # terminal nodes: 36		Sample 4: Accuracy: 79.3115 # terminal nodes: 36	
#cls	Acc.	#cls	Acc.	#cls	Acc.	#cls	Acc.
				6	79.3451	6	78.22
6	79.513	6	78.6566	12	84.0134	12	82.0991
12	82.0823	12	83.7615	18	84.5844	18	83.4761
18	83.6272	18	84.534	24	**84.6348**	24	83.6776
24	84.6683	24	**84.9034**	30	83.9798	30	84.0134
*30	83.8455	*30	84.3493	*36	83.6608	*36	83.8959
36	84.2653	36	83.7783	42	82.2334	42	83.2914
42	83.9295	42	84.2989	48	83.0227	48	83.5264
48	**84.6851**	48	83.2317	54	84.0638	54	84.2149
54	84.534	54	80.5542	60	83.5097	60	**84.4668**
60	83.5936	60	80.7389	66	84.0974	66	83.7615
66	83.3249			72	83.9463	72	84.1814
72	83.2914			78	82.4685	78	83.1906
78	82.1998			84	81.7128	84	83.1234

Table 8. Results of statlog data set with sample size of 480 (cont.)

Sample 5: Accuracy: 77.5819 # terminal nodes: 34		Sample 6: Accuracy: 80.1679 # terminal nodes: 36		Sample 7: Accuracy: 78.4587 # terminal nodes: 42	
#cls	Acc.	#cls	Acc.	#cls	Acc.
		6	78.6566	12	82.3707
6	78.2872	12	82.8212	18	82.7233
12	82.1998	18	84.6348	24	83.2774
18	83.2242	24	85.6591	30	82.6058
24	**83.7615**	30	85.3568	36	83.546
*30	83.5936	*36	**86.2972**	*42	83.865
36	83.0563	42	85.7935	48	82.6897
42	82.6868	48	85.8774	54	83.546
48	83.0898	54	84.9874	60	83.9154
54	83.0898	60	84.8027	66	79.7179
60	82.9891	66	84.5844	72	**84.1336**
				78	83.5628
				84	82.9248
				90	81.4641

If we give attention to sample set 2, 4, and 6 of sample size 480, the accuracy of RBF network with six clusters is worse than the accuracy of decision tree of C4.5. But we can avoid choosing this RBF network because of the repeated trials with different number of clusters. Table 9 and 10 show the result of experiment for statlog data set with sample size of 960.

Table 9. Results of statlog data set with sample size of 960

Sample 1: Accuracy: 81.9543 # terminal nodes: 79		Sample 2: Accuracy: 81.1872 # terminal nodes: 63		Sample 3: Accuracy: 81.9316 # terminal nodes: 76		Sample 4: Accuracy: 80.8584 # terminal nodes: 69	
#cls	Acc.	#cls	Acc.	#cls	Acc.	#cls	Acc.
6	79.2511						
18	84.0731	6	79.9269	12	79.5251	6	79.0685
30	84.7854	12	82.9589	24	**85.4064**	18	85.6073
42	85.6986	24	**86.6849**	36	84.2374	30	85.589
54	**86.8128**	36	85.1142	48	84.1096	42	**86.3379**
66	86.3379	48	85.4429	60	84.347	54	85.2968
*78	86.3014	*60	85.5525	*72	84.1826	*66	84.9315
90	85.2237	72	85.7149	84	82.9589	78	84.6027
102	86.6119	84	86.0091	96	83.7626	90	83.5982
114	85.6986	96	86.3927	108	82.6119	102	82.9406
126	86.3014	108	83.3151	120	84.3105	114	79.8904
138	86.5205	120	84.4018	132	83.8539	126	81.4429

Table 10. Results of statlog data set with sample size of 960 (cont.)

Sample 5: Accuracy: 81.4098 # terminal nodes: 71		Sample 6: Accuracy: 80.2557 # terminal nodes: 69		Sample 7: Accuracy: 82.3014 # terminal nodes: 76	
#cls	Acc.	#cls	Acc.	#cls	Acc.
6	79.5654	6	79.0137	12	83.0502
18	84.2586	18	**86.1005**	24	**86.1005**
30	86.176	30	85.3151	36	85.6256
42	85.1717	42	85.1142	48	85.2603
54	85.7378	54	83.9087	60	85.1142
*66	84.7699	*66	85.7717	*72	84.9315
78	**86.6508**	78	85.8265	84	85.8265
90	86.3769	90	85.1142	96	84.895
102	85.9021	102	85.9726	108	84.6758
114	85.4456	114	85.7169	120	84.968
126	82.9255	126	85.5525	132	84.3836
138	83.9664				
150	80.558				
162	80.9898				

If we look at table 9 and table 10, we can notice that the best accuracies have been found in smaller number of clusters than the number of terminal nodes of C4.5. But there is no such regularity for sample size of 480. Anyway, we can find better RBF networks with the algorithm.

5 Conclusions and Future Works

Decision tree algorithms have good property that makes it easy to cope with large-sized data sets, but the good property of decision trees for large-sized data sets can also be harmful in data mining tasks, because we often may not have complete or perfect data sets so that fragmenting the data sets could neglect minor classes, even the size of the data sets are large. Other good point of decision trees is understandability, because the structure of decision tree is represented in symbolic form.

RBF networks make approximation based on training data, and Gaussian functions are used mostly as the radial basis function. In order to train RBF networks, we may use some unsupervised learning algorithms like k-means clustering. Since RBF networks have different performance depending on the number of clusters and training data sets, we want to find better RBF networks based on some objective knowledge models like decision trees, where we can understand the structure of the decision trees easily. Most target data sets for data mining have some skewed distribution in class values, and this fact can be checked easily by inspecting the terminal nodes of decision trees. Moreover, we may also use the information of the number of terminal nodes in the trees as the initial number of clusters for k-means clustering of RBF networks.

The proposed procedure uses the class distribution information and the information of the number of terminal nodes in the generated tree for over-samplng and initialization for the number of clusters of RBF networks. Experiments with two real world data sets in business and scientific domain give us the possibility that we may find better RBF networks effectively.

Because oversampling can generate different class distribution in training data set, we can infer that trained RBF networks may have some different performance compared to original data set. Future work will be some detailed analysis for the effect of oversampling to utilize the RBF networks effectively.

References

1. Tan, P., Steinbach, M., Kumar, V.: Introduction to Data Mining. Addison Wesley, Reading (2006)
2. Bishop, C.M.: Neural networks for pattern recognition. Oxford University Press, Oxford (1995)
3. Heaton, J.: Introduction to Neural Networks for C#, 2nd edn. Heaton Research Inc. (2008)
4. Lippmann, R.P.: An Introduction to Computing with Neural Nets. IEEE ASSP Magazine 3(4), 4–22 (1987)
5. Howlett, R.J., Jain, L.C.: Radial Basis Function Networks I: recent developments in theory and applications. Physica-Verlag, Heidelberg (2001)
6. Russel, S., Novig, P.: Artificial Intelligence: a Modern Approach, 2nd edn. Prentice Hall, Englewood Cliffs (2002)
7. Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, Inc., San Francisco (1993)
8. Larose, D.T.: Data Mining Methods and Models. Wiley-Interscience, Hoboken (2006)
9. Shenouda, E.A.M.A.: A Quantitative Comparison of Different MLP Activation Functions in Classification. In: Wang, J., Yi, Z., Żurada, J.M., Lu, B.-L., Yin, H. (eds.) ISNN 2006. LNCS, vol. 3971, pp. 849–857. Springer, Heidelberg (2006)
10. Orr, M.J.L.: Introduction to Radial Basis Function Networks, `http://www.anc.ed.ac.uk/~mjo/intro.ps`
11. Kubat, M.: Decision Trees Can Initialize Radial-Basis Function Networks. IEEE Transactions on Neural Networks 9(5), 813–821 (1998)

12. Schwenker, F., Kestler, H.A., Palm, G.: Three learning phases for radial-basis-function networks. Neural Networks 14, 439–458 (2001)
13. Fukunaga, K., Hayes, R.R.: Effects of Sample Size in Classifier Design. IEEE Transactions on Pattern Analysis and Machine Intelligence 11(8), 873–885 (1989)
14. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research 16, 341–378 (2002)
15. Mazuro, M.A., Habas, P.A., Zurada, J.M., Lo, J.Y., Baker, J.A., Tourassi, G.D.: Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance. Neural Networks 21(2-3), 427–436 (2008)
16. Sug, H.: An Objective Method to Find Better RBF Networks in Classification. In: Proceedings of the 5th International Conference on Computer Sciences and Convergence Information Technology, vol. 1, pp. 373–376 (2010)
17. Suncion, A., Newman, D.J.: UCI Machine Learning Repository. University of California, School of Information and Computer Sciences, Irvine, CA (2007), `http://www.ics.uci.edu/~mlearn/MLRepository.html`
18. Kohavi, R.: Scaling up the accuracy of Naive-Bayes classifiers: a decision-tree hybrid. In: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, pp. 202–207 (1996)
19. Statlog (Landsat Satellite) Data Set, `http://archive.ics.uci.edu/ml/datasets/Statlog+%28Landsat+Satellite%29`
20. Witten, I.H., Frank, E.: Data Mining, 2nd edn. Morgan Kaufmann, San Francisco (2005)