Investigating Better Multi-layer Perceptrons for the Task of Classification

HYONTAI SUG
Division of Computer and Information Engineering
Dongseo University
Busan, 617-716
REPUBLIC OF KOREA
hyontai@yahoo.com http://kowon.dongseo.ac.kr/~sht

Abstract: - The task of deciding proper sample sizes for multi-layer perceptrons tends to be arbitrary so that, depending on sample data sets, the performance of trained multi-layer perceptrons has a tendency of some fluctuation. As sample size grows, multi-layer perceptrons have the property that performance in prediction accuracy becomes better slowly with some fluctuation. In order to exploit this property this paper suggests a progressive and repeated sampling technique for better multi-layer perceptrons to cope with the fluctuation of prediction accuracy that depend on samples as well as the size of samples. Experiments with six different data sets in UCI machine learning repository showed very good results.

Key-Words: - Multi-layer perceptrons, neural networks, data mining, classification

1 Introduction

Artificial neural networks are very often used for forecasting tasks of data mining like the tasks of classification and numerical prediction. Artificial neural networks are mostly favored, because their performance with small number of available data instances in the forecasting task is relatively good compared to other data mining or machine learning techniques. Therefore, finding neural networks with good accuracy for a given data set has been a major concern [1].

Many kinds of successful artificial neural network algorithms have been applied to a variety of tasks. For example, multi-layer perceptrons (MLPs) are used for various prediction tasks, Hopfield networks are used for associative memory and optimization problems, ART networks are used for autonomous learning systems, boltzman machines are used for optimization problems, etc.[2]. Among them we are interested in MLPs, because they are one of the mostly used artificial neural networks for prediction [3, 4, 5, 6]. But even though MLPs are one of the most successful data mining or machine learning methodologies, there are some points of improvement due to the fact that they are built based on greedy search method like backpropagation algorithms, and the structure of the neural networks is usually determined by the knowledge of experts [7, 8, 9].

Training the connection weights of MLPs use backpropagation algorithms that rely on some greedy search algorithms like gradient decent. Even though the gradient descent works well in most cases, there is still some possibility of considering local optima as global optima [10]. Moreover, because the backpropagation algorithms take a lot of computing time, small sized training data sets are preferred. Because most target databases for data mining are very large, most people rely on random sampling to the target databases to determine small sized training data sets. But the task of determining proper sample sizes is arbitrary and the found knowledge based on the random samples is prone to sampling errors so that the accuracy of the MLPs has the tendency of some fluctuation.

But, because the accuracy of the trained MLPs has the tendency of increase in some logarithmic way as the sample size grows, we want to exploit the property of MLPs by adapting larger and larger samples on the condition that other factors for further optimization are fixed.

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

2 Related Work

Artificial neural networks have drawn many researchers' attention for the task of machine learning since the pioneering neural network algorithm, the perceptron [11]. Because of the limited predictability of the perceptron, multi-layer perceptrons have been invented. A good point of MLPs is that they can be converged well. Moreover, because the weights that link nodes in the neural networks are adjusted slowly, MLPs are known to be robust against irrelevant features and to be able to tolerate erroneous data well [12, 13, 14].

There are many examples to exploit the good points of MLPs. For example, El-Fegh [15] used MLP to classify handwritten Arabic words. Sala-Burgos and Gil-Pita [16] used MLPs to detect microfossils from high resolution images of sediment.

The behavior of trained or inducted models also dependent on the training data sets. For example, Zuters [17] tried to get better set of training instances that are in uniform distribution. Randomly generated additional instances are tested with MLP to get better set of training instances. So, there is research on sample size as well as the property of samples. Chen [18] used random sampling based on ancient numeric concept called Lo-Shu to achieve almost three times faster training time for MLPs. Fukunaga and Hayes [19] discussed the effect of sample size for parameter estimates in a family of functions for classifiers. In [20] the authors showed that class imbalance in training data has effects in neural network development especially for medical domain. In [21] sampling techniques for relatively small sized data sets like cross-validation, the leave-one-out, etc. are investigated to see the effect of the sampling techniques in the performance of neural networks, and discovered that the sampling techniques has different results in accuracy depending on feature space and sample size. In [22] three sampling schemes, arithmetic, geometric, and dynamic sampling are investigated for decision tree algorithms. In arithmetic sampling and geometric sampling, the sample size grows in arithmetic and geometric manner respectively. Dynamic sampling method determines the sample size based on dynamic programming. The authors found that the accuracy of decision tree classifier increases as the sample size increases and the curve of accuracy is logarithmic, so they used the rate of increase in accuracy as stopping criteria for sampling. Domingo [23, 24] discussed the quantity of test data and showed that measuring the performance of underlying knowledge models based on relatively small testing data only is not enough, because the size of feature space is usually far greater than available data set.

3 Suggested Method

Because we have only limited number of data in the data set and the data set should be divided into two parts, training and testing, it is not easy to determine an appropriate sample size that is the best for the target data set. In order to overcome this problem we resort to repeated sampling scheme that considers various sizes of samples to find the best one among the samples.

We repeat sampling within a sample size several times, and MLPs are generated for each sample data set. Among them we call the best accuracy value 'better accuracy', and the worst accuracy value 'worse accuracy'. We do the sampling until the accuracy of trained neural networks has reached to a plateau on the condition that the sample size is less than or almost the half of the target data set, because we want to have enough test data also.

The following is a brief description of the procedure of the method.

- Input: a target data set,
  S_0: the initial sample size
- Output: A

  /* the array of accuracy values of trained MLPs */

  i := 0;

  Do while $S_i \approx$ half of the target data set

  For j := 1 to n **do**

  /* n: the number of repeat for each sample size. For our experiments in the following section n = 2 or 4 */

  Do random sampling of size s

  Train and test a MLP;

  Aij = the accuracy of the trained MLP; /* the best one among n is called better accuracy */

  End for;

  i++;

  If S_{i-1} = predefined_limit **Then**

  $S_i := S_{i-1}$ + predefined_increment;

  ELSE

  $S_i := S_0 \times 2^i$;

  End if

  Stop if the better accuracy has reached a plateau;

  End do while;

In the above procedure depending on the available data and the property of data set, we double the sample size or increase the sample size by some predefined increment, until the sample size reaches to about half of the data set size. In addition, we can stop the while loop, if the accuracy improvement in better accuracy has reached a plateau.

If we do random sampling several times within the given sample size, the accuracy values of the trained MLPs can be slightly different, because each random sampling generate different data set. This is the reason why we need the concept of 'better accuracy'. Using the better accuracy as a stopping criterion is a good criterion

for stopping, because we prefer better prediction models in practice.

In order to confirm a plateau we do additional sampling for the last two sample sizes. The total number of sample data sets for each sample size in the last two sample sizes is seven for the experiments in the following section.

In the experiments of following section we did random sampling twice for each sample size when available data set size is large. On the other hand, we did random sampling four times for each sample size when available data set size is relatively medium or small. In order to train MLPs the given number of hidden layers is the half of the number of attributes plus the number of classes, and the training time is 500. We used a desktop computer of pentium 4 processor with 2 GB main memory that has relatively weak computing capability. Depending on the available computing resources, one may do random sampling more within a given sample size.

4 Experiments

Six data sets in UCI machine learning repository [25] called 'forest cover types', 'adult', 'statlog', 'yeast', 'letter recognition', and 'ozone level detection' were used to see the effect of the method.

4.1 Experiments for 'forest cover types' data set

The forest cover types data set [26] includes forest information in four wilderness areas found in the Roosevelt National Forest of northern Colorado. It has twelve numerical attributes as conditional attributes, while seven major forest cover types were used as a class attribute. The number of instances in the forest cover types data set is 581,012.

Table 1 shows the result of training of MLPs for forest cover types data set. The initial sample size for training is 200, and two random sample sets are drawn for each sample size. The size of samples is doubled as the while loop runs, and we stop sampling when the accuracy reaches a plateau. The rest of the data set after sampling is used for testing, so we have bigger test set data when sample size is small. The values in the table are arranged to have the results of the better accuracy values first.

Table 1. The accuracy of MLPs for forest cover types data set

Sample size	Better accuracy (%)	Worse accuracy (%)
200	61.6938	58.5325
400	64.7122	62.6313
800	66.9400	65.8595
1,600	69.8988	67.5559
3,200	71.5321	69.9281
6,400	73.8785	72.4115
12,800	75.9382	75.4211
25,600	76.8910	76.8689
51,200	78.6211	77.2580
102,400	79.3894	78.4677
204,800	79.7059	79.3204

If we look at table 1, we can notice the fact that when sample size becomes larger and larger, the accuracy values of the MLPs become better and better, and the tendency of accuracy in better accuracy resembles a plateau as the sample size becomes bigger. Fig. 1 displays the trend of prediction accuracy of the MLPs for the forest cover types data set more clearly as the training data set size grows. Dotted line is the worse accuracy and solid line is better accuracy within the same sample size. In the figure axis X represents the sample size and axis Y represents prediction accuracy. As we can see in the graph, the better accuracy reaches a plauteau at the sample size of 204,800.

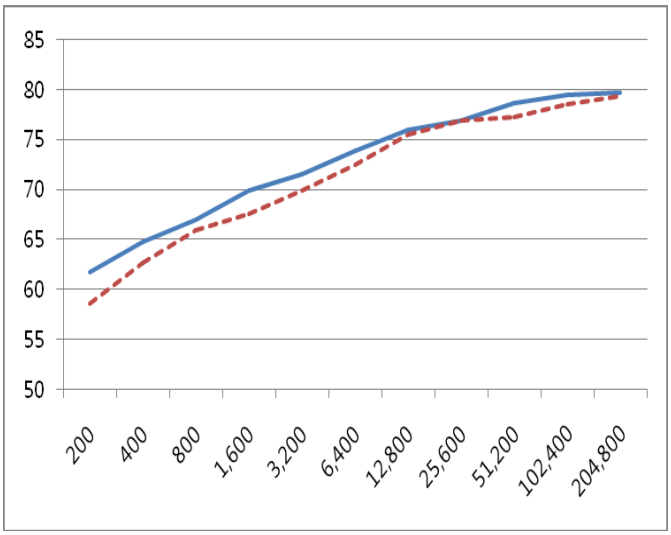

Fig. 1. The accuracy of MLPs for forest cover types data set

Five additional sampling for sample size 102,400 and 204,800 were done to make it sure that it has reached a plateau. Table 2 summerizes the result of the experiment. In the table the better accuracy is numbered sample number 1 and the worse accuracy is numbered sample number 2 for convenience. The difference of average accuracy between the two sample sizes is only 0.456386%, so this value proves that it has reached a plateau.

Table 2. The accuracy of MLPs for sample size 102,400 and 204,800 for forest cover types data set.

Sample number	Accuracy(%): Sample size 102,400	Accuracy(%): Sample size 204,800
1	79.3894	79.7059
2	78.4677	79.3204
3	78.5900	78.9903
4	78.0465	79.2781
5	79.3286	78.5790
6	78.6301	78.4595
7	78.8102	80.1240
Average:	78.75179	79.20817

4.2 Experiments for 'adult' data set

The adult data set [27] is a refined version of 'census income' data set. The census income data set is census in 1994. The census income data set is originated from the census bureau database. The number of instances in the adult data set is 48,842. The total number of attributes in the adult data set is fourteen, and among them six attributes are numerical attributes and one attribute is a class attribute where it has two classes, yearly income being greater than or equal to 50,000 and less than 50,000.

Let's see the result of experiment for adult data set. Table 3 shows the result. The initial sample size for training is 200, and two random sample sets are drawn for each sample size. The size of samples is doubled as the while loop runs, and we stop sampling when the sample size reaches about half of the data set. The rest of the data set after sampling is used for testing, so we have bigger test set data when sample size is small. The values in the table are arranged to have the results of the better accuracy values first.

Table 3. The accuracy of MLPs for adult data set.

Sample size	Better accuracy (%)	Worse accuracy (%)
200	80.7060	78.2143
400	79.9472	78.0149
800	80.3089	79.7927
1,600	81.9292	80.4708
3,200	82.6476	82.1677
6,400	83.1299	80.7690
12,800	83.5704	83.2236
25,600	83.6078	81.5505

If we look at table 3, we can notice the fact that when sample size becomes larger and larger, the accuracy values of the MLPs become better and better with some fluctuation of the accuracy values, and the tendency of accuracy in better accuracy resembles a plateau as the sample size becomes bigger. Fig. 2 displays the trend of prediction accuracy of the MLPs for the adult data set more clearly as the training data set size grows. Dotted line is the worse accuracy and solid line is better accuracy with the same sample size. In the figure axis X represents the sample size and axis Y represents prediction accuracy. As we can see in the graph, the better accuracy reaches almost a plateau also at the samples size of 25,600.

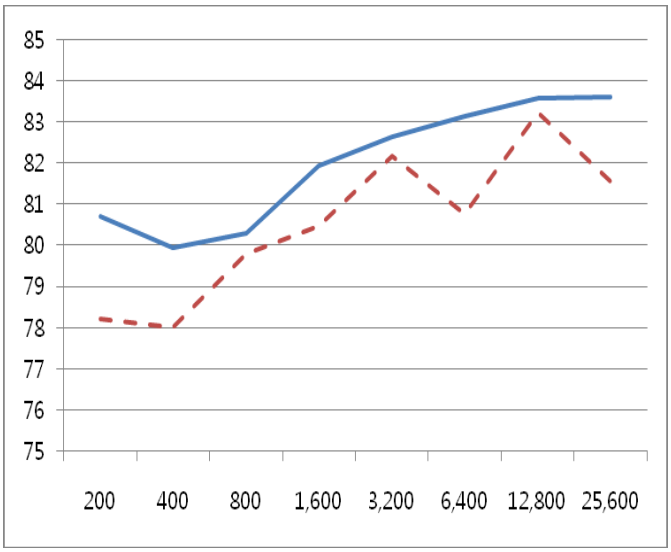

Fig. 2. The accuracy of MLPs for adult data set

Five additional sampling for sample size 12,800 and 25,600 were done to confirm that it has reached a plateau. Table 4 summerizes the result. In the table the better accuracy is numbered sample number 1 and the worse accuracy is numbered sample number 2 for convenience. So, the difference of average accuracy of the MLPs between the two sample size is only 0.011929%, and this value proves that it has reached a plateau.

Table 4. The accuracy of MLPs for sample size 12,800 and 25,600 for adult data set.

Sample number	Accuracy(%): Sample size 12,800	Accuracy(%): Sample size 25,600
1	83.5704	83.6078
2	83.2236	81.5505
3	82.9035	84.0796
4	82.9409	82.1388
5	81.6789	83.0087

6	81.7775	83.6420
7	82.9465	81.0974
Average:	82.72019	82.73211

4.3 Experiments for 'statlog' data set

The statlog data set [28] contains data of landsat satellite data of images. The data set was generated taking a small section from the original data of satellite data in binary form. The binary values were converted to numbers and there are 36 numerical attributes to represent the images. There are seven class lables like red soil, cotton crop, grey soil, damp grey soil, soil with vegetation stubbel, very deep grey soil, and a mixture class. But in the data set there is no mixture class, so there is only six class values in the data set. The total number of instances is 6,435.

Let's see the result of experiment for statlog data set. Table 5 shows the result. The initial sample size for training is 100, and four random sample sets are drawn for each sample size, since the data set size is relatively medium. The given predefined_limit and predefined_increment is 1,600 and 800 respectively. The size of samples is doubled until the sample size reaches the sample size of predefined_limit, and incremented by the predefined_increment after the predefined_limit. We stop sampling when the sample size reaches about half of the data set. The rest of the data set after sampling is used for testing, so we have bigger test set data when sample size is small. The values in the tables are arranged to have the results of the better accuracy values first.

Table 5. The accuracy of MLPs for statlog data set.

Sample size	Better accuracy (%)	Worse accuracy (%)
100	82.6835	79.3212
200	83.3841	81.0746
400	84.5236	83.5128
800	86.9210	85.8030
1,600	87.9628	86.4971
2,400	91.2785	87.5589
3,200	89.5518	88.1026

If we look at table 5, we can notice the fact that when sample size becomes larger and larger, the accuracy values of the MLPs become better and better with some fluctuation of the accuracy values, and the tendency of accuracy in better accuracy resembles a plateau as the sample size becomes bigger. Fig. 3 displays the trend of prediction accuracy of the MLPs for the data set more clearly as the training data set size grows. Dotted line is the worse accuracy and solid line is better accuracy with the same sample size. In the figure axis X represents the sample size and axis Y represents prediction accuracy. As we can see in the graph, the better accuracy reaches almost a plateau also at the samples size of 2,400.

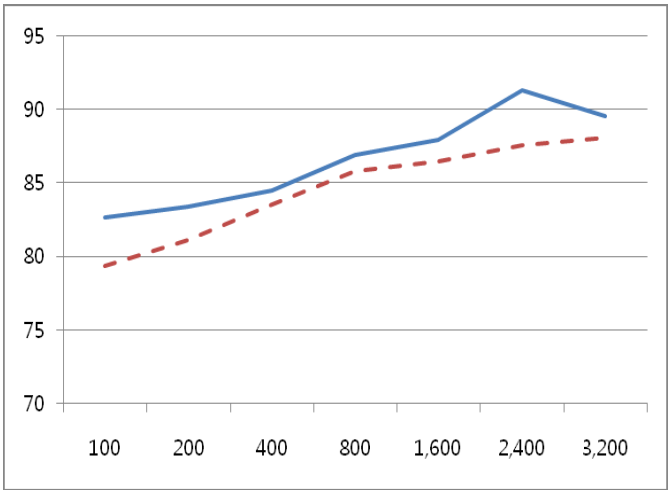

Fig. 3. The accuracy of MLPs for statlog data set

Three additional sampling for sample size 2,400 and 3,200 were done to confirm that it has reached a plateau. Table 6 summerizes the result. In the table the better accuracy is numbered sample number 1 and the worse accuracy is numbered sample number 2 for convenience. So, the difference of average accuracy between the two sample size is only 0.418071%, and this value proves that it has reached a plateau.

Table 6. The accuracy of MLPs for sample size 2,400 and 3,200 for statlog data set.

Sample Number	Accuracy(%): Sample size 2,400	Accuracy(%): Sample size 3,200
1	91.2785	89.5518
2	87.5589	88.1026
3	87.7819	89.4623
4	88.3796	88.2226
5	87.8810	87.8245
6	88.5035	89.2769
7	87.5589	89.4281
Average:	88.42033	88.83840

4.4 Experiments for 'yeast' data set

The yeast data set [29, 30, 31] is used to predict the cellular location sites of protein. There are eight numeric attributes and one nonnumeric attribute as a class

attribute having ten class values. The class attribute indicates the localization site of protein. The total number of instances is 1,484, and there are no missing values.

Let's see the result of experiment for yeast data set. Table 7 shows the result. The initial sample size for training is 100, and four random sample sets are drawn for each sample size, since the data set size is relatively small. The given predefined_limit and predefined_increment is 400 and 200 respectively. The size of samples is doubled until the sample size reaches the predefined_limit, and incremented by the predefined_increment after the predefined_limit. We stop sampling when the sample size reaches about half of the data set. The rest of the data set after sampling is used for testing, so we have bigger test set data when sample size is small. The values in the table are arranged to have the results of the better accuracy values first.

Table 7. The accuracy of MLPs for yeast data set.

Sample Size	Better accuracy (%)	Worse accuracy (%)
100	52.3121	47.4711
200	53.8941	52.3364
400	56.2731	51.6129
600	61.8788	54.7511
800	59.3567	56.7251

If we look at table 7, we can notice the fact that when sample size becomes larger and larger, the accuracy values of the MLPs become better and better with some fluctuation of the accuracy values, and the tendency of accuracy in better accuracy resembles a plateau as the sample size becomes bigger. Fig. 4 displays the trend of prediction accuracy of the MLPs for the data set more clearly as the training data set size grows. Dotted line is the worse accuracy and solid line is better accuracy with the same sample size. In the figure axis X represents the sample size and axis Y represents prediction accuracy. As we can see in the graph, the better accuracy reaches almost a plateau also at the samples size of 600.

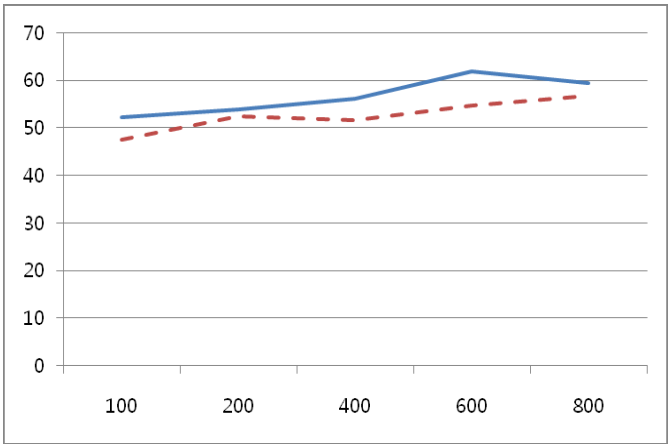

Fig. 4. The accuracy of MLPs for yeast data set

Three additional sampling for sample size 600 and 800 were done to confirm that it has reached a plateau. Table 8 summerizes the result. In the table the better accuracy is numbered sample number 1 and the worse accuracy is numbered sample number 2 for convenience. So, the difference of average accuracy between the two sample size is 0.579229%, and this value proves that it has reached a plateau.

Table 8. The accuracy of MLPs for sample size 600 and 800 for yeast data set.

Sample number	Accuracy(%): Sample size 600	Accuracy(%): Sample size 800
1	61.8778	59.3567
2	54.7511	56.7251
3	59.1629	57.1637
4	57.0136	59.2701
5	55.4299	57.8102
6	55.8824	55.9942
7	56.3348	58.1871
Average:	57.2075	57.78673

4.5 Experiments for 'letter recognition' data set

The letter recognition data set [32] is a data set having character images. The character images were come from black-and-white rectangular pixel displays. There is one class attribute having one of 26 capital letters in English alphabet. The number of attributes is 16 having numerical values, and the total number of instances is 20,000, and there are no missing values.

Let's see the result of experiment for letter data set. Table 9 shows the result. The initial sample size for training is 100, and two random sample sets are drawn for each sample size, because the size of the data set is

relatively large. The given predefined_limit and predefined_increment is 6,400 and 3,200 respectively. The size of samples is doubled until the sample size reaches the predefined_limit, and incremented by the predefined_increment after the predefined_limit. We stop sampling when the sample size reaches about half of the data set. The rest of the data set after sampling is used for testing, so we have bigger test set data when sample size is small. The values in the table are arranged to have the results of the better accuracy values first.

Table 9. The accuracy of MLPs for letter recognition data set.

Sample Size	Better accuracy (%)	Worse accuracy (%)
100	51.3367	41.6281
200	62.8889	55.5303
400	69.0765	68.2449
800	74.7383	73.2031
1,600	77.2023	75.8002
3,200	78.4416	77.4630
6,400	81.0991	80.7678
9,600	82.1193	80.5358
12,800	82.4745	81.5217

If we look at table 9, we can notice the fact that when sample size becomes larger and larger, the accuracy values of MLPs become better and better with some fluctuation of the accuracy values, and the tendency of accuracy in better accuracy resembles a plateau as the sample size becomes bigger. Fig. 5 displays the trend of prediction accuracy of MLPs for the data set more clearly as the training data set size grows. Dotted line is the worse accuracy and solid line is better accuracy with the same sample size. In the figure axis X represents the sample size and axis Y represents prediction accuracy. As we can see in the graph, the better accuracy reaches almost a plateau also at the samples size of 12,800.

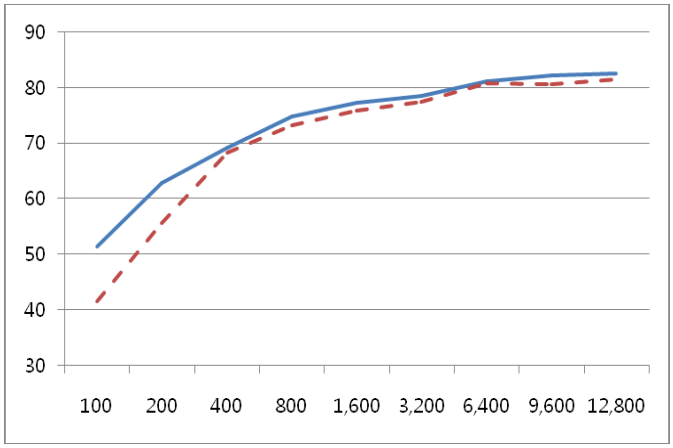

Fig. 5. The accuracy of MLPs for letter recognition data set

Five additional sampling for sample size 9,600 and 12,800 were done to confirm that it has reached a plateau. Table 10 summerizes the result. In the table the better accuracy is numbered sample number 1 and the worse accuracy is numbered sample number 2 for convenience. So, the difference of average accuracy between the two sample size is 0.58944%, and this value proves that it has reached a plateau.

Table 10. The accuracy of MLPs for sample size 9,600 and 12,800 for letter recognition data set.

Sample number	Accuracy(%): Sample size 9,600	Accuracy(%): Sample size 12,800
1	82.1193	82.4745
2	80.5358	81.5217
3	82.1415	81.2224
4	80.5913	81.2217
5	80.3534	82.0030
6	79.9873	81.3925
7	80.8598	80.8787
Average:	80.9412	81.53064

4.6 Experiments for 'ozone level detection' data set

The ozone level detection data set [33] has two ground ozone level data. One is eight hour peak data set, and the other is one hour peak data set. In the experiment eight hour peak data set was used. The ozone level data were collected at Houston, Galveston, and Brazoria area from 1998 to 2004. There are 73 attributes having numerical values, and there is a class attribute having two class values of ozone day and normal day. The total number of instances is 2,536.

Let's see the result of experiment for adult data set. Table 11 shows the result. The initial sample size for training is 100, and four random sample sets are drawn for each sample size, since the data set size is relatively small. The given predefined_limit and predefined_increment is 800 and 400 respectively. The size of samples is doubled until the sample size reaches the predefined_limit, and incremented by the predefined_increment after the predefined_limit. We stop sampling when the sample size reaches about half of the data set. The rest of the data set after sampling is used for testing, so we have bigger test set data when sample size is small. The values in the table are arranged to have the results of the better accuracy values first.

Table 11. The accuracy of MLPs for ozone level detection data set.

Sample size	Better accuracy (%)	Worse accuracy (%)
100	93.2210	88.9984
200	92.3736	91.7309
400	92.4405	91.0280
800	93.3779	92.2732
1,200	93.8172	92.7781
1,600	94.2835	92.3148

If we look at table 11, we can notice the fact that when sample size becomes larger and larger, the accuracy values of the MLPs become better and better with some fluctuation of the accuracy values, and the tendency of accuracy in better accuracy resembles a plateau as the sample size becomes bigger. Fig. 6 displays the trend of prediction accuracy of the MLPs for the data set more clearly as the training data set size grows. Dotted line is the worse accuracy and solid line is better accuracy with the same sample size. In the figure axis X represents the sample size and axis Y represents prediction accuracy. As we can see in the graph, the better accuracy reaches almost a plateau also at the samples size of 1,600.

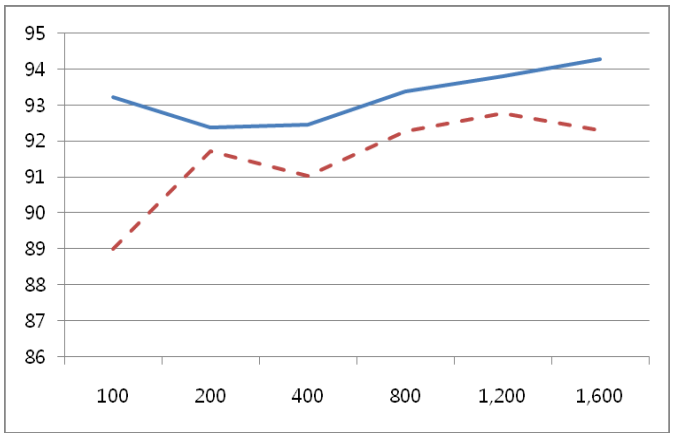

Fig. 6. The accuracy of MLPs for ozone level detection data set

Three additional sampling for sample size 1,200 and 1,600 were done to confirm that it has reached a plateau. Table 12 summerizes the result. In the table the better accuracy is numbered sample number 1 and the worse accuracy is numbered sample number 2 for convenience. So, the difference of average accuracy between the two sample size is only 0.458129%, and this value proves that it has reached a plateau.

Table 12. The accuracy of MLPs for sample size 1,200 and 1,600 for ozone level detection data set.

Sample number	Accuracy(%): Sample size 1,200	Accuracy(%): Sample size 1,600
1	93.8172	94.2835
2	92.7781	92.3148
3	92.9102	93.0016
4	92.8571	93.0872
5	94.2671	94.3503
6	92.2052	92.4731
7	92.4174	94.9487
Average:	93.03604	93.49417

5 Conclusion

For the task of data mining there are many artificial neural networks that are widely used. Among them multi-layer perceptrons(MLPs) are widely accepted for classification tasks because of their good performance and their property of convergence, even for the existence of irrelevant features and erroneous data. Robustness to irrelevant features and erroneous data is especially important in data mining field, because the target data sets of data mining often contain such characteristics. But, whatever artificial neural networks are used, the neural networks may not always be the best predictors due to the fact that they are trained based on some greedy algorithms with limited data sets and the structures are built based on the experience of human experts. So, some improvements may be possible.

Because the target data sets in data mining tasks contain a lot of data, in order to train MLPs random sampling has been considered a standard method to cope with large data sets that are very common in data mining tasks. But, simple random sampling might not generate perfect samples that are best for the used MLPs as well as for the available data sets. Moreover, the task of determining a proper sample size is arbitrary so that the reliability of the trained MLPs might not be good enough to be trusted due to the fluctuation of accuracy values of the trained MLPs.

In order to cope with the problem a progressive, and repeated sampling method within a sample size, which considers various sample sizes incrementally, is proposed to decide the best random samples for multi-layer perceptrons. Experiments with six real world data sets in various domain showed very good results.

References:

[1] D.T. Larose, *Data Mining Methods and Models*, Wiley-Interscience, 2006.
[2] C.M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University press, 1995.
[3] P.F. Marteau, V. Monbet, Conditional Prediction of Markov Processes Using Parametric Viterbi Algorithm – Comparison with MLP and GRNN Models, *WSEAS Transaction on Systems*, issue 2, vol. 3, 2004, pp. 346-351.
[4] R.L.S. Alves, A.C.M. Albuquerque, J.D. Melo, A.D. Dória Neto, Competitive Strategies for Multilayer Perceptrons' Training using Backpropagation and Parallel Processing, *WSEAS Transaction on Systems*, issue 2, vol. 3, 2004, pp. 352-357.
[5] G. Schlotthauer, M.E. Torres, Automatic Diagnosis of Pathological Voices, *Proceeding of the 6th WSEAS International Conference on Signal, Speech and Image Processing*, Lisbon, Portugal, September 22-24, 2006, pp. 150-155.
[6] S. Meghriche, M. Boulemden, A. Draa, Agreement Between Multi-Layer Perceptron and a Compound Neural Network on ECG Diagnosus of Aatrioventricular Blocks, *WSEAS Transactions on Biology and Biomedicine*, issue 1, vol. 5, 2008, pp.12-22.
[7] P. Tan, M. Steinbach, V. Kumar, *Introduction to Data Mining*, Addison Wesley, 2006.
[8] K.Z. Mao, K.C. Tan, W. Ser, Probabilistic Neural Network Structure Determination for Pattern Classification, *IEEE Transactions on Neural Networks*, vol. 11, issue 4, 2000, pp. 1009-1016.
[9] X. Yao, Evolving Artificial Neural Networks, *Proceedings of the IEEE*, vol. 87, no. 9, 1999, pp. 1423-1447.
[10] S. Russel, P. Novig, *Artificial Intelligence: a Modern Approach*, 2nd ed., Prentice Hall, 2002.
[11] M.L. Minsky, S.A. Papert, *Perceptrons – extended edition: an Introduction to Computational Geometry*, MIT press, 1987.
[12] J. Heaton, *Introduction to Neural Networks for C#*, 2nd ed., Heaton Research Inc., 2008.
[13] R.P. Lippmann, An Introduction to Computing with Neural Nets, *IEEE ASSP Magazine*, vol. 3, no. 4, 1987, pp. 4-22.
[14] M. Qu, F.Y. Shih, J. Jing, H. Wang, Automatic solar flare detection using MLP, RBF, and SVM, *Solar Physics*, vol. 27, no. 1, 2003, pp. 157-172.
[15] I. El-Fegh, Z.S. Zubi, A.A. Elrowayati, F.A. El-Mouadib, Handwritten Arabic Words Recognition using Multi Layer Perceptron and Zernik Moments, *Proceedings of the 10th WSEAS International Conference on Evolutionary Computing*, Prague, Czech Republic, 2009, pp. 46-51.
[16] N. Sara-Burgos, R. Gil-Pita, Automatic Microfossil Detection in Somosaguas Sur Paleontologic Site (Pozuelo de Alarcón, Madrid, Spain) using Multilayer Perceptrons, *WSEAS Transactions on Signal Processing*, issue 2, vol. 2, 2006, pp. 218-223.
[17] J. Zuters, Towards Multi-Layer Perceptron as an Evaluator Through Randomly Generated Training Patterns, *Proceedings of the 5th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases*, Madrid, Spain, 2006, pp. 15-17.
[18] H.H. Chen, Fast Training MLP Networks with Lo-Shu Data Sampling, *Proceedings of the 8th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases*, Cambridge, UK, 2009, pp. 165-169.
[19] K. Fukunaga, R.R. Hayes, Effects of Sample Size in Classifier Design, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, issue 8, 1989, pp. 873-885.
[20] M.A. Mazuro, P.A. Habas, J.M. Zurada, J.Y. Lo, J.A. Baker, G.D. Tourassi, Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance, *Neural Networks*, vol. 21, issues 2-3, 2008, pp. 427-436.
[21] S. Berkman, H. Chan, L. Hadjiiski, Classifier performance estimation under the constraint of a finite sample size: Resampling scheme applied to neural network classifiers, *Neural Networks*, vol. 21, issues 2-3, 2008, pp. 476 -483.
[22] F. Provost, T. Oates, D. Jensen, Efficient progressive sampling, *Proceedings of the Fifth International Conference on Knowledge Discovery and Data Mining*, 1999, pp. 23-32.
[23] P. Domingos, Occam's Two Razors: The Sharp and The Blunt, *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*, AAAI Press, 1998, pp. 37-43.
[24] P. Domingos, The Role of Occam's Razor in Knowledge Discovery, *Data mining and Knowledge Discovery*, vol.3, 1999, pp.409-425.
[25] A. Suncion, D.J. Newman, *UCI KDD Archive* [http://kdd.ics.uci.edu], Irvine, CA: University of California, Department of Information and Computer Science, 2007.
[26] J.A. Blackard, J.D. Denis, Comparative Accuracies of Artificial Neural Networks and Discriminant Analysis in Predicting Forest Cover Types from Cartographic Variables, *Computers and Electronics in Agriculture*, vol. 24, no. 3, 2000, pp. 131-151.
[27] R. Kohavi, Scaling up the accuracy of Naive-Bayes classifiers: a decision-tree hybrid, *Proceedings of the*

Second International Conference on Knowledge Discovery and Data Mining, 1996, pp. 202-207.

[28] Statlog (Landsat Satellite) Data Set, http://archive.ics.uci.edu/ml/datasets/Statlog+%28Landsat+Satellite%29

[29] P. Horton, K. Nakai, A Probablistic Classification System for Predicting the cellular Localization Sites of Proteins, *Intelligent Systems in Molecular Biology*, St. Louis, USA, 1996, pp. 109-115.

[30] K. Nakai, M. Kanehisa, Expert System for predicting Protein Localization Sites in Gram-Neggative Bacteria, *PROTEINS: Structure, Function, and Genetics*, vol. 11, 1991, pp. 95-110.

[31] K. Nakai, M. Kanehisa, A Knowledge Base for Predicting Protein Localization Sites in Eukaryotic cells, *Genomics*, vol. 14, 1992, pp. 897-911.

[32] P.W. Frey, D.J. Slate, Letter Recognition Using Holland-style Adaptive Classifiers, *Machine Learning*, vol. 6, no. 2, 1991, pp. 161-182.

[33] K. Zhang, W. Fan, Forecasting Skewed Biased Stochastic Ozone Days: Analyses, Solutions and Beyond, *Knowledge and Information Systems*, vol. 14, no. 3, 2008, pp. 299-326.