A Hybrid Approach Support Vector Machine (SVM) – Neuro Fuzzy For Fast Data Classification

Elsen Ronando[1], M. Isa Irawan[1], and Erna Apriliani[1]

Abstract-In recent decade, Support Vector Machine (SVM) was a machine learning method that widely used in several application domains. It was due to SVM has a good performance for solving data classification problems, particularly in non-linear case. Nevertheless, several studies indicated that SVM still has some inadequacies, especially the high time complexity in testing phase that is caused by increasing the number of support vector for high dimensional data. To address this problem, we propose a hybrid approach SVM – Neuro Fuzzy (SVMNF), which neuro fuzzy here is used to avoid influence of support vector in testing phase of SVM. Moreover, our approach is also equipped with a feature selection that can reduce data attributes in testing phase, so that it can improve the effectiveness of time computation. Based on our evaluation in real benchmark datasets, our approach outperformed SVM in testing phase for solving data classification problems without significantly affecting the accuracy of SVM.

Index Terms – Support Vector Machine (SVM), Neuro Fuzzy, Classification, Computation Time.

INTRODUCTION

Machine learning was one of many areas in artificial intelligence that widely applied for decision-making process. In general, the aim of machine learning technique was to analyze the data for making the model and knowledge that can be used to predict the future behavior of the data. Nowadays, most popular of machine learning technique is Support Vector Machine (SVM) method, where this method has a high accuracy [6] and widely applied in many application, such as text categorization, digit recognition, time series prediction, financial forecasting, pattern selection, and voice recognition.

The principle of Support Vector Machine (SVM) is searching optimal margin hyper plane to divide some different classes. It means that the dataset are classified according to its class based on the best hyper plane which built through the training phase. Although the results of SVM showed an accuracy higher than other methods in testing phase, such as neural networks. Yet, some studies literature indicated that SVM still has some problems, especially highly computational time of the testing phase [6]. It was due to SVM in the testing phase is strongly influenced by the increasing number of support vector if the data has huge dimensions.

Based on these problems, we proposed a hybrid approach SVM – Neuro Fuzzy (SVMNF), which Neuro Fuzzy that applied here has been integrated with feature selection. Thus, it can significantly reduce the data dimension for decreasing computation time in testing phase. The methods and the results of our evaluation will be explained more details in the next section.

METHODS

In this section, we describe our approach to improve the efficiency of the computing time in the testing phase of SVM. In our concept, the framework of our approach can be illustrated in Figure 1.

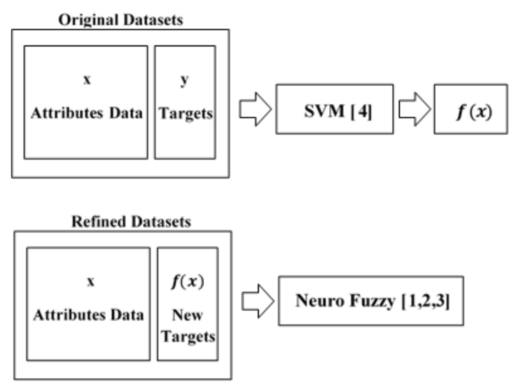

Figure 1. Framework Of Our Approach.

Based on the Figure 1 above, datasets are trained using SVM method that offered by [4]. Then, the value of the estimated function $f(x)$in the training phase of SVM is used to refine the first attributes data. Thus, it is obtained data attributes and a new target which is then used for the training process neuro fuzzy. Neuro fuzzy method used here is a neuro fuzzy method based on linguistic hedge integrated with the principles of feature selection [1,2,3]. The results of feature selection and weighting of neuro fuzzy training process then is used for predicting new datasets.

[1]Elsen Ronando, M. Isa Irawan, and Erna Apriliani are with Departement of Mathematics, Faculty of Mathematics and Science, Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia. E-mail: gimbz16@gmail.com; mii@its.ac.id; april@matematika.its.ac.id.

RESULTS

In our evaluation, we implemented our algorithm on a laptop 2,40 GHz Intel Core i5-520M with 4 GB of RAM using MATLAB R2012b. Moreover, we also conducted our experiments with 10 dataset for classification problems [5], such as iris, wine, sonar, glass, ionosphere, breast cancer, vehicle, vowel, yeast, and segment. Each dataset was divides into 70% for training phase and 30% for testing phase. In addition, some datasets were normalized into range [-1,1]. For kernel function, we used two types of kernel functions, such as Gaussian radial basis function and exponential radial basic function.

The following is a comparison of error rate between SVM and our approach that illustrate in Table 1.

TABLE 1. A COMPARISON OF ERROR RATE.

Dataset	Dimension Data	SVM Error (%)	SVMNF Error (%)	Difference Error (%)
Iris	[150x 4]	4	0	4
Wine	[178x13]	1,9231	1,9231	0
Sonar	[208x59]	54,8387	54,0323	0,8064
Glass	[214 x 9]	40	35,3846	4,6154
Ionosphere	[351x34]	5,7143	5,7143	0
Breast Cancer	[683x9]	0,9756	0,8293	0,1463
Vehicle	[846x18]	34,252	34,133	0,199
Vowel	[990x8]	58,2492	49,293	8,9562
Yeast	[1484x8]	39,72	48,72	9
Segment	[2310x19]	12,6984	17,922	5,2236
Average	-	-	-	3,3

Meanwhile, the results of the comparison of the two approaches for computation time in the testing phase can be shown in the Table 2.

TABLE 2. A COMPARISON OF COMPUTATION TIME.

Dataset	Dimension Data	SVM Runtime (second)	SVMNF Runtime (second)	Reduction Runtime (%)
Iris	[150x 4]	0,6208	0,039	93,71
Wine	[178x13]	0,6988	0,123	82,39
Sonar	[208x59]	0,5366	0,5194	3,205
Glass	[214 x 9]	2,6504	0,110761	95,82
Ionosphere	[351x34]	1,2074	0,0670	94,45
Breast Cancer	[683x9]	3,5349	0,145	95,89
Vehicle	[846x18]	15,2506	8,7766	42,45
Vowel	[990x8]	84,37	11,1775	86,75
Yeast	[1484x8]	140,3605	73,1566	47,87
Segment	[2310x19]	255,296	229,2256	10,21
Average	-	-	-	65,2745

CONCLUSIONS

Based on the results of our evaluation in several cases of classification datasets, our approach showed better performance than SVM for improving the effectiveness of testing computation time. Our approach also significantly is not affecting the accuracy of SVM itself. It is caused by SVM and Neuro Fuzzy in training phase which are capable of searching the best hyper plane and reduce overlapping class. Although our approach presented a good performance, but it still has some drawbacks, such as the high computation time of the training phase. It is due to two training process, such that SVM and Neuro fuzzy as illustrated in Figure 1. Moreover, our approach also still use random process for selecting input parameters. It means the required additional algorithm that is able to search the best parameters input for increasing their performance. Noise filtering method can also be added for future work in this approach to reduce the complexity of data in preprocessing phase.

ACKNOWLEDGEMENTS

The first author would like to thank Fresh Graduate Scholarship from Directorate General of Higher Education (DIKTI), and also Unggulan Biro Perencanaan Kerjasama Luar Negeri (BU-BPKLN) Scholarship from the Ministry of National Education Indonesia (KEMENDIKNAS) for their funding of this research and his study.

REFERENCES

[1] B. Cetisli, "Development of an adaptive neuro-fuzzy classifier using linguistic hedges: Part 1," Expert Syst. Appl., vol. 37, no. 8, pp. 6093–6101, 2010.

[2] ——, "The effect of linguistic hedges on feature selection: Part 2," Expert Syst. Appl., vol. 37, no. 8, pp. 6102–6108, 2010.

[3] B. Cetisli and A. Barkana, "Speeding up the scaled conjugate gradient algorithm and its application in neuro-fuzzy classifier training," Soft Comput., vol. 14, no. 4, pp. 365–378, 2010.

[4] C. Hsu and C. Lin, "A comparison of methods for multiclass support vector machines," IEEE Transactions on Neural Networks, vol. 13, no. 2, pp. 415–425, 2002.

[5] K. Bache and M. Lichman, "UCI machine learning repository," 2015. [Online]. Available: http://archive.ics.uci.edu/ml

[6] S. Kang and S. Cho, "Approximating support vector machine with artificial neural network for fast prediction," Expert Syst. Appl., vol. 41, no. 10, pp. 4989–4995, 2014.