How to cite this paper:

Abdullah Husin & Ku Ruhana Ku-Mahamud. (2017). Reservoir gate opening classification using multiple classifier system with ant system-based feature decomposition in Zulikha, J. & N. H. Zakaria (Eds.), Proceedings of the 6th International Conference of Computing & Informatics (pp 99-104). Sintok: School of Computing.

RESERVOIR GATE OPENING CLASSIFICATION USING MULTIPLE CLASSIFIER SYSTEM WITH ANT SYSTEM-BASED FEATURE DECOMPOSITION

[1]Abdullah Husin and [2]Ku Ruhana Ku-Mahamud

[1]Department of Information System, Universitas Islam Indragiri
Tembilahan, 29213, Indonesia
[2]School of Computing, Universiti Utara Malaysia
06010 Sintok, Kedah, Malaysia

ABSTRACT. Classification of reservoir gate opening (RGO) is an important task in flood management. Reservoir water level has been used to determine the number of gates to be opened when flood is imminent to prevent disaster. Predicting the number of gates to be opened is crucial to avoid any disaster. Multiple classifier system has been shown to provide better classification accuracy as compared to single classifier system. However, there is no guideline on the number of classifiers to be combined and no measurement was proposed to measure the compactness of the classifiers. This study proposes an ant system-based feature decomposition approach to develop a multiple classifier ensemble for classification of RGO. Experiments have been conducted using the k-nearest neighbour, decision tree, nearest mean classifier and linear discriminant analysis as base classifier, and performance of ant system has been compared with random subspace method. Based on the results, it can be concluded that the multiple classifier with ant system-based feature decomposition produced better classification accuracy than random subspace method. Best classification results were obtained when multiple decision tree is constructed to make predictions of RGO with an average accuracy of 89.17%. This method is expected to be useful to apply for RGO classification and future work can be done to include rainfall precipitation besides reservoir water level.

Keywords: reservoir gate opening, multiple classifier system, ant system

INTRODUCTION

Reservoir provides many benefits to human life. Multipurpose reservoirs are used for versatile operations such as water supply, flood control, hydropower generation, agriculture and recreation. Thus, reservoir management is complex due to conflict interest among the objectives (Lin & Rutten, 2016). Reservoir gate opening (RGO) classification is one of the important tasks for reservoir operators to determine the amount of water that will be released from the reservoir. Reservoir water release should be sufficient to ensure that the capacity of the reservoir is at a safe level and the water released does not lead to disaster (Mokhtar, Ishak, & Norwawi, 2014).

Combining multiple classifier has been used in solving pattern classification problems and has been proven in several application domains such as email classification (Chharia & Gupta, 2013), cancer classification (Margoosian & Abouei, 2013) and bankruptcy classification (Tsai, Hsu, & Yen, 2014). Previous studies show that effective ensemble must consist of a set of accurate and diverse classifier (Woźniak, Graña, & Corchado, 2014). Feature decomposition method is one of method that can be used to induce diversity in a set of classifier. The idea of feature decomposition is to train each classifier using different feature set. The advantages of this method are (1) training a classifier on different feature set will reduce correlation among classifiers (Rokach, 2010); (2) all available information in the original training set is used, since these features may contain important information; (3) can be used to overcome the high dimensionality problem, by dividing the original feature set into several subset of features (Ahn et al., 2007; Rokach, 2006; Yang, Yang, Zhou, & Zomaya, 2010); (4) the training process is faster because a set of classifier models is trained simultaneously using a substantially smaller feature set instead of training a classifier model using the whole feature set (Ming Ting, Wells, Chuan Tan, Wei Teng, & Webb, 2011); (5) decomposition features will improve the classification performance to the small sample size problem (Yang et al., 2010).

The difficulty in developing an ensemble is finding the optimal feature set partition for training process which will provide an accurate and diverse classifier. Commonly used methods for classification are genetic algorithm (GA) and simulated annealing algorithm. However, Ant-based algorithm shows better performance than two previous mentioned algorithms (Chicco, 2011). The Ant System (AS) algorithm which is the most popular variant of Ant-based algorithm. The use of Ant system for set partitioning problem has been applied in several studies (Crawford et al., 2013; Maniezzo & Milandri, 2002; Randall & Lewis, 2010).

The purpose of this study is to build ant system-based feature decomposition to train a set of classifier in constructing diverse and accurate classifier ensemble for RGO classification. This paper is structured as follows. The proposed method is presented in Section 2 while Section 3 presents the experiments and discussion on results. Concluding remarks and future work are provided in the last section.

PROPOSED METHOD

The AS algorithm and the majority voting technique are used to construct diverse and accurate classifier ensemble. Figure 1 shows the block diagram of the proposed method which consists of two components namely the ant system-based feature set partitioning and the majority voting technique. In combining several classification outputs of classifier ensemble, the majority voting has been seen as the optimal combiner (Ponti, 2011).

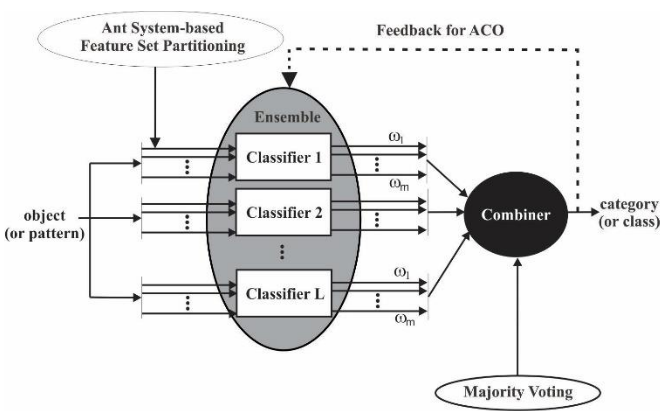


Figure 1: Block diagram of proposed method

The ant system-based algorithm is developed to form optimal feature set partition. A disjoint feature set decomposition is performed based on AS feedback. The best partition will be formed by AS feedback until classification accuracy reaches 100% or the maximum iteration limit has been reached. The classifier ensemble is trained on different feature set partition to induce diversity. There is no feature in the training set is eliminated. The output of classifier is combined by majority voting to produce combination output.

The compactness in a set of classifier is introduced with the aim of answering diversity-accuracy dilema (Li & Gao, 2010). Let D = {D1, …, DL } be a set of classifier, Ω = {ω1, …, ωc } be a set of class labels and x ∈ Rn be a vector with n features to be labeled in Ω. Let Z = {z1, …, zN} be a labeled data set, zj ∈ Rn be a feature vector with n features for data instance j. The output of a classifier Di can be represented as an N-dimensional vector v=[Di(z1), Di(z2), …, Di(zN)]T such that Di(zj) = 1 if Di recognize correctly zj and 0 otherwise, i = 1, …, L and j = 1, …, N. We denote $\bar{x}$ to be the mean accuracy of base classifiers as follows:

$$\bar{x} = \frac{1}{NL}\sum_{j=1}^{N}\sum_{i=1}^{L} D_i(z_j) \qquad (1)$$

Each classifier *Di* (*i* = 1, . . ., *L*) assigns an input feature vector z $\in R^n$ to one of the class label from Ω. i.e., $D_i : R^n \rightarrow \Omega$. The output of classifier ensemble is an *L* dimensional vector $r = [D_1(z), \ldots, D_L(z)]^T$ containing the decisions of each classifier. The compactness is the ratio between the number of observations on which all of classifier are correct to the total number of observations. In this way the compactness of a set of classifiers measured directly (non-pairwise). The compactness (s) is given as:

$$s = \frac{1}{N}\sum_{i=1}^{N} t_i \qquad (2)$$

where

$$t_{i=} \begin{cases} 1 : \text{ if } D1(zj) = D2(zj) = \cdots = DL(zj) = \omega j \\ 0 : otherwise \end{cases}$$

EXPERIMENT AND RESULTS

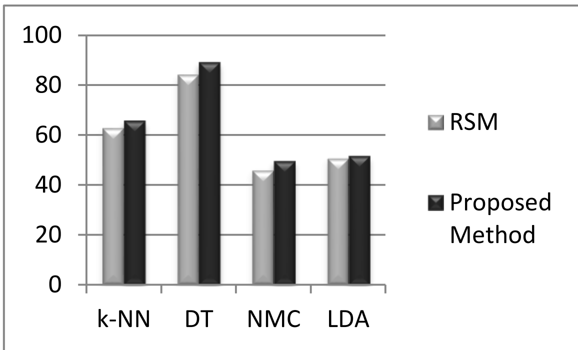
The experiments were conducted on RGO dataset collected from Timah Tasoh reservoir, situated in the northern part of Malaysia. The Department of Irrigation and Drainage monitors and manages the reservoir which serves for flood mitigation as well as water supply and recreation. The data consist of reservoir gate opening decisions in terms of water level, rainfall and number of gates. The hydrological data has the daily rainfall readings measured in milimetre (mm), recorded from five gauging stations along the upstream rivers. Data has been cleaned and represented where three class labels i.e. GO2, GO4 and GO6 are used to represent the opening of two, four and six gates respectively.

The performance of proposed method were evaluated by constructing four homogenous ensembles (e.g. *k*-nearest neighbor (*k*-NN) ensembles, decision treee (DT) ensembles, nearest mean classifier (NMC) ensembles and linear discriminant analysis (LDA) ensembles) to classify the RGO dataset. The classification accuracy is used to determine the performance of proposed method. The classification accuracy is the ratio of numbers of all correctly classified instances and the total number of instance. In order to validate the performace evaluation, all the experiments were repeated ten times, each using 10-fold cross validation method (Burton, Thomassen, Tan, & Kruse, 2012). The classification accuracy of the constructed ensembles are compared with random subspace method (RSM). RSM is chosen because it is popular and widely used by previous studies to build an ensemble classifier.

Table 1 shows the means and standard deviations of the classification accuracy of RSM and proposed method. The results are again displayed in Figure 2 for the classification accuracies and in Figure 3 to show the relationship between compactness and classification accuracy.

Table 1. Accuracy of RSM and Proposed Method

	RSM				Proposed Method							
Experiment #	k-NN	DT	NMC	LDA	k-NN	Compactness	DT	Compactness	NMC	Compactness	LDA	Compactness
1	62.07	83.33	46.51	50.10	**65.49**	18.40	**90.48**	38.88	**50.49**	16.94	**51.70**	31.31
2	62.08	84.52	46.11	51.30	**65.47**	18.24	**89.28**	38.78	**49.70**	16.54	**51.50**	31.17
3	63.07	84.52	44.71	50.89	**66.07**	17.80	**86.90**	36.99	**50.29**	16.47	**52.90**	30.64
4	63.07	83.33	46.31	51.30	**66.67**	18.72	**91.67**	40.89	**48.70**	17.54	**52.50**	30.27
5	62.48	83.33	47.11	50.30	**65.06**	19.00	**88.10**	36.66	**48.90**	16.95	**51.90**	31.02
6	63.67	84.52	46.31	51.30	**65.26**	19.28	**91.67**	39.90	**49.10**	17.36	**53.09**	31.23
7	63.27	83.33	45.31	50.10	**65.67**	18.36	**88.10**	36.30	**49.30**	16.06	**51.30**	31.41
8	62.48	85.71	45.11	49.90	**66.27**	19.72	**88.10**	37.09	**50.09**	16.36	**50.30**	30.88
9	63.07	85.71	45.91	50.90	**65.27**	18.60	**86.90**	35.05	**49.30**	16.45	**50.70**	30.03
10	62.87	83.33	46.31	51.30	**66.47**	18.64	**90.48**	38.56	**51.09**	15.98	**51.50**	30.65
Mean	62.81	84.16	45.97	50.74	**65.77**	18.68	**89.17**	37.91	**49.70**	16.67	**51.74**	30.86
Standard deviation	0.52	0.98	0.72	0.58	0.56	0.55	1.82	1.79	0.77	0.52	0.90	0.46



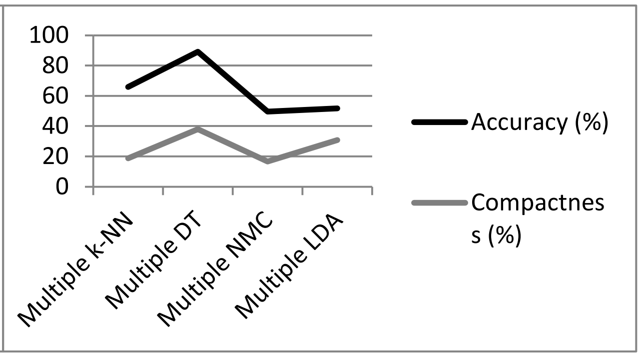


Figure 2. Accuracy % of RSM vs Proposed Method Figure 3.Compactness vs Classification accuracy

From the results, it can be seen that a small deviation of the classification accuracies was obtained which showed that the experiments were accurate and good. The proposed method has successfully delivered better results in developing four homogenous ensembles. Accuracy obtained by the proposed method always exceeded the RSM method (refer Figure 2). There is a positive relationship between compactness and classification accuracy as displayed in Figure 3. The compactness has been shown to be the factor that significantly influences the ensemble accuracy.

CONCLUSION

The proposed ant system-based feature set decomposition has been evaluated and compared to the random subspace in the context of RGO classification. A compactness measurement method in classifier ensemble has also been introduced. The ant system-based feature set decomposotion produced better classification results than the RSM. Furthermore the compactness measurement has been shown to have a positive relationship to classification accuracy. Additional issue to be further studied is how the method can be implemented with other classifiers and it is essentialto ensure that the compactness parameter can be generally accepted.

ACKNOWLEDGEMENTS

The authors wish to thank the Ministry of Higher Education Malaysia for funding this study under the Long Term Research Grant Scheme (LRGS/b-u/2012/UUM/Teknologi Komunikasi dan Informasi), and the Department of Irrigation and Drainage Malaysia for supplying hydrology and reservoir operational data.

REFERENCES

Ahn, H., Moon, H., Fazzari, M. J., Lim, N., Chen, J. J., & Kodell, R. L. (2007). Classification by ensembles from random partitions of high-dimensional data. *Computational Statistics and Data Analysis*, 51(12), 6166–6179. http://doi.org/10.1016/j.csda.2006.12.043

Burton, M., Thomassen, M., Tan, Q., & Kruse, T. A. (2012). Prediction of breast cancer metastasis by gene expression profiles: A comparison of metagenes and single genes. *Cancer Informatics, 11*, 193–217. http://doi.org/10.4137/CIN.S10375

Chharia, A., & Gupta, R. K. (2013). Email classifier: An ensemble using probability and rules. *2013 6th International Conference on Contemporary Computing, IC3 2013*, 130–136. http://doi.org/10.1109/IC3.2013.6612176

Chicco, G. (2011). Ant colony system-based applications to electrical distribution system optimization. In A. Otsfeld (Ed.), Ant Colony Optimization - Methods and Applications (pp. 237–262). Rijeka, Croatia: InTechOpen. http://doi.org/10.5772/13840

Crawford, B., Soto, R., Monfroy, E., Castro, C., Palma, W., & Paredes, F. (2013). A hybrid soft computing approach for subset problems. *Mathematical Problems in Engineering*, 2013. http://doi.org/10.1155/2013/716069

Li, K., & Gao, H. (2010). A novel measure of diversity for support vector machine ensemble. *3rd International Symposium on Intelligent Information Technology and Security Informatics, IITSI* 2010, 366–370. http://doi.org/10.1109/IITSI.2010.136

Lin, N. M., & Rutten, M. (2016). Optimal Operation of a Network of Multi-purpose Reservoir: A Review. *Procedia Engineering*, 154, 1376–1384. http://doi.org/10.1016/j.proeng.2016.07.504

Maniezzo, V., & Milandri, M. (2002). An ant-based framework for very strongly constrained problems. *Lecture Notes in Computer Science, vol. 2463. Berlin Heidelberg: Springer-Verlag*. http://doi.org/10.1007/3-540-45724-0_19

Margoosian, A., & Abouei, J. (2013). Ensemble-based classifiers for cancer classification using human tumor microarray data. *Proceedings of the 21st International Conference on Electrical Engineering (ICEE), 1–6*. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6599553

Ming Ting, K., Wells, J. R., Chuan Tan, S., Wei Teng, S., & Webb, G. I. (2011). Feature-subspace aggregating: Ensembles for stable and unstable learners. *Machine Learning*, 82(3), 375–397. http://doi.org/10.1007/s10994-010-5224-5

Mokhtar, S. A., Ishak, W. H. W., & Norwawi, N. M. (2014). Modelling of reservoir water release decision using neural network and temporal pattern of reservoir water level. *Proceedings - International Conference on Intelligent Systems, Modelling and Simulation*, 127–130. http://doi.org/10.1109/ISMS.2014.27

Ponti, M. P. (2011). Combining classifiers: From the creation of ensembles to the decision fusion. *Proceedings - 24th SIBGRAPI Conference on Graphics, Patterns, and Images Tutorials*, SIBGRAPI-T 2011, (Icmc), 1–10. http://doi.org/10.1109/SIBGRAPI-T.2011.9

Randall, M., & Lewis, A. (2010). Modifications and additions to ant colony optimisation to solve the set partitioning problem. *Proceedings - 6th IEEE International Conference on E-Science Workshops, E-ScienceW 2010*, 110–116. http://doi.org/10.1109/eScienceW.2010.27

Rokach, L. (2006). Decomposition methodology for classification tasks: A meta decomposer framework. *Pattern Analysis and Applications*, 9(2–3), 257–271. http://doi.org/10.1007/s10044-006-0041-y

Rokach, L. (2010). *Pattern Classification Using Ensemble Methods. Singapore*: World Scientific Publishing Co. Pte. Ltd.

Tsai, C. F., Hsu, Y. F., & Yen, D. C. (2014). A comparative study of classifier ensembles for bankruptcy prediction. *Applied Soft Computing Journal*, 24, 977–984. http://doi.org/10.1016/j.asoc.2014.08.047

Woźniak, M., Graña, M., & Corchado, E. (2014). A survey of multiple classifier systems as hybrid systems. *Information Fusion*, 16(1), 3–17. http://doi.org/10.1016/j.inffus.2013.04.006

Yang, P., Yang, Y. H., Zhou, B. B., & Zomaya, A. Y. (2010). A review of ensemble methods in bioinformatics. *Current Bioinformatics*, 5(4), 296–308. http://doi.org/10.2174/157489310794072508