Classification model for performance diagnosis of dry port by rail

Raywat Makkhongkaew, Stéphane Bonnevay, Alex Aussem, Khalid Benabdeslem

CLASSIFICATION MODEL FOR PERFORMANCE DIAGNOSIS OF DRY PORT BY RAIL

Raywat MAKKHONGKAEW
PhD Student
Université de Lyon
Université Lyon 1
LIRIS UMR CNRS 5205, F-69622, France
Fax: +33-4-72-43-15-37
E-mail: raywat.makkhongkaew@liris.cnrs.fr

Stéphane BONNEVAY
Associate Professor
Université de Lyon
Université Lyon 1
ERIC 5 PIERRE MENDES, 69676, France
Fax: +33-4-72-43-15-89
E-mail: stephane.bonnevay@univ-lyon1.fr

Alexandre AUSSEM
Professor
Université de Lyon
Université Lyon 1
LIRIS UMR CNRS 5205, F-69622, France
Fax: +33-4-72-43-15-37
E-mail: alexandre.aussem@univ-lyon1.fr

Khalid Benabdeslem
Associate Professor
Université de Lyon
Université Lyon 1
LIRIS UMR CNRS 5205, F-69622, France
Fax: +33-4-72-43-15-37
E-mail: khalid.benabdeslem@univ-lyon1.fr

Abstract: Dry port is an inland intermodal terminal directly linked by road or rail to a seaport as a center for the transshipment of sea freight to inland destinations. However roads contribute to pollution more than rail. For sustainable transportation systems, we need to improve the performance of rail dry port systems. The purpose of this research is to propose a diagnostic system for modeling and classification of rail dry port performance, by use of supervised ensemble learning approaches. We propose an empirical comparison of supervised ensemble learning approaches and application for performance diagnosis of dry port by rail. Our results demonstrate that the performance testing of proposed diagnostic system is found to be very satisfactory. The calculation methods demonstrating the results of the data are provided in this study diagnosis of dry port by rail. Finally, the conclusions for classification modeling for dry port by State Railway of Thailand (SRT) are given.

Key Words: Dry port, Ensemble learning, Classifier ensembles, Empirical performance comparison

1.INTRODUCTION

Dry ports or Inland ports is a part of the transport system, particularly in gateway regions having a high confidence on commerce. A dry port is an inland container terminal that has direct road, rail or other access to an adjacent sea port, and has export/import facilities. The dry port should in fact accept the container as if the container has reached the seaport itself. Dry ports offers a one stop service for cargo handling and a logistics solution for international export and import, as well as domestic distribution. It provides integrated port and logistics services with lots of logistics and supply chain businesses, such as exporters, importers, carriers, terminal operators, container freight station, bonded warehouse, transportation, third party logistics (3PL), empty container depot, as well as banks and other supporting facilities. Being the extension gate of international dry port, document formalities for port clearance and customs clearance will be completed in the dry port (United Nations, 2013). It is like taking the sea port to the industrial manufacturing regions that works both as a port of origin and as a port of destination. Dry ports normally have two types, Dry ports by road and rail. This study works on dry port by rail (The State Railway of Thailand) due to road makes pollution more than rail. For sustainable transport, we need to care about the pollution of environment. The rail transport is environmentally friendly with our globe by decrease the impact of pollution by physical of operation. Rail is a sustainable transport. The problem is the question about, How can we encourage the rail transport and competition with road?

The model of performance diagnosis for dry port by rail is the key of the competition performance of rail dry port. The performance diagnosis is a main thing for every problems. The main task of this study is modeling of the diagnosis system then classification the performance for dry port by rail to encourage the sustainable rail transport and competition with road. Supervised ensemble learning approaches are the machine learning method for performance diagnosis using a combination of statistical data and qualitative causal assumptions by expert. From the performance diagnosis by proposed method will yield the performance forecast model for dry port by rail.

Interdisciplinary is relating to more than one branch of knowledge. In this case is comparing and supplement statistical technique with expert knowledge for sustainable transport. Nowaday, It has no method can perform well without qualitative assumptions by expert. As our proposed method, It also has the problem. However the expert can give method about the solution. It's complement each other like a conductor guides the musician in orchestra. As this study the expert will guide the machine learning method to play the data for given the result to improve the performance of sustainable rail transport. No universal method or multifunctional method. That is why interdisciplinary is necessary to be carefully for proposing the new method.

The paper is organized as follows. Section 2 describes the background of SRT dry port. In Section 3, we describe the methodology of experiments that is used in our research work and Section 4 for our data set of experiments. We propose application to the performance diagnosis in Section 5 with the experimental results of our technique. And Section 6 concludes the paper.

2.BACKGROUND

We will describe the overview background to introduce the important of SRT dry port. In the future, the government plans the development to be the gateway dry port of Southeast Asia. The planning and construction, that would connect all the countries of mainland Southeast Asia as Lat Krabang Inland Container Depot Bangkok is center of the region.

2.1 Lat Krabang Inland Container Depot (LICD)

The concept of an inland container depot near Bangkok has been developed in conjunction with the new deep sea port at Laem Chabang on the eastern seaboard by Japan International Cooperation Agency (JICA) in 1989. The study concluded that the ICD would be needed as a back-up facility to serve a rapid growing of industrial expansion in Thailand; the recommended site was a Greenfield area near Lat Krabang Industrial estate which is approximately 30km east of Bangkok. The chosen location which has been scheduled for development was adjacent to the main eastern railway line and surrounded by new Chonburi highway and new international airport on the north. By early 1993, the government purchased a number of essential lands, and the State Railway of Thailand (SRT) had been authorized to commence filling the site and carrying out the design and facility construction.

A design of the facilities has been implemented in accordance with the layout outlined in the JICA report; however, a number of adjustments have been made by the consultants who closely work with SRT, relevant government bodies, and potential customers.

2.2 Time and Cost of Construction

As a terminal is required by the Government to get ready as soon as possible, a "fast-track" design and construction system has been instituted. The terminal had been built in two overlapping phases; the first phase included the first module (Module A) - common area including road, administration building, utilities supply, and partly railhead area. It was completed in February 9th, 1995. The entire site which included the rest of modules (Modules B-F), perimeter roadways, and the remaining infrastructures and railhead area were finished in October 25th, 1995. The total cost of this project is 2,943.543 Million Baht or 73.59 Million USD (1 USD = 40 Baht).

2.3 Location and Facilities at LICD

Inland Transport Links, Lat Krabang ICD is located approximately 30 km east of Bangkok by rail, northwest of Hue Take railway station and approximately 118 km from Laem Chabang Port. By road: Access to LICD via Chao Khun Tahan road from the north. In the future, the LICD will be directly entered from the south by the new Bangkok-Chonburi express way no. 7 which is now under construction. By rail: LICD is linked to the eastern main line through Hua Takae station. Railway infrastructure has been provided to the rail transfer area inside the LICD alongside the modules. LICD modules, A terminal is provided with a full range of facilities for standardized modern ICD operations. The facilities divided into three sections are listed. Railhead area, 4 track railhead area, laid on ballast, approx.1,200 m long overall, the layout provides 4x500 m berths for train standing, access area are fully paved in reinforced concrete, railhead operators' office and small workshop. Administration area, main office building, full facilities for customs, SRT, banks, etc. Modules office is available for commercial leasing, extensive car-parking area and large truck rest area, warehouse under customs administration, weigh-bridge. LICD modules, dimensions of a CY area, CFS shed length, and number of reefer points can be varied according to customer requirements. Each module contains the following facilities. Container Yard area is instantly adjacent to railhead operation area. 48 reefer points (380-440 v.3ph). Security fencing to customs standards around the bonded area. Terminal perimeter fencing. Warehouse from 5,800 to 8,440 square metres. Main office building : 1,736 square meters. Container Gate : 780 square meters. Workshop : 720 square meters. Canteen : 336 square meters. Washing area : 500 square meters. Container Yard : 48,000 - 97,600 square meters. Parking and miscellaneous : 14,654 - 25,998 square meters.

2.4 The Concession of LICD

It is the Government's policy to allocate private sector to participate in the container and cargo handling services to the greatest extent practicable in order to ensure economic efficiency. To this end, the Government has decided to grant concessions to private sector to develop, manage and operate the inland container depot. The concession fees are utilized for developing and managing the common facilities. SRT has been appointed by the Government to be the administrator of the concession with the Module Operators (MO) as concessionaires. SRT invited interested and qualified applicants to submit Tenders for the Concession to manage and operate the Modules at Lat Krabang in March 1995 and June 1996. The Concession Contracts were signed in March 6th, 1996 (Modules A,B,C & F) and December 19th , 1997 (Modules D & E) for the first and second tenders respectively.

Functions of SRT at LICD, Monitoring the activities of the MOs to ensure that the operations have been implemented in accordance with the law and regulations, giving recommendation and providing updated information to the statutory bodies for any necessary changes. Investigating any complaints from the public which have not been resolved by the MOs. Monitoring the MO's performances under the concession contract, especially with regard to operational efficiency, environmental protection, safety procedures and satisfactory maintenance of fixed assets. Requesting and receiving operating statistics from the MOs and preparing reports on the use of the ICD's assets. Controlling land side traffic by ensuring gate control at the main gate, not at the Module gates. Figure 1 presents the operation of LICD.

Figure 1. The operation of LICD

3.METHODOLOGY

The ensemble approaches in machine learning have great potential as a classifier model to significantly increase prediction accuracy over individual classifier models (Schwenker, 2013). Mohamed *et al*. (2013) compared twenty prototypical supervised ensemble learning algorithms. The experiments support the conclusion that the Rotation Forest (Rot) family of algorithms outperforms all other ensemble methods which is much in line with the results earned by Zang (2008). Rot is a method for generating classier ensembles based on feature extraction. To create the training data for a base classier (Rodrı´guez and Kuncheva, 2006). According to the original Rot, let $x = [x_1,...,x_n]^\top$ be a data point described by n features and let X be the data set containing the training objects in a form of an $N \times n$ matrix. Let Y be a vector with class labels for the data, $Y = [y_1,..., y_j]^\top$, where y_j takes a value from the set of class labels $\{w_1,...,w_c\}$. Denote by $D_{1,1},...,D_{s,L}$ the classifiers that we proposed in to the ensemble and by F, the feature set. As with most Ensemble methods, we need to pick L in advance. In order to be fair in term of ensemble size, we construct an ensemble consisting of 40 Rotation Forests which are learned by AdaBoost during 5 iterations. This ratio has been shown to be approximately the empirically best in (Zhang and Zhang, 2008). The number of trees was fixed to 200 in accordance with a recent empirical study (Lobato *et al*., 2013) which tends to show that ensembles of size less or equal to 100 are too small for approximating the infinite ensemble prediction. All classifiers can be trained in parallel, Parallel Computing (Almási and Gottlieb, 1989) will be benefit in this case. Rot method builds multiple classifiers on randomized projections of the original dataset. The feature set is randomly split into K subsets (K is a parameter of Rot) and Principal Component Analysis (PCA) (Han, 2001) is applied to each subset in order to create the training data for the base classifier. The idea of the rotation approach is to encourage simultaneously individual accuracy and diversity within the ensemble (Bibimoune *et al*., 2013). The size of each subsets of feature was fixed to 3 by Rodriguez *et al.* (2006). The number of sub classes randomly selected for the PCA was fixed to 1 as we focused on binary classification. The size of the bootstrap sample over the selected class was fixed to 75% of its size (Juan and Ludmila, 2006). Our proposed method investigated the performance $D_{1,1},...,D_{s,L}$ then selected via Mean Squared Normalized Error Performance Function (MSE). MSE is a network performance function. It measures the network's performance according to the mean of squared errors. Let $\hat{Y}$ is a vector of j predictions, and Y is the vector of the true values, then the (estimated) MSE of the predictor is

$$D_{s,L} = \frac{1}{n}\sum_{j=1}^{n}(\hat{Y}_j - Y_j)^2 \qquad (1)$$

Let

$$D = \{\text{Decision Trees, AdaBoost, ET}\} \qquad (2)$$

(Bibimoune *et al.*, 2013) (Breiman *et al.*, 1984) (Coppersmith *et al.*, 1999) (Freund, 2009) (Geurts *et al.*, 2006). Rot voted class prediction from L times to each recording from the best thing of D_s algorithms in Rot. According to the original Rot (Rodrı´guez and Kuncheva, 2006), algorithm 1 depicts view of the Training Phase of Rot then the Choosing Phase chooses only the best D_s by

$$D_s = \frac{1}{n}\sum_{c=1}^{n}(\mu_j(x) - \omega_c)^2,\ j = 1,...,c \qquad (3)$$

then denotes the best D_s by D'.

Classification Phase
Start
- Perform K-Fold Cross-Validation Phase and Training Phase by D' then feed each testing data set and classifier to next step below.
- For a given x, let $d'_{i,j}(\text{testing data set})$ be the probability assigned by the classifier D'_i to the hypothesis that x comes from class ω_j. Calculate the confidence for each class ω_j then assign x to the class with the largest confidence.
End

K-Fold Cross-Validation Phase
Given
  $K = 10$;
  training data set = Fold f;
  testing data set $\neq$ Fold f;
for $f = 1...K$ **do**
  Training Phase
  Given
  X : the objects in the training data set (an $N \times n$ matrix);
  Y : the labels of the training set (an $N \times 1$ matrix);
  L : the number of classifiers in the ensemble;
  $\{\omega_1, ..., \omega_c\}$: the number of subsets;
  for $s = 1, ..., n$ **do**
    for $i = 1...L$ **do**
      Prepare the rotation matrix R_i^a
      Split **F** (the feature set) into K' subsets: $\mathbf{F}_{i,j}$ (for $j = 1...K'$);
      for $j = 1...K'$ **do**
        Let $X_{i,j}$ be the data set X for the features in $\mathbf{F}_{i,j}$;
        Eliminate from $X_{i,j}$ a random subset of classes;
        Select a bootstrap sample from $X_{i,j}$ of size 75% of the number of objects in $X_{i,j}$. Denote the new set by $X'_{i,j}$;
        Apply PCA on $X'_{i,j}$ to obtain the coefficients in a matrix $C_{i,j}$;
      end
      Arrange the $C_{i,j}$ for $j = 1...K'$ in a rotation matrix R_i;
      Construct R_i^a by rearranging the columns of R_i so as to match the order of features in **F**;
      Build classifier $D_{s,i}$ using (XR_i^a, Y) as the training set
    end
  end
end

Algorithm 1. Training Phase

4.DATA DESCRIPTION

The data set used in our experiments was taken from SRT in the period of ten years. It represents a variety of features and four different types of attributes consist of Nominal, Ordinal, Interval and Ratio without selection or preprocessing by expert due to the experiments need to show the method ability. It's binary classification problem of small size (119 records), many features (152 features) and missing values. The prediction returns two scores to be class whether good or bad performance.

5.APPLICATION

From our database, we extract a training and a test base. First we must choose our training and test data. For this we use K-fold cross-validation method. We repeat the experiment 10 times and average the results, each of them is randomly generated indices for a K-fold cross-validation of N observations. Indices contains equal or approximately equal sections of the integers 1 through K that determine a partition of the N observations into K disjoint subsets. Repeated gives back different randomly generated partitions. K defaults to 10 (McLachlan *et al.*, 2004). In K-fold cross-validation, K-1 folds are used for training and the last fold is used for testing. This mechanism is iterated K times, leaving one different fold for testing each term. The method builds random blocks, which depend on the condition of the default random flow. Therefore, the results from the following experiments will change from those shown. To find the best learner, we compare the Rotation Forest family of algorithms, Rot with Decision Trees (Rot), AdaBoost (Rotb) and Extremely Randomized Tree (RotET) as based learner by Mohamed's literature (Bibimoune *et al*., 2013) that Rotation Forest family of algorithms outperforms all other ensemble methods. The comparison is performed based on three performance metrics: MSE, area under the ROC curve (AUC) and accuracy (ACC).

Rot builds an ensemble of decision trees according to the classical top-down procedure.

AdaBoost (Freund, 2009) (Hsu *et al*., 2010) (Friedman, 2001) (Freund and Schapire, 1997) trains learners sequentially. For every learner with index t and computes the weighted classification error as

$$\varepsilon_t = \sum_{i=1}^{n} d_i^{(t)} I(y_i \neq h_t(x_i)) \tag{4}$$

AdaBoost then increases weights for observations misclassified by learner t and reduces weights for observations correctly classified by learner t. The next learner $t+1$ is then trained on the data with updated weights $d_i^{(t+1)}$. After training finishes, AdaBoost computes prediction for new data using

$$f(x) = \sum_{i=1}^{n} a_i h_i(x) \tag{5}$$

where

$$a_i = \frac{1}{2}\log\frac{1-\varepsilon_i}{\varepsilon_i} \tag{6}$$

are weights of the weak hypotheses in the ensemble. The output from the prediction method of an AdaBoost classification ensemble is an N-by-2 matrix of classification scores for the two classes and N observations.

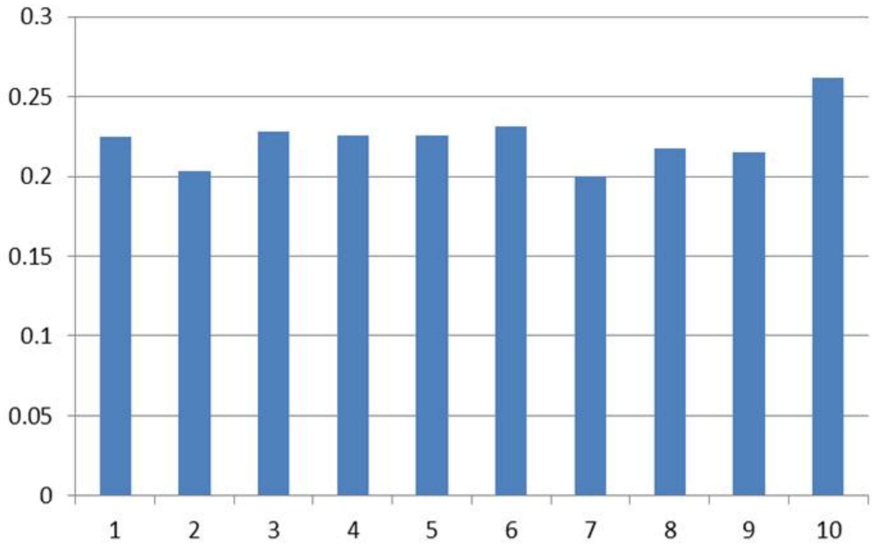
For ET, we use the regression tree package proposed in Geurts *et al*. (2006). The Extra-Trees algorithm builds an ensemble of un-pruned decision or regression trees according to the classical top-down procedure. Its two main differences with other tree based ensemble methods are that it splits nodes by choosing cut-points fully at random and that it uses the whole learning sample (rather than a bootstrap replica) to grow the trees.

Table 1. Mean of MSE, AUC and ACC

	MSE	AUC	ACC
Rot	0.272	0.729	0.729
Rotb	0.229	0.768	0.771
RotET	**0.223**	**0.775**	**0.777**

RotET is the best in three metrics of cost function as shown in table 1. We use the best significant method RotET to extend our experiment.

Figure 2, 3 and 4 show how the mean MSE, AUC and ACC are variation according to the experiment.



Figrue 2. RotET MSE variation according to experiment

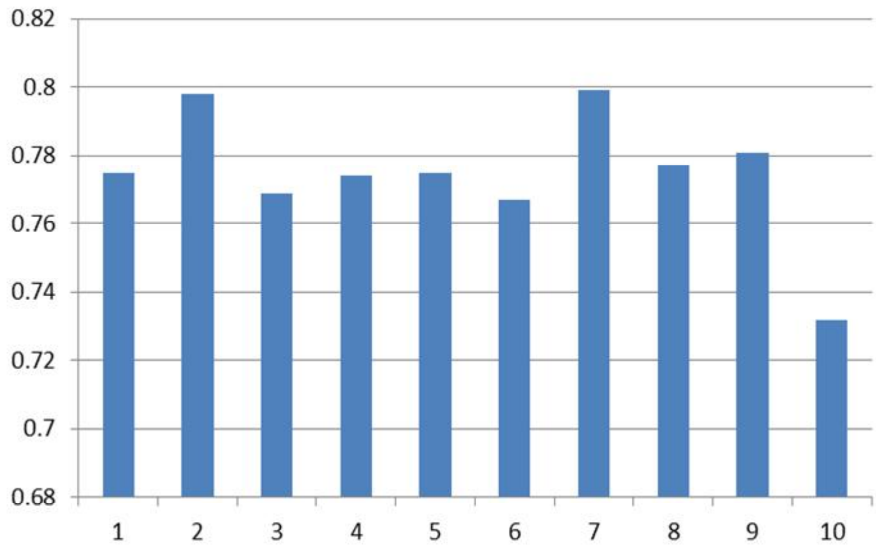


Figure 3. RotET AUC variation according to experiment

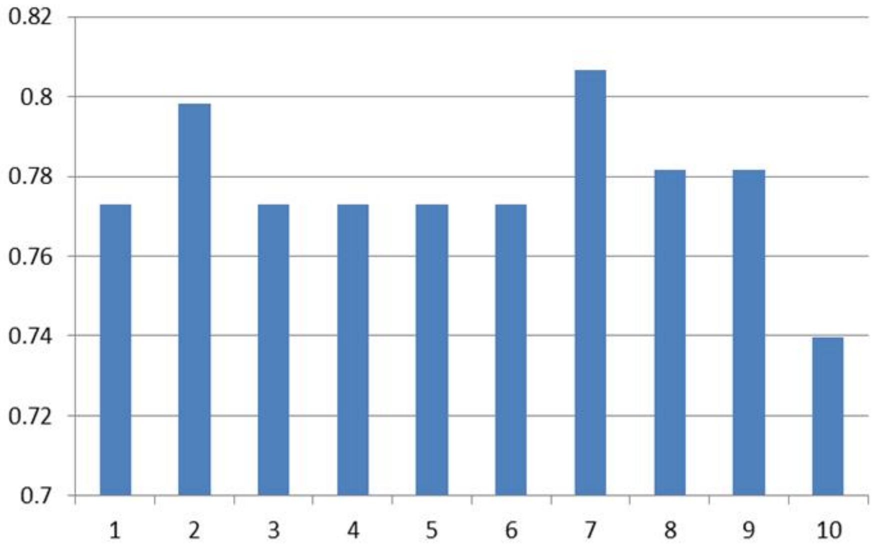


Figure 4. RotET ACC variation according to experiment

6.CONCLUSIONS

We described an empirical comparison between three outstanding prototypical supervised ensemble learning algorithms over SRT dataset with binary labels for SRT problem application. The experiments presented here support the conclusion that the success of RotET approach is closely tied to its ability to simultaneously encourage diversity and individual accuracy via rotating the feature space and keeping all principal components. The model will perform forecasting the performance in real time then the planning of operation and investment will forecast timely planning way by SRT operation unit to encourage the performance for SRT dry port.

ACKNOWLEDGEMENTS

The work was supported by State Railway of Thailand. I would like to thank Asst. Prof. Siripong Preutthipan and Asst. Prof. Sathit Nakkrasae for the useful comments. In addition, a thank you to Mr. Sahathat Narkkhongkam for consent to include copyrighted pictures as a part of our paper.

REFERENCES

Almási, G.S. and Gottlieb, A. (1989) **Highly parallel computing**. Benjamin-Cummings Publishing Co., Inc., Redwood City, CA.

Bibimoune, M., Elghazel, H. and Aussem, A. (2013) An Empirical Comparison of Supervised Ensemble Learning Approaches. **International Workshop on Complex Machine Learning Problems with Ensemble Methods**, Prague, 24, Sep 2013, pp. 123-138.

Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984) **Classification and Regression**

Trees. CHAPMAN & HALL/CRC, Washington, D.C.

Coppersmith, D., Hong, S.J. and Hosking, J.R.M. (1999) Partitioning Nominal Attributes in Decision Trees, **Data Mining and Knowledge Discovery, Vol. 3, Issue 2**, 197–217.

Freund, Y. (2009) A more robust boosting algorithm, **arXiv:0905.2138v1**.

Freund, Y. and Schapire, R.E. (1997) A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting, **Journal of Computer and System Sciences, Vol. 55, Issue 1**, 119–139.

Friedman, J.H. (2001) Greedy function approximation: A gradient boosting machine, **The Annals of Statistics, Volume 29, Number 5**, 1189–1232.

Geurts, P., Ernst, D. and Wehenkel, L. (2006) Extremely randomized trees, **Machine Learning, Vol. 63, Issue 1**, 3-42.

Han, J. (2001) **Data Mining: Concepts and Techniques**. Morgan Kaufmann Publishers, San Francisco.

Hsu, C.W., Chang, C.C. and Lin, C.J. (2010) A Practical Guide to Support Vector Classification. Tech. rep., Department of Computer Science, National Taiwan University.

Lobato, D.H., Muñoz, G.M. and Suárez, A. (2013) How large should ensembles of classifiers be?, **Pattern Recognition, Vol. 46, Issue 5**, 1323-1336.

McLachlan, G., Do, K.A. and Ambroise, C. (2004) **Analyzing microarray gene expression data**. Wiley, West Sussex.

Rodrı´guez, J.J. and Kuncheva, L.I. (2006) Rotation Forest: A New Classifier Ensemble Method, **IEEE Transactions on pattern analysis and machine intelligence, Vol. 28, No. 10**, 1619-1630.

Schwenker, F. (2013) Ensemble methods: Foundations and algorithms [book review], **Computational Intelligence Magazine, IEEE, Vol. 8, Issue 1**, 77-79.

United Nations (2013) Economic and Social Commission for Asia and the Pacific. Economic and Social Council, United Nations, Bangkok.

Zhang, C.X. and Zhang, J.S. (2008) Rotboost: A technique for combining rotation forest and adaboost, **Pattern Recognition Letters, Vol. 29, Issue 10**, 1524-1536.