Australasian Transport Research Forum 2023 Proceedings
29 November – 1 December, Perth, Australia
Publication website: http://www.atrf.info

Exploring the effectiveness of ensemble models in predicting driver route selection

Haotian Wang[1], Emily Moylan[2], David Levinson[2]

[1,2] School of Civil Engineering, The University of Sydney, Australia

Email for correspondence (presenting author): hwan6559@uni.sydney.edu.au

1. Introduction

Route choice concerns the selection of routes between origins and destinations in the network. Conventional statistical models such as the Multinomial Logit (MNL) model, Path-Size Logit model and other modified MNL models are widely used in route choice modelling (Prato, 2009). With the advancements in Global Positioning System (GPS) and computer technology, tracking travellers' journeys is now possible, and powerful artificial intelligence (AI) tools have been developed for various applications across different fields. Machine learning, a subset of AI, comprises a group of algorithms for diverse modelling tasks and gains more and more attention (Sun and Park, 2017).

For both statistical and most machine learning models, the aim is to find the 'best' model based on the given information. However, in a group of tested models, the 'best' model can vary if data includes different amounts of noise or when different training and testing sets are used. Moreover, the 'non-best' models still have a chance to make a correct prediction when the 'best' model makes mistakes. Therefore, including all the results from each individual model, rather than abandoning them, can provide an even more accurate result than the single 'best' model in theory. Ensemble models, which follow the same idea, have been proven to provide a more accurate result than individual models in weather forecasting (Palmer, 2019), economic prediction (McNees, 1990), and disease forecasting (Sharma et al., 2021), and the transport field shares similar uncertainty with those fields (Wu and Levinson, 2021). Therefore, this study aims to take advantage of the ensemble Model to predict travellers' route choices. To be specific:

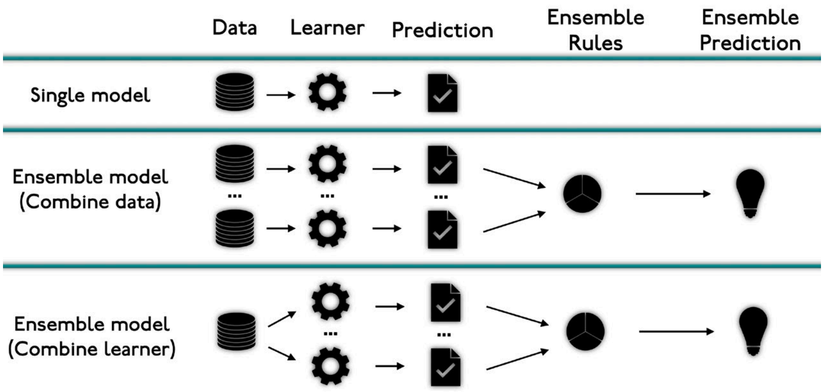
- Demonstrate that the 'best' model may vary as data changes.
- Show the improvement in model performance by applying ensemble techniques to individual models.
- Show the better performance of ensemble models in general.
- Evaluate a set of ensemble techniques and provide recommendations for future studies.

2. Theory of ensemble model

In general, ensemble learning includes the generation and combination of multiple individual learners to solve the learning task. Ensemble modelling is well-known in machine learning but does not have a fixed form or algorithm. As shown in Figure 1, it could be seen as a two-stage model, which firstly trains all individual learners to gain a set of results and then aggregates those results based on ensemble strategies to provide a final prediction. Based on the type of individual learners (Zhou, 2021), ensemble models could be classified as follows:

- Homogeneous ensembles: Comprises individual learners (base models) of the same type. For example, all individual learners in the 'decision tree ensemble' include only the decision tree as the base model.
- Heterogeneous ensembles: Contains individual learners (base models) and algorithms of different types.

Figure 1: Ensemble model vs single model



2.1. Homogeneous ensembles

2.1.1. Bagging

For Bagging (bootstrap aggregation), each individual learner is trained with a sample that is randomly taken from the original data set with replacement. Those samples are constituted differently for learners but have the same size. Since the sample is taken with replacement, some data in an origin data set may be included in different samples. Based on the parallel structure, each individual learner can be trained independently using different cores to reduce computational time. To combine the results from each learner, bagging adopts simple voting for the classification task and simple average for the regression task.

Random Forest (RF) is an extension of bagging. The base learners in RF are decision trees. Similar to bagging, each individual tree is trained based on a stochastic sample from the original training set, but each tree only randomly takes a part of all features (explanatory variables). The process of randomly taking partial of all features is called feature randomness. Results from each tree are then combined based on a simple voting rule.

2.1.2. Boosting

Boosting is a group of algorithms that convert weak learners to strong learners. The idea behind boosting is that one base learner is trained first, and then the distribution of the sample is adjusted based on the results of the first base learner before training the next learner. This adjustment allows subsequent learners to pay more attention to previously misclassified samples. There are various approaches to performing boosting, such as AdaBoost and XGboost. Each individual learner in Boosting algorithms is trained dependently, and thus, the time cost is normally higher than bagging.

2.1.3. Homogeneous ensembles set-up

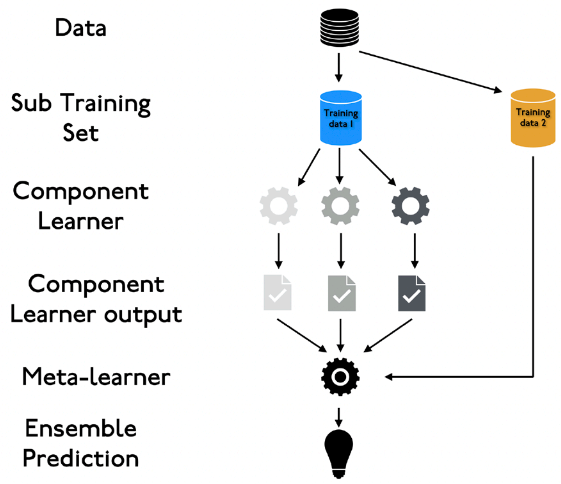
As described above, homogeneous ensembles can be set up in various ways. In this study, we only focus on using bagging and feature randomness techniques to form homogeneous ensemble models. The Tree-based AdaBoost model is included as a base model and a meta-learner for heterogeneous ensembles. Three pairs of base models and homogeneous ensemble models are evaluated:

- MNL model and MNL-based model with bagging and feature randomness techniques (RMNL)
- PSL model and PSL-based model with bagging and feature randomness techniques (RPSL)
- Decision tree and random forest

2.2. Heterogeneous ensembles

Stacking is a two-level ensemble learning model, as shown in Figure 2. Unlike Bagging and Boosting, where results from individual learners are combined based on ensemble rules such as simple voting to make ensemble predictions, stacking takes the outputs from individual learners as inputs for the meta-learner to produce final results. Data are normally split into training and test sets in machine learning approaches. In stacking, the training set is further split into two sub-training sets. The first sub-training set is used to train individual learners (first-level learners), and results gained from first-level learners are then taken as input to second-level learners (meta-learners). To reduce the risk of over-fitting, the second-level learner is trained based on the second sub-training set, and the validation is performed on the test set. The individual learners (first-level learners) in stacking are heterogeneous, while second-level learners could be one of the models in the first-layer learner or other kinds of models.

Figure 2: Stacking



2.2.1. Heterogeneous ensembles set-up

In this study, the base models for heterogeneous ensembles include MNL, RMNL, PSL, RPSL, decision tree (DT), random forest (RF), extra tree (ET), AdaBoost, support vector machine (SVM) and neural network (NN). All these base models are trained with the same training set and tested in the same testing set. To form heterogeneous ensembles, the results from individual learners (base models) are assembled by different ensemble rules. Meta-learners in stacking, including random forest, support vector machine, AdaBoost and logistic model, are tested in this study. For each machine learning model, the tested hyperparameters are listed in Table 1, and a grid search technique is applied to finalise the best hyperparameters. Ensemble rules will be described in detail in the following Sub-section 2.3.

Table 1: Tested hyperparameters of each Machine Learning model

Models	Hyperparameters
Decision tree	maximum depth of the tree, criterion, maximum feature
Extra tree	maximum depth of the tree, criterion, maximum feature
Random forest (RF)	the number of estimators, maximum depth of trees, criterion, maximum feature
AdaBoost	base estimator, the number of estimators, algorithm, learning rate
Support vector machine	regularization parameter (C), kernel, probability, decision function shape
Neural networks	Hidden layer sizes, Activation, solver, learning rate, initial learning rate

2.3. Ensemble rules

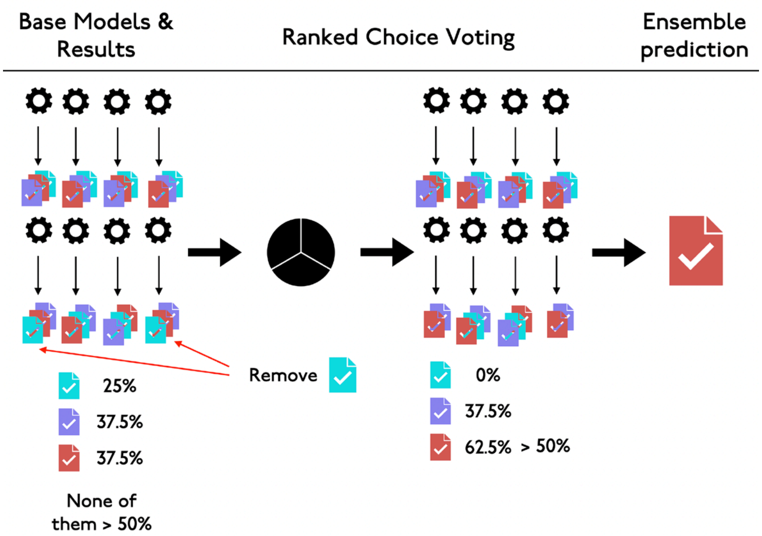
Ensemble rules are the strategies to combine results from individual learners.

2.3.1. Voting

Voting is commonly applied to classification tasks. It normally includes simple voting, which is based on majority rule, and weighted voting, which takes the probability as the weight for the results. For hard voting (simple voting), the result that appears most frequently is the ensemble result. Soft voting (weighted voting) considers the probability of the result given by base models. In that case, base models which have a higher probability for their result have a greater influence on the ensemble result.

In addition, we propose a Ranked Choice Voting strategy to aggregate results from individual learners, as shown in Figure 3. For multi-class classification problems, each individual learner includes the preference order of alternatives, and the first choice from each individual learner consists of the initial result set. If none of the alternatives in the initial set obtains more than 50% of the votes, the alternative with minimum votes will be removed for learners who come up with that alternative, and the votes will be recalculated for the rest alternatives. As shown in Figure 3, since all colours are below 50%, aqua blue is removed from the results in the left-bottom and right-bottom models. After that, as the red colour exceeds 50%, the ensemble prediction is then red. For binary choice problems, the Ranked Choice Voting strategy provides the same results as the hard voting.

Figure 3: Ranked choice voting



2.3.1. Combine by Learning

Combine by learning uses the second-level learner in stacking. As shown in Figure 2, after obtaining the predictions from individual learners, a meta-learner is applied to take results from individual learners as inputs to generate an output.

3. Case study

3.1 Data description

Part of the I-35 West Bridge data set, which includes GPS trajectories for travellers in Minneapolis - St. Paul (The Twin Cities) in 2008, is used for the case study. As this study only focuses on the morning commute trips, 4538 real trips from 131 travellers remain after the cleaning process. The Lawrence Group (TLG) road network for the Twin Cities is used, and it includes 108,561 nodes and 277,747 links.

A hybrid method that combines the labelling approach and the link penalty methods is applied to generate a choice set. Finally, on average, 40 unique alternative routes for each OD pair are included in the choice set. More details about the choice set generation methods can be found in our previous study (Wang et.al, 2023). Route attributes like route length, minimum legal travel time, freeway coverage, the number of left turns, the number of traffic lights, the number of bus stops and path size are included. In addition, travellers' age, gender and income are also included.

3.2 Model evaluation

Three criteria from statistics are used for evaluating the performance of all models in this study and are presented in the following:

- Sensitivity (Recall): Sensitivity $= \frac{\text{TP}}{TP+FN}$, which measures the percentage of true positive results in alternative routes which have been selected. TP refers to the chosen routes are correctly predicted. FN refers to the chosen alternative routes that are incorrectly identified as predicted-to-be-taken routes.
- Specificity: Specificity $= \frac{\text{TN}}{\text{TN} + \text{FP}}$, which measures the percentage of true negative results in alternative routes which have not been selected. TN refers to the untaken routes that are correctly identified as predicted-to-be-untaken routes. FP refers to the untaken routes which are incorrectly identified as predicted routes.

- Precision: $\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$, which measures the percentage of true positive results in alternative routes which are predicted to be selected.

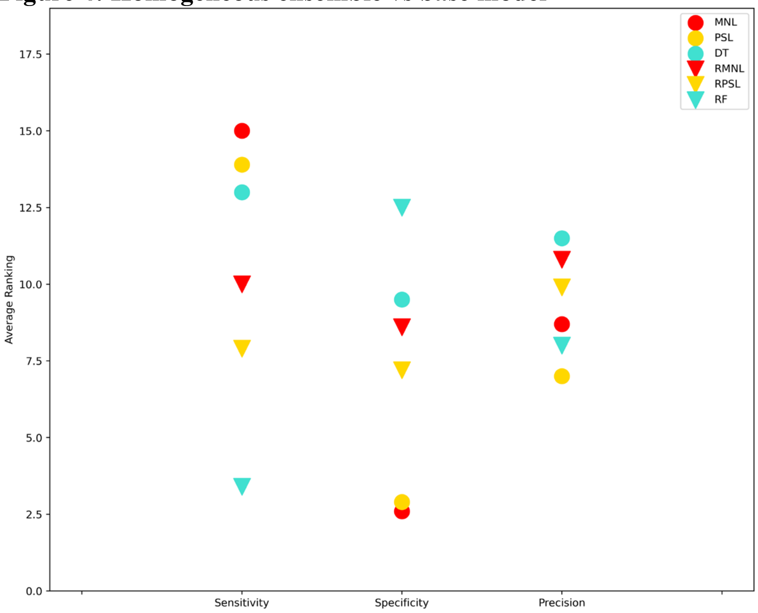
The priority of these metrics are as follows: Sensitivity, Precision and Specificity. Since machine learning approaches are highly dependent on the data itself, 10 validations are performed in this study to obtain a general conclusion. For each cross validation, the training set is randomly taken from the original data set and includes trips from 80% of travellers. At each cross validation, for each metric, all models are ranked independently based on their model performance, and the average rank across the 10 cross validations for each model under each metric is recorded. The higher average rank means this model generally performs better.

4. Preliminary results

4.1 Homogeneous ensemble vs base model

The average rank across the 10 validations for MNL, RMNL, PSL, RPSL, decision tree and random forest model are presented in Figure 4. Clearly, no model dominates in three criteria, and all three homogeneous ensemble models show better performance in sensitivity. The improvement in rank is great, especially for Tree-based models, and ranking order in three base models is consistent with that in their related homogeneous ensemble models. However, for specificity, homogeneous ensemble models do not perform well, and they all have lower ranks compared with their base models. Moving to precision, only RF shows better performance than its base model in general, and the difference between homogeneous ensemble and base models is smaller than that for sensitivity.

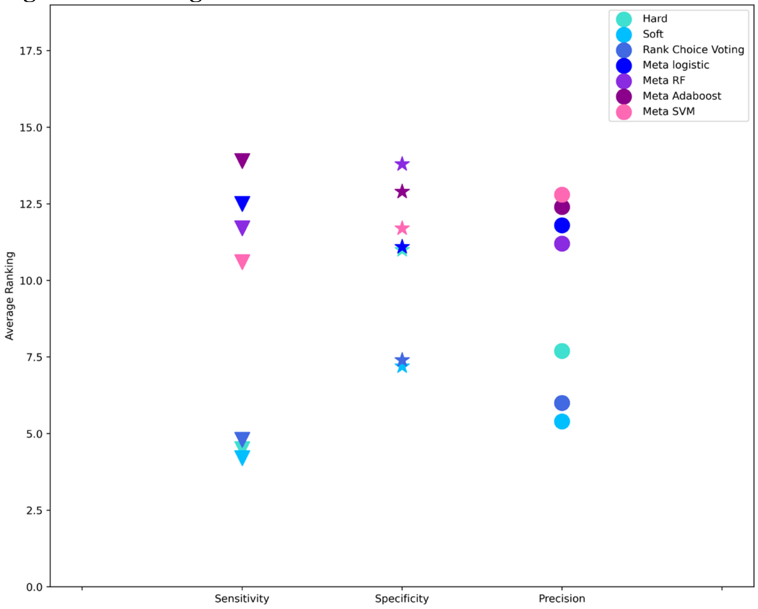
Figure 4: Homogeneous ensemble vs base model



4.2 Heterogeneous ensembles with different ensemble rules

As mentioned in Sub-section 2.2.1, all heterogeneous ensemble models include the same type and number of base models, the difference between them is the ensemble rules, which is the strategy for aggregating results from each individual model to a final ensemble prediction. According to Figure 5, for the heterogeneous ensemble model, soft voting is the best ensemble rule within the tested ensemble rules, which outperforms all three criteria. The voting strategies, in general, perform better than stacking strategies in this case.

Figure 5: Heterogeneous with different ensemble rules



References

McNees, S.K., 1990. The role of judgment in macroeconomic forecasting accuracy. *International Journal of Forecasting*, *6*(3), pp.287-299.

Palmer, T., 2019. The ECMWF ensemble prediction system: Looking back (more than) 25 years and projecting forward 25 years. *Quarterly Journal of the Royal Meteorological Society*, *145*, pp.12-24.

Prato, C.G., 2009. Route choice modeling: past, present and future research directions. Journal of choice modelling, 2(1), pp.65-100.

Sharma, N., Dev, J., Mangla, M., Wadhwa, V.M., Mohanty, S.N. and Kakkar, D., 2021. A heterogeneous ensemble forecasting model for disease prediction. *New Generation Computing*, pp.1-15.

Sun, B. and Park, B.B., 2017. Route choice modeling with support vector machine. Transportation research procedia, 25, pp.1806-1814.

Wang, H., Moylan, E. and Levinson, D., 2023. Route Choice Set Generation on High-Resolution Networks. Transportation Research Record, p.03611981231188775.

Wu, H. and Levinson, D., 2021. The ensemble approach to forecasting: a review and synthesis. Transportation Research Part C: Emerging Technologies, 132, p.103357.

Zhou, Z.H., 2021. Machine learning. Springer Nature.