UCLM Universidad de Castilla-La Mancha

esi Escuela Superior de Informática

A comparative study of machine learning, deep neural networks and random utility maximization models for travel mode choice modelling

ewgt2021
EURO Working Group on Transportation Meeting

José Carlos García-García
Ricardo García-Ródenas
Julio Alberto López-Gómez
José Ángel Martín Baos

Introduction

Discrete Choice Models

Random Utility Theory (RUM)

- Have dominated travel behaviour researches since 1970s.
- Have acquired a high degree of sophistication.
- Are highly interpretable.
- Requires to specify a functional expression beforehand.

Artificial Intelligence

Machine Learning (ML)

- Successful application to many areas.
- Alternative to RUM to model individual behaviour.
- More precision and no functional expression.
- Black-box models: Difficult to interpret.

Introduction

Hillel et al. (2021): The methodologies in the literature are highly fragmented and there are technical limitations which makes difficult to asses properly ML models in choice modelling.

In this study:

- Comprehensive comparison
- Systematic assessment
- Two dataset of a completely different nature
  - > 1.900 obs.
  - > 230.000 obs.

[1] Hillel et al. 2021. A systematic review of machine learning classification methodologies for modelling passenger mode choice. Journal of Choice Modelling.

Related work

In existing literature, RUM and ML methods are compared from two points of view:

- Behavioural interpretation in a context of discrete choice modelling.
- Assessment of the performance of the models.

Reference	Domain	RUM	Neural networks			Single classifiers				Ensembles			
		MNL	NN	DNN	CNN	KNN	SVM	NB	CART	BOOST	BAG	RF	KF
Zhao et al. (2018)	Travel mode choice	x	x				x	x	x	•	•	•	
Lhéritier et al. (2019)	Airline itinerary choice modeling	x										•	
Hagenauer and Helbich (2017)	Travel mode choice	x	x				•	x		x	x	•	
Omrani (2015)	Travel mode choice	x	•				x						
Ballings et al. (2015)	Stock price direction	x	x			x	•			x		•	x
This study	Travel mode choice	x	x	•	•		x					•	

[2] Zhao et al. 2018. Modeling Stated Preference for Mobility-on-Demand Transit: A Comparison of Machine Learning and Logit Models.
[3] Lheritier et al. 2019. Airline itinerary choice modeling using machine learning. Journal of Choice Modelling.
[4] Hagenauer and Helbich 2017. A comparative study of machine learning classifiers for modeling travel mode choice. Expert Systems with Applications.
[5] Omrani 2015. Predicting travel mode of individuals by machine learning. Transportation Research Procedia.
[6] Ballings et al. 2015. Evaluating multiple classifiers for stock price direction prediction. Expert Systems with Applications.

Methodology: Datasets

OPTIMA

- Revealed preferences survey to Swiss people from 2009 to 2010.
- 1124 surveys with 115 variables → 1906 trips.
- After pre-processing, 7 variables selected.

 28%  66%  6%

NTS

- ML focused dataset containing:
  - Data from a Dutch transport survey from 2010 to 2012.
  - Environmental data.
- 230.608 surveys with 16 variables.
- After pre-processing, 100.000 trips where randomly selected.

 4%  55%  24%  17%

Methodology: Methods

Multinomial Logit Model (MNL)

- Utility functions: $U_{in} = V_{in} + \epsilon_{in}$
- Stochastic part determines the probability of alternative i: $P_{in} = \frac{\exp(V_{in})}{\sum_{j \in C} \exp(V_{jn})}$
- Deterministic part:
  - For NTS dataset: $V_{in} = \beta_i^T \mathbf{x}_{in}$
  - For OPTIMA: Bierlaire (2018)

$$
\begin{aligned}
V_{PT} = {} & \beta_{\text{Time\_PT}} * \text{TimePT} + \beta_{\text{Cost\_PT}} * \text{MarginalCostPT} + \\
& \beta_{\text{Fulltime\_PT}} * \text{Fulltime} + \beta_{\text{Man\_PT}} * \text{Man} + \beta_{\text{Woman\_PT}} * \text{Woman} + \beta_{\text{Unreported\_PT}} * \text{Unreported}
\end{aligned}
$$

$$
\begin{aligned}
V_{Car} = {} & \beta_{\text{ASC\_Car}} + \beta_{\text{Time\_Car}} * \text{TimeCar} + \beta_{\text{Cost\_Car}} * \text{CostCarCHF} + \\
& \beta_{\text{Fulltime\_Car}} * \text{Fulltime} + \beta_{\text{Man\_Car}} * \text{Man} + \beta_{\text{Woman\_Car}} * \text{Woman} + \beta_{\text{Unreported\_Car}} * \text{Unreported}
\end{aligned}
$$

$$
\begin{aligned}
V_{SM} = {} & \beta_{\text{ASC\_SM}} + \beta_{\text{Dist}} * \text{distance\_km} + \\
& \beta_{\text{Fulltime\_SM}} * \text{Fulltime} + \beta_{\text{Man\_SM}} * \text{Man} + \beta_{\text{Woman\_SM}} * \text{Woman} + \beta_{\text{Unreported\_SM}} * \text{Unreported}
\end{aligned}
$$

[7] Bierlaire 2018. Mode Choice in Switzerland (Optima). Technical Report. Transportation Center (EPFL)

Methodology: Methods

Neural Network (NN)

- Multilayer Perceptron (MLP) is a widely used NN in classification problems.
- A MLP with 1 hidden layer can model any non-linear relationship between the input variables and the target.
- Backpropagation algorithm is used to minimise the log-loss function.
- The output of the model is a vector of probabilities per alternative:

$$P(y_k|x) = s'\left\{\sum_{j=1}^{n_2} \omega_{jk} \cdot s\left\{\sum_{i=1}^{n_1} \omega_{ij} \cdot x_i + b_{0j}\right\} + b_{0k}\right\}$$

- ω: weights
- s and s': activation functions (ReLU)
- n_1 and n_2: number of neurons in the layer

Methodology: Methods

Deep Neural Network (DNN)

- Consists in adding multiple hidden layers to a MLP.
- It improves the predictive capability on problems with non-structure data or high non-linearities.
- We have used a DNN model with 3 hidden layers of 64 neurons.

Convolutional Neural Network (CNN)

- They are based on the concept of filter (kernel).
- They are especially effective where features from different abstraction levels must be extracted.
- The kernels are two-dimensional in the case of images. In this study, since we work with matrix data, we apply a one-dimensional convolution.
- We have used a CNN model with 3 one-dimensional convolution layers of 64 kernels with a size of two units.

Methodology: Methods

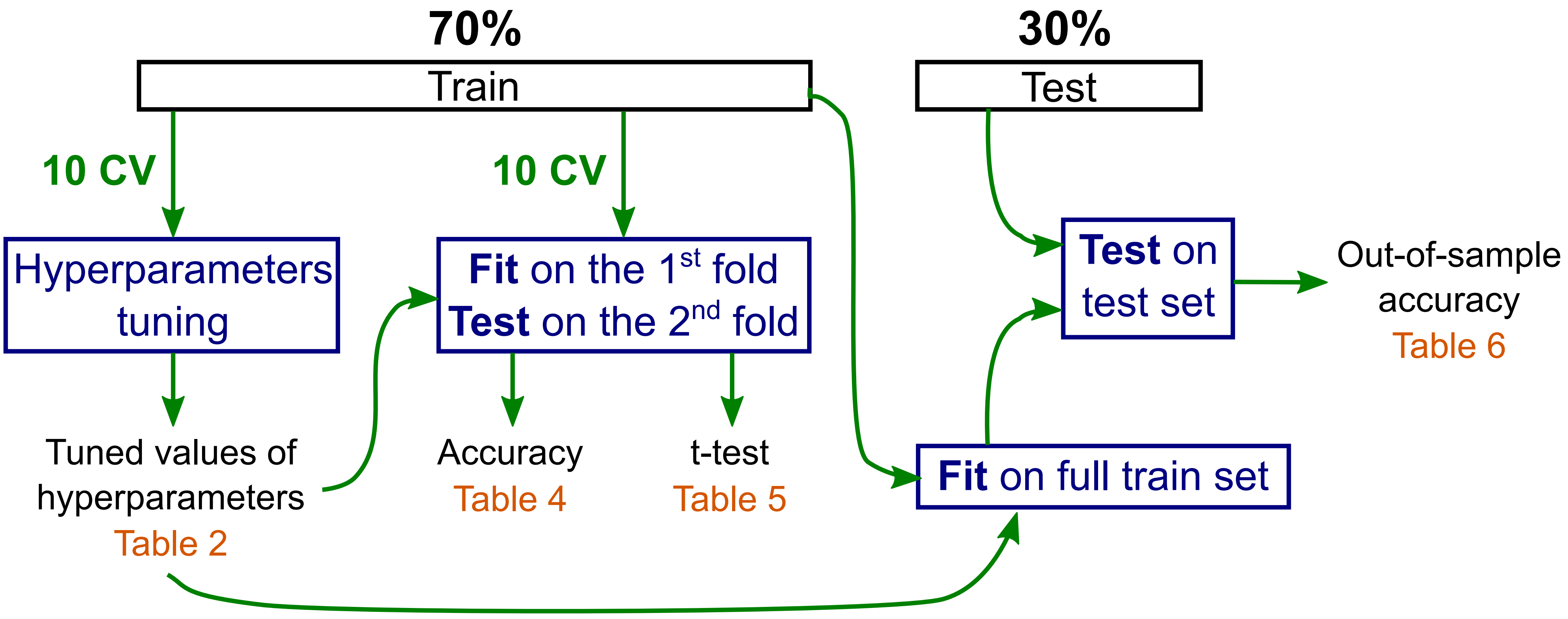
Random Forest (RF)

- Consists on tree-like data structures used for classification tasks.
- Each node of the tree represent a binary decision. The leaves are the alternatives.
- A RF is an ensemble composed of several different trees, using a subset of the features for each tree to improve the accuracy.

Support Vector Machine (SVM)

- A binary SVM assumes that the data can be labelled as $y_n \in \{-1,\ 1\}$.
- Then, it builds a decision function $f(x) = \mathrm{sgn}(\sum_{n=1}^{N} y_n \alpha_n K(x_n, x) + \rho)$ where $K(x_n, x)$ is a kernel function. We apply the RBF kernel.
- Then, the vector α_n is estimated.
- For multiclass problems (I alternatives), we estimate $\frac{I(I-1)}{2}$ binary SVM.

Methodology



Methodology: Hyperparameters

- We formulate a Hyperparameter Optimization problem:

$$\lambda^* = \operatorname*{argmin}_{\lambda \in \Lambda} \mathbb{E}_{(D_{train}, D_{valid}) \sim \mathcal{D}} \mathbf{V}\left(\mathcal{A}_\lambda, D_{train}, D_{valid}\right),$$

- Random search with 1.000 iterations with 10 CV to estimate $\mathbf{V}$.

Technique $\mathcal{A}$	Name of the hyperaparameter	Notation	Type	Domain	OPTIMA	NTS
SVM	The parameter of the Gaussian function	σ	Continuous	$[10^{-4}, 1]$	0.145	0.031
	The cost or also called soft margin constant	C	Continuous	$[1, 10^3]$	10.678	10.973
RF	Number of decision trees	B	Discrete	[2, 200]	192	186
	Max features	m	Discrete	[2, N features]	6	3
NN	Size of hidden layer	P	Discrete	[10, 500]	11	423
	Initial learning rate	η_n	Continuous	$[10^{-4}, 1]$	0.072	0.044
DNN	Epochs	*epochs*	Discrete	[50, 200]	186	198
	Batch size	*BS*	Discrete	[1, DatasetRows]	393	951
CNN	Epochs	*epochs*	Discrete	[50, 200]	190	120
	Batch size	*BS*	Discrete	[1, DatasetRows]	78	200

Methodology: Model comparison

- There is no golden standard for comparing classifiers.
- The most widely used index is classification accuracy (Demsar, 2006).
- In most of studies on transport mode choice, only one dataset is used.
- The standard way of assessing the classifiers on a single dataset is using cross-validation (CV).

- In this paper we follow Dietterich (1998) and propose a 5x2 CV over each dataset.
- Moreover, this methodology allows us to apply a t-test to the results (Dietterich, 1998).

---

[8] Demsar 2006. Statistical Comparisons of Classifiers over Multiple Data Sets. Journal of Machine Learning Research.
[9] Dietterich 1998. Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. Neural Computation.

Results

- All the experiments have been implemented in Python.
- RF, SVM, and NN methods have been executed using scikit-learn package.
- For DNN and CNN we have apply Keras Python library.
- MNL model has ben implemented on PyKernelLogit (Martín-Baos, 2019).
- Finally, hyperopt package was used to tune hyperparameters.

- To address class imbalance, a re-sampling procedure was applied to NTS dataset, which is a common procedure in ML.

[10] Martín-Baos 2019. Design and implementation of a software library for the estimation and analysis of non-parametric discrete choice models. Application to transport planning. Technical Report. Universidad de Castilla-La Mancha.

Results

Dataset		MNL	RF	SVM	NN	DNN	CNN
OPTIMA	Accuracy	0.713	0.753	0.746	0.742	0.757	0.744
	95% CI	[0.683 , 0.743]	[0.732 , 0.775]	[0.725 , 0.767]	[0.708 , 0.776]	[0.742 , 0.772]	[0.701 , 0.787]
	CPU time (s)	0.178	0.463	0.013	0.088	1.757	6.513
	95% CI	[0.159 , 0.196]	[0.402 , 0.524]	[0.012 , 0.015]	[0.037 , 0.138]	[1.45 , 2.064]	[6.037 , 6.99]
NTS	Accuracy	0.531	0.687	0.558	0.593	0.580	0.590
	95% CI	[0.526 , 0.536]	[0.682 , 0.691]	[0.551 , 0.565]	[0.527 , 0.658]	[0.556 ,0.604]	[0.572 , 0.607]
	CPU time (s)	16.905	1.167	48.922	10.742	35.127	134.065
	95% CI	[9.849 , 23.961]	[1.166 , 1.169]	[47.881 , 49.962]	[5.427 , 16.057]	[33.924 , 36.33]	[127.653 , 140.478]

Results: Significance t-test

P-value and significance t-test results

Dataset		MNL	RF	SVM	NN	DNN	CNN
OPTIMA	**MNL**		0.000	0.000	0.001	0.000	0.002
	RF	***		0.164	0.114	0.460	0.270
	SVM	***			0.553	0.024	0.810
	NN	**				0.032	0.822
	DNN	***		*	*		0.120
	CNN	**					
NTS	**MNL**		0.000	0.000	0.000	0.000	0.000
	RF	***		0.000	0.000	0.000	0.000
	SVM	***	***		0.007	0.000	0.000
	NN	***	***	**		0.300	0.806
	DNN	***	***	***			0.066
	CNN	***	***	***			

Note: ***<0.001, **<0.01, *<0.05

Results: Out-of-sample accuracy

70%	30%
Train	Test

Dataset	MNL	RF	SVM	NN	DNN	CNN
OPTIMA	0.739	0.762	0.752	0.746	0.768	0.771
NTS	0.531	0.721	0.566	0.566	0.568	0.585

Conclusions

- The ranking of models is similar in both dataset.
- The highest difference in accuracy in OPTIMA is between MNL and DNN (3.2%)
  - However, in NTS, the highest difference is between MNL and RF (19%).
  - This shows that on datasets designed for RUM models, MNL can achieve a better performance than on ML ones.

- We have shown than RF is the best classifier in terms of accuracy and computational cost.

- The classifiers act in a naive way when the data is not balanced on NTS dataset, predicting only the majority classes and achieving a fictious better accuracy.

- Finally, we evidence the need for other indicators such as the recall of the travel modes, as well as the capability of the model to provide behavioural insights.

Universidad de Castilla-La Mancha
UCLM

Escuela Superior de Informática
esi

A comparative study of machine learning, deep neural networks and random utility maximization models for travel mode choice modelling

ewgt2021
EURO Working Group on Transportation Meeting

José Carlos García-García
Ricardo García-Ródenas
Julio Alberto López-Gómez
José Ángel Martín Baos