Attitudes and Latent Class Choice Models using Machine Learning

Lorena Torres Lahoz*[1] and Francisco Camara Pereira[1], Georges Sfeir[1], Ioanna Arkoudi[1], Mayara Moraes Monteiro[1], Carlos Lima Azevedo[1]

[1]DTU Management, Technical University of Denmark

Short summary

We present a method of efficiently incorporating attitudinal indicators in the specification of Latent Class Choice Models (LCCM), extensions of Discrete Choice Models (DCMs) that segment populations based on the assumption of preference similarities. We introduce Artificial Neural Networks (ANN) to formulate the latent variables constructs. This formulation overcomes structural equations in its ability to explore the relationship between the attitudinal indicators and the decision choice, given the machine learning (ML) flexibility and power to capture unobserved and complex behavioural features, such as attitudes and beliefs. All of this, while maintaining the consistency of the theoretical assumptions presented in the Generalized Random Utility model and the interpretability of the estimated parameters. We test our proposed framework for estimating a car-sharing service subscription choice with stated preference data. The results show that our proposed approach provides a complete and realistic segmentation, which helps design better policies.
Keywords: Car-sharing, Discrete choice modelling, Machine learning, Psychometric Indicators.

1 Introduction

This study explores a new method of efficiently incorporating attitudinal indicators in the specification of LCCM by relying on ML techniques while preserving the benefits of the economic and behavioural interpretability of DCMs.
Walker & Ben-Akiva (2002) presented a practical generalized random utility model with extensions for latent variables and classes. They extended the Random Utility Model (RUM) to relax its assumptions and enrich the model's capabilities. They refer to latent classes as unobserved population groups, in which each individual has an associated probability of belonging to each group/class on the assumption of preference similarities. On the other hand, psychometric indicators measure the effect of unobserved attributes on individuals' preferences on topics related to the choice and they are additional information that helps specify and estimate latent classes.
Atasoy & Bierlaire (2011) estimated an LCCM where psychometric indicators are included in the maximum likelihood estimation to improve the model's accuracy. The psychometric indicators were modelled, conditional on the latent class, as parameters jointly estimated with the choice and the class membership model. The model showed that the psychometric indicators allow for richer analysis and generate significantly different class membership estimates. In another approach, Hurtubia et al. (2014) introduced psychometric indicators by computing the probability of giving an agreement level to an attitudinal statement as an ordinal logit, also dependent on the individual class. However, complex interactions between attitudinal variables and the decision-making process should be expected Bahamonde-Birke et al. (2017). We hypothesise that ML could be a good starting point to explore such interactions, given its flexibility and power in capturing unobserved and complex interactions.
In recent years, the use of ML techniques has increased, mainly due to their power to improve prediction accuracy. However, one of the main critiques of ML techniques in contrast to econometric models, is that they tend to generate less interpretable results. Thus, transportation researchers have focused on providing meaningful estimates from ML applications, that can be useful for travel analysis and policy decisions. For example, Arkoudi et al. (2021) proposed an embedding encoding for the socio-characteristic variables that provided a latent representation of these variables in concordance with individuals' choices. Han (2019) included a nonlinear LCCM using a neural network to specify the class membership model. Their model outperformed the traditional ones

in prediction accuracy with the trade-off of losing some interpretability. Sfeir et al. (2021, 2022) presented two model formulations for the construct of latent class choice models using Gaussian process and Mixture models. All these works employed ML in DCMs to allow for more flexibility in the definition of the latent constructs. However, there is still a lack of effective use of these techniques for incorporating attitudinal information into the model formulation.

2 Methodology

We follow the generalized RUM structure presented by Walker & Ben-Akiva (2002) for interpretability purposes and we include the information on the attitudinal indicators by employing an ANN to formulate with greater flexibility the latent variables. Figure 1 shows the graphical representation of the proposed formulation.

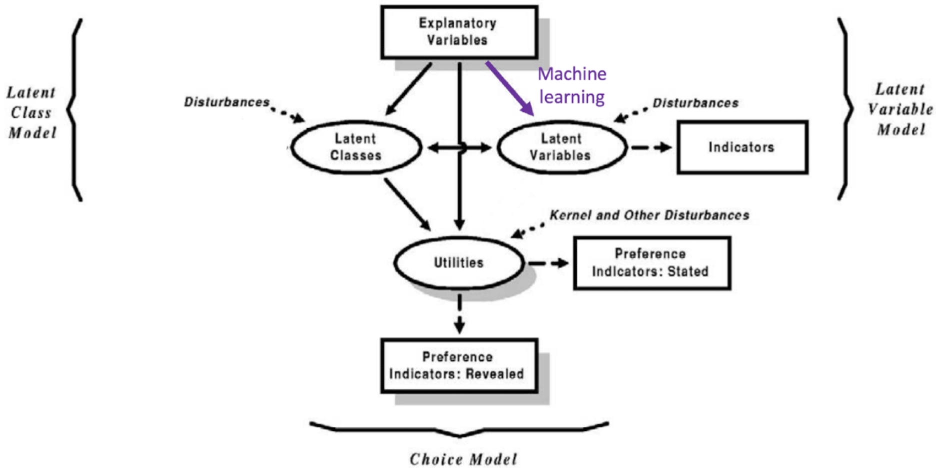

Figure 1: Graphical representation of the model formulation

LCCMs are composed of two sub-models: *a class membership model* and a *class-specific choice model*. The former computes the probability of an individual n belonging to a certain class, while the latter assigns the probability of choosing each alternative, given that individual n belongs to a certain class k.
The utility of the class membership model can be written as:

$$U_{nk} = V_{nk} + \upsilon_{nk} \tag{1}$$

where V_{nk} is the representative utility of individual n belonging to class k and υ_{nk} is the error term that is assumed to be independent and identically distributed (iid) Extreme Value Type I over individuals and classes. In this case, we define V_{nk} as:

$$V_{nk} = ASC_k + Q_n\gamma_k + r_n\delta_k + \omega_n b_k \tag{2}$$

where ASC_k is the alternative-specific value for class k, Q_n is the vector containing socio-characteristics of individual n, and γ_k the vector of unknown parameters that need to be estimated for each class k. In addition, r_n is a vector of length Z containing the latent variables for individual n and δ_k the corresponding vector of unknown parameters specific to class k. Finally, ω_n is an individual-specific constant with its corresponding coefficient b_k for each class k. It represents the individual variation of all the latent variables caused by the variance of their underlying distributions. It is formulated as a one-layer ANN that gets activated by the ID of each individual in the train set (Id_n is one for individual n and 0 otherwise),

$$\omega_n = \sum_{1}^{N} w_{1n}^{(1)} Id_n \tag{3}$$

where $w_{1n}^{(1)}$ are the weights of the layer.
Given the distribution of the error term (υ_{nk}), the probability $P(q_{nk}|Q_n, \gamma_k, r_n, b_k)$ can be expressed as:

$$P(q_{nk}|Q_n, \gamma_k, r_n, b_k) = \frac{e^{V_{nk}}}{\sum_{k'=1}^{K} e^{V_{nk'}}} \tag{4}$$

The novelty of this work is the employment of ANN for the construction of latent variables. We propose a non-linear relationship between the socio-characteristics of the individuals and the latent constructs by employing two densely connected layers:

$$r_{zn} = a_2(\sum_{h=0}^{H} w_{zh}^{(2)} a_1(\sum_{m=0}^{M} w_{hm}^{(1)} Q_{mn})) \tag{5}$$

where M is the number of socio-characteristic variables used to predict the answer to the indicators, and H is the number of hidden units in the hidden layer. $w_{hm}^{(1)}$ are the weights of the first layer, and a_1 represents the first activation function defined as a Rectified Linear Unit (ReLU) ($a_1(x) = max(0, x)$); for the second layer, a linear activation function is applied $a_2(x) = x$, and the weights are represented by $w_{zh}^{(2)}$. By adding an extra input Q_{0n}, which is set to one and extending the sum to go from zero, we avoid writing the intercept term.
The number of latent variables Z, the number of hidden neurons in the hidden layer H, and the number of densely connected layers should be tuned since they are not observed in the data.
The formulation presented is based on the hypothesis that the socio-characteristics of the individuals define the latent variables. Moreover, these latent constructs influence the response to specific attitudinal indicators. We focus on the case where indicators take the form of statements that receive an ordered response, in Likert (1932) scale. Thus, we define the utility of individual n for indicator p, as a measurement of the level of agreement with the statement, and we formulated it as:

$$U_{pn} = V_{pn} + \nu_{pn} = r_n \alpha_p + c_p \omega_n + \nu_{pn} \tag{6}$$

where V_{pn} is the representative utility of individual n to indicator p and ν_{pn} is the error term that is assumed to be iid Extreme Value Type I over individuals and indicators. r_n is a vector of length Z containing the latent variables of individual n, α_p is the vector of corresponding parameters to be estimated. ω_n is the individual-specific parameter estimated together with the latent class model, and c_p is its corresponding coefficient for each indicator p .
Therefore, the probability that individual n answers with a certain level of agreement l to indicator p is expressed as:

$$P(I_{pln} = 1|r_n, \alpha_p, c_p, \omega_n) = P(\tau_{l-1}^p < U_{pn} < \tau_l^p) \tag{7}$$

where we define I_{pln} as 1 if individual n answers with a level of agreement l to indicator p and 0 otherwise. τ_l^p are strictly increasing class-specific thresholds that define an ordinal relation between the utility U_{pn} and the level of agreement to indicator p.
The probability of individual n providing an answer l to indicator p can be computed as an ordinal softmax:

$$\begin{aligned} P(I_{pln} = 1|r_n, \alpha_p, c_p, \omega_n) = P(\tau_{l-1}^p < U_{pn} < \tau_l^p) =& P(\tau_{l-1}^p < V_{pn} + \nu_{vp} < \tau_l^p) = \\ = Prob(\nu_{vp} < \tau_l^p - V_{pn}) - P(\nu_{vp} < \tau_{l-1}^p - V_{pn}) = =& \frac{e^{\tau_l^p - V_{pn}}}{1 + e^{\tau_l^p - V_{pn}}} - \frac{e^{\tau_{l-1}^p - V_{pn}}}{1 + e^{\tau_{l-1}^p - V_{pn}}} \end{aligned} \tag{8}$$

where one threshold per indicator is set to zero, as only the difference between them matters.
We estimate all components of the proposed model simultaneously by employing the EM Dempster et al. (1977) algorithm, which combines an expectation step with a maximization one until convergence is reached. The final model architecture is presented in Figure 2.

3 Results and discussion

We test the model on a dataset from a 2020 tailor-made online survey in Copenhagen (CPH). Respondents needed to be at least 18 years old and have a valid driver's license. The sample consists of 542 complete answers from which 80% are used for training and 20% for testing. The relevant parts employed in the estimation include:

1. A survey on the respondent's socio-characteristic characteristics

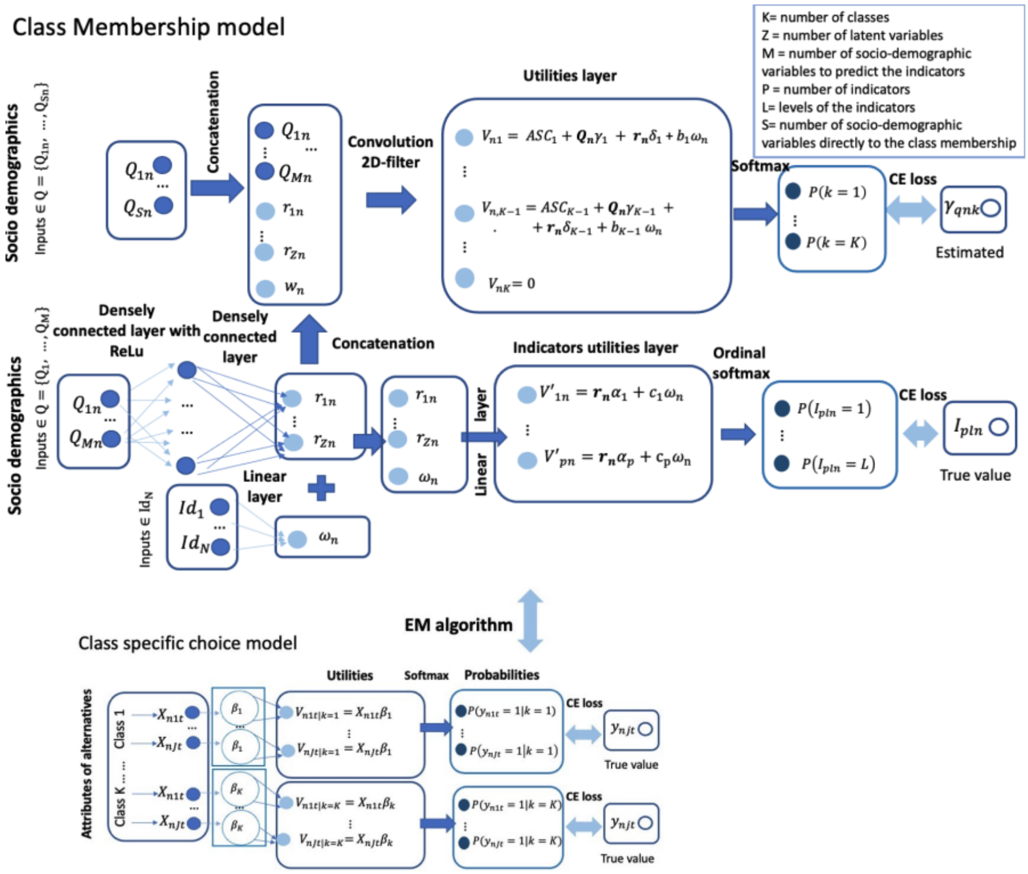

Figure 2: Model arquitecture

2. A survey addressing questions regarding respondents' attitudes toward private and car-sharing (CS) using a 5-point Likert (1932) scale

3. A Stated Preference (SP) experiment with different options for CS plans

For further details on the data, the reader is referred to Frenkel et al. (2021).

Baseline Results

To benchmark the proposed model, we tried to follow the formulation from Walker & Ben-Akiva (2002) which can be shown as a graphical representation (Figure 2), with the difference that the relation between the indicators' utility and the individuals socio-characteristic characteristics is linear. However, the class membership formulation's simplicity made the model unable to converge, resulting in a non-invertible Hessian matrix. Therefore, to get a comparable magnitude for the likelihood, we estimated the model with a traditional LCCM as a baseline, where the representative part of the class membership utility is just a linear combination of the socio-characteristic variables. Results are presented in Table 1.

Table 1: LCCM results without attitudinal variables

Model	Nº Classes	Nº parameters	Null LL	LL	AIC	BIC	R-squared	Test null LL	Test LL
LCCM	2	30	-2047.21	-1599.41	3258.81	3413	0.22	-515.02	-400.52
LCCM	3	43	-2047.21	-1568.14	3234.29	3487	0.23	-515.02	-398.73

We employ the same socio-characteristics for constructing the class membership as in our proposed formulation. More specifically, we include age, binary variables that indicate if the individual has a bike, a car or kids at home and if they are students, retired or CS members. However, we found that having a car at home, being retired or being a student, were not statistically significant under

the LCCM formulation. Instead, our proposed formulation allows us to significantly include this information in the model, improving our characterization of the latent variables.
Given the probability of each individual in the sample and its corresponding socio-characteristics, we have represented the classes from the baseline model in Figure 3, by using the Bayes Theorem to compute:

$$P_n(socio - characteristic|K = k) \tag{9}$$

Figure 3 is compared in the next subsection with the proposed model results.

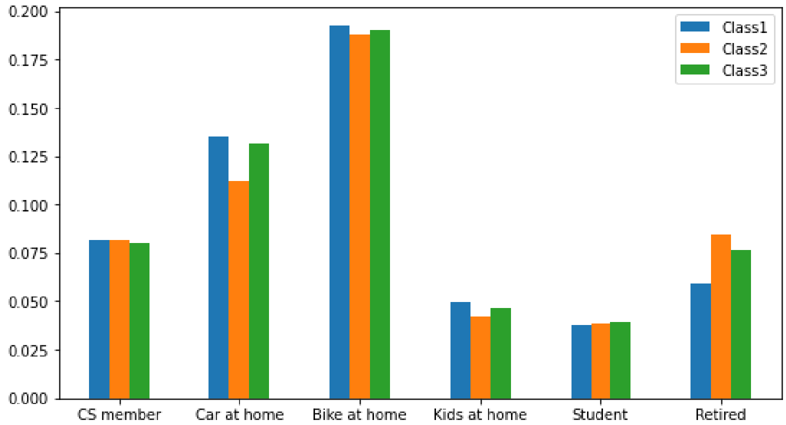

Figure 3: Representation of the class membership of the LCCM model

Proposed Model Results

We have employed the same train/test split as for the baseline model. The EM process has been estimated multiple times with random initializations. We have computed the likelihood variance between the different model estimates to check for stability. The results are summarized in Table 2.

Table 2: Model results

Nº Classes	Nº latent variables	Iterations	Null LL	LL	Variance LL	R-squared	Test null LL	Test LL	Variance Test LL
2	2	15	-2047.21	-1575.32	14.62	0.23	-515.02	-404.72	0.28
3	**2**	**30**	**-2047.21**	**-1539.56**	**22.01**	**0.25**	**-515.02**	**402.72**	**9.8**
3	3	25	-2047.21	-1531.41	26.36	0.25	-515.02	-400.73	9.36

The model with three latent classes and two latent variables is selected as the best model. The one with three classes and three latent classes has a slightly better fit, however, its corresponding latent variables parameter estimates were not statistically significant.
Comparing the results from tables 1 and 2, we observed an increase in the training likelihood for our formulation. We do not provide better results for the test data, but just comparable ones. This could be due to the small size of the test sample and/or to the fact that we don't have access to attitudinal information or w_n values in the test stage, which affect prediction accuracy.
Table 3 shows the estimated parameters of the class-specific choice model with their corresponding standard deviations, where all the data has been used for the estimation. The utility for not choosing any of the CS services is set to zero due to parameters' identification.
Based on the values and signs of the estimated beta parameters, we observe that class 1 and class 2 are more negatively affected by the subscription cost, while class 3 is less influenced by this cost, but more negatively affected by the usage cost. Moreover, individuals with a high probability of belonging to class 3 are the most concerned if the type of engine is combustion. Thus, CS could be seen as an electric alternative for them. Given the beta values for displaying the cost in hours ($\beta_{Usage\ cost\ per\ hour}$) or days ($\beta_{Usage\ cost\ per\ day}$), there is a bias towards displaying the price per minute (baseline), related to the fact that CS users tend to drive for short time periods. Regarding the probability of finding a car, it is a more important feature for classes 2 and 3, which make them more dependent on the availability of the service. Overall, class 2 seems to be less prone to use any CS (including P2P), given all its estimated parameters.

Table 3: Estimate and standard deviation of the parameters of the class-specific choice model

Variable	Class specific choice model		
	Class 1	Class 2	Class 3
$ASC_{CS\ free-floating}$	3.50(0.46)	-2.71(1.34)	-1.68(0.89)
$ASC_{CS\ station-based}$	3.04(0.48)	-2.64(1.34)	0.07(0.80)
$ASC_{CS\ peer\ to\ peer}$	4.12(0.52)	0.47(1.45)	1.50(0.83)
$ASC_{roundtrip}$	2.97(0.48)	-3.50(0.40)	0.15(0.80)
$\beta_{One\ time\ subscription\ cost}$	-1.00(0.17)	-1.12(0.51)	-0.35(0.22)
$\beta_{Usage\ cost(OWFF,OWST,RT)}$	0.05(0.04)	-0.16(0.09)	-0.34(0.08)
$\beta_{Usage\ cost(P2P)}$	-1.30(0.39)	-5.57(1.31)	-3.86(0.80)
$\beta_{Usage\ cost\ per\ day}$	-0.35(0.24)	-2.40(0.74)	-1.09(0.39)
$\beta_{Usage\ cost\ per\ hour}$	-0.10(0.21)	-0.99(0.50)	-0.43(0.34)
$\beta_{Only\ combustion\ cars}$	-0.23(0.12)	0.09(0.33)	-0.66(0.20)
$\beta_{Probability\ of\ finding\ a\ shared\ car}$	0.13(0.44)	1.38(1.35)	2.70(0.75)
$\beta_{Walking\ time\ from\ parking\ to\ destination}$	-0.05(0.01)	0.02(0.04)	0.03(0.02)

Table 4: Parameters of the class membership model

Variable	Parameter	St error	P-value
ASC_{class_1}	0.97	0.45	0.031
ASC_{class_2}	-1.40	0.51	0.0057
$\gamma_{kidsathome,class_1}$	0.56	0.29	0.050
$\gamma_{kidsathome,class_2}$	-0.45	0.37	0.22
$\delta_{r_1,class_1}$	0.18	0.13	0.17
$\delta_{r_1,class_2}$	-0.48	0.14	0.0009
$\delta_{r_2,class_1}$	0.12	0.10	0.27
$\delta_{r_2,class_2}$	-0.44	0.11	0.00
b_{class_1}	0.61	0.44	0.17
b_{class_2}	-4.014	0.54	0.00

The parameters of the class membership model are summarised in Table 4. Given the probability of each individual in the sample, we have characterised the classes in Figure 4. Individuals with a higher probability of belonging to class 1 have around 20% probability of being a CS member, a bit above the sample average (17.5%). They also tend to have more kids at home, as well as bikes than other classes. Studies like Uteng et al. (2019) have shown that when there are significant life changes (e.g., birth of a child), people become more inclined to use CS. In opposition, class 2 presents the lowest probability of being a CS member and having kids or bikes at home. Retired people tend to have more predisposition for this class, while students have less. This is aligned with Prieto et al. (2017) which suggested that young people are more prone to use this service. Finally, class 3 has the same probability of being a CS member as class 1, but it also has a lower probability of owning a car, making people more reliant on the service's availability. Comparison between Figures 3 and 4, shows that the configuration of the classes changes when we include attitudinal information, as it is expected.

By analysing the parameters for the latent variables in Table 4 and looking at their distributions over individuals in Figure 5, we notice that the values of the first latent variable (r_1) are always negative. The more negative value of r1, the more probable is to belong to class 2, and therefore, the less inclined people are to use CS services. A negative value of r_2 seems to have the same effect. Thus, individuals with a more negative combination of r_1 and r_2 tend to be less inclined about CS and the other way around. Figure 5 suggests that students are more prone to use the service while retired people are the least predisposed. Moreover, having or not having a car seems to determine the clusters in which the r_s values are structured. Finally, Figures 6 and 5 show that people with a car at home agree more with the statement that the car is a status symbol. For indicator 15, people with a more positive value of r_2 seem to agree more with the statement that

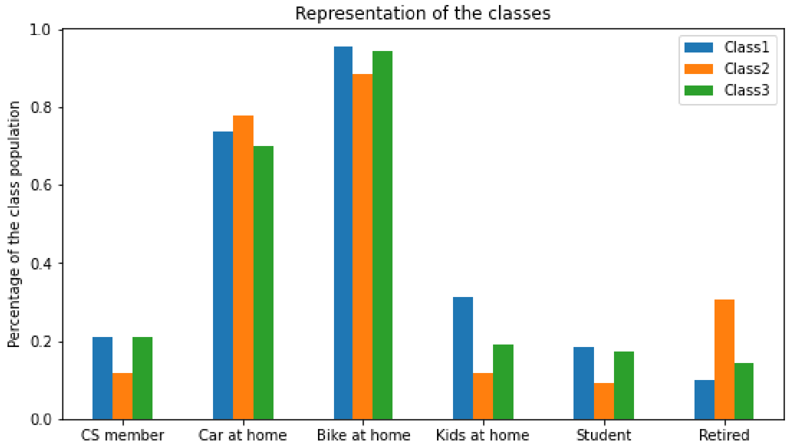

Figure 4: Representation of the class membership of our proposed model

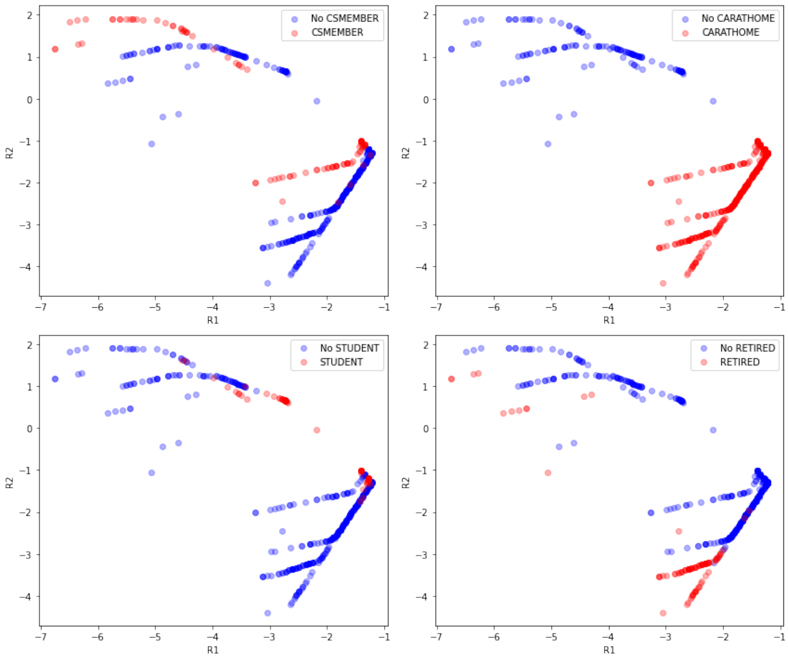

Figure 5: Latent variables representation

they wouldn't need a car if they have CS, as we would expect given the r_2 coefficients of Table 4.

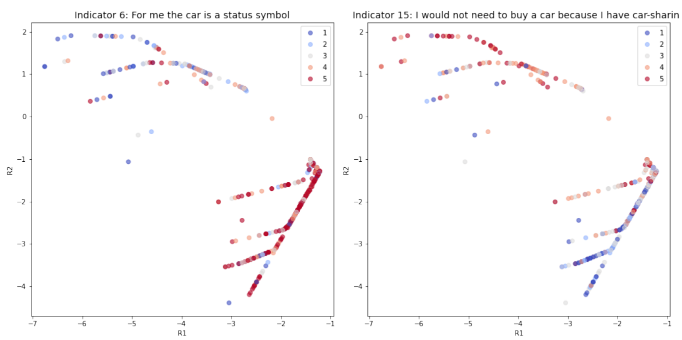

Figure 6: Latent variables representation characterized by the answers to indicators 6 and 15

4 Conclusions

Our results suggest that the inclusion of attitudinal variables provides a DCM that is more behaviorally realistic. For example, individuals who are more inclined towards the concept of CS tend to be grouped together in clusters with higher parameter estimates of the utility of choosing CS plans. This indicates that beliefs and attitudes play a key role in decision-making, and including this information allows for more accurate estimation and a better understanding of the classes that help design better policies.
Within the limitations, convergence is defined empirically by setting the number of iterations due to small fluctuations in the convergence of the EM algorithm. In addition, given the small sample size, we could not divide the dataset in training, validation, and testing; therefore, the hyperparameters of the ANN were not tuned according to the validation samples. This could be solved by employing a bigger dataset. Moreover, to improve the prediction performance, other types of explainable AI (e.g., SHAP) could be explored.
Although the limitations, we are optimistic that this analysis has opened the door to future research on integrating attitudinal variables in DCMs through ML techniques.

Acknowledgements

We thank Share-More project (https://eitum-sharemore.net.technion.ac.il) for providing the data used in this study

References

Arkoudi, I., Azevedo, C. L., & Pereira, F. C. (2021). Combining discrete choice models and neural networks through embeddings: Formulation, interpretability and performance. Retrieved from `https://arxiv.org/abs/2109.12042`

Atasoy, G. A., B., & Bierlaire, M. (2011). Mode choice with attitudinal latent class: a swiss case-study. *Proceedings of the Second International Choice Modeling Conference (ICMC).*

Bahamonde-Birke, F. J., Kunert, U., Link, H., & Ortúzar, J. d. D. (2017). About attitudes and perceptions: finding the proper way to consider latent variables in discrete choice models. *Transportation*, *44*, 475–493. doi: https://doi.org/10.1007/s11116-015-9663-5

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, *39*(1), 1–38. Retrieved 2023-02-08, from `http://www.jstor.org/stable/2984875`

Frenkel, A., Shiftan, Y., Gal-Tzur, A., Tavory, S., Lerner, O., Antoniou, C., ... Peretz, D. (2021). Share more: Shared mobility rewards - summary report.

Han, Y. (2019). *Neural-embedded discrete choice models* (Unpublished doctoral dissertation).

Hurtubia, R., Nguyen, M. H., Glerum, A., & Bierlaire, M. (2014). Integrating psychometric indicators in latent class choice models. *Transportation Research Part A: Policy and Practice*, *64*, 135-146. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0965856414000755` doi: https://doi.org/10.1016/j.tra.2014.03.010

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, *22*(140).

Prieto, M., Baltas, G., & Stan, V. (2017). Car sharing adoption intention in urban areas: What are the key sociodemographic drivers? *Transportation Research Part A: Policy and Practice*, *101*, 218-227. doi: https://doi.org/10.1016/j.tra.2017.05.012

Sfeir, G., Abou-Zeid, M., Rodrigues, F., Pereira, F., & Kaysi, I. (2021). Latent class choice model with a flexible class membership component: A mixture model approach. *Journal of Choice Modelling*, *41*.

Sfeir, G., Rodrigues, F., & Abou-Zeid, M. (2022). Gaussian process latent class choice models. *Transportation Research. Part C: Emerging Technologies*, *136*.

Uteng, T. P., Julsrud, T. E., & George, C. (2019). The role of life events and context in type of car share uptake: Comparing users of peer-to-peer and cooperative programs in oslo, norway. *Transportation Research Part D: Transport and Environment*, *71*, 186-206. doi: https://doi.org/10.1016/j.trd.2019.01.009

Walker, J., & Ben-Akiva, M. (2002). Generalized random utility model. *Mathematical social sciences*, *43*(3), 303–343.