DTU

Interpretable Embeddings for Representing Categorical Variables within Discrete Choice Models

Ioanna Arkoudi, Carlos Lima Azevedo, Francisco C. Pereira
MLSM Group - DTU Technical University of Denmark

Outline

- Motivation
- Embeddings representations
- Model Formulation & Network Architecture
- Experiment: Swissmetro Dataset - Mode choice
- Results and Comparison to Benchmark models
- Visualizing and Interpreting the embeddings representations
- Conclusions and Future Directions

Motivation: issues with dummy encoding

(1) Conceptual challenges

Do not reflect human perception and intuition

Orthogonal and equidistant representations

e.g.

Self-Employed, Employee, Unemployed, Student, Pupil

Self-Employed, Employee, Unemployed, Student, Pupil

(2) Challenges related to data and estimation requirements

Including richer data → crucial for promoting transport research

(Ben-Akiva et al., 2002)

- **Modeling perspective:**

Model's complexity increases proportionally to the cardinality of the variables → increased sample size requirements to avoid overfitting or poor parameter estimation

(Rose et al., 2013)(Giron et al., 2010)(Cherchi et al., 2012)

- **Survey design perspective:**

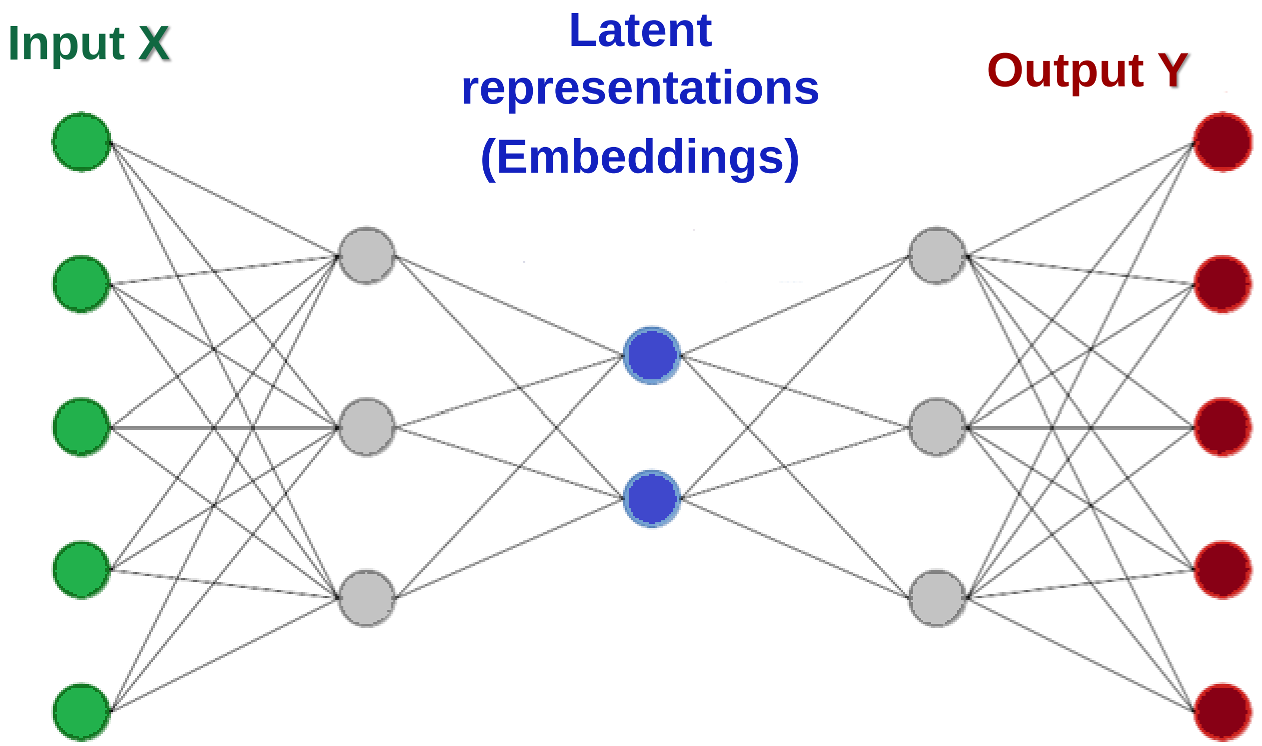
Travel data collection → most expensive and time-consuming part of the model development process (Kulpa,& Szarata, 2016)

Data-driven (NN) approach for encoding categorical variables:

- More intuitive for human analyzer
- More compact yet informative representations of the data

→ Embeddings encoding

Embeddings representations



	one-hot encoding	embeddings encoding
Self-Employed	[1 0 0 0 …]	[0.5 -0.2 0.7 0.9]
High-Income	[0 1 0 0 …]	[-0.3 0.4 0.5 -0.8]
>65 age	[0 0 1 0 …]	[0.6 -0.1 -0.7 0.2]
….	[0 0 0 1 …]	[-0.9 0.5 -0.6 0.7]
	sparse – high dimensional	dense – low dimensional / fixed n of dims

- Latent ***lower space representations*** (vectors) of the input data
- Their weights are learned in a **supervised** manner using a (shallow) ***NN architecture***
- Performing ***dimensionality reduction*** while preserving the ***relevant information to a given predictive task***

Travel behavior embeddings

- **Previous work:**

  Pereira, F. C. (2019). Rethinking travel behavior modeling representations through embeddings.

  Sifringer, B., Lurkin, V., & Alahi, A. (2020). Enhancing discrete choice models with representation learning.

- **Current work:** Embeddings Multinomial Logit (E-MNL) model
  - ➢ NN-based DCM
  - ➢ Embeddings for encoding sociodemographic and travel variables

Compared with previous work:

1. estimate jointly the embedding weights and the parameters (betas) of the DCM
2. all categories across variables embedded in the same continuous space
3. constraints for enforcing interpretability → model with fully interpretable params

E-MNL Model Formulation

We consider a choice set C with J number of alternatives,

$$X = \{X_1, X_2, ..X_K\} \rightarrow \text{continuous}$$

$$Q = \{Q_1, Q_2, \dots, Q_M\} \rightarrow \text{categorical}$$

such that the systematic part of the utility of that individual n associates with alternative $i = 1, \dots, J$, is given by:

$$V_{i,n}(X_n, Q_n) = \boldsymbol{f}\left(X_{i,n}; \boldsymbol{B}\right) + \boldsymbol{f}\ (\boldsymbol{g}_i(Q_n; \boldsymbol{W}); \boldsymbol{B}')$$

$\boldsymbol{B}, \boldsymbol{B}'$: trainable preference parameters (betas)

$\boldsymbol{W}$: trainable weights of the embedding matrix (values of embedding vectors)

$\boldsymbol{g}$: embedding function→ mapping each category in Q_n to an embedding representation Q'_n, such that:

$$g\ (Q_n; \boldsymbol{W}) = Q'_n$$

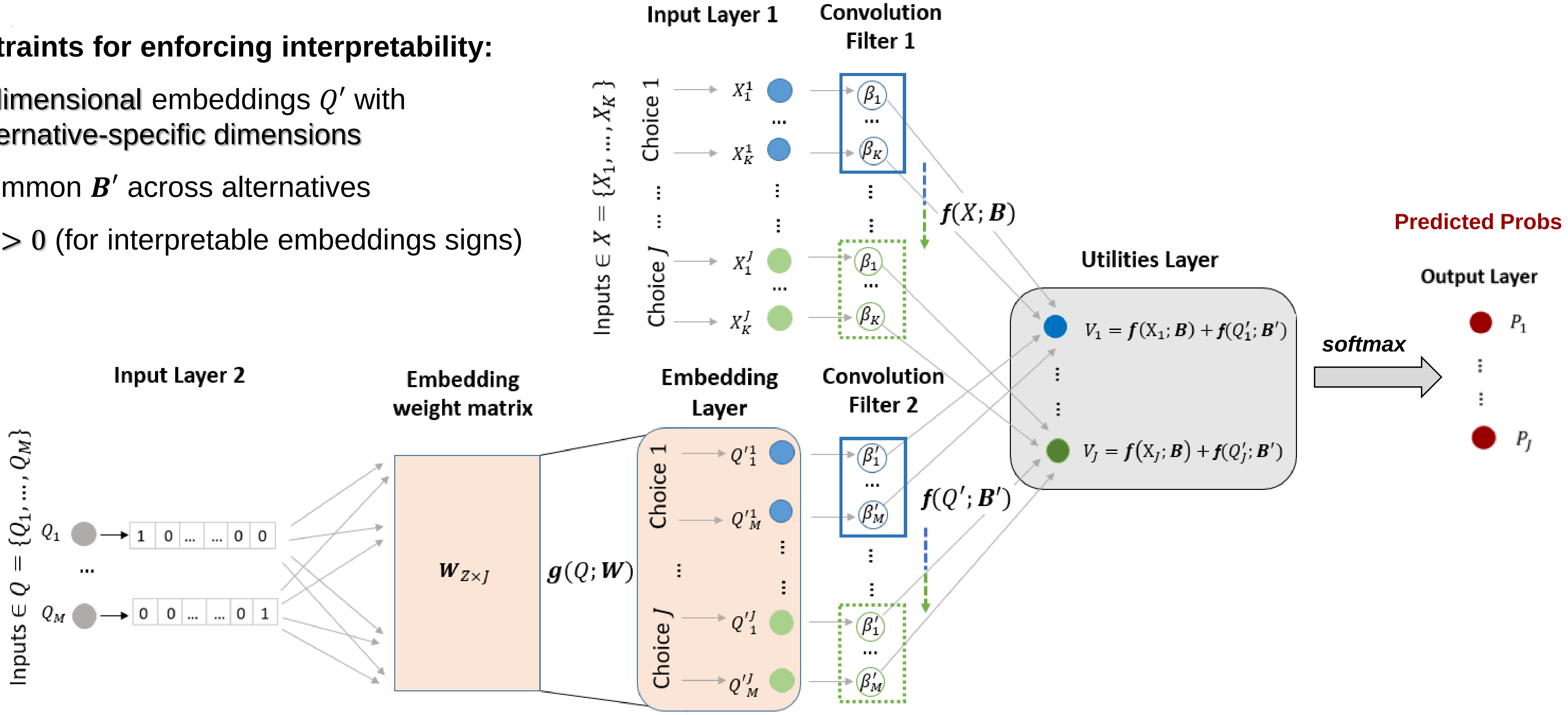
$\boldsymbol{f}$: linear function for the utilities such that: $f\left(X_{i,n}; \boldsymbol{B}\right) = \boldsymbol{B}X_{i,n}$ and $f\ (g(Q_n; \boldsymbol{W_i}); \boldsymbol{B}') = \boldsymbol{B}'Q'_{i,n}$

$$V_{i,n}(X_n, Q_n) = \boldsymbol{B}X_{i,n} + \boldsymbol{B}'Q'_{i,n}$$

E-MNL– model architecture

Constraints for enforcing interpretability:

- $\boldsymbol{J}$ dimensional embeddings Q' with alternative-specific dimensions
- Common $\boldsymbol{B}'$ across alternatives
- $\boldsymbol{B}' > 0$ (for interpretable embeddings signs)



Swissmetro Dataset - Predicting Mode Choice

- Swissmetro dataset (Bierlaire et al., 2001): openly available dataset for mode choice / SP data.
- Three alternatives (*J=3*): Train, Swissmetro , Car
- 9,036 obs → split into training (7,234 obs) and test set (1,802 obs)

Variable	Description	Type
PURPOSE	Trip purpose (business, leisure etc.)	Categorical
FIRST	First class traveler	Binary
TICKET	ticket type (one-way, half-day, etc.)	Categorical
WHO	Who pays (self, employer etc.)	Categorical
MALE	Traveler's gender	Binary
INCOME	Traveler's income level per year	Ordinal
LUGGAGE	Traveler's luggage pcs	Ordinal
AGE	Traveler's age class	Ordinal
SEATS	Airline Seat configuration in the Swissmetro	Binary
GA	Annual season ticket	Binary
ORIGIN	Travel origin corresponding to a canton	Categorical
DEST	Travel destination corresponding to a canton	Categorical
TT	Door-to-door travel time in minutes, scaled by 1/100	Continuous
TC	Travel cost in CHF, scaled by 1/100	Continous
HEADWAY	Transportation headway in minutes, scaled by 1/100	Continous

Benchmark Models	Specification
MNL$orig$	original utility specification from (Bierlaire et al., 2001)
MNLdum	all variables / categorical as dummies
MNL dum - (NO OD)	same as above but excluding ORIGIN and DEST

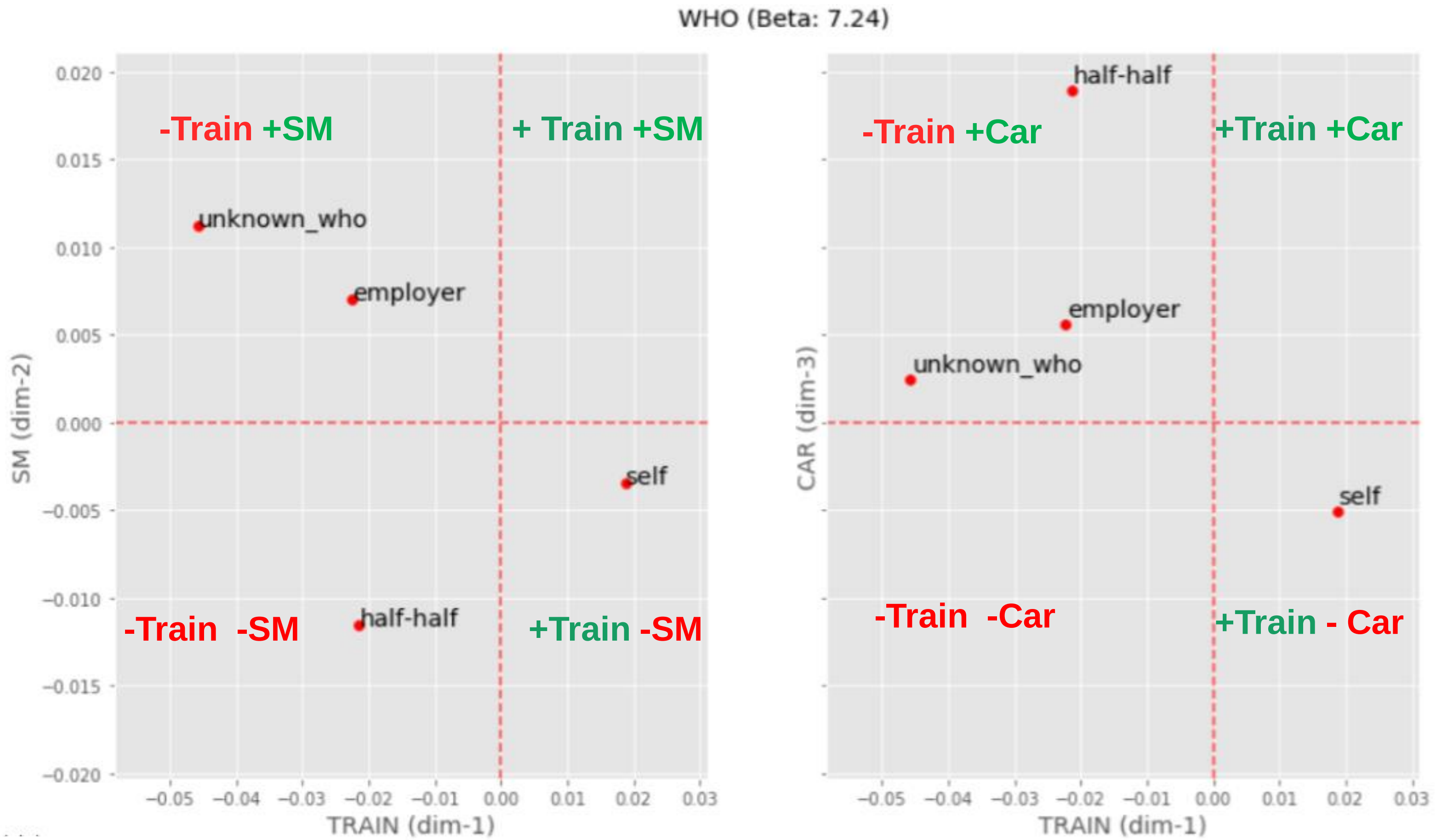
E-MNL → all variables / categorical as embeddings

Results – Comparison to Benchmark Models

Suggested Model	$\overline{LL}_{train}$ (std.)	$\overline{LL}_{test}$ (std.)	$\overline{\text{AIC}}$ (std.)
E-MNL	-4912.7 (3.0)	$\mathbf{-1231.1}$ (1.1)	10345.5 (6.0)
Benchmark Models	LL_{train}	LL_{test}	AIC
MNL_{orig}	-5764	-1433	11546
MNL_{dum} (No OD)	-5243	-1292	10604
MNL_{dum}	$\mathbf{-4903}$	-1928	10066

Model	Params	Estimates	Std errors	t-stat	p-value
E-MNL LL_{train}: -4911.4 LL_{test}: -1229.2	ASC_{Car}	0.252	0.093	2.722	0.006
	ASC_{SM}	0.476	0.079	6.022	0.000
	TT	-1.536	0.054	-28.437	0.000
	TC	-1.129	0.043	-26.099	0.000
	$\text{HEAD}_{Train,SM}$	-0.808	0.122	-6.647	0.000
	PURPOSE	14.074	1.052	13.377	0.000
	FIRST	2.059	1.142	1.802	0.072
	TICKET	12.952	0.676	19.167	0.000
	WHO	8.256	1.392	5.929	0.000
	LUGGAGE	7.118	2.428	2.931	0.003
	AGE	8.484	0.894	9.494	0.000
	MALE	3.752	2.078	1.806	0.071
	INCOME	7.417	1.124	6.599	0.000
	GA	7.429	1.479	5.022	0.000
	ORIGIN	15.325	1.949	7.862	0.000
	DEST	15.425	1.572	9.814	0.000
	SEATS	7.526	1.614	4.662	0.000

Swissmetro embeddings: comparing categories within variables



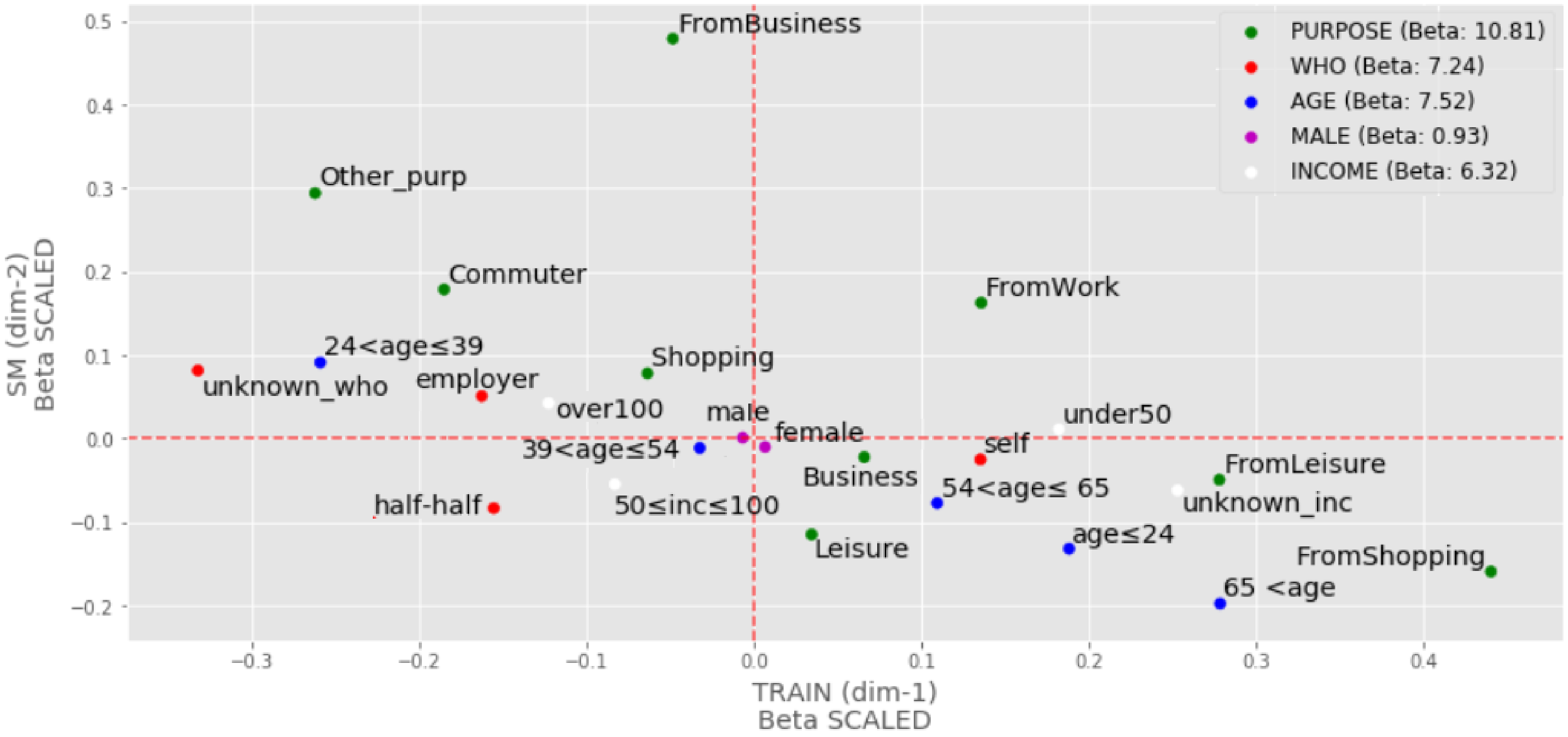
Alternative-specific dimensions

	Train	Car	SM
e.g. *self*	[0.018,	- 0.004	-0.005]

Signs $(B' > 0)$ & Magnitude of each dimension

Embeddings for mode choice trained on the Swissmetro dataset.

Swissmetro embeddings: comparing categories across variables



- **All** categories are **embedded into the same** continuous **space**
- **Similar categories** have **similar vectors** – Categories close in space

  → related to the study behavior in a similar manner

- Formation of **meaningful clusters**

Embeddings for mode choice trained on the Swissmetro dataset.

Conclusions

(1) Embedding representations → behaviorally meaningful outputs → diagnostic and analysis purposes

(2) Increased predictive performance

(3) Efficiently incorporating categorical variables with high cardinality without overfitting

(4) Preserving direct interpretability for all the model's parameters despite being based on NN principles:

- Utilities coefficients → measuring contribution of all the input variables to the models' predictions

- Alternative-specific embedding values→ reflecting internal variable relationships / further insights into understanding the observed behavior

Future directions

- Extending E-MNL for Nested, Mixed, Latent Class Logit models

- Use embeddings for modeling the dynamics of behavior → how do these representations change across time and space?

- Estimating embeddings for multiple outputs - Multitask Learning → generalized embeddings representations → describe travel-behavior as a whole?

Thank you!

Contact:

ioaar@dtu.dk , climaz@dtu.dk , camara@dtu.dk

Find us:

https://mlsm.man.dtu.dk/

Code:

https://github.com/ioaar/Interpretable-Embeddings-MNL

References

Ben-Akiva, M., McFadden, D., Train, K., Walker, J., Bhat, C., Bierlaire, M., ... & Munizaga, M. A. (2002). Hybrid choice models: Progress and challenges. Marketing Letters, 13(3), 163-175.

Rose, J. M., & Bliemer, M. C. (2013). Sample size requirements for stated choice experiments. Transportation, 40(5), 1021-1041.

Giron, M. D. D., & Barrios, E. B. (2010, October). High Dimensional Nonparametric Discrete Choice Model. In Proceedings of the 11th National Convention on Statistics (NCS), Mandaluyong City, Philippines.

Cherchi, E., & Guevara, C. A. (2012). A Monte Carlo experiment to analyze the curse of dimensionality in estimating random coefficients models with a full variance–covariance matrix. Transportation Research Part B: Methodological, 46(2), 321-33

Kulpa, T., & Szarata, A. (2016). Analysis of household survey sample size in trip modelling process. Transportation Research Procedia, 14, 1753-1761.

Pereira, F. C. (2019). Rethinking travel behavior modeling representations through embeddings. arXiv preprint arXiv:1909.00154

Sifringer, B., Lurkin, V., & Alahi, A. (2020). Enhancing discrete choice models with representation learning. Transportation Research Part B: Methodological, 140, 236-261

Bierlaire, M., Axhausen, K., & Abay, G. (2001). The acceptance of modal innovation: The case of Swissmetro. In Swiss Transport Research Conference (No. CONF).