Discrete choice and machine learning: Thomson and Thompson?

Michel Bierlaire

Transport and Mobility Laboratory School of Architecture, Civil and Environmental Engineering Ecole Polytechnique Fédérale de Lausanne

June 11, 2019



EPFL

Michel Bierlaire (EPFL)

Discrete Choice/Machine Learning

June 11, 2019 1 / 48

Outline

Introduction

A little exercise

- ML and discrete choice
- Looking back

5 Data collection
6 Model output
7 Estimation
8 Cross-validation
9 Conclusions





イロト イポト イヨト イヨト

Future mobility

Trends

- Mobility as a service
- Shared mobility
- Demand patterns are more and more complex
- New sources of data

Travel demand

TRANSP-OR

- Traditional methodology: discrete choice
- Emergence of machine learning





IATBR 2018



Session 3E: Machine Learning – Fundamentals Session 6E: More Machine Learning





Michel Bierlaire (EPFL)

Discrete Choice/Machine Learning

June 11, 2019 4 / 48

A (10) < A (10) < A (10) </p>

Interest from young researchers

My PhD topic is "Understanding Multi-Modal Passenger Behaviour at City Scale." I have used trip diary data to compare the performance of multiple discrete choice models, including various multinomial logistic regression models, random forests, support vector machines and neural networks."





Michel Bierlaire (EPFL)

Outline

Introduction

A little exercise

ML and discrete choice

Looking back







イロト イポト イヨト イヨト

Journal of choice modelling

Issues

- 22 to 29
- 2017 and 2018

Procedure

- Download HTML
- Write Python script to extract words.
- Calculate the occurrences of words.





Some counts

logit	4948
utility	8326
machine	99
learning	1459
statistics	634
pattern	383
classification	0



Michel Bierlaire (EPFL)

Visual representation: first attempt



Manual cleaning

Remove common words

the is in and of with to for a that are as each et al on by this we be can from it has where such also may pp not all an their one other was than two at only when use table our how new at or they but using both were using if three no more which these have then given into while over used because section based there will about you some many been did between who same would its any among under could

Remove patterns

Keep only real words - no digits, no special character



Visual representation: second attempt



Michel Bierlaire (EPFL)

Discrete Choice/Machine Learning

June 11, 2019 11 / 48

Manual cleaning

Remove obvious words model models choice data





Michel Bierlaire (EPFL)

Discrete Choice/Machine Learning

June 11, 2019 12 / 48

Visual representation: third attempt



Machine learning

Manual intervention is common

- "A great deal of manual work goes into building and training intelligent machine learning algorithms." Sascha Schubert, business solutions manager at SAS, May 22, 2017.
- "Whenever new learning is involved in ML, the human programmer has to intervene and adapt the programming algorithm to make the learning happen." Paramita Ghosh, Dataversity.net, April 13, 2017.
- Hyperparameter tuning.
- Learning rate tuning.



Outline

Introduction

A little exercise

ML and discrete choice

Looking back

3

5 Data collection
6 Model output
7 Estimation
8 Cross-validation
9 Conclusions





< □ > < 同 > < 回 > < 回 > < 回 >

Definitions

Machine learning is

- an interdisciplinary field
- that uses statistical techniques
- to give computer systems the ability to "learn" from data,
- without being explicitly programmed.

[Wikipedia]





Definitions

Applications of machine learning

- classification
- regression
- clustering
- density estimation
- dimensionality reduction

[Wikipedia]



▲ 同 ▶ → 三 ▶





June 11, 2019 17 / 48

Discrete choice and classification

Discrete choice from a ML perspective

- dependent variable is discrete
- supervised learning
- Iogistic regression



- **→** ∃ →

Introduction to Discrete Choice Models www.edx.org



Michel Bierlaire (EPFL)



Outline

1 Introduction

2 A little exercise

ML and discrete choice

Looking back

5	Data collection
6	Model output
7	Estimation
8	Cross-validatio
9	Conclusions





June 11, 2019 19 / 48

イロト イポト イヨト イヨト

10 years ago: Automatic Facial Expression Recognition

"The face is the most extraordinary communicator, capable of accurately signaling emotion in a bare blink of a second, capable of concealing emotion equally well"

Deborah Blum

Typical machine learning application





















Discrete Choice/Machine Learning

June 11, 2019 20 / 48

Choice experiment

	1/10	
01:08:53:15		
	Happiness Suprise Orear Disgust Oxager OAnger ONeutral	
(Other OI don't know (Validate Survey)	(->)



イロト イポト イミト イミト 一日

Choice model



Ingredients

- Facial Action Coding System (FACS) [Ekman and Friesen, 1978]
- Expression Descriptive Units (EDU) [Antonini et al., 2006]
- Active Appearance Model (AAM) [Edwards et al., 1998]

Main conclusions of this work



- Quality of classification similar to neural networks and Bayesian networks.
- Behavioral insights of the discrete choice model.
- Interpretation of the parameters.
- Possibility to exploit know-how in the specification.



Outline

Introduction

- 2 A little exercise
- 3 ML and discrete choice
- 4 Looking back





Michel Bierlaire (EPFL)



イロト イポト イヨト イヨト

Data universe

Machine Learning: data processing

- Dataset is the universe
- Data generation process is usually ignored
- Representativity is assumed
- Main argument: the size of the dataset is very large

Discrete choice: inference

- A population is identified
- Data collection strategies are designed
- Data sets are rebalanced to represent the population





Potential implications



Classification

- Results from statistics: bias of the parameters
- Not necessarily an issue if cross-validation is applied

Aggregation

- Counting
- Aggregation biases may be severe





June 11, 2019 26 / 48

Example

City of Geneva

- Data for March 2, 2017.
- Phone data: boundary flows, between adjacent zones.
- ML: results of the ML learning algorithm of the phone company.
- Compared with loop detectors: flows of cars



Results



Source: Montesinos Ferrer, Lamotte, Geroliminis

イロト イポト イヨト イヨト

June 11, 2019 28 / 48

3

Outline

1 Introduction

- 2 A little exercise
- 3 ML and discrete choice
- 4 Looking back







June 11, 2019 29 / 48

< □ > < 同 > < 回 > < 回 > < 回 >

Model output

Probability that an item n belongs to a class i



Choice models

Probability is used in applications

Classification

Class with highest probability is selected





Severe aggregation bias

Example: classify 1000 items in two classes.

Data generation process

51% class 1 / 49% class 2

Perfect ML model

After projection: always predicts class 1

Total number of items in class 1

- In reality: 510
- Predicted: 1000



< (日) × (日) × (1)

Aggregation bias increases with the number of classes

Example: classify N items in K classes.

Data generation process

 $\frac{1+\varepsilon}{K}$ class 1 / $\frac{K-1-\varepsilon}{K(K-1)}$ class i

Perfect ML model

After projection: always predicts class 1



Outline

Introduction

- 2 A little exercise
- 3 ML and discrete choice
- 4 Looking back





Michel Bierlaire (EPFL)



June 11, 2019 33 / 48

イロト イポト イヨト イヨト

Loss function/goodness of fit

Formalism

- Class *i*, item *n*
- Independent variables/features : $x_n = (x_{in})_{i=1}^J$
- Choice / class: $y_n = (y_{in})_{i=1}^J \in \{0,1\}^J$
- Unknown parameters: $\boldsymbol{\beta} \in \mathbb{R}^{K}$
- Model: $f(x_n; \beta) \in [0, 1]$
- Loss function: finite sums

$$L(\beta) = \sum_{n=1}^{N} L(f(x_n; \beta), y_n)$$



Loss function/goodness of fit

 $L(f(x_n;\beta),y_n) =$

-Log likelihood / cross entropy

$$-\sum_{i=1}^{J} y_{in} \ln f(x_n;\beta)$$

Square loss

$$\sum_{i=1}^{J} (1 - y_{in}f(x_n;\beta))^2$$

Hinge loss

$$\sum_{i=1}^{J} |1-y_{in}f(x_n;\beta)|_+$$

Exponential loss

 $\sum_{n=1}^{\infty} \exp(-\gamma y_{in} f(x_n;\beta))$ i=1

< □ > < 同 > < 回 > < 回 > < 回 >

June 11, 2019 35 / 48

3

Stochastic gradient descent

Loss function

$$L(\beta) = \sum_{n=1}^{N} L(f(x_n; \beta), y_n)$$

Key ingredient for optimization

Gradient: $\nabla L(\beta) = \sum_{n \in \{1,...,N\}} \nabla L(f(x_n; \beta), y_n)$



イロト イポト イヨト イヨ

Big data

Approx.:
$$\sum_{n\in B\subseteq\{1,\ldots,N\}} \nabla L(f(x_n;\beta),y_n).$$



TRANSP-OR

Stochastic gradient on choice data [Lederrey et al., 2019]



Outline

1 Introduction

- 2 A little exercise
- 3 ML and discrete choice
- 4 Looking back







June 11, 2019 38 / 48

< □ > < 同 > < 回 > < 回 > < 回 >

Cross-validation



Main ideas

- How to select the best model?
- It should be the one that predicts best

Example: leave-one-out

$$I_f = \frac{1}{N} \sum_{n=1}^{N} L(f(x_n; \widehat{\beta}_{n-}), y_n)$$

Image: A match a ma



June 11, 2019 39 / 48

Outline

Introduction

- 2 A little exercise
- 3 ML and discrete choice
- 4 Looking back







June 11, 2019 40 / 48

イロト イポト イヨト イヨト

Summary







Michel Bierlaire (EPFL)

Discrete Choice/Machine Learning

June 11, 2019 41 / 48

メロト メポト メヨト メヨト

Summary

	DCM	ML
Manual intervention	Model spec.	Algorithm
Interpretability	Yes	Not quite
Sampling issues	Handled	Mainly Ignored
Model output	Probability	Mostly $0/1$
Estimation	standard NL opt.	stochastic gradient
Cross-validation	Mainly ignored	Yes





3

<ロト < 四ト < 三ト < 三ト

Conclusions

- Two different communities
- Two different state-of-practice
- Similar objectives

Research agenda

Bring the best from each world



Some recent examples

Application of ML models to choice data [Wong et al., 2018], [Lhéritier et al., forthcoming]

Application of ML algorithms to DC

[Lederrey et al., 2019]

Economic interpretation of ML

[Wang and Zhao, 2018]

Choice model including a neural network [Sifringer et al., 2018]

Assisted specification

[Paz et al., 2019], [Ortelli, 2019], [Hillel et al., forthcoming].

Michel Bierlaire (EPFL)

Discrete Choice/Machine Learning

Bibliography I

Antonini, G., Sorci, M., Bierlaire, M., and Thiran, J.-P. (2006).
 Discrete choice models for static facial expression recognition.
 In Blanc-Talon, J., Philips, W., Popescu, D., and Scheunders, P., editors, Advanced Concepts for Intelligent Vision Systems, volume 4179 of Lecture Notes in Computer Science, pages 710–721. Springer Berlin / Heidelberg.
 ISBN:978-3-540-44630-9.

- Edwards, G. J., Cootes, T. F., and Taylor, C. J. (1998).
 Face recognition using active appearance models.
 In European conference on computer vision, pages 581–595. Springer.
- Ekman, P. and Friesen, W. (1978).
 Facial coding action system (facs): A technique for the measurement of facial actions.

イロト 不得 トイヨト イヨト 二日

Bibliography II

Hillel, T., Bierlaire, M., Elshafie, M., and Jin, Y. (forthcoming).
 Weak teachers: Assisted specification of discrete choice models using ensemble learning.
 In *hEART 2019 - 8th Symposium of the European Association for*

Research in Transportation, Budapest, Hungary.

Lederrey, G., Lurkin, V., Hillel, T., and Bierlaire, M. (2019). Stochastic optimization with adaptive batch size: Discrete choice models as a case study.

In Proceedings of the 19th Swiss Transportation Research Conference.

 Lhéritier, A., Bocamazo, M., Delahaye, T., and Acuna-Agost, R. (forthcoming).
 Airline itinerary choice modeling using machine learning. *Journal of Choice Modelling*.

イロト イボト イヨト イヨト

Bibliography III

Ortelli, N. (2019).

Automatic utility specification in discrete choice models. Master's thesis, Ecole Polytechnique Fédérale de Lausanne (EPFL).

Paz, A., Arteaga, C., and Cobos, C. (2019). Specification of mixed logit models assisted by an optimization framework.

Journal of Choice Modelling, 30:50 – 60.

Sifringer, B., Lurkin, V., and Alahi, A. (2018).
 Enhancing discrete choice models with neural networks.
 In Proceedings of the 18th Swiss Transportation Research Conference.

< ロ > < 同 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ >

Bibliography IV

Sorci, M., Antonini, G., Cruz, J., Robin, T., Bierlaire, M., and Thiran, J.-P. (2010).
 Modelling human perception of static facial expressions.
 Image and Vision Computing, 28(5):790–806.

Wang, S. and Zhao, J. (2018). Framing discrete choice model as deep neural network with utility interpretation.

arXiv preprint arXiv:1810.10465.

Wong, M., Farooq, B., and Bilodeau, G.-A. (2018). Discriminative conditional restricted Boltzmann machine for discrete choice and latent variable modelling. *Journal of Choice Modelling*, 29:152 – 168.

イロト 不得 トイヨト イヨト 二日