

Comparaison de performance de modèles de choix discret de consommation

État d'avancement

Nikita Gusarov

Master 2, Chargé d'études économiques et statistiques

Iragaël Joly (GAEL), Béatrice Roussillon (GAEL),
Sihem Amer-Yahia (LIG), Laurent Muller (GAEL)

06 juillet 2020

Introduction

Problématique

La performance des modèles peut être étudiée de différentes façons

:

- ▶ Qualité des ajustements aux données
- ▶ Capacité prédictive en termes de précision
- ▶ Qualité d'inférence
- ▶ Capacité de produire des indicateurs économiques et comportementaux :
 - ▶ les préférences des consommateurs
 - ▶ les consentements à payer
 - ▶ les parts de marché

Rappel sur les objectifs

Les objectifs principaux du stage

- ▶ Proposer un panorama des outils DCM du ML, économétrie et de l'IA
- ▶ Construire une grille d'analyse des ces outils
- ▶ Réaliser une étude empirique pour tester et évaluer la grille d'analyse sur un jeu de données :
 - ▶ données réelles proposées par TOTAL
 - ▶ données simulées (Japkowicz and Shah 2011)

Éléments clés

- ▶ Les objectifs de la recherche
- ▶ Données
- ▶ Modélisation
- ▶ Évaluation des performances
- ▶ Comparaison des résultats des évaluations

Les objectifs de la recherche

Les objectifs et questionnement

- ▶ Quels sont les objectifs de la modélisation ?
- ▶ Quelles sont les questions posées ?
- ▶ Quel est le champ étudié ?

Données

Questions basiques

Questions concernant la provenance des données

- ▶ Collecte des données
- ▶ Génération et simulation des données

Questions sur le format des données

- ▶ Quelles sont les propriétés des variables utilisées ?
- ▶ Quel format de variables est utilisé ?

Questions plus complexes

- ▶ Quelles données utiliser à quelles étapes de l'étude pour garantir :
 - ▶ la puissance statistique ?
 - ▶ la réplicabilité ?
 - ▶ la reproductibilité ?
 - ▶ la validité externe des résultats obtenus ?

Modélisation

Les différentes idéologies (Breiman and others 2001)

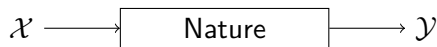


Figure 1: L'état naturel

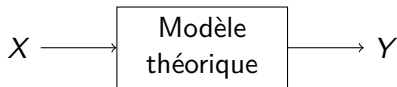


Figure 2: Économétrie

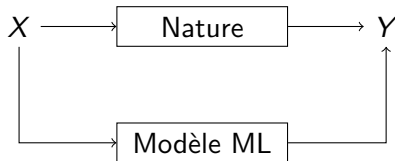
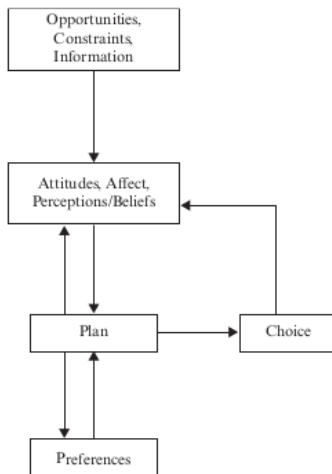


Figure 3: *Machine Learning*

Processus de la prise des décisions (De Palma et al. 2011)



Fondements algorithmiques

- ▶ Comment estimer le modèle ?
- ▶ Quelles techniques liées à la reproductibilité des résultats et leur validité d'estimation utiliser ?
- ▶ Quels algorithmes utiliser ?
- ▶ Quelle est l'implémentation technique d'algorithme sélectionné ?
- ▶ Comment choisir les variables à inclure dans le modèle ?
- ▶ Comment sélectionner les valeurs des hyper-paramètres pour des modèles complexes ?

Définition du modèle

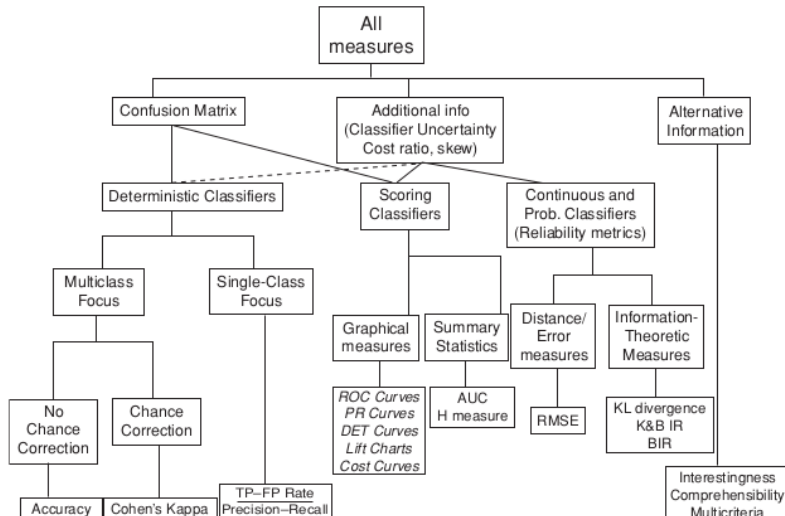
- ▶ Qu'est-ce qu'on considère comme un modèle ?
 - ▶ La théorie étant à la base
 - ▶ Le modèle mathématique
 - ▶ Le processus théorique en combinaison avec un modèle théorique
 - ▶ Le modèle mathématique et l'algorithme utilisé pour l'estimer
 - ▶ La combinaison du modèle mathématique, l'algorithme et sa implémentation technique
 - ▶ *etc.*

Les mesures de performances

Choix des mesures

- ▶ Mesures générales
- ▶ Mesures spécifiques
 - ▶ Faisant le focus sur les résultats (*outcomes*) précis
 - ▶ Faisant le focus sur les effets des variables précises

Multiplicité des mesures (Japkowicz and Shah 2011)



Merci de votre attention

Bibliographie

Agresti, Alan. 2013. *Categorical Data Analysis, Third Edition*.

Athey, Susan, and Guido W. Imbens. 2019. "Machine Learning Methods That Economists Should Know About." *Annual Review of Economics* 11 (1): 685–725. <https://doi.org/10.1146/annurev-economics-080217-053433>.

Breiman, Leo, and others. 2001. "Statistical Modeling: The Two Cultures (with Comments and a Rejoinder by the Author)." *Statistical Science* 16 (3). Institute of Mathematical Statistics: 199–231.

De Palma, André, Robin Lindsey, Emile Quinet, and Roger Vickerman. 2011. *A Handbook of Transport Economics*. Edward Elgar Publishing.

Drummond, Chris. 2006. "Machine Learning as an Experimental Science (Revisited)." In *Proceedings of the Twenty-First National Conference on Artificial Intelligence: Workshop on Evaluation Methods for Machine Learning*, 1–5.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science & Business Media.

Japkowicz, Nathalie, and Mohak Shah. 2011. *Evaluating Learning Algorithms: A Classification Perspective*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511921803>.

Mullainathan, Sendhil, and Jann Spiess. 2017. "Machine Learning: An Applied Econometric Approach." *Journal of Economic Perspectives* 31 (2): 87–106. <https://doi.org/10.1257/jep.31.2.87>.

Varian, Hal R. 2014. "Big Data: New Tricks for Econometrics." *Journal of Economic Perspectives* 28 (2): 3–28. <https://doi.org/10.1257/jep.28.2.3>.