

Apprentissage automatique pour optimiser les ventes de services ancillaires

Benoit Lardeux¹, Mourad Boudia¹, Ahmed Abdelli²

¹ Amadeus Innovation and Research, 485 route du Pin Montard, 06560 Sophia-Antipolis, France
`{benoit.lardeux,mourad.boudia}@amadeus.com`

² Eurecom, SophiaTech, 450 route des Chappes, 06410 Biot, France
`ahmed.abdelli@eurecom.fr`

Mots-clés : *apprentissage automatique, arbre de décision, optimisation*

1 Introduction

Depuis quelques années la part du revenu des compagnies aériennes qui provient de la vente de produits ancillaires, comme le bagage, un accès au salon ou des repas spéciaux à bord de l'avion ne cessent d'augmenter. Cependant, une large partie des potentiels bénéfiques dûs à cette nouvelle source de revenu sont encore à exploiter. Dans la plupart des cas, ces services ancillaires sont vendus à la demande des passagers. Nous présentons dans ce papier des méthodes d'optimisation, utilisées à partir d'un large volume de données historiques, permettant de reproduire le comportement des passagers et de proposer les services ancillaires qui ont le plus de chance d'être achetés. Différentes méthodes d'apprentissage automatique basées sur des arbres de décision et des algorithmes de régression logistique ont été ajustées à ce problème et comparées. Après avoir expliqué le contexte métier, nous détaillerons les ajustements effectués sur ces méthodes standards ainsi que les résultats numériques obtenus sur des données réelles provenant des systèmes d'information d'Amadeus.

2 Un gros volume de données disponibles

Le revenu ancillaire est défini comme le revenu en plus du prix du billet pour une place dans un avion. Deux sources principales de données sont utilisées pour cette étude : les billets émis et les recus associés aux services ancillaires, qui sont appelés des EMDs (Electronic Miscellaneous Document) [3]. Les systèmes d'émission de billets d'Amadeus gèrent ce service pour un grand nombre de compagnies aériennes.

L'apprentissage automatique a été effectué sur des millions de tickets émis pour trois compagnies aériennes européennes de taille conséquente pendant trois années.

3 Recherche des méthodes optimales de classification

L'objectif de cette étude est, à terme, de construire un module de recommandation de service ancillaire approprié aux besoins du voyageur. Grâce aux méthodes d'apprentissage automatique, nous avons ainsi pu définir trois modèles de prévisions permettant d'optimiser les chances de vendre ses services en fonction du contexte de la réservation du billet d'avion.

- Le modèle 1 calcule la probabilité qu'un service soit accepté
- Le modèle 2 détermine le moment le plus approprié pour proposer ce service
- Le modèle 3 recherche quel type de service a le plus de chance d'être accepté

Pour ces trois problèmes, des méthodes de classification supervisées ont été utilisées. L'apprentissage de ses trois modèles s'effectue en fonction des caractéristiques du voyage (lieu d'origine, la destination, le prix...) et du voyageurs (lieux et dates de l'achat...) indiqués dans les billets d'avion.

Le premier modèle a pour objectif de prévoir si un service additionnel doit être proposé ou non, selon donc un choix binaire. Les paramètres des modèles de régression logistique ([4]) sont optimisés à partir des données réelles. Le modèle est entraîné sur 70% de l'ensemble de ses données, qui regroupe à la fois les billets vendus seuls et les billets auxquels sont associés des EMDs. Ces modèles ont été évalués en fonction d'indicateurs bien connus en apprentissage automatique qui sont la précision, le rappel et le F1-score, qui n'est autre que la moyenne harmonique de la précision et du rappel. Les résultats des approches les plus performantes seront présentés ainsi que les caractéristiques des billets qui ont le plus d'importance.

Le deuxième modèle permet de prévoir le moment opportun pour proposer un service. Trois intervalles de temps ont été définis : le moment de l'achat du billet, le jour du départ de l'avion et le temps entre ses deux dates.

Pour le troisième modèle qui concerne le type de service qui a le plus de chance d'être accepté. L'ensemble des possibilités a été partagé en neuf catégories présentés dans le tableau 1

4 Résultats numériques et conclusion

	Precision	Rappel	F1-score
Bagages additionnels	0.78	0.79	0.78
Choix du siège	0.57	0.65	0.61
Surclassement	0.60	0.66	0.63
Repas spéciaux	0.50	0.41	0.45
Animaux	0.70	0.48	0.57
Passager non accompagnés	0.75	0.51	0.61
Accès au salon	0.72	0.41	0.52
Service financier	0.84	0.47	0.60
Autres	0.70	0.74	0.72
average total	0.68	0.57	0.61

TAB. 1 – Resultats obtenus avec un algorithme de random forest

Comme indiqué dans le tableau ci-dessus, les résultats obtenus à partir de forêt d'arbres de décision sont de bonnes qualités en moyenne malgré quelques variations notables entre les différents types de services. Ces résultats ouvrent des perspectives intéressantes pour la développement de systèmes de recommandations embarqués dans les systèmes informatiques des agents de voyages ou autre distributeur en ligne de billets d'avion.

Références

- [1] *Breiman, Leo*. Random Forests. Machine Learning, Springer. 45(1) :5-32. 2001.
- [2] *Breiman, Leo ; Friedman, J.H. ; Olshen, R.A. ; Stone, C.J.*. Classification and regression trees. Wadsworth and Brooks/Cole Advanced Books and Software. ISBN 978-0-412-04841-8. Monterey, CA. 1984.
- [3] *IATA*. Electronic Miscellaneous Document. <http://www.iata.org/whatwedo/airline-distribution/e-services/Pages/index.aspx>
- [4] *Menard, Scott*. Applied logistic regression analysis. SAGE. ISBN 978-0-7619-2208-7. Denver, CO. 2002.