Prédiction de temps de parcours de bus par chaînage des données d’entraînement

Gabriel Ferrettini*, François Queyroi*, Mounira Harzallah*

* Nantes Université, Ecole Centrale Nantes, CNRS, LS2N, UMR 6004
F-44000 Nantes, France
prenom.nom@univ-nantes.fr

Résumé. Dans le cadre de la planification de lignes de bus, il est intéressant de prédire à la fois la durée d’une course (du départ au terminus) et le temps de parcours de chaque inter-arrêt. Une solution est de chaîner les prédictions de modèles construits pour chaque inter-arrêt et de les sommer pour obtenir le temps total de la course. Cependant, ce modèle dit “chaîné” ne prend pas en compte la dégradation des données prédites d’une étape pour l’étape suivante. Nous proposons d’améliorer ce modèle en intégrant le chaînage des prédictions dans l’entraînement des modèles de chaque étape. L’évaluation de notre proposition montre que notre méthode améliore la prédiction de la durée totale d’une course. Cependant, l’analyse comparative des performances des modèles inter-arrêts et du modèle chaîné montre qu’une amélioration des premiers n’améliore pas forcément le second.

1 Introduction

Dans le contexte actuel, il est d’une importance croissante de développer les transports en commun et d’en promouvoir l’utilisation. À cet effet, de nombreuses compagnies visent à prédire de manière efficace le temps de trajet de leurs véhicules, afin de mieux en planifier les trajets et identifier les sources d’éventuels retards. Cet article se place dans le contexte d’un réseau de lignes de bus, dans le but de simuler le comportement d’un bus au cours de son trajet. Dans cette optique, de nombreuses informations sont intéressantes. En effet, le temps de trajet total du bus de son départ à son terminus est important, mais le détail des événements survenant au cours de ce trajet l’est également, afin d’identifier les éventuelles causes de ralentissements et donc de retards du service et de modifier l’aménagement des lignes de bus en conséquence.

Nous étudions donc des modèles de régression visant non seulement à prédire la durée totale du trajet d’un bus, mais également la durée de ses différentes étapes. Pour cela, nous proposons d’utiliser un modèle prédictif distinct pour prédire le trajet du bus pour chaque inter-arrêt (i.e. entre deux arrêts successifs) d’une ligne, et chaîner les prédictions de ces modèles afin de prédire la durée de son trajet total. Enfin, nous proposons une nouvelle méthode d’entraînement visant à améliorer la performance de modèles chaînés en modifiant les données d’entraînement afin d’y inclure le fait que le