

# APPROCHE NON-PARAMÉTRIQUE POUR LA PRÉDICTION D'HEURE D'ARRIVÉE DANS LES TRANSPORTS

Chandesris Maguelonne <sup>1</sup>

<sup>1</sup> *SNCF Innovation & Recherche*  
*40 avenue des terroirs de France, 75611 PARIS CEDEX 12*

**Résumé.** La régularité est un objectif essentiel pour un opérateur de transports planifiés<sup>1</sup>. La prédiction de l'heure d'arrivée des mobiles en circulation aux prochains points d'arrêts est un élément important tant pour les régulateurs du trafic que pour l'information au public. Différents travaux et approches ont fait l'objet de nombreux développements scientifiques, notamment pour le trafic urbain des bus. Cette communication propose une nouvelle approche non-paramétrique pour réaliser cette prédiction, ici dans le contexte ferroviaire. L'idée est de tirer parti de la capacité d'un estimateur non-paramétrique à s'appuyer sur des similarités spatio-temporelles passées pour fournir une prédiction adaptée au contexte.

**Mots-clés.** non-paramétrique, prédiction, spatio-temporel, transports

**Abstract.** Punctuality is a key issue for a public transportation operator. Arrival time prediction for the next stations is an important information for both traffic managers and people in transportation or waiting for it. Numerous works and scientific methods have been developed for that case, particularly for buses. In this paper, we propose a new non-parametric approach to achieve the prediction. Our idea is to advantage of the non-parametric estimator specificity to catch spatio-temporal similarities, in order to provide good predictions.

**Keywords.** non-parametric, prediction, spatio-temporal, transportation

## 1 Prédiction de l'heure d'arrivée à destination dans les transports

Pour un opérateur de transport tel que la SNCF, la prédiction de qualité des heures d'arrivées aux prochains arrêts des mobiles en circulation constitue un enjeu important tant pour alimenter la régulation du trafic (personnel mais aussi outils d'aide à la décision) que pour informer le public (transporté ou en attente). De tels systèmes existent déjà en exploitation. Nous avons ici souhaité élaborer et tester une nouvelle approche pour

---

<sup>1</sup>s'engageant sur des horaires et/ou une fréquence de passage

améliorer la qualité des prédictions quelque soit l’horizon (prochains arrêts jusqu’au terminus) et le contexte de circulation.

L’information disponible pour effectuer cette prédiction provient de la connaissance des horaires théoriques et du suivi temps réel des mobiles (balises, suivi GPS, crowd-sourcing, etc.). Le développement important du recueil de données temps réel a permis le développement de nombreuses approches, en particulier pour les bus [1]. Pour cela, différentes approches sont envisageables :

**Approche inductive** Dans ce cas, il s’agit au préalable d’établir un modèle qui résume la structure et qui sera réappliqué par la suite. Les modèles peuvent être de différents types :

- modèles d’inspiration mécanique (proche du fonctionnement réel) cf. par exemple [2].
- modèles statistiques (approche par apprentissage) construit à partir d’observations passées, de nombreuses approches existent : modèles de régression, filtre de Kalman, réseaux de neurones, SVM, etc.

**Approche transductive** Dans ce cas, il n’y a pas de modèle explicite, le calcul se fait directement à partir des données disponibles, par exemple utilisation des derniers temps moyens/vitesse de parcours.

Dans les approches inductives, une étape de modélisation ou d’apprentissage est nécessaire, cette étape peut s’avérer complexe en fonction des systèmes étudiés<sup>2</sup>. L’approche proposée ici relève de la seconde approche. L’idée est d’étudier dans quelle mesure on peut construire, par une approche non-paramétrique un prédicteur robuste mais suffisamment générique pour qu’il puisse à terme être déployé sur différents types de trafic. Les techniques non-paramétriques pour la prévision ont montré leur intérêt et efficacité [3], on a souhaité ici les mettre en œuvre afin de capter à la volée la dynamique du système.

## 2 Approche non-paramétrique pour la prédiction de l’heure d’arrivée à destination

Le principe de l’approche retenue est relativement simple : à un instant  $t_0$ , pour prédire l’heure d’arrivée à tous les prochains points d’arrêts que doit desservir le mobile considéré dans sa mission, on cherche dans l’historique les situations les plus semblables qui ont déjà eu lieu (cf. figure 1), avant d’en effectuer le barycentre pondéré par un coefficient de similarité. On se place ici dans un contexte en deux dimensions spatio-temporelle, puisque l’on considère des mobiles se déplaçant dans l’espace et le temps. La trame générale peut se décomposer en quatre étapes :

---

<sup>2</sup>en particulier dans le contexte ferroviaire constitué d’un contrôle commande

- 1. Analyse de la situation considérée** Dans cette première étape, on détermine la taille  $r$  de la fenêtre spatio-temporelle qui va permettre les comparaisons avec les situations passées. Cette fenêtre inclut le mobile considéré mais également les mobiles environnants (typiquement le précédent) afin de prendre en compte le contexte dans lequel le mobile étudié se déplace.
- 2. Sélection de l'historique à comparer** La sélection de l'historique permet d'éviter de calculer un trop grand nombre de similarités. On peut par exemple sélectionner les historiques pour lesquels les horaires théoriques relatifs des fenêtres sont très proches de la situation considérée.

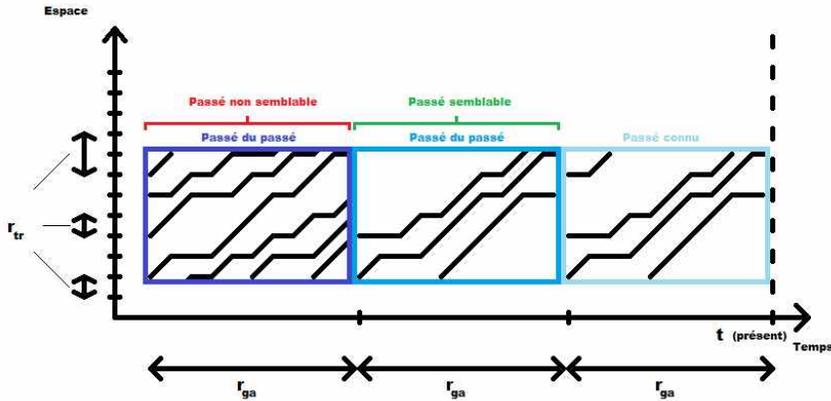


Figure 1 : Comparaison du passé immédiat avec des passés plus lointains

- 3. Calcul des similarités** Le calcul des similarités entre le passé du présent  $X_{t_0,(r)}$  et les passés historiques  $X_{t,(r)}$ ,  $t < t_0$  se fait de manière classique à l'aide du noyau gaussien  $K$  et de la norme euclidienne  $\|\cdot\|$  :

$$sim_{t,r}(X_{t_0,(r)}) = \frac{1}{Z} K \left( \frac{\|X_{t_0,(r)} - X_{t,(r)}\|}{h_n} \right)$$

où :

- $X_{t_0,(r)} = (X_{t_0}, X_{t_0-1}, \dots, X_{t_0-r+1})$  est la fenêtre de largeur  $r$  sur les observations passées du présent  $t_0$  ;
- $X_{t,(r)} = (X_t, X_{t-1}, \dots, X_{t-r+1})$  est la fenêtre glissante de largeur  $r$  sur les observations passées ;
- $Z$  est le coefficient de normalisation, tel que  $\sum sim_{t,r}(x) = 1$  ;
- $h_n$  est la constante de lissage fixée par validation croisée.

**4. Prévision** Pour la prévision d’horizon  $s$ , on effectue le barycentre des futurs du passé pondéré par les similarités que l’on vient de calculer :

$$\forall x \in \mathbb{R}^r, f(x) = \sum sim_{t,r}(x) \times \begin{pmatrix} X_{t+1} \\ \vdots \\ X_{t+s} \end{pmatrix}$$

### 3 Résultats et perspectives

Les travaux réalisés ont permis de concevoir l’approche proposée ci-dessus mais aussi de la mettre en œuvre et de la tester sur des données réelles du RER D<sup>3</sup>. Des comparaisons avec d’autres méthodes plus classiques ont pu être menées sur une journée (soit plus de 15 000 prédictions à fournir par méthode).

Les résultats démontrent la faisabilité de l’approche proposée et incitent à en poursuivre le développement. L’ambition est de pouvoir fournir en temps réel des prédictions de bonne qualité quelque soit l’horizon. Les prochaines étapes de recherche pourraient être :

- dans le calcul des similarités, inclusion d’une norme qui prennent en compte les déformations de séries spatio-temporelles (généralisation du Dynamic Time Warping (DTW) [4]),
- prise en compte de la contrainte du calcul temps réel,
- définition de zones d’efficacité selon les méthodes pour définir une méta-méthode ou méthode hybride.

**Remerciements** Cette nouvelle approche a pu être testée grâce au travail de Xavier Chapuis que l’auteur tient ici à remercier pour la qualité de sa collaboration.

### Bibliographie

- [1] Altinkaya, M. et Zontul, M. (septembre 2013), *Urban bus arrival time prediction : a review of computational models*, International journal of recent Technology and Engineering (IJRTE), Vol. 2, Issue 4.
- [2] Kecman, P. et Goverde, R. M. P. (2013), *An online railway traffic prediction model*, RailCopenhagen2013: 5th International Conference on Railway Operations Modelling and Analysis, Copenhagen, Denmark, 13-15 May 2013

---

<sup>3</sup>l’implémentation a été réalisée à l’aide du logiciel R

- [3] Poggi, J. M. (1994), *Prévision non-paramétrique de la consommation électrique*, Revue de Statistique Appliquée, XLII (4), 83-98.
- [4] Berndt, D. J. et Clifford, J. , *Using Dynamic Time Warping to Find Patterns in Time Series*, Knowledge Discovery in Databases workshop, Vol. 10, No 16, pp. 359-370, Seattle, Washington, USA, juillet 1994