

Prévision à court terme des flux de voyageurs du réseau ferré urbain : une approche par les réseaux bayésiens dynamiques

Jérémy Roos^{*,**} Stéphane Bonnevey^{**} Gérald Gavin^{**}

*RATP, F75012 Paris, France
jeremy.roos@ratp.fr

**Univ Lyon, ERIC EA3083, F69100 Villeurbanne, France
{stephane.bonnevey, gerald.gavin}@univ-lyon1.fr

Résumé. Nous proposons une approche de prévision à court terme des flux de voyageurs du réseau ferré d'Île-de-France basée sur les réseaux bayésiens dynamiques. La structure du modèle repose sur les relations de causalité entre les flux adjacents et permet d'intégrer l'offre de transport. En présence de données manquantes, l'apprentissage est réalisé via l'algorithme espérance-maximisation (EM) structurel. En appliquant notre approche sur une ligne de métro, les résultats obtenus sont globalement supérieurs à ceux des autres méthodes testées.

1 Introduction

Principal opérateur de transport public d'Île-de-France, la RATP dispose de plusieurs outils pour évaluer l'impact à long terme de changements d'infrastructures ou de politiques de transport sur les flux de voyageurs. Ces outils n'étant pas destinés à la prévision à court terme, il ne peuvent pas tenir compte des événements imprévus ou non récurrents tels que les perturbations d'exploitation, les fermetures de stations ou les événements générateurs d'affluence.

Dans cet article, nous proposons un modèle de prévision à court terme des flux de voyageurs du réseau ferré (métro et RER). Basée sur les réseaux bayésiens dynamiques, cette approche permet de gérer les données manquantes. En fournissant des prédictions en temps réel, elle répond à une grande diversité d'applications relatives à l'information voyageurs, la régulation des flux ou encore la planification de l'offre de transport.

Après un bref état de l'art de la prévision du trafic à court terme, nous introduisons les réseaux bayésiens, puis détaillons la démarche de modélisation. Nous testons finalement notre approche sur une ligne du métro de Paris, avant de conclure l'article en proposant des pistes d'amélioration du modèle.

2 Prévision du trafic à court terme

Il existe une vaste littérature sur la prévision du trafic à court terme. Les diverses méthodes proposées peuvent être classées en trois catégories : les méthodes naïves (telles que la moyenne historique), fréquemment utilisées en raison de leur facilité d'implémentation, les méthodes paramétriques (ARIMA, filtre de Kalman...) et les méthodes non paramétriques (k plus proches

voisins, réseaux de neurones...), ces dernières ayant une meilleure capacité à modéliser les processus non linéaires (van Hinsbergen et al., 2007).

La plupart des recherches ont été menées sur des réseaux de transport routier. À l'inverse, peu de travaux ont été consacrés jusqu'à présent aux flux de voyageurs de réseaux de transport public, les modèles existants étant principalement destinés à la planification à long terme (Ma et al., 2014). Quelques auteurs se sont toutefois penchés sur la question, notamment à travers des approches par les réseaux de neurones (Wei et Chen, 2012; Li et al., 2013).

Le caractère incomplet des données est un problème courant en situation réelle. S'il existe des méthodes d'imputation des données de trafic, peu d'entre elles sont conçues pour opérer en temps réel (Haworth et Cheng, 2012). Ainsi, la plupart des modèles sont mal équipés pour faire face aux données manquantes. Plusieurs approches ont été proposées afin de pallier ce problème, notamment par les réseaux bayésiens (Sun et al., 2006; Whitlock et Queen, 2000).

3 Réseaux Bayésiens

Les réseaux bayésiens (Pearl, 1988) représentent les relations de dépendance (et d'indépendance) entre des variables par un graphe orienté acyclique. Ces relations sont décrites par une loi de probabilité jointe qui se factorise en un produit de distributions conditionnelles locales :

$$p(X_1, \dots, X_n) = \prod_{i=1}^n p(X_i | Pa(X_i)) \quad (1)$$

où $Pa(X_i)$ est l'ensemble des parents de X_i . Le mécanisme de propagation de l'information inhérent à ces modèles permet d'inférer en présence de données manquantes.

Si les variables sont continues, les distributions conditionnelles d'un réseau bayésien peuvent être décrites par des modèles linéaires gaussiens (Shachter et Kenley, 1989) :

$$p(X_i | Pa(X_i)) = \mathcal{N}(\beta_{i,0} + \beta_i^\top Pa(X_i), \sigma_i^2) \quad (2)$$

où $\beta_{i,0}$, β_i et σ_i sont les paramètres à estimer (pour chaque i). Le choix de ce type de distribution implique de faire l'hypothèse que les relations entre les variables sont linéaires.

Les réseaux bayésiens dynamiques étendent les réseaux bayésiens à la modélisation des relations temporelles entre les variables (Dean et Kanazawa, 1989). Dans ces modèles, chaque nœud $X_i^{(t)}$ correspond à une instantiation de la variable X_i à un intervalle de temps t .

4 Démarche de modélisation

4.1 Construction du modèle

En suivant le raisonnement de Sun et al. (2006), il existe une relation de causalité entre les flux de voyageurs observés à un endroit du réseau et ceux observés plus tard en aval. Ainsi, nous définissons un paramètre d tel que chaque flux à l'intervalle de temps t dépend de ses flux amont adjacents à $t - 1, \dots, t - d$. Étant donné que les valeurs historiques des flux nous informent sur leur tendance, nous définissons également un paramètre m tel que chaque flux à t dépend de ses propres valeurs à $t - 1, \dots, t - m$ (Sun et al., 2006).

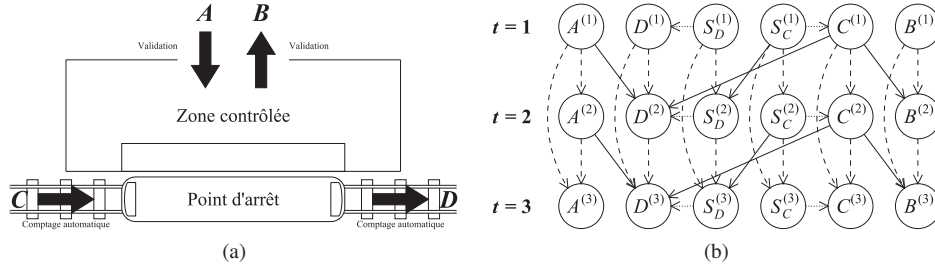


FIG. 1 – (a) Exemple de flux de voyageurs du réseau ferré. (b) Réseau bayésien dynamique correspondant ($d = 1$ et $m = 2$) déroulé sur 3 intervalles de temps.

Dans un réseau de transport public, les flux de voyageurs sont étroitement liés à l’offre de transport. Intuitivement, le nombre de voyageurs à bord d’un train dépend du temps d’attente entre ce train et le précédent. Plus ce temps est long, plus les voyageurs s’accumulent sur le quai avant d’embarquer. Soit X un flux de voyageurs circulant en train entre deux points d’arrêt et $\mathcal{D}_X^{(t)}$ l’ensemble des horaires de départ depuis le point d’arrêt d’origine de X pendant l’intervalle de temps t . Nous définissons la “variable d’offre de transport” associée à X à t :

$$S_X^{(t)} = \begin{cases} \max \mathcal{D}_X^{(t)} - \max \bigcup_{k < t} \mathcal{D}_X^{(k)}, & \text{si } \mathcal{D}_X^{(t)} \neq \emptyset \\ 0, & \text{sinon} \end{cases} \quad (3)$$

où $\max \mathcal{D}_X^{(t)}$ et $\max \bigcup_{k < t} \mathcal{D}_X^{(k)}$ sont les derniers horaires de départ respectivement pendant et avant t . $S_X^{(t)}$ est équivalente à la somme des temps d’attente précédant les trains dont le départ a lieu pendant t et impacte donc directement $X^{(t)}$. Elle dépend également de ses variables d’offre de transport amont adjacentes et de ses propres valeurs aux intervalles de temps précédents.

La figure 1 présente un exemple de flux de voyageurs du réseau ferré ainsi que le réseau bayésien dynamique correspondant, avec $d = 1$ (arcs continus dans la figure 1b) et $m = 2$ (arcs à tirets). Dans cet exemple, A est en amont de D et C est en amont de B et D . C et D dépendent directement de l’offre de transport (arcs à points).

En supposant que les interactions entre les variables sont linéaires, nous décrivons les distributions conditionnelles du réseau bayésien dynamique par des modèles linéaires gaussiens.

4.2 Apprentissage et inférence

Selon les valeurs de d et m , le nombre d’arcs du réseau bayésien dynamique peut être très élevé. Outre la complexité de calcul, cette situation favorise les risques de surapprentissage et donc de dégradation des performances du modèle. Afin de remédier à ce problème, l’idée est de sélectionner le meilleur sous-ensemble d’arcs parmi ceux décrits précédemment. En présence de données manquantes, cette sélection ainsi que l’estimation des paramètres peuvent être réalisées via l’algorithme espérance-maximisation (EM) structurel (Friedman et al., 1998).

La structure du réseau bayésien dynamique est initialisée sans arcs. En pratique, chaque itération de l’algorithme EM structurel consiste à réaliser l’algorithme EM paramétrique pour

compléter les données à partir de la précédente estimation du modèle (étape E), puis à mettre à jour cette estimation en maximisant une fonction de score (étape M). Dans le cas présent, le score utilisé est le critère d'information bayésien (BIC), qui tient compte de la vraisemblance tout en pénalisant les modèles trop complexes. Une fois les données complétées, la propriété de décomposition du BIC peut être exploitée afin de mettre en œuvre des procédures locales de recherche gloutonne permettant d'améliorer graduellement la structure du modèle (en ajoutant ou retirant des arcs). La valeur du BIC augmente à chaque itération de l'algorithme EM structurel jusqu'à convergence vers un maximum local (Friedman et al., 1998).

La prévision à court terme (c'est-à-dire à l'intervalle de temps suivant) des flux de voyageurs est un problème d'inférence dans le réseau bayésien dynamique. Afin de prédire les flux en temps réel, il est préférable de recourir aux méthodes d'inférence approchée, généralement moins coûteuses en temps de calcul. Parmi elles, le filtre bootstrap (Gordon et al., 1993) est un algorithme de simulation stochastique particulièrement efficace. Son principe consiste à générer un certain nombre de particules en échantillonnant les valeurs non observées, puis à les propager en fonction de leur vraisemblance avec les mesures collectées au fil du temps.

5 Expérimentation

5.1 Méthode expérimentale

Nous appliquons notre approche aux flux de voyageurs relatifs à la ligne 2 du métro de Paris. Dans le cadre de cette expérimentation, trois types de données sont collectées durant 33 jours ouvrés de mars et avril 2015, de 7 h 30 à 9 h 30, et agrégées par tranche de 2 minutes :

- les données de validation des titres de transport, en entrée (28 flux), en correspondance (2 flux) et en sortie (5 flux) des zones contrôlées (c'est-à-dire accessibles uniquement aux voyageurs munis d'un titre valide) ;
- les données de comptage automatique, au départ des trains par des systèmes de pesage embarqués (60 flux) ;
- les données d'offre de transport (114 variables d'offre de transport).

En raison de dysfonctionnements ou d'absences de dispositifs de collecte, 80 des 95 flux enregistrés sont incomplets. Au global, ce taux d'incomplétude est de 4.8 %, mais peut dépasser 50 % pour certains flux. Les données manquantes connaissent également une forte dispersion temporelle, couvrant 99.7 % des tranches collectées.

Nous divisons le jeu de données en un échantillon d'apprentissage et un échantillon de test, composés respectivement des 24 premières journées et des 9 journées restantes. Afin d'évaluer les performances de prédiction, nous adoptons l'erreur absolue moyenne pondérée (WMAPE). Facilement interprétable, la WMAPE pondère les erreurs par les valeurs des flux et favorise ainsi les modèles qui prédisent le mieux les fortes affluences.

Lors de la construction du réseau bayésien dynamique, nous choisissons les paramètres $d = 2$ et $m = 3$, qui donnent empiriquement de bons résultats de prédiction. Afin d'évaluer la contribution individuelle de chaque type de relation, nous comparons la version complète du modèle à une version sans l'offre de transport et une version sans les relations entre les flux adjacents. Les résultats sont également comparés à ceux de deux autres méthodes : la moyenne historique et le report de la dernière valeur observée (LOCF).

Flux de voyageurs	Réseau bayésien dynamique			MH	LOCF
	complet	sans offre	sans adj.		
Au départ des trains	17.8	37.3	21.1	40.3	63.7
Entrants	19.0	19.0	19.0	16.9	24.0
Correspondants/sortants	22.6	23.7	24.8	22.2	31.6
Total	18.5	30.9	20.7	32.1	49.6

TAB. 1 – Comparaison des résultats de prédiction (WMAPE moyenne en %)

5.2 Résultats

Les résultats de l'expérimentation sont détaillés dans le tableau 1. Au global, notre approche obtient une erreur moyenne de 18.5 %, contre 32.1 % pour la moyenne historique (MH dans le tableau 1) et 49.6 % pour LOCF. L'intégration de l'offre de transport contribue largement à ces bons résultats. Elle améliore considérablement la prédiction des flux de voyageurs au départ des trains, dont l'erreur moyenne diminue de 30.9 % à 17.8 %.

Les relations entre les flux adjacents jouent également un rôle important dans la modélisation. Leur contribution est particulièrement visible pour les flux au départ des trains, dont l'erreur moyenne diminue de 21.1 % à 17.8 %, et, dans une moindre mesure, pour les flux en correspondance ou en sortie des zones contrôlées (de 24.8 % à 22.6 %). Ces résultats aident à comprendre pourquoi notre approche se révèle moins performante que la moyenne historique pour les flux en entrée des zones contrôlées (respectivement 19.0 % et 16.9 %). Situés en bordure de la zone d'étude, ces derniers ne possèdent pas de flux amont. De ce fait, ils dépendent uniquement de leurs valeurs historiques et n'exploitent donc pas pleinement le potentiel du modèle. Les bons résultats de la moyenne historique peuvent quant à eux s'expliquer par une certaine régularité des flux de jour en jour.

6 Conclusion

Dans cet article, nous proposons une approche de prévision à court terme des flux de voyageurs du réseau ferré d'Île-de-France basée sur les réseaux bayésiens dynamiques. À travers l'utilisation de l'algorithme EM structurel (pour l'apprentissage) et du filtre bootstrap (pour l'inférence), cette approche permet de prédire les flux en temps réel en présence de données manquantes. L'expérimentation menée sur la ligne 2 du métro de Paris témoigne de sa supériorité globale sur les autres méthodes et met en évidence le rôle clé de l'offre de transport.

Malgré ces résultats encourageants, notre approche rencontre un certain nombre de limites. En particulier, le choix de modèles linéaires gaussiens implique la linéarité des relations entre les variables, ce qui est une hypothèse discutable. Nous pouvons en effet nous demander si l'utilisation de distributions capables de modéliser des processus non linéaires, telles que des modèles de mélanges gaussiens, peut améliorer les performances de prédiction.

Comme le montrent les résultats, le modèle doit encore être amélioré, notamment pour les flux situés en bordure de la zone d'étude. Dans les travaux à venir, nous pourrions exploiter la modularité des réseaux bayésiens pour introduire de nouvelles sources d'information, telles que des facteurs temporels (jour de la semaine, mois, vacances...) ou des événements externes (conditions météorologiques, événements sportifs ou culturels...) potentiellement impactants.

Références

- Dean, T. et K. Kanazawa (1989). A Model for Reasoning About Persistence and Causation. *Computational Intelligence* 5(3), 142–150.
- Friedman, N., K. Murphy, et S. Russel (1998). Learning the Structure of Dynamic Probabilistic Networks. In *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, Madison, WI, USA, pp. 139–147.
- Gordon, N. J., D. J. Salmond, et A. F. M. Smith (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings-F - Radar and Signal Processing* 140(2), 107–113.
- Haworth, J. et T. Cheng (2012). Non-parametric regression for space-time forecasting under missing data. *Computers, Environment and Urban Systems* 36(6), 538–550.
- Li, Q., Y. Qin, Z. Wang, Z. Zhao, M. Zhan, Y. Liu, et Z. Li (2013). The Research of Urban Rail Transit Sectional Passenger Flow Prediction Method. *Journal of Intelligent Learning Systems and Applications* 5(4), 227–231.
- Ma, Z., J. Xing, M. Mesbah, et L. Ferreira (2014). Predicting short-term bus passenger demand using a pattern hybrid approach. *Transportation Research Part C: Emerging Technologies* 39, 148–163.
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco: Morgan Kaufmann.
- Shachter, R. D. et R. C. Kenley (1989). Gaussian influence diagrams. *Management Science* 35(5), 527–550.
- Sun, S., C. Zhang, et G. Yu (2006). A Bayesian Network Approach to Traffic Flow Forecasting. *IEEE Transactions on Intelligent Transportation Systems* 7(1), 124–132.
- van Hinsbergen, C. P. I. J., J. W. C. van Lint, et F. M. Sanders (2007). Short Term Traffic Prediction Models. In *Proceedings of the 14th World Congress on Intelligent Transport Systems*, Beijing, China, pp. 5013–5030.
- Wei, Y. et M.-C. Chen (2012). Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks. *Transportation Research Part C: Emerging Technologies* 21(1), 148–162.
- Whitlock, M. E. et C. M. Queen (2000). Modelling a Traffic Network with Missing Data. *Journal of Forecasting* 19(7), 561–574.

Summary

We propose a dynamic Bayesian network approach to forecast the short-term passenger flows of the urban rail network of the Paris region. The structure of the model is based on the causal relationships between the adjacent flows and allows to integrate the transport service. In the presence of missing data, the learning is performed using the structural expectation-maximization (EM) algorithm. When applying our approach on a metro line, the obtained results overall outperform those of the other tested methods.