

Vers l'utilisation de graphes de liens causaux pour l'amélioration de la prévision court-terme du trafic routier

Julien Salotti¹ Romain Billot^{2,3} Nour-Eddin El Faouzi² Serge Fenet^{1,4} Christine Solnon¹

¹ Univ Lyon, Université Lyon 1/INSA de Lyon, LIRIS, F-69622, Lyon, France

² LICIT, Université de Lyon, F-69000, Lyon, France,

IFSTTAR, LICIT, F-69675, Bron, ENTPE, LICIT, F-69518, Vaulx En Velin

³ Institut Mines-Télécom, Télécom Bretagne UMR CNRS 6285 Lab-STICC, Technopôle Brest Iroise, CS 83818, 29238 Brest Cedex 3, France

⁴ STEEP, Inria Grenoble - Rhône-Alpes

julien.salotti@insa-lyon.fr

Résumé

La prévision à court terme du trafic routier est un exercice complexe en milieu urbain du fait de la forte variabilité des conditions de circulation et de la difficulté de tenir compte des structures spatiales et temporelles d'un tel système complexe. Nous présentons une contribution à l'analyse des dépendances spatio-temporelles entre les différents capteurs de trafic routier d'un réseau urbain. La méthodologie proposée s'inspire des travaux de Runge et al. [11, 8] pour réduire la dimension du problème et reconstruire les dépendances causales entre les différents capteurs du réseau. Les premiers résultats montrent la cohérence des graphes de liens causaux obtenus avec les dynamiques spatio-temporelles du trafic. Une telle méthode permet d'identifier un sous-ensemble de capteurs critiques qui portent en eux un maximum d'information sur la propagation des états de trafic dans un réseau urbain tout en assurant des performances satisfaisantes en généralisation pour la prévision à court terme des conditions de circulation.

Mots Clef

Trafic routier, prévision, dépendances causales, séries temporelles, information mutuelle

1 Contexte et motivation

La prévision court terme de l'état du trafic routier à l'échelle d'une ville concentre un certain nombre de difficultés que l'on retrouve dans bon nombre de systèmes complexes (réseaux d'énergie, internet, climat, événements naturels). Le problème consiste à prévoir l'état du réseau au temps $t + k$ à partir d'un historique de mesures sur différents capteurs jusqu'au temps t , avec k petit (≤ 30 minutes dans notre cas de réseau routier). Les réseaux de trafic routier sont caractérisés par une grande variabilité

d'un pas de temps à l'autre [15] qui rend la prévision difficile pour des horizons courts ou moyens [5]. De plus, ces réseaux sont très sensibles à des événements non récurrents comme les accidents, travaux, ou autres événements météorologiques [7]. Les exigences actuelles des gestionnaires de réseaux urbains et la concurrence de quelques géants du monde des mégadonnées (*i.e.* voir en temps réel le trafic sur google maps est déjà possible) créent un engouement certain pour le développement de méthodes de prévision court-terme aptes à passer à l'échelle en présence de données spatiotemporelles complexes, tout en évitant, d'un point de vue de l'apprentissage statistique, le fléau de la dimension qui s'applique aux grands réseaux de capteurs : l'information utile est diluée et seuls quelques capteurs sont utiles pour obtenir de bonnes performances en généralisation. Ainsi, dans notre contexte urbain les données ne peuvent pas tout, et des méthodes d'analyse des dépendances spatiotemporelles sont nécessaires pour mieux identifier ces capteurs critiques et réduire la dimension de façon intelligente [6, 3].

Dans l'optique d'améliorer la prévision court-terme de trafic, notre contribution se concentre sur l'adaptation d'une méthode capable d'identifier des graphes de liens causaux entre les différents capteurs du réseau. Cette méthode s'appuie sur la théorie de l'information et ne fait pas d'hypothèse sur le modèle générant les données. Elle est donc moins restrictive que les méthodes linéaires s'appuyant sur la corrélation, ce qui est important pour des données complexes de trafic. Elle capture également plus précisément l'aspect temporel de la dépendance pour des données fortement auto-corrélées. Enfin, la complexité de cette méthode dépend plus du degré moyen des nœuds du réseau que du nombre de nœuds, ce qui la rend compatible avec l'étude du réseau routier d'une grande ville.

2 Méthodologie

Cadre théorique. En théorie de l'information, l'*information mutuelle* est une mesure permettant de caractériser la dépendance entre deux variables aléatoires X et Y . Il est parfois utile de qualifier la dépendance entre X et Y conditionnée à une troisième variable Z . On parle alors d'*information mutuelle conditionnelle*.

L'*entropie de transfert* [13] est une mesure de la dépendance entre deux processus stochastiques X_t et Y_t . Cette mesure est asymétrique : $I_{X \rightarrow Y}^{TE}$ désigne la diminution de l'incertitude sur le futur de Y apportée par la connaissance du passé de X , alors que le passé de Y est déjà connu. Il s'agit donc de l'information mutuelle entre Y_t et le passé de X_t conditionnée au passé de Y_t . La définition classique de cette mesure entraîne des problèmes d'estimation liés au *fléau de la dimension*, ce qui impacte négativement la fiabilité des causalités découvertes.

Pour résoudre ce problème d'estimation, [10] introduit une décomposition de la version multivariée de l'entropie de transfert, nommée *entropie de transfert décomposée*. Il s'agit d'intégrer l'entropie de transfert dans le cadre théorique des *modèles graphiques probabilistes*, ce cadre ayant été adapté aux séries temporelles en introduisant la notion de graphe de série temporelle multivariée [1].

La proposition de [10] permet d'extraire un graphe caractérisant les dépendances causales (au sens de Granger [2]) entre les différentes séries temporelles du jeu de données. L'estimation de ce graphe repose sur une adaptation de l'algorithme PC [14], souvent utilisé pour l'inférence de la structure de dépendance d'un réseau bayésien.

Application à l'étude du trafic. Des tronçons de route appartenant au réseau sont équipés de capteurs permettant de qualifier l'état du trafic à cet endroit. Chacun des capteurs mesure à intervalles réguliers des variables physiques telles que le débit (nombre de véhicules/heure) ou le taux d'occupation (lié à la densité du trafic). La séquence des mesures successives sur un tronçon constitue une série temporelle. Notre jeu de données est donc un ensemble de séries temporelles, décrivant chacune l'évolution de l'état du trafic sur un tronçon du réseau.

Après application de l'algorithme de [10], on obtient un graphe de dépendance causale. Soient X_t, Y_t des variables décrivant l'état de deux tronçons. La présence de l'arc orienté $X_{t-\tau} \rightarrow Y_t$ dans le graphe indique que l'état de trafic sur le tronçon X a une influence causale sur l'état du tronçon Y avec un décalage temporel τ . On peut aussi observer une dépendance contemporaine (non orientée) $X_t - Y_t$. L'absence de décalage temporel vient du fait que les valeurs mesurées ne sont pas instantanées ; elles sont le résultat d'une agrégation sur un pas de temps.

3 Premiers résultats

Pour nos premières expérimentations, nous disposons d'un jeu de données d'environ 600 capteurs installés sur les

axes principaux du réseau routier de la métropole de Lyon. Chaque capteur relève sur un tronçon les valeurs du débit et du taux d'occupation, par tranches de 6 minutes.

Afin de qualifier la pertinence des résultats obtenus, deux échantillons composés respectivement de 10 et 20 capteurs ont été prélevés sur une durée d'un mois. L'objectif est de vérifier que les sorties de l'algorithme, obtenues sur des tronçons connus, sont conformes à la vérité terrain.

Pour effectuer l'estimation du graphe, nous avons utilisé TiGraMITE [9, 10, 12]. Les premiers résultats sont encourageants car le graphe de dépendance est conforme à la dynamique du réseau : la proximité physique entre deux capteurs peut être représentée par un arc contemporain non-orienté dans le graphe causal tandis qu'avec un retard d'un pas de temps, on observe des dépendances entre des tronçons adjacents dans le graphe physique, orientées dans le sens d'écoulement du trafic. Avec un retard de deux pas de temps (12 mn), on observe certaines dépendances dans le sens inverse de l'écoulement du trafic : cela est cohérent avec les phénomènes d'ondes de congestion qui remontent le flux de véhicules. Enfin, on observe certains arcs qui ne sont pas directement interprétables physiquement.

Certains résultats, comme le niveau de significativité d'un arc ou la densité du graphe causal, sont sensibles au paramétrage du modèle. Nous explorerons donc plus en profondeur ce paramétrage, en vue d'utiliser le graphe de dépendance pour sélectionner les variables d'intérêt [8] fournies aux modèles de prévision, et ainsi contribuer à une sélection des segments critiques sur un réseau routier.

Remerciements

Ce travail est financé par le LabEx IMU (ANR-10-LABX-0088) de l'Université de Lyon, programme "Investissements d'Avenir" (ANR-11- IDEX-0007) de l'ANR. Les données ont été fournies par la Métropole de Lyon.

Références

- [1] Eichler, M. Graphical modelling of multivariate time series. *Probability Theory and Related Fields*, 153, 233-268, 2011.
- [2] Granger, C.W.J. Investigating causal relation by econometric and cross-sectional method, *Econometrica*, 37 : 424-438, 1969.
- [3] Haworth, J., Shawe-Taylor, J., Cheng, T. and Wang, J. Local online kernel ridge regression for forecasting of urban travel times. *Transportation Research Part C : Emerging Technologies*, 46, 151-178, 2014.
- [4] Hofleitner, A., Herring, R., Abbeel, P., and Bayen, A. Learning the dynamics of arterial traffic from probe data using a dynamic Bayesian network. *Intelligent Transportation Systems, IEEE Transactions on*, 13(4), 1679-1693, 2012.
- [5] Kamarianakis, Y., Shen, W., and Wynter, L. Real-time road traffic forecasting using regime-switching space-time models and adaptive LASSO. *Applied stochastic models in business and industry*, 28(4), 297-315, 2012.

- [6] Li, L., Su, X., Wang, Y., Lin, Y., Li, Z., and Li, Y. Robust causal dependence mining in big data network and its application to traffic flow predictions. *Transportation Research Part C : Emerging Technologies*, 58, 292-307, 2015.
- [7] Pan, B., Demiryurek, U., and Shahabi, C. Utilizing real-world transportation data for accurate traffic prediction. In *Data Mining (ICDM), 2012 IEEE 12th International Conference on* (pp. 595-604). IEEE, 2012.
- [8] Runge, J., Donner, R., and Kurths, J. Optimal model-free prediction from multivariate time series. *Phys. Rev. E*, 91, 052909, 2015.
- [9] Runge, J., Heitzig, J., Marwan, N., Kurths, J. Quantifying Causal Coupling Strength : A Lag-specific Measure For Multivariate Time Series Related To Transfer Entropy. *Physical Review E*, 86, 06112, 2012.
- [10] Runge, J., Heitzig, J., Petoukhov, V., and Kurths, J. Escaping the Curse of Dimensionality in Estimating Multivariate Transfer Entropy. *Phys. Rev. Lett.* 108, 258701, 2012.
- [11] Runge, J., Petoukhov, V., Donges, J. F., Hlinka, J., Jajcay, N., Vejmelka, M. and Kurths, J. Identifying causal gateways and mediators in complex spatio-temporal systems. *Nature communications*, 6, 2015.
- [12] Runge, J., Petoukhov, V., Kurths, J. Quantifying the strength and delay of climatic interactions. *Journal of Climate*, 27, 720-739, 2014
- [13] Schreiber, T. Measuring Information Transfer. *Phys. Rev. Lett.* 85, 461-464, 2000.
- [14] Spirtes, P., Glymour, C., and Scheines, R. *Causation, Prediction, and Search*, MIT, Cambridge, MA, 2000.
- [15] Vlahogianni, E. I., Karlaftis, M. G., and Golias, J. C. Short-term traffic forecasting : Where we are and where we're going. *Transportation Research Part C : Emerging Technologies*, 43, 3-19, 2014.