

PRÉVISION NON PARAMÉTRIQUE. QUELQUES RÉFLEXIONS ISSUES DE CAS PRATIQUES DÉVELOPPÉS À SNCF

Maguelonne Chandesris¹

¹ *SNCF Innovation & Recherche*
40 avenue des terroirs de France, 75611 PARIS CEDEX 12
prenom.nom@sncf.fr

Résumé. Des approches de modélisation non-paramétriques ont été développées pour différents cas d'étude de prédiction à SNCF. Ces cas d'études seront brièvement présentés et quelques réflexions suscitées par ces cas discutées.

Mots-clés. non-paramétrique, prédiction, cas d'étude, transports

Abstract. Some non-parametric models have been developed to cope with SNCF prediction use cases. After a brief description of the use cases, some thoughts about it will be then discuss.

Keywords. non-parametric, prediction, use case, transportation

1 Prévoir par transduction

La prévision non-paramétrique (dont les approches en régression sont remarquablement synthétisés dans Dreesbeke et Saporta (2011)) peut être interprétée comme une approche transductive : le futur prédit $\hat{Y}_p^{np}(t)$ pour la date $t = p + s$ étant estimé à la date p à l'aide d'une pondération des futurs des passés similaires choisis dans l'historique de $i = p - l$ à p à l'aide du vecteur $X(i)$, cf. par exemple p262 de l'ouvrage qui vient d'être cité.

$$\hat{Y}_p^{np}(t) = \sum_{i=p-l}^p \omega(i, t) Y(i)$$

avec la pondération entre les dates i et j donnée par la fonction

$$\omega(i, j) = \frac{K\left(\frac{\Delta(i, j)}{h}\right)}{\sum_{i=p-l}^p K\left(\frac{\Delta(i, j)}{h}\right)}$$

La fonction ω faisant elle-même appel au noyau K , à la fenêtre h et à une fonction de similarité entre les moments i et j , $\Delta(i, j)$, par exemple la distance euclidienne entre les vecteurs $X(i)$ et $X(j)$. À noter qu'ici, on ne souhaite pas se restreindre au cas où l'on

compare des portions du passé $X(i) = (Y(i), Y(i - 1), \dots, Y(i - u))$, dans la suite le vecteur X peut contenir des portions du passé de Y mais pas uniquement.

Cette approche généralise l'approche des k -plus proches voisins, le noyau K et la similarité Δ permettant d'adapter la manière dont les passés proches au sens de la similarité sont pris en compte et pondérés.

2 Cas d'étude

Ce type d'approche a été mis en oeuvre à la direction Innovation & Recherche de SNCF sur différents cas d'étude que nous présentons ici brièvement.

2.1 Prévision court terme de séries chronologiques

Le premier type de cas d'étude est la prévision de série chronologique journalières ou horaires, à la manière de Poggi (1994). La variable Y représentée sur la figure 1 peut être un chiffre d'affaire, une fréquentation, un nombre de requêtes, etc.

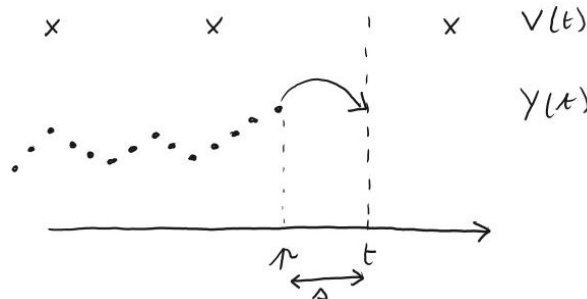


FIGURE 1 – Prévision d'une série chronologique à horizon s

La spécificité de nos cas d'étude est que l'on souhaite prévoir des (dizaines de) milliers de séries à l'aide de ces techniques. Nous souhaitons également pouvoir prendre en compte :

- des événements particuliers dont l'occurrence est connue à l'avance (cf. par exemple Chandèsris et Riss (2009)), par exemple sous la forme d'un vecteur $V(t)$, connu quelque soit t (y compris dans le futur).
- une information connue au fur et à mesure que l'on se rapproche de l'échéance : la variable prédite résulte d'un processus cumulatif connue progressivement et on cherche à prévoir le point d'arrivée à date (cf. par exemple Chandèsris et Rémy (2017)) $Y(t) = \sum_{i=t-q}^t Z(i)$, cf. figure 2.

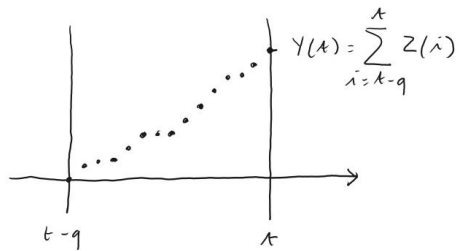


FIGURE 2 – Processus cumulatif connu progressivement conduisant à $Y(t)$

2.2 Prédiction temps réel de séries discrètes

Un autre cas d'application est de prédire l'évolution d'une variable pour une série discrète, typiquement Y est l'affluence et/ou le retard d'un train au cours de son parcours (cf. à ce sujet les premières expérimentations menées dans le domaine dans Chandresris (2014)).

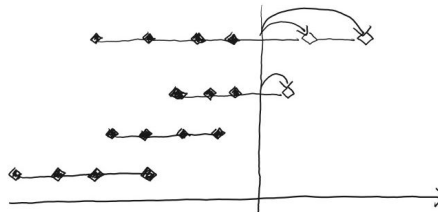


FIGURE 3 – Prédiction de l'affluence et/ou le retard de trains pour leurs prochains arrêts

L'ambition est de pouvoir fournir des prédictions de bonne qualité en temps réel, là encore sur des volumes importants. À titre d'illustration, le réseau Transilien d'Île-de-France opère plus de 7 500 trains par jour, chacun d'eux s'arrêtant plus d'une dizaine de fois.

3 Quelques réflexions

Le développement de ces différents cas pratiques avec une visée opérationnelle (mise en production dans un contexte industriel) nous conduit à différentes réflexions que nous souhaitons partager et soumettre à discussion.

3.1 Principes de fonctionnement

Le **mode de fonctionnement** de la méthode de prévision non-paramétrique présente une originalité intéressante par rapport aux autres méthodes "classiques" d'apprentissage décrites par exemple dans Hastie, Tibshirani et Friedman (2009).

Les méthodes "classiques" se déroulent en deux étapes :

1. une étape dite d'"apprentissage" qui permet, à partir de la connaissance des historiques $\{(X(i), Y(i))\}_{i=1}^n$, d'estimer le modèle $F_{\hat{\theta}_{(1,n)}}^c$ dépendant du paramètre $\hat{\theta}_{(1,n)}$ ajustant au mieux le modèle sur l'historique $(1, n)$.
2. la deuxième étape permet de réaliser la prévision pour une nouvelle observation $X(t) : F_{\hat{\theta}_{(1,n)}}^c(X(t)) = \hat{Y}_{\hat{\theta}_{(1,n)}}^c(t)$. Cette étape nécessite le calcul de $F_{\hat{\theta}_{(1,n)}}^c(X(t))$, donc l'accès à $X(t)$ et à $\hat{\theta}_{(1,n)}$ (qui "résume" l'historique $\{(X(i), Y(i))\}_{i=1}^n$) et aux opérations nécessaires au calcul de F^c . Ces opérations sont généralement simples (additions et multiplications) en fonction des types de modèles retenus (régression, arbre de décision, réseaux de neurones, etc.).

Dans l'approche non-paramétrique, il n'existe qu'une seule étape de calcul de la prévision : $F^{np}(\{(X(i), Y(i))\}_{i=p-l}^p, h, K, \Delta, X(t)) = \hat{Y}_p^{np}(t)$. Cette étape est certes plus complexe que l'étape de prévision des méthodes "classiques" : elle nécessite d'avoir accès aux historiques des valeurs (ici de $(i = p - l, p)$) et de calculer (un grand nombre de) comparaisons, puis une somme pondérée. Pour autant, cette méthode ne nécessite pas d'"apprentissage" ou de ré-apprentissage : l'historique pris en compte évolue au fur et à mesure du temps (par exemple $p = t - 1$) et les prévisions vont automatiquement s'adapter aux évolutions du phénomène.

Dans le cas d'une mise en oeuvre sur des milliers de prédictions, le procédé permet donc de s'abstraire de milliers de modèles et de paramètres qu'il faudrait remettre à jour de manière périodique.

Non-cumul des erreurs Dans le cas de calcul de prévisions à des horizons différents plus ou moins lointains, il est à noter que les erreurs ne sont pas cumulatives : les deux prévisions calculées au temps p , $\hat{Y}_p^{np}(t + s)$ et $\hat{Y}_p^{np}(t)$ sont indépendantes.

L'**interprétation** du fonctionnement ce type d'approche est très simple en exploitant la notion de similarité, par exemple à partir du cas dégénéré des k -plus proches voisins. Le fait que les prévisions soient directement calculées à partir de valeurs précédentes de la variable ne nécessite pas de recourir à la notion abstraite (et parfois mal comprise) de modèle. Dans les modèles "classiques", la prévision est calculée à partir d'autres variables et du paramètre $\hat{\theta}^1$. Dans certains cas, l'interprétation peut s'avérer difficile mais dans tous les cas, on fait appel à la notion abstraite de modèle qui n'est pas toujours bien saisie.

1. faisant ici respectivement appel à un raisonnement de déduction pour le calcul et d'induction pour l'estimation des paramètres du modèle

Dans le cadre non-paramétrique, le mécanisme de prévision est compréhensible par tous les utilisateurs² qui peuvent si nécessaire, analyser les résultats en accédant aux historiques utilisés (et mieux appréhender que le mécanisme "se trompe" parfois).

3.2 Choix des variables et de la fonction de similarité

Le choix du vecteur X pour le calcul de la similarité est bien évidemment une question clé pour la performance de l'approche. Plusieurs questions se posent :

- Comme dit en introduction, on ne souhaite pas se restreindre au cas où l'on compare uniquement des portions du passé entre elles, c'est-à-dire $X(i) = (Y(i), Y(i-1), \dots, Y(i-u))$. La question est donc de savoir comment introduire d'autres variables dans le vecteur X tout en évitant le fléau de la dimension :
- on peut bien entendu avoir recours aux techniques de réduction de la dimension, en particulier les splines (comme dans Antoniadis et al. (2012)) ou les modèles à directions révélatrices (cf. le chapitre 7 de Droesbeke et Saporta (2011)), mais ces approches ont le défaut de restreindre l'interprétation de la similarité.
- le choix du type de variable à introduire dans le vecteur de comparaison est à effectuer en fonction du type de l'approche retenue : dans un modèle "classique", nous aurons tendance à choisir des variables dites "explicatives" qui vont aider à "résumer" le phénomène dans un modèle. Mais ici, le principe va plutôt dans le sens de privilégier des observations de configurations caractéristiques qui pourront être rapprochées dans l'historique. On cherchera donc à donner la priorité aux observations Y , Z ou autres selon les cas.
- la connaissance d'un contexte a priori V n'est bien évidemment pas à négliger et on peut très bien imaginer inclure dans le vecteur $X(i)$ la connaissance de $V(i+s)$.
- Dans quelle mesure doit-on favoriser les proximités temporelles (et/ou spatiales) ? C'est-à-dire pour un même écart entre les vecteurs $(X(i), X(j_1))$ et $(X(i), X(j_2))$ pour lesquels $j_1 < j_2 < i$, faut-il favoriser en le surpondérant $Y(j_2+s)$?
- Quelle taille d'historique glissant est à considérer ? Augmenter l permet de prendre en compte un historique de référence plus important, accroissant d'autant le nombre de comparaisons à effectuer mais en rendant sans doute plus robuste l'estimation finale.

3.3 Performance, limites et perspectives

Dans les cas d'étude qui ont fait l'objet de travaux et qui ne peuvent être tous détaillés pour des raisons de confidentialité, les modèles non-paramétriques ont permis d'améliorer la qualité des prévisions tant en biais qu'en variance par rapport aux méthodes traditionnellement utilisées.

2. les enfants eux-mêmes recourent au raisonnement de type transduction

Le gain en terme de simplification des processus industriels de prévision est également à souligner : diminution drastique du nombre de modèles et de paramètres. La fréquence de mise à jour (recalibration) est elle aussi un effet notable. Pour autant, il ne nous semble pas qu'il s'agisse de la "fin des modèles" telle que décrite par Anderson (2008) : il s'agit d'une forme particulière de modèles pour lesquels le choix des variables et des méta-paramètres (choix de la distance en particulier) reste un travail d'expertise déterminant pour la qualité des prévisions.

En dépit des moyens supplémentaires nécessaires au moment du calcul de la prédiction par rapport aux méthodes traditionnelles, les temps de calcul actuels permettent d'envisager un fonctionnement "temps réel" (alimentation au fil de l'eau en $X(t)$ et calcul à la volée des $\hat{Y}_{t-1}^{np}(t)$), ce qui fait actuellement l'objet de travaux de recherche complémentaires.

Remerciements L'auteur tient à remercier Perrine Bouche, Xavier Chapuis, Sonia Pelloux et Anaïs Rémy pour les travaux qu'ils ont menés dans le domaine et les échanges riches qui ont nourri cette synthèse.

Bibliographie

- [1] Anderson, C. (2008), *The data deluge makes the scientific method obsolete*, WIRED.
- [2] Antoniadis, A. , Brosat, X., Cugliari, J. et Poggi, J.-M. (2012), *Prévision d'un processus à valeurs fonctionnelles en présence de non stationnarités*, Rapport de recherche, RR-7982, INRIA.
- [3] (à venir) Chandesris, M. et Rémy, A. (2017), *Passive data to help the traffic demand anticipation*, *3rd International Workshop and Symposium Transitdata*, Santiago, Chili.
- [4] Chandesris, M. et Riss, M. (2009), *Anticiper les pics et creux de trafic*, *41ème journées de Statistique de la SFDS*, Bordeaux.
- [5] Chandesris, M. (2014), *Approche non-paramétrique pour la prédiction d'heure d'arrivée dans les transports*, *46ème journées de Statistique de la SFDS*, Rennes.
- [6] Droesbeke, J.-J. et Saporta, G. (2011), *Approches non paramétriques en régression*, Editions TECHNIP, Paris.
- [7] Hastie, T., Tibshirani, R. et Friedman, J. (2009), *The elements of statistical learning. Datamining, inference and prediction*, Springer, Paris.
- [8] Poggi, J.-M. (1994), *Prévision non-paramétrique de la consommation électrique*, *Revue de Statistique Appliquée*, XLII (4), 83-98.