



HAL
open science

Anticiper les pics et creux de trafic

Maguelonne Chandesris, Mariane Riss

► **To cite this version:**

Maguelonne Chandesris, Mariane Riss. Anticiper les pics et creux de trafic. 41èmes Journées de Statistique, SFdS, Bordeaux, 2009, Bordeaux, France, France. inria-00386616

HAL Id: inria-00386616

<https://inria.hal.science/inria-00386616>

Submitted on 22 May 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ANTICIPER LES PICS ET CREUX DE TRAFIC

Maguelonne CHANDESRIS & Mariane RISS

prenom.nom@sncf.fr

SNCF, Direction de l'Innovation et de la Recherche
45, rue de Londres
75379 PARIS CEDEX 08

MOTS CLEFS

Analyse des données - data mining, Apprentissage

RESUME

Anticiper le trafic voyageur constitue un enjeu déterminant pour un exploitant ferroviaire, cela lui permet en effet d'ajuster au mieux son offre capacitaire et tarifaire. Or un événement à une date particulière peut induire une augmentation (induction) ou diminution (désinduction) du trafic voyageur par rapport à "la normale" sur une région, sur un marché (trajet entre une gare origine et gare destination), une gare, un matin ou après-midi... Ces phénomènes peuvent être observés au niveau d'un périmètre spatial (national, un marché, un ensemble de marchés, une gare...) et d'un périmètre temporel (journée, semaine, mois, etc.). Citons par exemple :

- le lundi de Pentecôte induit une désinduction de demande le dimanche précédent.
- le festival de la BD à Angoulême provoque des inductions de demande à destination et en origine d'Angoulême.

La Direction Innovation & Recherche de la SNCF a développé une méthodologie statistique originale nommée "Calendrier Commercial" permettant d'anticiper l'influence des événements qui marquent l'année sur le trafic voyageurs. Cette méthodologie permet de mettre en évidence et de prévoir les jours sur lesquels la demande de voyage sera plus élevée ou moins élevée qu'à "la normale", en raison d'un événement particulier (jour férié, vacances, manifestation culturelle, etc.).

L'objet de cette communication est de présenter la démarche et les méthodes statistiques développées dans le "Calendrier Commercial".

ABSTRACT

Anticipating passenger traffic is a key issue for a rail operator, it allows to best adjust both capability and price offer. One specific event at a particular date may induce an increase (induction) or decrease (disinduction) on passenger traffic compared to "normal" on a region, in a market (a travel between origin and destination stations), a station, one morning or afternoon ... These phenomena can be observed at a space level (national, market, all markets, a station ...) and a time level (day, week, month, etc.). Examples:

- Whit Monday causes disinduction on demand on the previous Sunday.
- The comics festival in Angoulême causes both induction and disinduction from and to Angoulême city.

The Innovation & Research department of SNCF has developed an original statistical methodology

called "Calendrier Commercial" to anticipate the impact of events that occur during the year on passenger traffic. This methodology allows to identify and anticipate days on which travel demand will be higher or lower than "normal", due to a specific event (holiday, vacation, cultural event, etc.).

The issue of this paper is to present the approach and statistical methods developed in the "Calendrier Commercial".

ARTICLE

Contexte : prévisions de volume de production

Anticiper des volumes de production liés à des événements connus n'est pas une problématique spécifique à la SNCF et de très nombreux travaux, y compris actuels, abordent ce problème de prévision aussi bien du point de vue théorique que pratique, cf. à ce sujet S. Makridakis, S. C. Wheelwright et R. J. Hyndman (1998). Citons par exemple les travaux sur les prévisions de consommation électrique V. Lefieux (2007) ou les prévisions de trafic aérien ou routier M. Danech-Pajouh (2003). Ces travaux proposent différentes approches pour résoudre le problème de prévision quantitative de manière exhaustive. Mais quelque soient les méthodes proposées, les "jours exceptionnels", pouvant représenter un pourcentage important des valeurs à prévoir, sont souvent mal prévus par des méthodes automatiques, mêmes sophistiquées. En pratique, ces jours sont traités par des méthodes "ad hoc" et/ou par des experts métiers.

L'objectif des travaux menés par la Direction Innovation & Recherche de la SNCF était justement de se concentrer sur ces jours particuliers et d'évaluer dans quelle mesure il était possible de les identifier et de les prévoir automatiquement. En effet, suivant les niveaux d'agrégation spatiaux retenus les volumes de données à traiter peuvent être relativement conséquent.

Préambule : formalisation et approche retenue

Les volumes de demande de voyages journaliers sont sensibles aux événements calendaires du moment. Si l'on appelle "normal" le volume de réservations pour un jour donné (non influencé par un événement particulier), alors on note :

$$\text{réservations} = \text{réservations "normales"} \pm \text{réservations induites par les événements calendaires (1)}$$

Si le second terme n'est pas nul, on parle d'induction ou de désinduction de demande de voyage liée à des événements calendaires.

L'approche retenue pour, à terme, anticiper les pics et creux de trafic futurs est de modéliser, par apprentissage statistique, le phénomène d'inductions/désinductions. La démarche se décompose donc en trois phases :

1. Détection des pics et creux sur les historiques
2. Apprentissage des règles d'induction à partir des pics et creux et des historiques d'événements
3. Application des règles pertinentes aux événements futurs

Détection de pics et creux

Pour identifier les pics et creux de trafic, l'idée est d'utiliser une décomposition de la série étudiée en différentes composantes. Classiquement, la décomposition retenue peut être additive ou multiplicative selon le contexte :

$$X = T + S + I \text{ (décomposition additive)}$$
$$X = T \times S \times I \text{ (décomposition multiplicative)}$$

T est la tendance, S la saisonnalité, I l'irrégularité restante c'est-à-dire l'induction que l'on cherche à identifier.

En s'appuyant sur les nombreuses techniques existantes dans le domaine – citons notamment D. Ladiray, B. Quenneville (2000) - et sur les spécificités de la problématique étudiée, la méthode de décomposition retenue est la suivante :

1. Désaisonnalisation permettant de prendre en compte les effets liés au jour de la semaine
2. Estimation de la tendance par lissage polynômial local
3. Evaluation d'une distance locale sans dimension
4. Détection par seuil sur cette distance

Les étapes 2 à 4 sont réitérées plusieurs fois de manière à identifier plusieurs catégories de pics/creux. Au final, la méthode permet d'identifier plusieurs catégories de d'inductions sous réserve de fixer un certain nombre de paramètres : paramètre de lissage, largeur de la fenêtre, seuils de détection.

Règles d'induction

En s'inspirant des travaux menés dans le domaine des règles d'association – cf. G. Piatetsky-Shapiro (1991) ou J. Blanchard (2005), la modélisation adoptée pour les inductions s'énonce de la manière suivante :

« Si un événement de type T a lieu le jour J qui est un JS,
il peut créer de l'induction/désinduction les jours $[J - f, J + f]$ »

f est la taille de la "fenêtre d'influence d'un événement, par exemple $f = 5$.

L'application de l'une de ces règles constitue une "induction élémentaire" :

PLAN DE TRANSPORT + EVENEMENTS + REGLES => INDUCTION PREVUES

Le plan de transport définit ici l'offre de transport disponible, suivant les caractéristiques spatiales et temporelles.

Une fois ce modèle postulé, il faut en faire l'apprentissage. A partir des historiques de pics et creux et de la liste des événements qui ont lieu sur la même période, on génère l'ensemble des règles applicables :

INDUCTIONS OBSERVEES + EVENEMENTS PASSES => REGLES APPLICABLES

Puis, par comparaison entre les inductions observées et prévues par les règles applicables, on ne conserve que les règles plus efficace que le hasard, pour une règle donnée il faut que :

$$\frac{\#(\text{INDUC_PREV} = \text{INDUC_OBS})}{\#(\text{INDUC_PREV})} > \frac{1}{3}$$

(exemple donnée ici pour trois niveaux d'inductions -/0/+)

Une fois les règles pertinentes sélectionnées, on peut les appliquer au plan de transport et événements futurs :

PDT FUTUR + EVENEMENTS FUTURS + REGLES PERTINENTES => INDUCTION FUTURES

La validation est réalisée en découpant les historiques en période d'apprentissage et de validation.

Exemple d'application

Cette méthodologie a été développée et appliquée avec succès à l'ensemble du trafic TGV France et ce, pour différents niveaux d'agrégation spatiaux (national, marchés, gares, etc.) et temporels.

Bibliographie

- [1] J. Blanchard (2005) *Un système de visualisation pour l'extraction, l'évaluation, et l'exploration interactives des règles d'association*. Thèse de doctorat en informatique, Ecole polytechnique de l'université de Nantes.
- [2] M. Danech-Pajouh (2003) *Les modèles de prévision du dispositif Bison futé et leur évolution*, Recherche Transports Sécurité n°78 janvier-mars 2003
- [3] D. Ladiray, B. Quenneville (2000) *Désaisonnaliser avec la méthode X-11*, US CENSUS bureau, http://www.census.gov/ts/papers/x11_french.pdf
- [4] V. Lefieux (2007). *Modèles semi-paramétriques appliquées à la prévision de séries temporelles. Cas de la consommation d'électricité*. Thèse de doctorat en statistiques, Université de Rennes 2
- [5] S. Makridakis, S. C. Wheelwright et R. J. Hyndman (1998), *Forecasting methods and application*. Third edition. Wiley.
- [6] G. Piatetsky-Shapiro (1991), *Discovery, analysis, and presentation of strong rules*, in G. Piatetsky-Shapiro & W. J. Frawley, eds, 'Knowledge Discovery in Databases', AAAI/MIT Press, Cambridge, MA.