

# Aide à la décision pour la maintenance ferroviaire préventive

Julien Rabatel<sup>\*,\*\*</sup>, Sandra Bringay<sup>\*,\*\*\*</sup>, Pascal Poncelet<sup>\*</sup>

<sup>\*</sup>LIRMM, Université Montpellier 2, CNRS

161 rue Ada, 34392 Montpellier Cedex 5, France

<sup>\*\*</sup>Fatronik France Tecnalía, Cap Omega, Rond-point Benjamin Franklin - CS 39521

34960 Montpellier, France

<sup>\*\*\*</sup>Dpt MIAP, Université Montpellier 3, Route de Mende

34199 Montpellier Cedex 5, France

{rabatel,bringay,poncelet}@lirmm.fr

**Résumé.** La maintenance de trains est un problème particulièrement délicat lié à de nombreux enjeux à la fois financiers, sécuritaires et énergétiques. Nous nous intéressons à la mise en place d'une maintenance préventive basée sur la détection et la correction de tout comportement anormal susceptible de provoquer un problème majeur dans un futur proche. Nous proposons ainsi un outil d'aide à la décision afin de (i) dégager des connaissances utiles sur l'historique des trains, et (ii) détecter et étudier les anomalies comportementales, dans le but de prendre des décisions optimales en termes de maintenance ferroviaire.

## 1 Introduction

Minimiser les coûts liés à la maintenance d'équipements complexes est un enjeu important pour les entreprises du secteur industriel. En effet, les coûts sont continuellement à la hausse. Wireman (1990) a montré qu'aux États-Unis, les frais de maintenance de telles entreprises ont augmenté de 10-15% tous les ans entre 1979 et 1990. Les stratégies classiques de maintenance ne sont pas suffisantes. Par exemple, la maintenance curative, consistant à remplacer ou réparer les équipements après l'apparition d'une panne n'est pas satisfaisante. Elle est onéreuse et pose des problèmes liés à la sécurité et à la consommation d'énergie. De même, une maintenance systématique et planifiée des équipements avant l'émergence d'une panne est trop coûteuse (le matériel est remplacé sans tenir compte de son état réel).

Il est donc important de mettre en œuvre une stratégie de maintenance préventive en anticipant les pannes et dysfonctionnements. Dans cet article, nous nous intéressons à la maintenance ferroviaire préventive, et proposons un outil d'aide à la décision afin de (i) dégager des connaissances utiles pour les utilisateurs à propos des trajets de trains passés, et (ii) détecter les anomalies comportementales pouvant être employées pour alerter l'expert. De plus, cet outil apporte de nombreuses informations pour permettre à l'expert d'analyser les causes possibles ou la gravité des anomalies détectées, afin de prendre des décisions optimales en termes de maintenance.

La maintenance préventive se décompose en trois étapes (Lee et al. (2004)) : (i) caractériser le comportement du système étudié, (ii) traiter les données pour l'analyse des comportements,

et (iii) prendre les décisions pour préconiser une politique de maintenance adaptée à l'état de l'équipement. C'est sur cette dernière étape que se focalisent nos travaux. Les techniques pour assister la prise de décision dans le domaine de la maintenance préventive peuvent traditionnellement être divisées en deux catégories principales : le *diagnostic* et le *pronostic* (Jardine et al. (2006)). La première traite la détection et l'identification de comportements anormaux, tandis que la seconde vise à prédire ces comportements avant leur apparition. L'établissement de diagnostics s'apparente au problème de la détection d'anomalies dans des données. Défini par Grubbs (1969), une anomalie ou outlier est une observation qui dévie nettement des autres membres de l'échantillon dans lequel elle apparaît. Plus récemment, différentes définitions ont été proposées et de nombreuses approches ont été développées pour répondre aux besoins de domaines d'applications divers, tels que le suivi de matériel industriel (Keogh et al. (2006)), mais également la détection d'intrusion (Shaikh et al. (2008)), etc. L'élaboration de pronostics a reçu moins d'attention dans la littérature. Il s'agit généralement de prédire le temps restant avant l'apparition d'une panne, en tenant compte du comportement actuel de l'équipement et de son âge (Li et Nilkitsaranont (2009), Samanta et Nataraj (2008)).

La problématique que nous abordons s'appuie sur l'établissement de diagnostics de comportements anormaux. Néanmoins, nous proposons de repérer ces comportements déviants suffisamment tôt pour anticiper les dysfonctionnements importants. De plus, il est indispensable dans notre cas de proposer une approche qui fournit des résultats interprétables par les utilisateurs.

La suite de l'article est organisée de la manière suivante. La section 2 décrit l'approche développée pour extraire des connaissances et détecter des anomalies. La section 3 décrit l'outil implémentant cette approche. Enfin, nous concluons dans la section 4.

## 2 Manipulation et caractérisation des données ferroviaires

**Description et manipulation des données.** Le système de surveillance des trains utilisé dans nos travaux s'appuie sur une grande quantité de capteurs répartis sur les trains étudiés. Chaque train est composé de 8 bogies, à raison de 2 par wagon, et sur chacun d'eux, 32 capteurs mesurent des informations diverses telles que des températures, des accélérations et la vitesse. Toutes les 5 minutes lorsqu'un train est en marche, chaque capteur collecte une valeur qui est stockée dans une base de données centralisée. Une description plus complète de ces données est fournie par Carrascal et al. (2009).

Les données issues de capteurs sont complexes pour deux raisons : (i) des erreurs diverses dégradent fréquemment les données (e.g., du bruit, des valeurs manquantes, etc.) (ii) des informations de types différents doivent être traitées simultanément (e.g., températures, vitesses, etc.).

De telles données sont présentées dans la figure 1. Chaque **capteur** décrit une propriété du comportement global d'un train. Une **mesure** est une valeur numérique collectée par un capteur à une date donnée. Notons que les valeurs sont discrétisées afin d'obtenir un jeu de données mieux adapté à l'étape de fouille de données décrite par la suite. Un **relevé**, i.e., l'ensemble des mesures pour une date donnée, décrit l'état global du train pour cette date.

**Classes de données.** Les données que nous manipulons sont associées à des informations contextuelles. Dans le cas de la maintenance ferroviaire, il faut tenir compte de l'influence de

| DATES                 | CAPTEURS |     |     |     |
|-----------------------|----------|-----|-----|-----|
|                       | A        | B   | C   | ... |
| 2008/03/27 06: 36: 39 | 0        | 16  | 16  | ... |
| 2008/03/27 06: 41: 39 | 82.5     | 16  | 16  | ... |
| 2008/03/27 06: 46: 38 | 135.6    | 19  | 21  | ... |
| ...                   | ...      | ... | ... | ... |

FIG. 1: Extrait de données brutes.

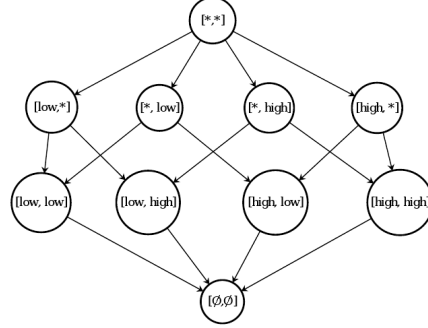


FIG. 2: Treillis de classes.

paramètres environnants sur le comportement d’un train. Par exemple, un train se comportera différemment en fonction des conditions climatiques. Plus formellement, nous définissons des classes contextuelles qui décrivent tous les contextes disponibles. Une telle classe est notée  $[c_1, \dots, c_i, \dots, c_k]$ , où  $c_i$  est la valeur associée à  $c$  pour la dimension  $D_i$ . Par exemple, considérons les dimensions *Température Extérieure* et *Humidité*. La classe  $[low, high]$  désigne les données enregistrées avec une température extérieure basse et un taux d’humidité élevé. La classe notée  $[high, *]$ , plus générale, contient l’ensemble des données enregistrées avec une température extérieure élevée, pour n’importe quel taux d’humidité. La classe  $[*, *]$  est la classe la plus générale. Ainsi, il existe un ordre de spécialisation/généralisation entre les classes. Cet ordre est utilisé pour construire un treillis (cf. figure 2).

**Caractérisation des comportements normaux.** L’objectif de la caractérisation des comportements normaux est, à partir d’une base de données de capteurs, de fournir une liste de motifs décrivant au mieux les comportements normaux.

Dans le cadre de nos travaux, nous considérons que les motifs peuvent être représentés sous la forme de motifs séquentiels (Agrawal et Srikant (1995)). Le but de la recherche de motifs séquentiels est d’extraire les ensembles d’items qui sont fréquemment associés au cours du temps. Par exemple, dans le cas de données liées à l’achat de produits dans un supermarché, un motif séquentiel est : “40% des clients achètent une télévision, puis achètent un lecteur DVD”. Ces motifs ont l’avantage d’être compréhensibles et sont donc particulièrement appréciables pour assister l’analyse d’un expert. Nous adaptons ci-dessous ce problème aux données ferroviaires. Dans les définitions suivantes, nous considérerons un ensemble de capteurs, noté  $\Omega$ . Un **item**  $A_v$  est une paire  $\{A, v\}$ , où  $A \in \Omega$  et  $v \in dom(A)$ . Il représente la valeur  $v$  collectée par un capteur à un temps donné. Soit  $\omega = \{A^1, \dots, A^i, \dots, A^p\}$  un ensemble de capteurs tel que  $\omega \subseteq \Omega$ . Un **itemset**  $I$  est un relevé ou une partie d’un relevé, i.e., un ensemble non ordonné d’items, noté  $I = (A_{v_1}^1, A_{v_2}^2, \dots, A_{v_i}^i, \dots, A_{v_p}^p)$ . Une **séquence**  $s$  est une liste d’itemsets notée  $s = \langle I_1 I_2 \dots I_i \dots I_q \rangle$ , où l’itemset  $I_i$  désigne le  $i^{ème}$  itemset de  $s$ . Par exemple, les données brutes décrites dans la figure 1 peuvent être traduites par la séquence  $\langle (A_{low} B_{low} C_{low}) (A_{avg} B_{low} C_{low}) (A_{high} B_{avg} C_{avg}) \rangle$ , une fois ces données discrétisées.

Soient deux séquences  $s = \langle I_1 I_2 \dots I_m \rangle$  et  $s' = \langle I'_1 I'_2 \dots I'_n \rangle$ . S'il existe des entiers  $1 \leq i_1 < i_2 < \dots < i_m \leq n$  tels que  $I_1 \subseteq I'_{i_1}$ ,  $I_2 \subseteq I'_{i_2}$ , ...,  $I_m \subseteq I'_{i_m}$ , alors la séquence  $s$  est une **sous-séquence** de la séquence  $s'$ , notée  $s \sqsubseteq s'$ . On dit que  $s'$  *supporte*  $s$ . Si une séquence  $s$  n'est pas une sous-séquence d'une autre séquence, alors  $s$  est dite **maximale**.

Le **support** d'une séquence est la fraction de séquences dans une base de séquences  $DB$  qui la supportent. Une séquence est **fréquente** si son support est plus grand ou égal à un seuil  $minSupp$ , spécifié par l'utilisateur.

Le problème de la recherche de motifs séquentiels pour la maintenance ferroviaire est finalement, pour un seuil  $minSupp$  et une base de séquences  $DB$ , de trouver toutes les séquences fréquentes dans  $DB$ . De plus, nous souhaitons prendre en compte les différentes informations contextuelles afin de caractériser au mieux les comportements normaux. Pour ce faire, nous recherchons les motifs correspondant à chacune des classes définies précédemment.

**Définition 1** Une séquence  $s$  est une **séquence c-générale** si  $s$  est fréquente dans toutes les classes filles de  $c$ . Si  $c$  est une classe spécialisée, i.e., dont la seule classe fille est  $[\emptyset, \emptyset]$ , alors l'ensemble des séquences c-générales dans  $c$  est l'ensemble des séquences fréquentes de  $c$ .

Chaque classe  $c$  est décrite par un triplet  $c = \langle \mathcal{D}, \mathcal{E}, \mathcal{S} \rangle$ , où  $\mathcal{D}$  est la **description** de  $c$ ,  $\mathcal{E}$  est l'**ensemble des séquences** correspondant aux trajets effectués dans le contexte associé,  $\mathcal{S}$  est l'**ensemble des séquences c-générales** de  $c$  décrivant les comportements caractérisant  $c$ .

### 3 Aide à la décision

Dans cette section, nous nous intéressons à l'analyse de nouveaux trajets par rapport à la représentation des comportements normaux décrits par des motifs séquentiels. Pour cela, nous décrivons un outil permettant d'aider le décideur lorsque des comportements inattendus interviennent. Son objectif est donc (i) de restituer les données pour l'utilisateur et (ii) d'aider l'utilisateur à prendre les décisions pertinentes de maintenance.

Parmi ses fonctionnalités, l'outil alerte les preneurs de décision en cas d'anomalie et associe des informations à ces alertes pour en faciliter l'interprétation. Ainsi, ils pourront répondre à des questions comme : *Quel composant est en cause ? Quand l'anomalie est elle apparue ? S'agit-il réellement d'une anomalie ?*

Pour ce faire, l'approche s'appuie sur le calcul d'un score de conformité qui, à partir d'une liste de motifs séquentiels et d'une séquence représentant un nouveau trajet, fournit un score pour chaque item (i.e., chaque capteur) à chaque relevé (i.e., chaque itemset) défini entre -1 et 1. Cette valeur est telle que : le comportement du capteur au relevé donné est considéré comme normal (respectivement anormal) si sa mesure de conformité est proche de 1 (respectivement -1). Plus de détails sur la définition de cette mesure sont disponibles dans Rabatel et al. (2009).

**Visualisation des anomalies.** Le score de conformité permet de détecter des anomalies liées à un capteur. Or, l'état d'un composant est décrit par plusieurs capteurs. Par exemple, une roue possède 4 capteurs de température physiquement très proches et relevant des informations similaires. Le score peut donc être généralisé au niveau du composant en considérant la moyenne des scores des capteurs. Nous pouvons alors distinguer deux types d'anomalies illustrés dans la figure 3. Dans le premier cas, 3 des 4 capteurs (cercles) ont un score élevé (traduit par la

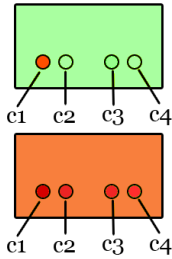


FIG. 3: Cas d'anomalies.

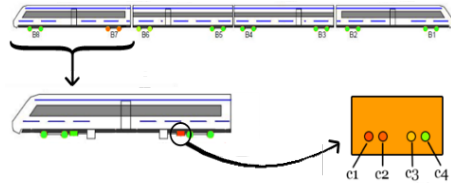


FIG. 4: Différents niveaux de granularité pour la visualisation des anomalies.

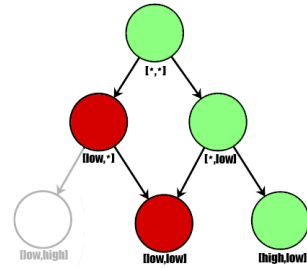


FIG. 5: Propagation des scores de conformité.

couleur vert clair) tandis que le dernier (le capteur  $c1$ ) a un score bas (rouge foncé), témoin d'une anomalie. Cependant, le score moyen du composant est élevé. Il s'agit donc d'un problème de capteur, et le comportement du composant lui-même n'est pas en cause. En revanche, le second cas est différent : le score moyen du composant est bas. Par conséquent, l'anomalie est liée au composant, et non à un capteur défectueux. L'outil rend donc possible l'analyse des anomalies suivant différents niveaux de granularité : au *niveau capteur*, au *niveau composant*, mais également au niveau d'un *wagon* ou d'un *train* (voir la figure 4).

**Informations complémentaires : exploiter les informations contextuelles.** Dans un deuxième temps, nous visons à donner plus d'informations sur l'anomalie détectée en tirant parti de la contextualisation des comportements (voir Section 2). Nous exploitons le principe suivant : plus le comportement d'un train est peu conforme aux comportements généraux (i.e., son score est bas dans les classes les plus générales) et plus son comportement est problématique. Ainsi, lorsqu'une anomalie est détectée, l'outil calcule le score dans les classes plus générales du treillis de classes. Par exemple, considérons le cas de la figure 5. Le score de conformité est bas dans la classe spécialisée et irrégulier dans les classes générales. Ici, le score dans la classe  $[low, low]$  est bas, mais les scores dans  $[*, low]$  et  $[high, low]$  sont élevés. C'est une nouvelle information pour l'expert. Le capteur se comporte comme si la température extérieure était haute, il peut s'agir d'une surchauffe.

## 4 Conclusion

La maintenance de systèmes complexes comme les trains est un problème particulièrement important. Une politique de maintenance visant à prévenir les problèmes les plus sérieux (pannes, immobilisation des trains, etc.) par la mise en œuvre d'opérations préventives est nécessaire. Cependant, il s'agit d'un problème difficile. Tout d'abord, la maintenance préventive requiert un système de suivi du comportement des trains s'appuyant sur une grande quantité de capteurs rassemblant au cours du temps diverses informations : températures, vitesses, accélérations, etc. Afin de proposer une approche automatique d'émission d'alertes lorsque une anomalie comportementale est repérée (i.e., lorsqu'un comportement dévie du comportement attendu), ces données sont traitées en tenant compte de leurs spécificités (e.g., données

temporelles, multi-sources, etc.). Nous proposons une méthode de détection d'anomalies dans les données comportementales, exploitant les motifs séquentiels préalablement extraits pour décrire le comportement normal d'un train. De plus, l'outil présenté offre aux experts les informations utiles pour procéder au diagnostic des anomalies détectées, et pour assister la prise de décisions de maintenance. Les travaux présentés ouvrent de nombreuses perspectives. En premier lieu, compléter l'approche proposée en y ajoutant un aspect prédictif constitue une piste intéressante. De plus, nous pouvons en ajustant les méthodes utilisées, adapter l'approche à des données traitées en temps réel.

## Références

- Agrawal, R. et R. Srikant (1995). Mining sequential patterns. In P. S. Yu et A. S. P. Chen (Eds.), *Eleventh International Conference on Data Engineering*. IEEE Computer Society Press.
- Carrascal, A., A. Díez, et A. Azpeitia (2009). Unsupervised methods for anomalies detection through intelligent monitoring systems. In *HAIIS '09 : Proceedings of the 4th International Conference on Hybrid Artificial Intelligence Systems*, Berlin, Heidelberg. Springer-Verlag.
- Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples. *Technometrics* 11.
- Jardine, A. K. S., D. Lin, et D. Banjevic (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing* 20.
- Keogh, E., J. Lin, S.-H. Lee, et H. Van Herle (2006). Finding the most unusual time series subsequence : algorithms and applications. *Knowl. Inf. Syst.* 11(1).
- Lee, J., R. Abujamra, A. Jardine, D. Lin, et D. Banjevic (2004). An integrated platform for diagnostics, prognostics and maintenance optimization. *Proceedings of the Intelligent Maintenance Systems*.
- Li, Y. et P. Nilkitsaranont (2009). Gas turbine performance prognostic for condition-based maintenance. *Applied Energy* 86(10).
- Rabatel, J., S. Bringay, et P. Poncelet (2009). Détection d'anomalies pour la maintenance ferroviaire. Technical report.
- Samanta, B. et C. Nataraj (2008). Prognostics of machine condition using soft computing. *Robotics and Computer-Integrated Manufacturing* 24(6). FAIM 2007, 17th International Conference on Flexible Automation and Intelligent Manufacturing.
- Shaikh, R. A., H. Jameel, B. J. d'Auriol, S. Lee, Y.-J. Song, et H. Lee (2008). Trusting anomaly and intrusion claims for cooperative distributed intrusion detection schemes of wireless sensor networks. In *ICYCS '08 : Proceedings of the 2008 The 9th International Conference for Young Computer Scientists*, Washington, DC, USA. IEEE Computer Society.
- Wireman, T. (1990). World Class Maintenance Management. *Industrial Press Inc., USA*.

## Summary

Train maintenance is a difficult problem associated with many security, financial and energy resources challenges. We are interested in developing a preventive maintenance by detecting and correcting any anomalous behavior that could cause a major problem in the near future. We thus provide a decision support tool to (i) extract useful knowledge in the history of trains, and (ii) detect and study the behavioral anomalies in order to make optimal decisions in terms of maintenance.