

PRIVATE LEARNING POUR LA MAINTENANCE PRÉDICTIVE : POURQUOI ET COMMENT ?

Chaumette, Serge, Dir. Innovation, serge.chaumette@{labri.fr, predition.com}, LaBRI/Université de Bordeaux - Preditic, Bordeaux (Orateur)

Dumora, Christophe, Data Scientist, christophe.dumora@preditic.com, Preditic, Bordeaux (Orateur)

Ouoba, Jonathan, Dir. Technique, jonathan.ouoba@preditic.com, Preditic, Bordeaux (Orateur)

Sissoko, Gaoussou, Data Scientist, gaoussou.sissoko@preditic.com, Preditic, Bordeaux

Thématique principale : Analyse des données

Résumé : Dans le milieu industriel, la transition de la maintenance corrective à la maintenance conditionnelle (ou prédictive) nécessite l'analyse de données qui peuvent revêtir un caractère sensible. Il est donc nécessaire d'utiliser des méthodes d'apprentissage qui préservent la confidentialité des données. Sur cette base et en utilisant un cas d'usage SNCF Réseau, nous proposons une méthodologie de mise en œuvre du Private Learning pour faire du calcul d'indicateurs métier, de la détection d'anomalies et de la prédiction sur des données chiffrées.

Mots clés : *Maintenance prédictive, Apprentissage, Private Learning, Time Series, Confidentialité, Fonctions homomorphes*

1. Introduction

Les enjeux, pour les acteurs industriels, en ce qui concerne l'analyse des données pour opérer la transition de la maintenance corrective à la maintenance conditionnelle ne sont plus à démontrer. Cette analyse de données (souvent réalisée par un prestataire extérieur) a pour objectif de prédire l'évolution de l'état d'un système pour anticiper les éventuelles pannes et ainsi planifier au mieux les opérations de maintenance. Pour cela, le prestataire doit disposer d'un volume suffisamment significatif de données dont l'historique de survenue des états à prédire. Toutefois, ces données peuvent être particulièrement sensibles et doivent donc rester confidentielles car elles représentent une ressource critique. Dans ce contexte, comment gérer le partage d'information pour préserver la confidentialité des données et des résultats tout en assurant la possibilité de faire les calculs nécessaires à la fourniture des services attendus sur ces données ? Pour répondre à cette problématique nous présentons une procédure de mise en œuvre de méthodes de *Private Learning* sur un cas d'usage lié aux solutions numériques d'optimisations des opérations de maintenance que nous développons pour le compte de SCNF Réseau.

2. Méthodologie

La méthodologie mise en œuvre est extrêmement pragmatique. La stratégie que nous proposons repose sur l'idée de travailler sur des données chiffrées pour réaliser l'apprentissage nécessaire aux prédictions et de produire des résultats eux-mêmes chiffrés. De façon concrète, les technologies sous-jacentes que nous utilisons sont liées à la cryptographie homomorphe (permutation chiffrement-opération). Il s'agit d'exploiter

l'homomorphie additive ou l'homomorphie multiplicative (comme avec RSA par exemple) voire l'homomorphie complète pour réaliser l'ensemble des calculs ce qui permet ainsi de préserver la confidentialité des données.

Cette première mise en œuvre se base sur les données collectées dans le cadre de la solution d'automatisation du contrôle de l'isolement des câbles alimentant les équipements de signalisation de SNCF Réseau (que nous avons développée). Le capteur déployé permet d'effectuer des mesures toutes les 15-30 secondes afin de connaître précisément les variations d'isolement des câbles considérés et les paramètres associés (température, humidité). Ces informations permettent d'identifier de potentielles fuites de courant qui peuvent générer des incidents sur les circuits. Nous exploitons l'approche présentée plus haut sur ce cas d'usage réel, constitué d'un grand volume de données de type séries temporelles. Plus précisément, nous appliquons des fonctions homomorphes ad hoc permettant le calcul d'indicateurs métier, la détection d'anomalies et l'anticipation des tendances de l'isolement avec des données chiffrées.

3. Originalité / perspective

Le caractère différenciant de notre approche repose sur le fait que la méthodologie proposée sert de base à l'étude, que nous souhaitons mener, des questions de recherche suivantes qui, à notre connaissance, demeurent ouvertes :

- Peut-on combiner dans l'apprentissage deux sources de données chiffrées différemment et produire un résultat chiffré différemment ?
- Peut-on découper à un autre niveau que simplement données/calcul, par exemple données-partie calculs/calculs ou données/partie-données calculs ?
- Comment introduire des paramètres eux-mêmes chiffrés (des seuils par exemple) dans l'analyse des données ?

Nous utilisons pour nos expérimentations un cas d'usage réel pour lequel nous avons déjà implémenté et déployé des solutions d'analyse de données dans un contexte industriel.

Nous proposons donc une démarche pragmatique permettant d'appliquer le chiffrement de données au calcul d'indicateurs métier, à la détection d'anomalies et à l'anticipation des tendances sur des séries temporelles dans un cadre industriel. Cette approche est une brique technologique de base du projet de R&D (soutenu en particulier par la Région Nouvelle-Aquitaine et BPI France) que nous menons dans le but d'améliorer les outils d'aide à la décision que nous proposons à nos clients. En termes de perspectives, il faudra continuer à réfléchir à la possibilité de restreindre le *Machine Learning* à de l'additif/multiplicatif ou encore à la résistance à la cryptanalyse face à une grande masse de données tout en poursuivant l'étude des questions ouvertes mentionnées ci-dessus.

Annexe - Présentation de l'entreprise

Créée en 2014, PREDITIC est une entreprise issue du Laboratoire Bordelais de Recherche en Informatique, experte en collecte, exploitation et valorisation des données en milieu industriel. Nous avons développé une plateforme logicielle et data permettant d'analyser et de valoriser les données métiers de nos clients. Nous partons de la collecte des données (capteurs, fiches de maintenance, logiciels, données externes) jusqu'à la mise en place d'une maintenance prédictive (avec des approches de type *machine learning*) en passant par l'administration et l'hébergement des données ainsi que la visualisation des données.

PREDITIC se différencie par son savoir-faire sur les questions d'analyse et de traitement de données en milieu complexe. Nous travaillons entre autres avec de grands groupes comme SNCF Réseau, le groupe Vinci (Mesea), Keolis Bordeaux, GT Logistics, et Enedis.

L'optimisation des processus et la maintenance prédictive sont des enjeux incontournables pour le développement et la pérennité des entreprises du milieu industriel. En lien avec ces enjeux, nous avons initié un projet de R&D (soutenu par la Région Nouvelle-Aquitaine et BPI France) dans le but d'étoffer les fonctionnalités des outils d'aide à la décision que nous proposons à nos clients. L'objectif est principalement de travailler sur les thématiques de l'extraction automatique d'indicateurs dans les jeux de données disponibles, la prédiction basée sur l'IA explicative, la simulation de nouvelles sources de données et l'étude de leur impact sur les analyses.