

Stage « Anomaly Detection in time series data »

Durée du stage : 6 mois (2018)

Lieu du Stage: Worldline, Seclin (59).

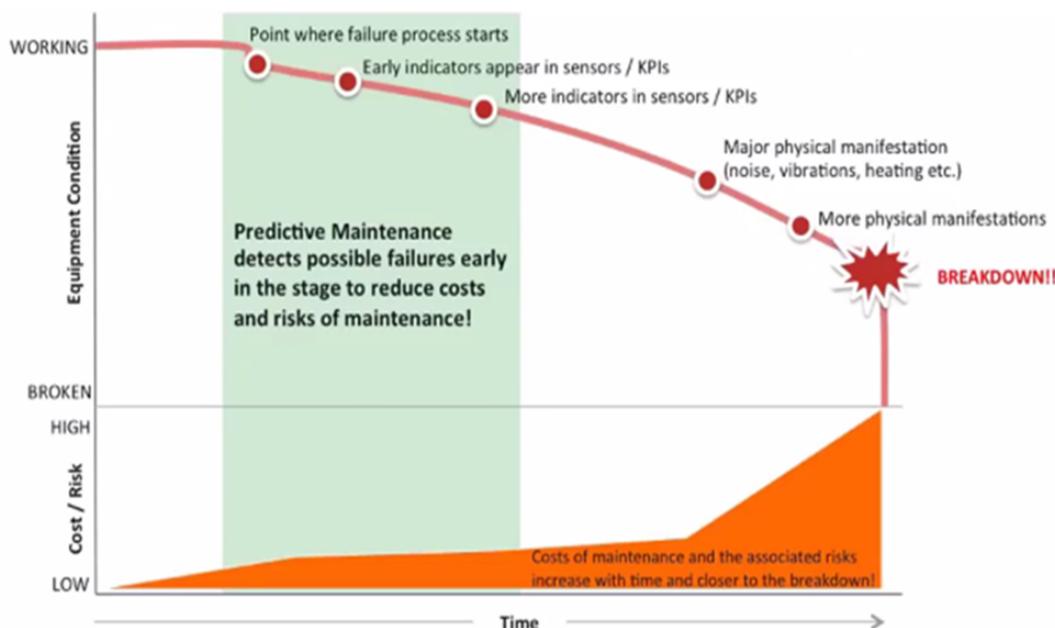
Profil : M2 ou Bac + 5, Data Science

*Contacts : Dalila Hattab (dalila.hattab@worldline.com) &
Pascal Yim (pascal.yim@centraledlille.fr)*

Contexte :

Prévoir les défaillances avant qu'elles ne surviennent, tel est l'enjeu du monitoring prédictif. Pour le satisfaire, il existe différentes techniques de mesures qui permettent d'évaluer l'état de nos serveurs et nos applications et de détecter les signes avant-coureurs de pannes.

Le monitoring prédictif est une forme de suivi de notre production nettement plus sophistiquée. Elle consiste à surveiller en continu les performances de nos serveurs et applications afin de détecter tout signe de défaillance éventuelle, de comportements déviants et de ruptures dans nos métriques métiers et systèmes. Cette stratégie permet d'avoir une plus grande maîtrise du fonctionnement général, tout en évitant les coûts liés aux temps d'indisponibilité et à la maintenance.



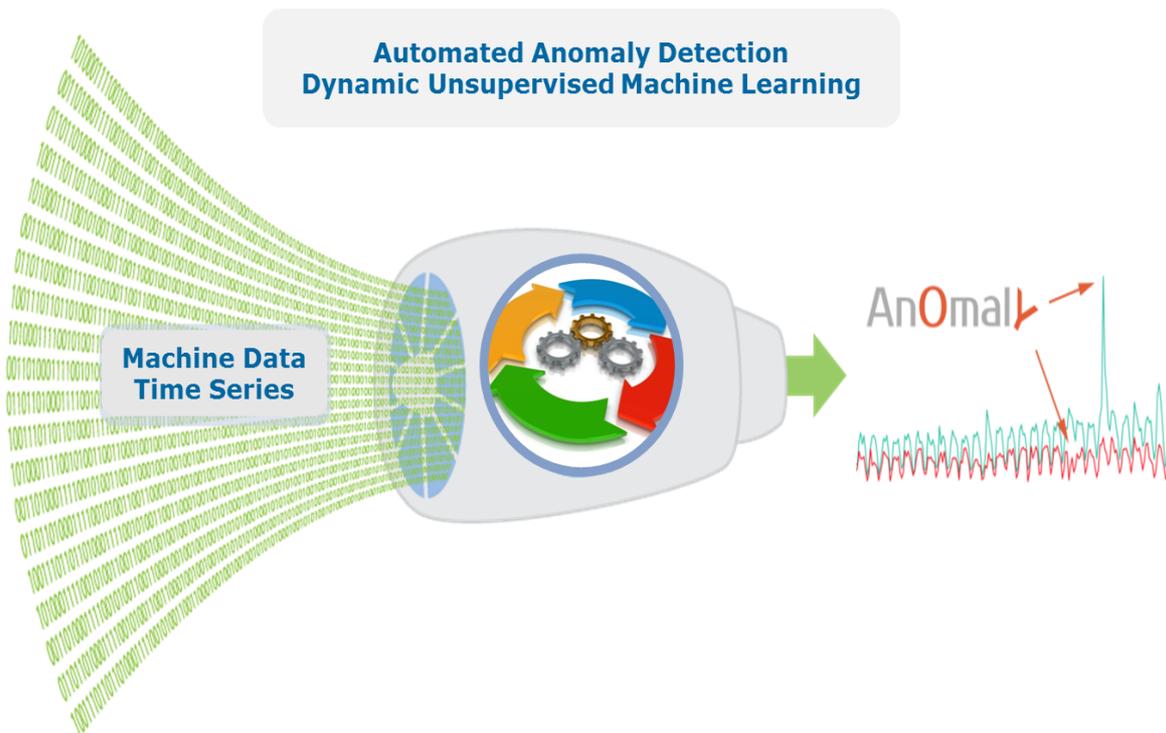
Au lieu de générer une alerte parce qu'une métrique a dépassé un seuil fixé, la détection d'anomalies, basée sur une analyse avancée, permet de surveiller en temps réel un grand nombre d'indicateurs de performance et de générer des alertes plus précises.

Des technologies d'apprentissage automatique (machine learning) sont particulièrement pertinentes pour que les algorithmes « apprennent » à reconnaître les nouveaux

événements et défaillances lorsqu'ils surviennent. Cette étape permet de ne pas se limiter aux seuls schémas d'anomalies et de défaillances connus mais d'en ajouter de nouveaux aux systèmes de détection, au fur et à mesure de la collecte d'informations sur le système.

Description du sujet :

Nous souhaitons enrichir un composant d'analyse comportementale de nos applications et systèmes en utilisant l'apprentissage automatique tel que neuronal Network (LSTM) ou le Deep Learning, pour détecter des anomalies sur des métriques collectées massivement sur nos serveurs.



Apprentissage non supervisé

Ce composant doit être construit sur la base d'algorithmes d'apprentissage qui automatisent l'analyse des données massives collectées sur les serveurs, et qui construisent et maintiennent des modèles statistiques et de machine Learning précis de ces données.

Détection d'anomalie plus fiable

Ce composant doit réaliser une détection d'anomalies très fiable en réduisant les fausses alertes afin de pouvoir rapidement, étudier et répondre à une activité anormale. L'analyse automatisée réduit l'usage de règles manuelles pouvant produire de fausses alertes.

Performance

Ce composant doit être conçu et supporté par une plateforme capable d'analyser une quantité massive de données, et de remonter rapidement les points de ruptures qui devraient attirer l'attention afin d'être rapidement prise en charge.

Attendus :

- *Etude bibliographique*
- *Expérimentation des techniques avancées d'apprentissage statistiques, des réseaux de neurones ou du deep learning sur des extractions de données disponibles (métriques systèmes et métiers). La problématique de prise en compte d'un certain niveau de bruit et de son traitement est à considérer.*
- *Mise en œuvre de ces nouvelles méthodes dans la solution existante.*
- *Rédaction d'un rapport de stage.*

Moyens Utilisés : *R, python, Java, Hadoop.*

Références bibliographiques :

Mahendra Kumar Ahirwar, Manish Kumar Ahirwar and Uday Chourasia Anomaly Detection in the Services Provided by MultiCloud Architectures: A Survey, IJRET, 2014.

Varun Chandola and Arindam Banerjee. Anomaly detection: a Survey. ACM Computing Surveys, September 2009.

David Goldberg and Yinan Shan The importance of Features for Statistical Anomaly Detection 7th Usenix Workshop on hot topics in Cloud Computing, 2015.

Pankaj Malhobtra, Lovekesh Vig, Gautam Shroff, Puneet Agarwal, Long Short Term Memory Networks for Anomaly Detection in Time series, ESANN 2015 .

Netflix, <http://techblog.netflix.com/2015/02/rad-outlier-detection-on-big-data.html>

PreAlert, <http://info.preAlert.com/>