



Université Claude Bernard



DIPLÔME NATIONAL DE DOCTORAT

(Arrêté du 25 mai 2016)

Date de la soutenance : **21 mars 2018**

Nom de famille et prénom de l'auteur : **BENKABOU Seif-Eddine**

Titre de la thèse : « Détection d'anomalies dans les séries temporelles : application aux masses de données sur les pneumatiques ».



Résumé

La détection d'anomalies est une tâche cruciale qui a suscité l'intérêt de plusieurs travaux de recherche dans les communautés d'apprentissage automatique et fouille de données. La complexité de cette tâche dépend de la nature des données, de la disponibilité de leur étiquetage et du cadre applicatif dont elles s'inscrivent.

Dans le cadre de cette thèse, nous nous intéressons à cette problématique pour les données complexes et particulièrement pour les séries temporelles uni et multi-variées. Le terme "anomalie" peut désigner une observation qui s'écarte des autres observations au point d'éveiller des soupçons. De façon plus générale, la problématique sous-jacente (aussi appelée détection de nouveautés ou détection des valeurs aberrantes) vise à identifier, dans un ensemble de données, celles qui diffèrent significativement des autres, qui ne se conforment pas à un « comportement attendu » (à définir ou à apprendre automatiquement), et qui indiquent un processus de génération différent. Les motifs « anormaux » ainsi détectés se traduisent souvent par de l'information critique.

Nous nous focalisons plus précisément sur deux aspects particuliers de la détection d'anomalies à partir de séries temporelles dans un mode non-supervisé. Le premier est global et consiste à ressortir des séries relativement anormales par rapport une base entière. Le second est dit contextuel et vise détecter localement, les points anormaux par rapport à la structure de la série étudiée. Pour ce faire, nous proposons des approches d'optimisation à base de clustering pondéré et de déformation temporelle pour la détection globale; et des mécanismes à base de modélisation matricielle pour la détection contextuelle.

Enfin, nous présentons une série d'études empiriques sur des données publiques pour valider les approches proposées et les comparer avec d'autres approches connues dans la littérature. De plus, une validation expérimentale est fournie sur un problème réel, concernant la détection de séries de prix aberrants sur les pneumatiques, pour répondre aux besoins exprimés par le partenaire industriel de cette thèse

Mots clés -- détection d'anomalies, séries temporelles, clustering, optimisation.