
Détection d'anomalies dans les données fonctionnelles multivariées.

Messan Martial Amovin-Assagba^{*1,2}, Irène Gannaz^{3,4}, and Julien Jacques²

¹Arpege Master K – Université Lumière - Lyon II – France

²Entrepôts, Représentation et Ingénierie des Connaissances – Université Lumière - Lyon 2 : EA3083,
Université Claude Bernard Lyon 1 – France

³Université Lyon 1 – Université Claude Bernard - Lyon I – France

⁴Institut Camille Jordan [Villeurbanne] – Ecole Centrale de Lyon, Université de Lyon, Université Claude Bernard Lyon 1, Institut National des Sciences Appliquées de Lyon, Institut National des Sciences Appliquées : UMR5208, Université Jean Monnet [Saint-Etienne], Centre National de la Recherche Scientifique – France

Résumé

Les récentes innovations réalisées sur les appareils de mesure et l'utilisation intensive des ressources informatiques facilitent la collecte de données massives avec une fréquence d'acquisition de plus en plus grande. Ces données peuvent être considérées comme des données fonctionnelles : des observations discrètes de fonctions aléatoires évoluant au cours du temps. Un des problèmes majeurs dans l'analyse de données fonctionnelles est la détection d'anomalies ou la détection de courbes anormales. Pour ce faire, nous proposons une méthode robuste basée sur les modèles de mélange gaussiens contaminés, permettant à la fois de grouper et de détecter des anomalies dans les données fonctionnelles multivariées. Un modèle de mélange gaussien contaminé, est un mélange gaussien où chaque cluster a deux composants : une première avec une probabilité à priori élevée, représente les observations normales et l'autre, avec une faible probabilité à priori, représente les observations aberrantes. Une parcimonie est introduite à partir des méthodes de réduction de dimension basées sur les modèles de mélange latent fonctionnel. L'algorithme ECM (Expectation-Conditional Maximization) est proposé pour l'inférence du modèle. Des expérimentations sur les données simulées prouvent l'efficacité de la méthode. Le modèle a également permis de détecter correctement les anomalies dans les jeux de données industrielles qui ont motivé ce travail et qui illustrent son intérêt.

*Intervenant