

Avis de Soutenance

Monsieur Trung Thanh LE

Informatique

Soutiendra publiquement ses travaux de thèse intitulés

Analyse des flux de données de signaux et d'images: du sous-espace au suivi tensoriel

dirigés par Monsieur KARIM ABED-MERAIM et Monsieur Adel HAFIANE

Ecole doctorale : Mathématiques, Informatique, Physique Théorique et Ingénierie des Systèmes - MIPTIS

Unité de recherche : PRISME - Laboratoire Pluridisciplinaire de Recherche en Ingénierie des Systèmes et Mécanique
Energétique

Soutenance prévue le **jeudi 20 octobre 2022** à 14h00

Lieu : 12 rue de Blois - 45100 Orléans

Salle : Amphi Turing

Composition du jury proposé

M. KARIM ABED-MERAIM	Université d'Orléans	Directeur de thèse
M. LAURENT ALBERA	Université de Rennes 1	Rapporteur
M. EL KORSO MOHAMMED NABIL	Université Paris Nanterre	Rapporteur
M. ADEL HAFIANE	INSA Centre Val de Loire	Co-directeur de thèse
M. LINH TRUNG NGUYEN	Vietnam National University, Hanoi	Co-encadrant de thèse
M. ROLAND BADEAU	Institut Mines-Télécom, Télécom ParisTech	Examineur
M. Rémy BOYER	Université de Lille	Examineur
M. PHILIPPE RAVIER	Université d'Orléans	Examineur

Mots-clés : flux de données, analyse de sous-espace, décomposition du tenseur,,

Résumé :

Le traitement des flux a récemment attiré l'attention du monde universitaire et de l'industrie, car les flux de données massives ont été de plus en plus collectés au fil des ans. Cette thèse se concentre principalement sur l'étude de l'un des problèmes les plus fondamentaux du traitement des flux, l'approximation de rang inférieur (LRA) des flux de données en ligne. Lorsque les échantillons de données arrivant à chaque pas de temps sont unidimensionnels, le problème de LRA en ligne est techniquement appelé suivi de sous-espace. Il s'agit d'un suivi tensoriel lorsque le flux de données est multidimensionnel. Pour le suivi du sous-espace, nous avons proposé deux nouveaux algorithmes pour suivre le sous-espace sous-jacent des flux de données dans deux scénarios spécifiques. Pour traiter les valeurs aberrantes clairsemées et les données manquantes, un algorithme efficace de suivi de sous-espace en deux étapes a été développé, à savoir PETRELS-ADMM. L'algorithme proposé est basé sur la méthode de direction alternée des multiplicateurs et des techniques de filtrage récursif des moindres carrés. Le deuxième algorithme appelé OPIT a été spécifiquement conçu pour suivre le sous-espace principal clairsemé dans les grandes dimensions. Plus précisément, OPIT introduit une nouvelle variante adaptative d'itération de puissance et un nouvel opérateur de seuillage basé sur des colonnes. Les deux algorithmes proposés appartiennent à la classe des méthodes de suivi prouvables avec une garantie de convergence. Pour le suivi des tenseurs, nous avons développé plusieurs nouveaux algorithmes pour suivre le LRA en ligne des tenseurs de streaming au fil du temps. Sous le format CP/PARAFAC, nous exploitons les techniques alternatives de minimisation et d'esquisse aléatoire pour développer ACP et RACP qui sont capables de factoriser des tenseurs incomplets et des tenseurs corrompus, respectivement. Sous le format Tucker, nous avons proposé un autre algorithme en ligne appelé ATD. ATD suit d'abord les sous-espaces de faible dimension sous-jacents couvrant les facteurs tensoriels, puis estime le tenseur central à l'aide d'une approximation stochastique $\{a\}$. Une analyse de convergence unifiée a été présentée pour justifier leur performance. En parallèle, il est bien connu que le format tenseur-train offre une représentation économe en mémoire et peut briser la malédiction de la dimensionnalité. En conséquence, nous avons conçu trois nouveaux algorithmes pour la décomposition en continu de trains de tenseurs. Le premier algorithme appelé TT-FOA est capable de suivre avec une grande précision les composantes de rang inférieur des tenseurs d'ordre élevé à partir de données bruitées et de grande dimension, même lorsqu'elles proviennent d'observations dépendant du temps. Le deuxième algorithme appelé ATT est spécifiquement conçu pour gérer les tenseurs de flux incomplets. Enfin, pour traiter les valeurs aberrantes éparses, nous avons proposé le soi-disant ROBOT. Techniquement, ROBOT a la capacité de suivre les tenseurs de flux à partir de flux imparfaits (c'est-à-dire en raison du bruit, des valeurs aberrantes et des données manquantes) ainsi que d'estimer leur variation temporelle dans des environnements dynamiques. En conclusion, notre étude apporte plusieurs nouvelles contributions à l'analyse des flux de données massives en général et au problème des LRA en ligne. Il s'agit de nouveaux outils d'analyse permettant de suivre efficacement les flux de données de LRA en ligne, des observations unidimensionnelles aux observations multidimensionnelles dans différents contextes. Par conséquent, ils devraient faire un pas en avant dans les applications en ligne du monde réel.