AVIS DE SOUTENANCE DE THESE

En Co tutelle avec l'Université Sorbone Paris Nord-villetaneuse

Le Doyen de la Faculté des Sciences Dhar El Mahraz –Fès – annonce que

Mme (elle) **: BEN BOUAZZA Fatima Ezzahraa**

Soutiendra **: le 26/11/2020** à **14h**

Lieu : **Centre Visio Conférence**

Une thèse intitulée:

Multi –Models clusteringth through Optimal Transport theory

En vue d'obtenir le Doctorat

FD : Mathématiques et Applications (MA)

Spécialité: Science des données

Devant le jury composé comme suit :

	NOM ET PRENOM	GRADE	ETABLISSEMENT
Président	Pr. STEPHANO Guerrini	PES	Université Sorbone Paris Nord-villetaneuse
Directeurs de thèse (cotutelle)	Pr. BENNANI Younès	PES	Université Sorbone Paris Nord-villetaneuse
	Pr. TOUZANI Abdelfattah	PES	Faculté des Sciences Dhar El Mahraz - Fès
Rapporteurs	Pr. ROSANNA Verde	PES	Universia della Campania-Naple
	Pr.YOUSFI Noura	PES	Faculté des Sciences Ben M'sik - Casablanca
	Pr. YOUSSFI Ahmed	PES	Ecole Nationale des Sciences Appliquées - Fès
Membres	Pr. LEVGEN Redko	PH	Université de Saint –Etienne- Saint –Etienne
	Pr.WARDI Souad	PES	Faculté des Sciences Dhar El Mahraz - Fès

Résumé :
Le travail de recherche exposé dans cette thèse concerne le développement d'approches de clustering multi-modèles à partir de données distribuées. Nous travaillons sur deux volets principaux en apprentissage automatique multi-modèles. Le premier concerne le clustering multi-vues où nous visons à apprendre un modèle optimal global à partir des modèles locaux des différentes vues. Le second volet porte sur le clustering collaboratif où l'idée principale est de trouver modèle d'échange de connaissances entre les différents collaborateurs afin d'améliorer leurs propres modèles.
Pour le clustering multi-vues, nous avons proposé deux approches basées sur la théorie du transport optimal. La première approche (PCA) consiste à trouver un modèle de consensus à partir des modèles locaux, en projetant les distributions des différentes vues sur l'espace global. La seconde approche (CNR) vise à apprendre une nouvelle représentation consensuelle à partir de la représentation locale des distributions.
Dans le cas du clustering collaboratif, nous introduisons une nouvelle approche basée sur la théorie du transport optimal (Co-OT) qui vise à améliorer le mécanisme de la collaboration et la manière de transporter les informations entre les collaborateurs avec un coût minimum. Pour ce faire, nous proposons une fonction objective pour la collaboration basée sur la distance de Wasserstein. Nous proposons une solution pour choisir les meilleurs collaborateurs par comparaison de la distribution locale des prototypes et l'analyse de la diversité entre les collaborateurs.
Pour une autre approche dans ce cadre, nous présentons un nouveau modèle de collaboration guidé par la sélection des caractéristiques, où l'idée principale est de choisir les caractéristiques qui donnent la meilleure représentation pour chaque Le travail de recherche exposé dans cette thèse concerne le développement d'approches de clustering multi-modèles à partir de données distribuées. Nous travaillons sur deux volets principaux en apprentissage automatique multi-modèles. Le premier concerne le clustering multi-vues où nous visons à apprendre un modèle optimal global à partir des modèles locaux des différentes vues. Le second volet porte sur le clustering collaboratif où l'idée principale est de trouver modèle d'échange de connaissances entre les différents collaborateurs afin d'améliorer leurs propres modèles.
Pour le clustering multi-vues, nous avons proposé deux approches basées sur la théorie du transport optimal. La première approche (PCA) consiste à trouver un modèle de consensus à partir des modèles locaux, en projetant les distributions des différentes vues sur l'espace global. La seconde approche (CNR) vise à apprendre une nouvelle représentation consensuelle à partir de la représentation locale des distributions.
Dans le cas du clustering collaboratif, nous introduisons une nouvelle approche basée sur la théorie du transport optimal (Co-OT) qui vise à améliorer le mécanisme de la collaboration et la manière de transporter les informations entre les collaborateurs avec un coût minimum. Pour ce faire, nous proposons une fonction objective pour la collaboration basée sur la distance de Wasserstein. Nous proposons une solution pour choisir les meilleurs collaborateurs par comparaison de la distribution locale des prototypes et l'analyse de la diversité entre les collaborateurs.
Pour une autre approche dans ce cadre, nous présentons un nouveau modèle de collaboration guidé par la sélection des caractéristiques, où l'idée principale est de choisir les caractéristiques qui donnent la meilleure représentation pour chaque collaborateur et garantissant la meilleure communication entre eux, tout en préservant la confidentialité des données de de chaque collaborateur. Cette dernière approche collaborative est aussi développée dans le cadre de la théorie du transport optimal.
Enfin, plusieurs expériences approfondies sur de multiples ensembles de données réelles sont proposées pour évaluer les approches développées et démontrent leur utilité et efficacité dans le cas des données distribuées.

Mots clés :
Transport optimal, clustering multi-modéles , clustering multi-view, apprentissage collaborative, la distance de Wasserstein, Sinkhorn-means, la selection des variable

MULT-MODELS CLUSTERING THROUGH OPTIMAL TRANSPORT THEORY

Abstract:
The research work presented in this Ph.D. thesis concerns the development of multi-model clustering approaches based on distributed data. We are working on two main aspects of multi-model machine learning. The first concerns multi-view clustering where we aim to learn an optimal global model from the local models of the different views. The second is collaborative clustering where the main idea is to find a model for knowledge exchange between different collaborators in order to improve their own models.

For multi-view clustering, we proposed two approaches based on the optimal transport theory. The first approach (PCA) consists in finding a consensus model from local models, by projecting the distributions of the different views on the global space. The second approach (CNR) aims at learning a new consensus representation from the local representation of the distributions.

In the case of collaborative clustering, we introduce a new approach based on optimal transport theory (Co-OT) which aims to improve the mechanism of collaboration and the way of transporting information between collaborators with minimum cost. To do so, we propose an objective function for collaboration based on the Wasserstein distance. We offer a solution for selecting the best collaborators by comparing the local distribution of prototypes and analysing the diversity between collaborators.

For another approach within this framework, we present a new model of feature-driven collaboration, where the main idea is to choose the features that give the best representation for each collaborator and guarantee the best communication between them, while preserving the confidentiality of each collaborator's data. This last collaborative approach is also developed within the framework of the theory of optimal transport.

Finally, several in-depth experiments on multiple real data sets are proposed to evaluate the approaches developed and demonstrate their usefulness and effectiveness in the case of distributed data.

Key Words:
Optimal Transport, Multi-Models Clustering, collaborative clustering, Multi-view clustering, Feature selection, Wasserstein distance, Sinkhorn-means.