Demande n°3171 - ARED

Acronyme :LevNoEuc

Identification

Identification

N°	3171
Année	2023
Statut	Brouillon
Date de création	30/01/2023
Date de modification	30/01/2023
Date de soumission	30/01/2023
Acronyme	LevNoEuc
Intitulé du projet	Leveraging non-Euclidean spaces for learning on structured data
Dispositif	ARED
Civilité du porteur	
Nom du porteur	Chapel
Prénom du porteur	Laetitia
Courriel	laetitia.chapel@univ-ubs.fr

Priorité régionale

Priorité régionale

Priorité régionale	Atténuation et/ou adaptation au changement climatique
Regional priority	Atténuation et/ou adaptation au changement climatique

Stratégie régionale de recherche et d’innovation (S3)

Stratégie régionale de recherche et d’innovation (S3)

Axe transversal	Transitions numérique et industrielle
Domaine d'innovation stratégique	Hors DIS
Domaine d'innovation stratégique	.
Domaine d'innovation statégique Secondaire	Hors DIS
Domaine d'innovation statégique Secondaire	.

Organisme de tutelle

Organisme de tutelle

Etablissement	UNIVERSITE BRETAGNE SUD
Commune	LORIENT
Civilité	Madame
Nom	CHAPEL
Prénom	Laetitia

Unité de recherche

Unité de recherche

Nom de l'unité	IRISA - équipe Obelix
Civilité	Monsieur
Nom	GRAVIER
Prénom	Guillaume
Courriel	dir@irisa.fr

Données complémentaires de l'unité de recherche

Libellé	Valeur
Statut et numéro de l'unité de recherche (UMR, USR, EA?)	UMR 6074
Le cas échéant, nom de l'équipe de recherche	Obelix
Le cas échéant, nom du responsable de l'équipe de recherche	Nicolas Courty nicolas.courty@univ-ubs.fr
Activités de recherche et points forts de l'équipe	OBELIX : Observation de l'environnement par imagerie complexe L'objectif global de l'équipe est le traitement d'images complexes à des fins environnementales. Elle est composée d'enseignants chercheurs reconnus dans le domaine de l'intelligence artificielle au service de l'observation de la Terre
Si implication dans un/plusieurs programmes européens ou internationaux, préciser	

Co-directeur de thèse

Co-directeur de thèse

Civilité	Madame
Nom	CHAPEL
Prénom	Laetitia
Unité de recherche	UMR 6074 - IRISA
Etablissement de tutelle	UNIVERSITE BRETAGNE SUD

Description

Description du projet

Libellé	Valeur
Résumé synthétique du projet (2000 caractères max)	L'objectif général de la thèse est de proposer de nouveaux développements en intelligence artificielle pour l'apprentissage sur des données issues de l'observation de la Terre. Les méthodes sont très souvent développées sur des espaces dits Euclidien alors que les données possèdent une structure intrinsèque (comme des relations hiérarchiques entre les différents éléments par exemple) qui nécessitent de travailler dans des espaces dédiés. Ci-dessous, plus de détails sur les objectifs scientifiques. L'objectif de la thèse est de tirer pleinement parti de la géométrie riemannienne pour apprendre sur des données structurées. Bien que l'intégration des données dans un espace approprié conduit à des performances d'apprentissage améliorées, la plupart des travaux existants se sont concentrés sur la conception métriques calquées sur les développements euclidiens. Très peu d'entre eux se sont concentrés sur des fonctions de perte spécifiques qui tirent pleinement parti de la géométrie à disposition : par exemple, les fonctions de pertes standard telles que l'entropie croisée sont utilisées dans les espaces structurés, avec peu de prise en compte de l'espace (c'est-à-dire le soft max angulaire pour les espaces sphériques ou le soft max hiérarchique pour les espaces hyperboliques). Ils ignorent très souvent certaines connaissances préalables explicites disponibles telles que la classe, la description du texte, les étiquettes ou la hiérarchie des classes. Les propriétés intrinsèques des espaces telles que les invariances de permutation ou de rotation ne sont pas non plus prises en compte. De plus, lors de la comparaison de différentes vues des mêmes données (p. ex. texte/image ou images prises par différents capteurs en télédétection), intégrées dans des géométries différentes, la comparaison de ces espaces dits incomparables est difficile. Ces questions difficiles seront d'abord abordées en considérant les pertes de type Gromov-Wasserstein, qui se sont avérées être une métrique efficace pour comparer des données de type graphe dans les espaces euclidiens. En particulier, elle permet de comparer des données vivant dans des espaces différents et conserve certaines invariances importantes telles que les invariances de translation ou de rotation. Elle a également montré son intérêt dans le traitement des données structurées en graphes. D'un point de vue applicatif, les données de télédétection, qui sont intrinsèquement organisées de manière hiérarchique, multimodales, avec des données spectrales et corrélées, serviront de terrain de jeu pour valider les méthodes proposées. Le problème de l'apprentissage en few-shot, c'est à dire avec peu d'exemple d'apprentissage, sera spécifiquement ciblé afin d'évaluer la capacité de généralisation et la performance en petite dimension. Les applications ciblées sont l'apprentissage dans des environnement adaptatifs, avec la réalisation de tâches d'apprentissage réalisées sur des images satellites impactées par des changements liées à des différences de climat notamment.
Porteur du projet : date d'obtention de l'HDR ou dérogation	Laetitia Chapel HDR : mai 2022
Hypothèses, questions posées, points de blocage, approche méthodologique, technique (4000 car. max)	La plupart des algorithmes d'apprentissage automatique exigent que les données d'entrée soient intégrées dans un espace euclidien, c'est-à-dire représentées sous une forme vectorielle comme un ensemble de flottants. Dans de nombreux domaines cependant, les données (par ex. données hiérarchiques, graphiques ou cycliques) peuvent être décomposées en un ensemble d'entités qui possèdent une structure interne complexe, généralement sous la forme d'une ou plusieurs relations entre elles. Il est donc nécessaire d'intégrer ces données structurées dans des représentations significatives et de définir une similarité entre elles. Néanmoins, ces données ne possèdent pas une structure euclidienne mais plutôt une structure hiérarchique ou cyclique. Dans ce cas, elles ne peuvent pas être intégrées dans un espace euclidien avec une faible distorsion. Ces dernières années, plusieurs travaux ont montré la supériorité des représentations non euclidiennes telles que la géométrie riemannienne, dont le choix détermine la manière dont les relations entre les points

	choix détermine la manière dont les relations entre les points sont quantifiées. Dans ce contexte, les espaces hyperboliques se sont révélés bien adaptés aux données hiérarchiques, avec des performances de pointe dans les tâches d'intégration de graphes ou de mots, de segmentation d'images ou d'apprentissage en quelques points. D'autre part, les espaces sphériques traitent avec succès le problème de la reconnaissance des visages ou des points situés sur la surface de la Terre. Ces espaces riemanniens sont de courbure constante positive (sphérique) ou négative (hyperbolique), par rapport à l'espace euclidien qui est de courbure nulle. courbure nulle. D'un domaine présentant un intérêt marginal dans la communauté ML il y a environ 5 ans (Google Scholar indique 56 résultats lors d'une recherche sur "Hyperbolic spaces "+"Machine Learning" en 2017), il s'agit aujourd'hui d'un domaine de recherche en plein essor (avec près de 500 résultats pour 2022). Au-delà des espaces de courbure constante, les espaces produits combinent plusieurs copies d'espaces euclidiens, hyperbolique et sphérique, en supposant que les données ont une structure complexe, dans le but d'obtenir le meilleur des mondes.Bien que ces espaces aient donné des résultats prometteurs, la combinaison optimale des espaces est toujours choisie par essais-erreurs, ce qui n'est pas acceptable dans les applications à grande échelle.
Environnement scientifique, positionnement dans contexte régional/national/international (2000 car.)	La recherche se déroulera au sein de l'équipe de recherche Obelix (www.irisa.fr/obelix) de l'IRISA situé sur le campus de l'UBS (Université Bretagne Sud) à Vannes. La thèse sera dirigée par Laetitia Chapel, qui a obtenu son Habilitation à Diriger des Recherches au printemps 2022. Elle supervise actuellement un étudiant qui soutiendra son doctorat au printemps 2023 (financée à 100% par le projet ANR Multiscale) et elle encadrera un seul doctorant de troisième année en octobre 2023 (co-encadrement avec Gilles Gasso, INSA de Rouen, financement : chaire AI de l'ANR).
Collaborations scientifiques (nature/partenariat/pays) et partenariat socio-économique envisagé	Du coté méthodologique, l'apprentissage dans des espaces hyperboliques et/ou sphériques est une thématique relativement jeune au sein de l'équipe de recherche et dans la communauté internationale en général. Du coté applicatif, le doctorant pourra bénéficier des compétences de l'équipe dans le domaine de l'observation de la Terre, ainsi que de ses collaborations (ESA, CNES etc.)
Adéquation du projet avec le DIS de rattachement (3000 caractères maximum)	
Si priorité régionale, préciser (200 caractères maximum)	Les applications ciblées sont l'apprentissage dans des environnement adaptatifs, avec la réalisation de tâches d'apprentissage réalisées sur des images satellites impactées par des changements liées à des différences de climat notamment.
Cotutelle internationale (indiquer "oui" ou "non"; si oui, préciser le pays et l'établissement)	
Nom(s) du(des) cofinanceur(s) envisagé(s) ou confirmé(s)	Co-financement CDE UBS
Si cofinancement industriel, justifier le choix du dispositif ARED au lieu du dispositif CIFRE	
État du cofinancement (indiquer "acquis" ou "demandé")	Demandé
Si demandé, date prévisionnelle de réponse du cofinanceur	février-mars 2023
Si cofinancement refusé, autre(s) source(s) de cofinancement identifiée(s)	Fonds propres de l'équipe (notamment grâce à un contrat d'accompagnement lié à une thèse CIFRE portée par un membre de l'équipe qui sera impliqué dans le travail de thèse)
Si reversement prévu, préciser (partenaire, montant)	

Candidat

Collaboration internationale

Collaboration internationale

Pays	
Etablissement partenaire à l'étranger	

Pièces jointes

Date de mi	Type de pièce	Commentaire	Document	Valide	Commentaire
30/01/2023	ARED - Dépôt - CV du porteur	Obligatoire - Le CV doit préciser si le porteur est titulaire de l'HDR et, le cas échéant, depuis quand, et mentionner la liste des doctorant·e·s déjà encadré·e·s et des publications associées.	resume.pdf	Non traité	
	ARED - Dépôt - Avis directeur·rice d'unité	15 lignes maximum		Non traité	
	ARED - Dépôt - Justificatif cofinancement	Facultatif - Dépôt de demande, justificatif d'obtention, courrier du cofinanceur pressenti...		Non traité	
	ARED - Dépôt - Lettre partenaire étranger	Facultatif - Lettre d'engagement de l'établissement partenaire étranger en cas de cotutelle internationale		Non traité	
	ARED - Dépôt - Pièce complémentaire	Facultatif		Non traité	
	ARED - Dépôt - Pièce complémentaire	Facultatif		Non traité	
	ARED - Instruction - Lettre de saisine	Lettre de l'établissement à transmettre à la Région (utiliser le modèle disponible dans la rubrique "Documents à télécharger")		Non traité	
	ARED - Versement - Convention de partenariat	Le cas échéant (si autorisation de reversement)		Non traité	
	ARED - Versement 1/2 - CV doctorant.e	Pour versement tranche 1 (cohortes jusqu'en 2018) Pour versement tranche 2 (cohortes à partir de 2019)		Non traité	
	ARED - Versement 1/2 - Copie du contrat doctoral (ou CDD ou attestation))	Pour versement tranche 1 (cohortes jusqu'en 2018) Pour versement tranche 2 (cohortes à partir de 2019)		Non traité	
	ARED - Versement 1 - Lettre d'approbation école doctorale	Pour versement tranche 1 (cohortes jusqu'en 2018)		Non traité	
	ARED - Versement 2 - Attestation bon déroulement travaux année 1	Pour versement tranche 2		Non traité	
	ARED - Versement 2 - Bilan financier année 1	Pour versement tranche 2 ARED		Non traité	
	ARED - Versement 3 - Attestation bon déroulement travaux année 2	Pour versement 50% tranche 3 ARED		Non traité	
	ARED - Versement 3 - Bilan financier année 2	Pour versement 50% tranche 3 ARED		Non traité	

	ARED - Versement 4 - Bilan financier année 3	Pour versement solde année 3 ARED		Non traité	
	ARED - Versement 4 - Certificat administratif précisant statut thèse	Pour versement du solde de la tranche 3 (à accompagner des pièces justificatives correspondantes pour les cohortes jusqu?en 2018 : avis du jury, attestation de réinscription en ED, justificatif de prolongation ou d'abandon)		Non traité	
	ARED - Versement 4 - Avis du jury si thèse soutenue	Pour versement solde tranche 3 ARED (cohortes jusqu'en 2018)		Non traité	
	ARED - Versement 4 - Attestation de réinscription en ED	Pour versement solde tranche 3 ARED (cohortes jusqu'en 2018)		Non traité	
	ARED - Versement 4 - Justificatif de prolongation	Pour versement solde tranche 3 ARED (cohortes jusqu'en 2018)		Non traité	
	ARED - Versement 4 - Justificatif d'abandon de thèse	Pour versement solde tranche 3 ARED (cohortes jusqu'en 2018)		Non traité	