Demande n°3516 - ARED

Acronyme :Miss2DA

Identification

Identification

N°	3516
Année	2024
Statut	Soumis à l'établissement
Date de création	26/01/2024
Date de modification	30/01/2024
Date de soumission	30/01/2024
Acronyme	Miss2DA
Intitulé du projet	Imputation of missing data in a domain adaptation context
Dispositif	ARED
Civilité du porteur	Madame
Nom du porteur	Chapel
Prénom du porteur	Laetitia
Courriel	laetitia.chapel@univ-ubs.fr

Priorité régionale

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	Économie numérique sécurisée et responsable
Domaine d'innovation stratégique	.

Organisme de tutelle

Organisme de tutelle

Etablissement	INSTITUT AGRO - ECOLE INTERNE AGROCAMPUS OUEST
Commune	RENNES

Civilité	Monsieur
Nom	Vollet
Prénom	Dominique

Unité de recherche

Unité de recherche

Nom de l'unité	UMR6074
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?)	
Le cas échéant, nom de l'équipe de recherche	
Le cas échéant, nom du responsable de l'équipe de recherche	
Activités de recherche et points forts de l'équipe	
Si implication dans un/plusieurs programmes européens ou internationaux, préciser	

Co-directeur de thèse

Co-directeur de thèse

Civilité	Monsieur
Nom	Tavenard
Prénom	Romain
Unité de recherche	LETG

Description

Description du projet

Libellé	Valeur
Résumé synthétique du projet (2000 caractères max)	Les méthodologies d'Intelligence Artificielle dépendent généralement de vastes ensembles de données qui peuvent être entachés de bruit ou de valeurs manquantes, ou qui peuvent être collectés dans des environnements hétérogènes mais apparentés. Les données manquantes sont omniprésentes dans de nombreuses applications ; néanmoins, les algorithmes d'apprentissage conventionnels supposent souvent que les données sont complètes et indépendantes et identiquement distribuées, c'est-à-dire qu'elles ont été tirées au hasard d'une distribution unique. L'imputation des données vise à remplacer les données manquantes par des valeurs plausibles, par exemple en les remplaçant par la valeur de l'échantillon le plus proche ou en les imputant à l'aide de statistiques pertinentes. L'imputation

Libellé	Valeur
	peut avoir un impact important sur les performances de la tâche d'apprentissage en question, conduisant à des résultats biaisés ou à des performances dégradées. La plupart des méthodes d'imputation reposent sur l'hypothèse d'une absence (totale) au hasard, ce qui signifie qu'il n'y a pas de modèle entre l'absence de données et les valeurs, ou qu'elle ne peut dépendre que des valeurs observées. Dans la pratique, les données sont souvent collectées dans des domaines différents mais apparentés, ce qui permet d'améliorer la capacité de généralisation de l'algorithme d'apprentissage. Par exemple, dans le domaine de l'observation de la Terre, on sait que les différences de conditions météorologiques, de sol ou de pratiques agricoles entre les sites d'étude induisent des décalages temporels qui peuvent être corrigés pour améliorer la performance de la tâche. L'adaptation au domaine vise à transférer les connaissances d'un domaine à un autre et a démontré des améliorations significatives dans les tâches d'apprentissage. L'objectif de la thèse est de proposer des méthodes d'imputation de données manquantes dans un contexte d'adaptation de domaine, avec comme objectif i) de pouvoir avoir des imputations de meilleures qualités grâce à la prise en compte d'informations complémentaires ii) de pouvoir disposer de données imputées fiables afin de résoudre des tâches d'apprentissage avec des performances améliorées.
Caractère structurant/stratégique du projet (niveau régional/établissement)	La thèse est co-dirigée par Laetitia Chapel (institut Agro Rennes-Angeres / IRISA) et Romain Tavenard (Université Rennes 2 / LETG). Elle permet de consolider une collaboration de longue date sur des thématiques en IA, avec des applications en observation de l'environnement. Les porteurs du projet pourront s'appuyer sur des expertises de leurs établissements et laboratoires respectifs (Institut Agro et LETG)
Porteur du projet : date d'obtention de l'HDR ou dérogation	Laetitia Chapel - HDR 2022
Hypothèses, questions posées, points de blocage, approche méthodologique, technique (4000 car. max)	L'objectif du projet de doctorat est de concevoir une méthode d'imputation des données dans le contexte de l'adaptation au domaine. Les approches existantes traitent généralement les valeurs manquantes dans un cadre inférentiel, où elles sont remplacées par des valeurs dérivées des statistiques de l'ensemble des données, en s'appuyant sur des hypothèses paramétriques robustes. Toutefois, lorsqu'il existe un décalage entre les ensembles de données, cette stratégie devient inadéquate. Au lieu de cela, nous proposons d'aborder les tâches d'imputation et d'apprentissage simultanément, en introduisant la complexité supplémentaire que les données peuvent provenir de domaines différents. Les principaux objectifs de la thèse sont les suivants : i) analyser théoriquement l'impact des changements de domaine sur la tâche d'apprentissage, à l'instar du cadre établi pour l'adaptation au domaine dans un contexte de classification ; ii) introduire de nouveaux schémas d'imputation dans des environnements hétérogènes en alignant les distributions dans une étape préliminaire et en appliquant ensuite des tâches d'apprentissage (par exemple, l'apprentissage supervisé) ; iii) proposer un cadre intégré pour l'imputation et l'apprentissage dans des environnements hétérogènes. Une attention particulière sera accordée au traitement des données de séries temporelles, où le défi est accru par la nécessité de tenir compte des corrélations temporelles entre les observations. Cela permettra de considérer le cas "blackout missing" (où toute une partie des données est manquante) comme un cas particulier de prédiction de séries temporelles. Les premiers axes de recherche exploreront les solutions basées sur le transport optimal, connues pour leur succès dans l'imputation des valeurs manquantes et l'alignement des distributions dans un contexte d'adaptation de domaine, en particulier lorsqu'il s'agit de données temporelles. D'autres pistes de recherche seront envisagées, telles que l'apprentissage profond adversaire et d'autres stratégies standard d'adaptation de domaine. D'un point de vue applicatif, l'accent sera mis sur la résolution des problèmes environnementaux à l'aide de données de télédétection.

Environnement scientifique, positionnement dans contexte régional/national/international (2000 car.)	
Collaborations scientifiques (nature/partenariat/pays) et partenariat socio-économique envisagé	Les porteurs du projet pourront s'appuyer sur un réseau national et international en ce qui concerne les collaborations scientiques : - ils co-encadrent une thèse CIFRE (avec Hawei) sur l'adaptation de domaine sur les séries temporelles (avec Ievgen Redko, un chercheur spécialiste en apprentissage dans des environnements hétérogènes) - le doctorant pourra bénéficier des collaborations multiples avec des chercheurs spécialistes du transport optimal (G. Gasso de l'Université de Rouen, Felipe Tobar de l'Université de Santiago du Chili) et de son application en adaptation de domaine (R. Flamary, Ecole polytechnique) ainsi que l'environnement dans une équipe de recherche en création à l'IRISA dont les séries temporelles sont un objet d'étude
Adéquation du projet avec le DIS de rattachement (3000 caractères maximum)	
Si priorité régionale, préciser (200 caractères maximum)	Projets dont la finalité contribuera à l'atténuation et/ou à l'adaptation au changement climatique Le projet est avant tout un projet méthodologique. Les applications portent sur la résolution des problèmes environnementaux à l'aide de données de télédétection dans un contexte évolutif : il permettra d'adapter des méthodes d'intelligence artificielle dans un contexte évolutif comme le changement climatique, en permettant une meilleure prise en compte de la dérive observée dans les données. Des exemples identifiés sont : - la prévision du rendement des cultures dans un climat qui évolue, afin d'assurer la sécurité alimentaire, la stabilité économique et des pratiques agricoles durables. - la conservation de la faune et de la flore, qui permet d'évaluer à grande échelle l'évolution des habitats de la faune et de la flore, les schémas de migration et l'évolution des populations. être évalués à grande échelle - la cartographie fine de l'occupation et de l'utilisation des sols, qui peut ensuite être utilisée pour la planification urbaine, la gestion de l'agriculture ou l'identification des activités illégales d'utilisation des sols, par exemple.
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)	Collège doctoral
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	mars 2023
Si cofinancement refusé, autre(s) source(s) de cofinancement identifiée(s)	ressources propres, notamment sur un contrat d'accompagnement avec un industriel
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
	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.		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	