



Projet ANR (2020-2024) « Hérèlles » : un cadre collaboratif unifié pour l'analyse interactive de données temporelles

Pierre Gançarski¹, Antoine Cornuejols², Bruno
Cremilleux³, Thi-Bich-Hanh Dao⁴, Maguelonne Teisseire⁵

Introduction

Face à la surabondance des données temporelles arrivant de façon quasi-continue, mais aussi à leur complexité et à celle des phénomènes étudiés, définir les informations nécessaires à une analyse de qualité de celles-ci s'avère de plus en plus difficile voire impossible à effectuer a priori. C'est le cas particulièrement lorsque l'on cherche à apprendre un modèle d'attribution de classes à des observations ou à des événements temporels. D'une part, la phase d'étiquetage de l'apprentissage supervisé ne peut plus être assurée par les experts, car trop fastidieuse et chronophage ; de surcroît, dans de nombreux domaines, du fait de la nouveauté de telles données, les champs sémantiques appropriés (principalement les classes à apprendre) peuvent n'avoir pas encore été parfaitement définis. D'autre part, les méthodes aptes à analyser des données temporelles sont elles aussi nombreuses, chacune traitant un type de données de façon privilégiée avec sa propre stratégie d'analyse, et donc ses propres biais.

-
1. Coordinateur du projet, ICube – université de Strasbourg, gancarski@unistra.fr.
 2. AgroParisTech – université Paris-Saclay, antoine.cornuejols@agroparistech.fr.
 3. GREYC - université de Normandie, bruno.cremilleux@unicaen.fr.
 4. LIFO – université d'Orléans, thi-bich-hanh.dao@univ-orleans.fr.
 5. TETIS – INRAE Montpellier, maguelonne.teisseire@irstea.fr.

Cette situation rend difficile le choix de la méthode « optimale » et de son paramétrage. Et même si des solutions existent pour combiner de telles méthodes (*boosting*, *co-training*, *ensemble learning*, *clustering* collaboratif [2]) et ainsi réduire l'impact de ces choix sur la qualité des résultats, ces approches se limitent majoritairement à la combinaison d'algorithmes partageant le même paradigme (e.g. supervisé ou non supervisé) ce qui limite leur intérêt.

L'objectif du projet HÉRELLES (*Hétérogénéité des données - Hétérogénéité des méthodes*) : un cadre collaboratif unifié pour l'analyse interactive de données temporelles) est d'étudier les concepts entrant dans la définition d'une approche fondée sur une collaboration entre des méthodes supervisées et non supervisées, autorisant une découverte progressive de ces informations à travers un cycle d'interactions entre l'expert et le système d'apprentissage. Ainsi, partant d'informations très lacunaires (peu ou pas de classes définies a priori, peu ou pas d'exemples...), l'expert pourra injecter à la volée des connaissances externes non encore exprimées et ce en fonction de l'avancement de l'analyse. Ceci devrait limiter fortement son implication directe et permettre une sémantisation plus aisée. Par ailleurs, cette interaction met en jeu des savoir-faire et des explicitations de connaissances propres au domaine d'application et à l'expert qui sont des biens précieux que le projet ambitionne de capitaliser et analyser pour améliorer les usages futurs du système.

Le consortium formé présente toutes les compétences nécessaires au succès du projet. Ainsi, l'équipe SDC d'ICube (Strasbourg) spécialisée en fouille de données, apporte son expertise en classification collaborative multistratégie de données temporelles guidée par des connaissances ainsi que ses compétences en sciences de l'environnement. Le laboratoire TETIS (Montpellier) spécialisé dans la compréhension des dynamiques des espaces ruraux et des territoires apporte son expertise en géographie, en télédétection et en fouille de données spatio-temporelles. L'équipe CODAG du Greyc (Caen) spécialisée sur les apports mutuels entre fouille de données et approches déclaratives apporte ses compétences sur la prise en compte de préférences explicites et l'éllicitation de préférences implicites de l'utilisateur. L'équipe CA du LIFO (Orléans) spécialisée en apprentissage automatique et fouille de données apporte ses compétences en *clustering* sous contraintes et en apprentissage supervisé actif. L'équipe Ekinocs d'AgroParisTech (Paris) particulièrement concernée par l'étude des systèmes complexes multiéchelles spatiales et temporelles apporte son expertise sur les méthodes collaboratives en apprentissage automatique et en fouille de données temporelles.

Hypothèses de recherche

En rupture avec les approches actuelles basées chacune sur un seul paradigme d'analyse, le projet scientifique d'HÉRELLES propose d'étudier et définir une architecture générique (Fig. 1) :

- permettant de faire collaborer des méthodes multiparadigmes travaillant potentiellement sur des données différentes et de définir les conditions optimales de son utilisation [10];
- basée sur des mécanismes d'interaction avec l'utilisateur lui offrant la possibilité d'injecter de nouvelles informations et de réduire le fossé sémantique entre les résultats et ses propres intuitions;
- offrant des méthodes d'extraction et de capitalisation des connaissances directement ou indirectement produites lors du processus d'analyse;
- possédant une mise en œuvre effective et une validation dans le cadre de l'analyse de séries temporelles hétérogènes.

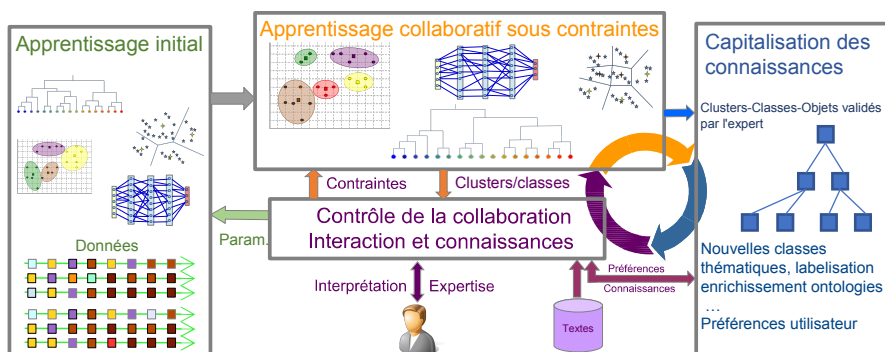


FIGURE 1. Une approche interactive pour la découverte de connaissances

Une approche collaborative. Concrètement, il s'agira de concevoir une approche autorisant la collaboration entre des algorithmes d'apprentissage différents à travers des transferts d'informations extraites à partir de données temporelles complémentaires. Par exemple, les clusters seront utilisés pour étiqueter des données ou comme classes thématiques potentielles, les données d'apprentissage serviront à la validation des clusters ou pour générer des contraintes, des informations sur des clusters pourront être partagées par toutes les méthodes... Cette collaboration sera contrôlée par des critères internes (convergence des résultats, adéquation avec des connaissances a priori...) et par l'expert qui interviendra en évaluant des résultats intermédiaires et en injectant de nouvelles informations (contraintes entre des objets, labélisation d'objets...).

Une approche active. L'expérience montre que demander à l'expert de fournir a priori toutes les informations nécessaires au processus de traitement des données n'est pas réaliste. Rendre incrémentales les méthodes d'apprentissage permettra à

l'utilisateur de guider le processus d'analyse en injectant de nouvelles données mais aussi, et surtout, de nouvelles connaissances. Cependant, devant la complexité ou la richesse des résultats, l'expert est souvent démuni et ne peut généralement pas proposer des modifications de ceux-ci ou l'ajout de données ou d'informations qui soient réellement pertinents. L'apprentissage actif permet de guider le processus d'analyse de données grâce à des phases de questions et de réponses dans lesquelles l'algorithme soumet à l'expert des propositions à infirmer ou confirmer. En général, ces questions découlent de l'analyse du résultat courant et se concentrent sur les objets difficiles à classer ou à affecter à un cluster. Dans ce projet, nous proposons que cette approche soit étendue à la découverte d'informations de plus haut niveau comme par exemple, les classes manquantes (existantes ou à créer dans la nomenclature) ou les liens sémantiques entre ces classes et les clusters en définissant des mécanismes d'analyse des résultats au regard de connaissances thématiques (ontologies, nomenclature...) ou de données complémentaires (textes, page web...). Parallèlement à l'étude de ces extensions, une part importante du projet sera consacrée au développement d'une interface d'interaction à la fois conviviale et efficace permettant de réduire l'effort cognitif de l'expert pour d'une part, comprendre les résultats et les questions posées par le système et d'autre part, y répondre de façon « optimale ».

Une approche générique. La méthode proposée sera générique et ne dépendra pas des domaines potentiels d'application. Afin d'en valider l'opérabilité, nous nous intéresserons à la compréhension de phénomènes complexes de notre environnement (artificialisation des sols, urbanisation, construction d'infrastructures...) majoritairement via des données hétérogènes majoritairement satellitaires et textuelles domaines dans lesquels l'expertise de TETIS et d'ICube est avérée. Pour la mise en œuvre concrète, nous nous appuierons sur la plateforme existante FODOMUST [6] développée par l'équipe SDC. Celle-ci sera étendue pour d'une part, intégrer toutes les avancées méthodologiques et algorithmiques et d'autre part, compléter l'interface MULTICUBE permettant une utilisation distante via un accès au serveur hébergé par l'université de Strasbourg.

Application à l'observation de la Terre

Notre volonté est d'aller plus loin qu'une simple preuve de concept. Pour cela, la plateforme FODOMUST sera le support d'une validation méthodologique via des mécanismes génériques et d'une validation thématique via des interfaces et développements dédiés à l'observation de la Terre. En effet, pour appréhender, observer, comprendre les dynamiques des paysages, les experts en sciences de la Terre et de l'environnement disposent de volumes considérables et toujours croissants de données d'observation satellitaires arrivant de façon massive en flux quasi continu. Avec le lancement et la mise en production des satellites européens de la constellation

Sentinel, ou franco-israélien *Venµs*, ces images de la Terre sont acquises à une fréquence jamais encore atteinte. De plus, elles sont de plus en plus souvent mises librement à la disposition des scientifiques, des décideurs ou des utilisateurs du domaine. Néanmoins, utiliser de façon optimale ces données reste un défi. En effet, peu d'utilisateurs sont capables d'exploiter tout le potentiel offert par des séries temporelles d'images satellitaires sans une aide leur permettant de relier les informations qu'elles fournissent entre elles ou à celles issues d'autres sources et surtout à leurs propres connaissances. Appliquer les propositions faites par HÉRELLES permettra de confronter l'approche à la réalité et ainsi de la valider concrètement, mais aussi de profiter des retours d'expériences afin d'en améliorer l'interface et le mode de fonctionnement.

Un exemple de cas d'étude type. L'artificialisation des sols est un enjeu important, notamment pour la métropole de Montpellier Méditerranée (3M) car cette agglomération se développe de façon très active. La croissance démographique s'y illustre par un accroissement des surfaces bâties et des infrastructures aux dépens des zones naturelles ou agricoles. Pour appréhender et comprendre les déterminants de cette artificialisation, que la seule démographie ne suffit pas à cerner, le cas d'étude se fondera sur des analyses spatio-temporelles à travers des jeux de données hétérogènes d'imagerie satellitaire associés à d'autres sources de données (textes, bases de données, etc.). L'objectif est de fournir des clés qui permettent à la collectivité d'accueillir les nouveaux citoyens tout en maîtrisant son artificialisation et en étant responsable face aux objectifs de développement durable. Ce cas d'étude s'inscrit dans les quatre hypothèses formulées dans le cadre du projet HÉRELLES. Une méthodologie collaborative multiparadigme permettra de répondre aux analyses demandées par les acteurs sur le territoire car celles-ci sont variées et fortement dépendantes des données et connaissances disponibles. Une approche incrémentale et interactive se justifie pleinement pour une co-construction des connaissances des dynamiques territoriales. La sémantisation est indispensable à l'appropriation et la capitalisation des résultats obtenus. Enfin, la dimension temporelle est un élément incontournable de l'analyse des dynamiques.

Programme scientifique

Outre une tâche dédiée à la coordination du projet, nous avons structuré le programme scientifique en quatre tâches, chacune étant portée par deux partenaires.

Collaboration multiparadigme (T1). L'objectif est de proposer une nouvelle architecture opérationnelle et un cadre théorique pour faire collaborer des techniques d'apprentissage supervisé et non supervisé dans une procédure itérative, d'étudier les conditions de la mise en œuvre d'une telle architecture et enfin de définir des mécanismes et conditions d'optimisation du processus. Pour cela, dans un premier temps, nous nous appuierons sur les principaux concepts mis en œuvre dans la méthode de

clustering collaboratif SAMARAH [8]. Nous y reprendrons l'hypothèse de base à savoir que la qualité d'une solution globale optimale ou satisfaisante pour l'expert est fortement corrélée à la qualité individuelle de chacun des résultats locaux produits par les méthodes impliquées dans la collaboration, leur similarité ainsi que le respect des connaissances et contraintes imposées par l'expert (labels initiaux des objets, contraintes de clustering...). Nous nous inspirons du processus de raffinement des résultats locaux basé sur des interactions (détection de divergences et tentative de résolution de celles-ci) entre les méthodes prises deux à deux. Étendre cette approche vers une architecture multiparadigme autorisant la collaboration entre des méthodes supervisées et non supervisées nécessite de lever quatre verrous :

- quelles mesures de compatibilité ou de désaccord entre méthodes supervisées, non supervisées et mixtes faut-il définir ?
- quelle est la stratégie optimale permettant aux agents impliqués de modifier localement leurs résultats en vue de maximiser la compatibilité ou de réduire ces désaccords ?
- comment trouver un formalisme commun de communication entre les différentes méthodes pour leurs échanges de données et de résultats ?
- quelle stratégie globale innovante doit être mise en œuvre afin de garantir une convergence de l'ensemble de classifieurs vers une solution satisfaisante pour l'expert ?

Dans un deuxième temps, si la durée du projet nous le permet, nous chercherons à optimiser la stratégie mise en œuvre en tentant de caractériser l'effet de la diversité et de la décorrélation des informations apportées par les méthodes [10] afin de garantir la convergence des processus et la prévention des collaborations négatives pouvant amenant une dégradation du résultat global ou des résultats individuels des méthodes.

Incrémentalité et interactions (T2). L'objectif est, d'une part, d'intégrer dans des méthodes d'apprentissage (principalement sous contraintes [1, 7, 3]) des mécanismes efficaces leur permettant d'exploiter de nouvelles données ou informations (nouvelles classes, labélisations, contraintes...) fournies par l'expert à la volée et d'autre part, de concevoir une méthode d'interaction active en proposant des mécanismes de sélection des informations ou questions à soumettre à l'expert et d'optimisation des retours de celui-ci.

Pour cela, dans un premier temps, nous rendrons incrémentales les méthodes d'apprentissage tout en optimisant la prise en compte des nouvelles informations. En effet, si l'expert ne voit pas l'amélioration notable de la solution grâce à son aide, il perdra rapidement confiance dans le système. Paradoxalement, il est préférable que tout nouveau résultat ne nécessite pas un effort cognitif important de la part de l'expert, du fait de la perturbation consécutive à cette information (autrement dit, le

nouveau résultat ne devra pas trop s'éloigner du précédent) au risque de le dérouter. Nous ferons en sorte de lever trois verrous principaux :

- comment tirer parti des indications venant de l'expert afin que celles apportant le plus d'information ou considérées comme les plus importantes par l'expert soient satisfaites en priorité ?
- comment limiter les perturbations de la solution courante issue de la collaboration en cours ?
- quels mécanismes faut-il mettre en place pour détecter et résoudre les incohérences potentielles induites par les contraintes ajoutées à la volée ?

Dans un deuxième temps, nous nous intéresserons à une méthode d'apprentissage actif générant automatiquement les questions sur des classes, labels, contraintes... à soumettre à l'expert. Les informations validées seront alors transmises au processus collaboratif. Il s'agit ici, là encore, de lever trois verrous :

- les différences entre les résultats dues à l'hétérogénéité des méthodes peuvent-elles être exploitées ?
- peut-on se baser sur un historique des retours des experts pour tenter d'identifier les contraintes les plus informatives et évaluer leur impact sur le modèle et ainsi en limiter les perturbations ?
- comment présenter les résultats et ces propositions afin d'aider l'expert à valider ou infirmer les propositions.

Il apparaît de façon évidente que l'interface nécessaire à cette interaction est cruciale et dépend fortement des domaines d'application visés. Sa conception sera étudiée dans la tâche T4.

Capitalisation des connaissances (T3). L'objectif est de concevoir des mécanismes originaux favorisant la sémantisation des résultats (nommage des entités et des classes, labélisation ou étiquetage d'objets ou de clusters...) et des profils d'évolution détectés, de proposer de nouveaux moyens d'identification et d'apprentissage des besoins et préférences de l'utilisateur et de mise en place de mécanismes de génération à partir de connaissances du domaine (ontologies, textes...) de contraintes exploitables par le processus collaboratif. Pour cela, dans un premier temps, nous nous concentrerons sur des mécanismes permettant d'associer une sémantique à des objets d'intérêt (éléments géographiques, classes ou clusters...) en s'appuyant sur une mise en correspondance avec des éléments textuels extraits de données exogènes (journaux, textes réglementaires...) [4] analysés selon leurs dimensions spatiales, temporelles et thématiques [9, 7, 5]. Il s'agit toujours de lever trois verrous :

- quel mécanisme d'extraction des éléments textuels exogènes potentiellement intégrables dans le processus de sémantisation doit-on définir ?
- comment ancrer spatialement et temporellement ces éléments aux résultats à partir des métadonnées disponibles sur ces derniers et des descripteurs spatio-temporels issus de textes ?

— quand et comment faire intervenir l'expert dans la définition et le choix des descripteurs caractérisant les clusterings et contribuant à leur sémantisation.

Dans un deuxième temps, si la durée du projet nous le permet, nous nous efforcerons de concevoir une méthodologie pour éliciter les préférences exprimant l'intérêt de l'utilisateur. En effet si l'utilisateur a une idée de ce qu'il cherche, il a souvent du mal à exprimer ses besoins et à les traduire en des mécanismes opérationnels par les méthodes de clustering.

Plateforme FoDOMUST et validation (T4). Il s'agit ici de valider les avancées méthodologiques et thématiques dans le domaine particulier de l'observation de la Terre, de les pérenniser dans une plateforme opérationnelle munie d'une interface conviviale permettant d'interagir avec le processus. Or, évaluer un projet de recherche exploratoire, bien que présentant de fortes connexions avec des domaines applicatifs, est par nature difficile. Les objectifs visés par HÉRELLES sont méthodologiques, voire théoriques, et non totalement garantis et de fait difficilement quantifiables. Pour les apports méthodologiques, nous nous appuierons principalement sur des validations statistiques via les critères classiques d'évaluation des méthodes d'apprentissage. Nous les compléterons par des critères propres aux approches collaboratives telles que la convergence des agents ou l'influence de la diversité sur la qualité des solutions. Pareillement, nous nous intéresserons à une validation thématique pour laquelle, il n'existe malheureusement pour l'instant que très peu de données de référence. Or, en l'absence de vérité terrain, voire de connaissances du domaine formalisées, il a été démontré que les critères classiques peuvent être faillibles et ne sont que peu utiles dans ce contexte. Ainsi, la validation thématique ne pourra se faire qu'à dire d'expert dans un premier temps. Néanmoins, nous essaierons, dès que cela est possible et si cela présente un intérêt scientifique, de rattacher ces résultats à des données de référence monodates. Nous nous efforcerons de faire évaluer par les communautés concernées les apports de la méthode dans leur domaine d'expertise. Nous nous intéresserons principalement aux capacités de notre méthode à proposer des interactions les plus efficaces possibles tout en imitant l'implication de l'expert. Il s'agira aussi, tout au long du projet, d'apprécier la facilité d'usage et d'analyser la performance des méthodes via des enquêtes, ateliers dédiées, hackathons... Une définition participative des critères d'évaluation ou de validation associée à une co-conception de la méthode nous semble le moyen le plus sûr d'atteindre cet objectif de validation.

Une architecture innovante pour une approche innovante

La première année de projet, en plus d'avancées ponctuelles mais néanmoins très

intéressantes dans chacune des tâches, nous a permis de formaliser le processus global d'analyse de séries temporelles (d'images satellitaires dans notre cas) et d'extraction de connaissances à partir de données exogènes. La figure 2 présente visuellement ce processus et son lien avec les différentes tâches (T_i). L'expert est au cœur du processus. En validant (ou non) les propositions faites par le système ou en injectant de nouvelles connaissances exogènes, il influe fortement sur le processus. Les premières expérimentations faites dans le cadre du suivi de coupes forestières par l'approche collaborative incrémentale semblent valider notre hypothèse en donnant des résultats très encourageants.

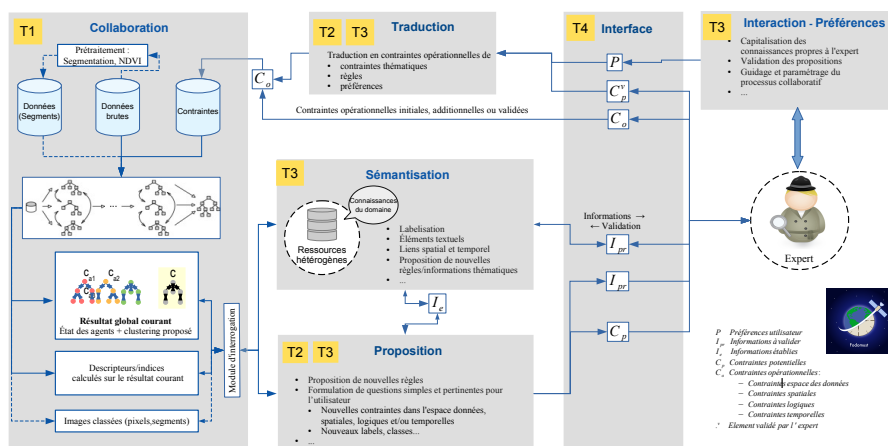


FIGURE 2. Architecture générale et découpage en tâches.

Impacts et retombées du projet

Avancées méthodologiques. Ce projet induira des retombées scientifiques multiples que nous tâcherons de diffuser massivement. On peut citer, au niveau global, la définition, l'intégration et l'optimisation d'une nouvelle méthode collaborative multi-paradigme dans une plateforme ouverte à des usages dans des domaines applicatifs variés ; le développement d'outils et de méthodes d'interaction entre le système et l'expert l'autorisant à intervenir sur le processus incrémental afin de limiter fortement ses interventions ou encore l'étude de l'applicabilité de contraintes thématiques à des processus non supervisés de fouille de données spatio-temporelles. L'usage et la validation thématique en observation de la Terre depuis cette plateforme devraient

permettre la découverte incrémentale de nouvelles classes d'évolution, voire de taxonomies nouvelles, indispensables à la compréhension des dynamiques des paysages. Nous veillerons à les présenter sous une forme diffusable et compréhensible par les différents acteurs et décideurs dans ce domaine. Concrètement, le projet débouchera sur une solution logicielle ouverte et libre de droit pérennisée via la plateforme FODOMUST ; ce qui devrait permettre un essor de ces approches.

Jeux de données et résultats. En plus de la valorisation des résultats scientifiques par des publications et communications, nous diffuserons des jeux de données de référence (*benchmark*) vers la communauté scientifique, action devenue à la fois classique et obligatoire dans le domaine de la fouille de données et de l'apprentissage. Ainsi, à l'issue du projet, les bases de données initiales, de même que celles constituées à partir des résultats obtenus, seront ouvertes et librement utilisables.

Dynamique des paysages. Enfin nous encouragerons les publications interdisciplinaires, en particulier autour de la géomatique et de la télédétection voire du TALN (revues RIG, IJGIS et conférences SAGEO, AGILE et SIG Spatial...). De même, des communications auront lieu, par exemple, dans le cadre des journées et ateliers des GdR MAGIS ou dans le cadre d'un atelier organisé dans la communauté d'agglomération de Montpellier, en collaboration avec le syndicat mixte du bassin de Thau (SMBT), ainsi qu'avec les participants à THEIA (en particulier ceux du CES Détection de changement).

Conclusion

Le projet HÉRELLES est ambitieux par sa volonté d'apporter des solutions dans l'analyse de données temporelles pour lesquelles l'expertise est encore incomplète et insuffisante pour une approche supervisée. En premier lieu, il s'adresse à des experts de domaines pour lesquels il n'existe pas encore de nomenclature et de bases d'apprentissage complètement définies ou opérables. Pour pallier ce manque, tout en limitant l'investissement de l'expert, la solution proposée se fonde sur une découverte progressive pendant l'exploitation et l'analyse des données à travers un cycle d'interactions entre l'expert et le système d'apprentissage. Cette interaction vise à réduire le fossé entre les résultats produits par les algorithmes et les intuitions thématiques de l'expert et rendre les résultats plus compréhensibles pour celui-ci, donc plus aisés à associer à une sémantique du domaine d'application. De plus, afin de limiter fortement les choix initiaux (données à analyser, données d'apprentissages, méthodes et paramètres) à toute analyse, il s'appuie sur les concepts mis en œuvre dans l'approche de clustering collaboratif incrémental SAMARAH qui a fait ses preuves dans de nombreux domaines. Son principal domaine d'expérimentation et de validation est celui des sciences de la Terre et de l'environnement. En effet si l'usage de l'imagerie satellitaire y est fortement répandu en analyse monodate, il est encore très

limité, dès qu'il s'agit d'analyses temporelles. En effet, la mise à disposition de données temporelles à haute fréquence d'acquisition est encore relativement récente. De fait, excepté quelques domaines tel que, par exemple, l'agriculture, il n'existe pas de telles nomenclatures et donc de données d'apprentissage opérables. Concevoir une plateforme complète et totalement utilisable en 4 ans est un défi. Néanmoins, nous sommes confiants sur notre capacité à produire des résultats scientifiques et des prototypes qui pourront servir de base à des travaux plus approfondis tant théoriquement qu'en termes de logiciels de qualité. Nous invitons toute personne intéressée par ces perspectives à nous contacter.

Remerciements

Ces travaux sont soutenus par l'Agence nationale de la recherche dans le cadre du projet Hérelles (ANR-20-C23-0022).

Références

- [1] Aymeric Beauchamp, Thi-Bich-Hanh Dao, Samir Loudni, and Christel Vrain. Intégration incrémentale de contraintes pour le clustering avec la programmation par contraintes. In *PFIA - Journées Franco-phones de Programmation par Contraintes, Saint-Étienne, 2022*.
- [2] Antoine Cornuéjols, Cédric Wemmert, Pierre Gançarski, and Younes Bennani. Collaborative clustering : Why, when, what and how. *Information Fusion*, 2018, volume = 39, pages = 81–95..
- [3] Thi-Bich-Hanh Dao, Khanh-Chuong Duong, and Christel Vrain. Constrained clustering by constraint programmin. *Artificial Intelligence*, 244 :70–94, 2017.
- [4] Roche Mathieu et al. Animitex : Analyse d'images fondée sur des informations textuelles. *Revue des Sciences et Technologies de l'Information*, 19(13) :163–167, 2014.
- [5] Jérôme Fize, Mathieu Roche, and Maguelonne Teisseire. Could spatial features help the matching of textual data ? *Intelligent Data Analysis*, 24(5) :1043–1064, 2020.
- [6] SDC FoDoMuST. Version interface déportée - client léger. <https://sdc.icube.unistra.fr/en/index.php?title=FODOMUST>, Août 2022.
- [7] Pierre Gançarski, Thi-Bich-Hanh Dao, Cremilleux Crémilleux, Germain Forestier, and Thomas Lampert. *A Guided Tour of AI Research*, chapter Constrained Clustering : Current and New Trends. Springer, 2020.
- [8] Pierre Gançarski and Cédric Wemmert. Collaborative multi-step mono-level multi-strategy classification. *Multimedia Tools and Applications*, 35(1) :1–27, 2007.
- [9] Urcel Kalenga-Tshingomba, Magali Jouven, Lucile Sautot, Imad Shaqura, and Maguelonne Teisseire. Cartographie des surfaces pastorales à l'aide des données sentinel 2 l3a et des données ouvertes : promesses et réalités. *Revue Internationale de Géomatique*, À paraître, 2022.
- [10] Parisa Rastin, Barasab Matei, Guénaël Cabanes, and Younès Grozavu, Nistor et Bennani. Impact of learners' quality and diversity in collaborative clustering. *Journal of Artificial Intelligence and Soft Computing Research*, 9(12) :149–165, 2019.