

Annexe 4

MODELE DE RAPPORT D'ACTIVITE ATLANSTIC 2020

A. IDENTIFICATION DU PROJET

Type d'action financée	Challenge post-doc international
Titre du projet	Learning Adequate Semantic from Transition Observations with Functional Uncertainty and Stochasticity
Acronyme	LAST OF US
Porteur	Morgan Magnin
Partenaires	National Institute of Informatics (Tôkyô, Japan)
Unité de rattachement	LS2N (Laboratoire des Sciences du Numérique à Nantes)

Période du projet	01/03/2020-31/12/2021
Période faisant l'objet du rapport d'activité	01/03/2020-31/12/2021

Budget total	71k€
Financement A2020	50k€
Co-financement(s) obtenus	21k€ par le National Institute of Informatics

B. REALISATION DU PROGRAMME SCIENTIFIQUE

B1. Description des travaux effectués

- Phase 1 (6 mois, mars-septembre 2020 au National Institute of Informatics à Tôkyô) : grâce au séjour de Tony Ribeiro au NII pendant la première partie du projet, les questions théoriques ont fait l'objet d'un échange soutenu avec le partenaire international du projet (Katsumi Inoue au NII), collaborateur reconnu pour son expertise en programmation logique inductive et en apprentissage automatique.

o La première tâche a consisté en une étude de l'état de l'art de l'apprentissage de la sémantique dans les domaines de l'apprentissage automatique et de la programmation logique.

o Nous nous sommes concentrés sur la définition et la modélisation théorique de la sémantique de dynamique dans le contexte de la programmation logique.

o Enfin, nous avons conçu un algorithme nouveau d'apprentissage de la sémantique de la dynamique par programmation logique inductive.

o Écriture d'un article de revue pour le Machine Learning journal soumis en septembre 2020, révisé deux fois pendant la phase 2 en France et finalement accepté et publié en novembre 2021. Présenté à la 30e Conférence internationale sur la programmation logique inductive ILP 2020-2021 et récompensé du *best paper award*.

o Implémentation Python3 des nouveaux algorithmes dans le dépôt open source GitHub de LFIT : <https://github.com/Tony-sama/pylfit>

- **Phase 2 (12 mois, janvier-décembre 2021 au LS2N à Nantes)** : Tony a dans un premier temps consolidé le travail théorique du projet, soumis à la revue "*Machine Learning Journal*" (soumis en septembre 2020) proposant une version améliorée de la première soumission. Cette révision majeure inclut une nouvelle approche heuristique probabiliste permettant d'appliquer nos méthodes sur des jeux de données réalistes avec une évaluation à la fois des pronostics et diagnostics des modèles appris par nos algorithmes. Cette version a été acceptée en août 2021 et a été lauréate du "*best paper award*" à la conférence internationale sur la programmation logique ILP 2020-2021.

En parallèle, Tony a développé une extension de l'Implémentation en une librairie open source Python3, incluant une API utilisateur facilitant l'utilisation de nos algorithmes par des tiers et disponible sur PyPI.org : <https://pypi.org/project/pylfit/>

Cet effort dans l'accessibilité de nos travaux nous à permis d'entamer une première collaboration avec une équipe de l'école polytechnique supérieur de l'université autonome de Madrid, où LFIT est utilisé pour expliquer les décisions d'un model « boîte noire » de classification. Une preuve de concept a été développée sur un jeu de données de "*metadata*" de C.V. et résultats de recrutement, publiée dans le workshop WACV 2021, puis étendue dans un article de revue de MDPI Computers.

Le but initial de cette phase 2 était de mettre en valeur le travail théorique conçu en montrant son apport sur des applications réelles, issues de collaboration avec des collègues biologistes. Trois voies fructueuses ont été explorées :

(A) l'une avec l'équipe de François Paris (équipe PETRY, du Centre de Recherche en Cancérologie et Immunologie Nantes Angers) sur la modélisation du comportement dynamique de cellules soumises à différents protocoles de radiothérapie. L'analyse des données est en cours de soumission, ainsi qu'une demande de financement pour continuer les travaux. Cette demande de financement a été déposée à l'appel à projets « Approches interdisciplinaires des processus oncogéniques et perspectives thérapeutiques : apports à l'oncologie des mathématiques et de l'informatique » organisé par l'INSERM (notification des résultats en mai 2022).

(B) une seconde avec une équipe de Boulogne-sur-Mer sur les interactions dynamiques dans les écosystèmes marins, dans une collaboration impliquant l'équipe BioComputing du CRISTAL. Ce travail a été présenté à ILP 2020-2021 dans un *late-breaking abstract* et un poster ;

(C) une dernière avec l'équipe SPARKS du laboratoire I3S de l'Université Côte d'Azur de Nice sur l'accélération de modélisation en utilisant LFIT pour généraliser les résultats du solveur TotemBionet.

B2. Résultats obtenus

Nous avons conçu de nouveaux développements méthodologies sur le framework LFIT, répartis comme suit :

- **GULA (General Usage LFIT Algorithm)** :
 - o Extension permettant une distinction *feature/target* pour la caractérisation des sémantiques capturées (les sémantiques concernées doivent être pseudo-idempotentes).
- **Synchronizer** :
 - o Nouveau algorithme qui capture le comportement de n'importe quelle sémantique sans mémoire.
 - o Formalisation de l'apprentissage de nouvelles règles empêchant des comportements sous la forme de contraintes
- **WDMVLP (Weighted Dynamic Multi-Valued Logic Program)** :
 - o Nouvel model prédictif heuristique permettant pronostique et diagnostique dans des situations non observées.
 - o Nouvel algorithme utilisant GULA pour apprendre un WDMVLP
 - o Évaluation sur des modèles venant de la littérature bio-informatique
- **PRIDE** :
 - o Nouvel algorithme heuristique
 - o Approximation de GULA
 - o Polynomial et donc pouvant être mis à l'échelle sur de grands jeux de données
- Mise en œuvre d'une bibliothèque codée Python, sous licence open source, de l'ensemble des algorithmes inclus dans le framework LFIT :

- Utilisation dans le cadre d'une collaboration avec l'université autonome de Madrid (UAM) sur l'explication du comportement de model statistique black-box sur un jeu de données de CV/recrutement (GULA/PRIDE).
- Utilisation dans le cadre d'une collaboration avec l'équipe SPARKS de l'I3S à l'Université Côte d'Azur sur l'extraction de connaissances depuis les sorties de leur logiciel TotemBionet (GULA/PRIDE).
- Utilisation dans le cadre d'une collaboration avec l'équipe Biocomputing au sein du laboratoire lillois CRISAL sur l'extraction de connaissances sur un jeu de données de biologie marine (PRIDE).
- Première démarche visant à intégrer la bibliothèque dans le consortium CoLoMoTo (travail en cours)

Ces résultats ont donné lieu à plusieurs publications (voir la partie C pour les références complètes) :

- Deux revues internationales avec comité de lecture :
 - Machine learning journal, portant sur GULA/Synchronizer/WDMVLP
 - MDPI Computers, portant sur l'utilisation de LFIT pour l'IA explicable (collaboration avec l'Université de Madrid)
- Un award: ILP 2020-21: *Best paper award* pour l'article accepté par *Machine Learning Journal*.
- Deux abstracts étendus et posters :
 - ILP 2020-21 : PRIDE, correspondant à la version polynomiale de GULA
 - ILP 2020-21 : modélisation des écosystèmes marins avec LFIT (collaboration avec le laboratoire CRISAL)
- Deux articles en cours de rédaction :
 - Une revue internationale sur l'analyse des données de radiothérapie avec l'équipe de François Paris
 - Un article de conférence internationale sur l'interfaçage avec TotemBionet dans le cadre de la collaboration avec l'équipe SPARKS de l'I3S à l'Université Côte d'Azur

B3. Difficultés rencontrées

La difficulté majeure de ce séjour a été l'**impact de la crise du coronavirus dans l'organisation de la recherche au Japon**.

Le projet a démarré en mars 2020 avec le déplacement de Tony Ribeiro au Japon pour 6 mois. Si ce dernier a pu arriver dans l'équipe de Katsumi Inoue à Tôkyô quelques jours avant la fermeture des frontières due à la crise COVID, les conditions de travail avec les collègues japonais ont été très rapidement bouleversées : elles sont ainsi passées de réunions présentiels au tout distanciel, entraînant un ralentissement dans la conception du socle théorique du projet (phase 1 du projet). Qui plus est, le lieu de publication visé pour la validation de ce socle théorique (la conférence ILP2020) a été annulé et remplacé par un "*special issue*" de la revue "*Machine Learning Journal*". Or le rythme de *reviewing* n'est évidemment pas le même entre une conférence et un journal, ni le temps de travail dessus. Un article a ainsi été soumis à la revue MLJ en septembre 2020, les premières *reviews* sont arrivées en janvier 2021, pour une acceptation finale en août 2021.

L'impact de la crise COVID dans la phase 1, ainsi que l'impossibilité de réaliser les déplacements courts envisagés en phase 2 (venue de Katsumi Inoue en France + visites au NII) ont entraîné un décalage dans la consolidation des résultats obtenus sur ces applications réelles et la valorisation de ceux-ci. Cette valorisation démarre tout juste :

- Dans le cas de la collaboration avec François Paris de l'INSERM sur la modélisation du comportement dynamique de cellules soumises à différents protocoles de radiothérapie, perspective de co-écriture d'un article de revue ;
- Dans le cas de la collaboration avec l'équipe de Boulogne-sur-Mer sur les interactions dynamiques dans les écosystèmes marins, acceptation d'un "*late-breaking abstract*" à ILP2021, mais qui gagnerait à être consolidé pour être soumis dans un *track* régulier de conférence.

B4. Frais engagés

	<i>Atlantc 2020</i>	<i>National Institute of Informatics</i>
<i>Frais de missions</i>		1k€
<i>Salaire (doctorat, post-doc)</i>	50k€	20k€
<i>Gratification de stage</i>		
<i>Equipements</i>		
<i>Autres (préciser)</i>		

C. LIVRABLES EXTERNES REALISES

Pendant la durée de ce projet, nous avons travaillé sur la liste suivante de publications pour rendre compte de nos avancées sur le framework LFIT :

Publication acceptée :

Article de revue avec comité de lecture :

T. Ribeiro, M. Folschette, M. Magnin and K. Inoue. Learning any memory-less discrete semantics for dynamical systems represented by logic programs. Machine Learning Journal, Springer, 2021. <https://doi.org/10.1007/s10994-021-06105-4>

A. Ortega, J. Fierrez, A. Morales, Z. Wang, M. de la Cruz, C. L. Alonso, T. Ribeiro. Symbolic AI for XAI: Evaluating LFIT Inductive Programming for Explaining Biases in Machine Learning. MDPI Computers, volume 10, issue 11. <https://doi.org/10.3390/computers10110154>

Conférence internationale avec comité de lecture :

T. Ribeiro, M. Folschette, M. Magnin and K. Inoue. Learning any memory-less discrete semantics for dynamical systems represented by logic programs. Publié dans le Machine Learning Journal, Springer, 2021. Présenté dans la conférence ILP 2020-2021, best paper award.

Chapitre de livre dont la parution est prévue dans les prochains mois :

T. Ribeiro, M. Folschette, L. Trilling, N. Glade, K. Inoue, M. Magnin and O.F. Roux. Les enjeux de l'inférence de modèles dynamiques des systèmes biologiques à partir de séries temporelles. Dans "Approches symboliques de la modélisation et de l'analyse des systèmes biologiques". Éditeurs : Élisabeth RÉMY et Cédric LHOUSSAINE. 40 pages. À paraître chez ISTE Sciences, 2022.

Workshop :

O. Iken, M. Folschette and T. Ribeiro. Automatic Modeling of Dynamical Interactions Within Marine Ecosystems. Late-breaking abstracts et poster, ILP 2020-2021.

T. Ribeiro, M. Folschette, M. Magnin and K. Inoue. Polynomial Algorithm For Learning From Interpretation Transition. Late-breaking abstracts et poster, ILP 2020-2021.

A. Ortega, J. Fierrez, A. Morales, Z. Wang, T. Ribeiro. Symbolic AI for XAI: Evaluating LFIT Inductive Programming for Fair and Explainable Automatic Recruitment, (WACV 2021), Waikola, HI, USA.

Médiation :

Participation aux Journées Portes ouvertes du National Institute of Informatics (12 et 13 juin 2020) qui se sont tenues en ligne (environ 400 personnes « connectées » pendant ces deux jours). Chaque équipe disposait d'un environnement numérique présentant les travaux menés.

- Environnement spécifique pour l'équipe de Katsumi INOUE : <https://openhouse.nii.ac.jp/wp/C02/>
- Environnement général par lequel les participants pouvaient accéder (sur ordinateur ou avec des lunettes de réalité virtuelle de type Google Cardboard) : <http://openhouse.nii.ac.jp>
- Programme complet des Portes Ouvertes : <https://www.nii.ac.jp/openhouse/>

Logiciels :

- Pylfit : bibliothèque python open source de LFIT - <https://github.com/Tony-sama/pylfit>
 - o Principaux algorithmes LFIT disponibles
 - GULA/PRIDE/Synchronizer (voir plus haut)
 - LFkT/LUST (capture des délais / non-déterminisme)
 - ACEDIA (variables continues)
 - o API utilisateur
 - Interface jeu de données/LFIT/model/prédiction
 - Facilite l'utilisation et l'interfaçage avec d'autre bibliothèque de machine learning/analyse

D. INDICATEURS DE VISIBILITE

Les principaux indicateurs de visibilité du projet sont (et seront) :

- les publications référencées au paragraphe C.
- Best paper award à la conférence ILP 2020-2021 pour l'article de revue "*Learning any memory-less discrete semantics for dynamical systems represented by logic programs*"
- la mise en œuvre d'une co-tutelle de thèse à partir d'avril 2021 (thèse de doctorat de Mitsuhiro Odaka entre Centrale Nantes et SOKENDAI)
- la mise à disposition publique, en licence libre, des implémentations du framework LFIT : <https://github.com/Tony-sama/pylfit>

E. PERSPECTIVES

Les travaux théoriques menés ont conduit à de premiers résultats biologiques encourageants, comme évoqué dans la partie B.1. Cette valorisation démarre tout juste et gagnerait à pouvoir être prolongée avec un travail plus approfondi. C'est pourquoi nous sommes désormais dans la recherche active de financements complémentaires pour envisager cette étape.

En France :

- Dépôt d'un projet à l'AAP "Approches interdisciplinaires des processus oncogéniques et perspectives thérapeutiques : apports à l'oncologie des mathématiques et de l'informatique". Projet intitulé "*Modeling and Inference of Dynamical Group for Application to Radiotherapy Data*", avec l'équipe de François Paris comme partenaire (329.412€ demandés)

A l'international :

- Visite scientifique au National Institute of Informatics lorsque les frontières rouvriront (frontières fermées jusqu'à fin février 2022) dans le cadre de la participation de T. Ribeiro et M. Magnin au nouveau projet Kakenhi KIBAN-A de Katsumi INOUE (2021-2025), "*Robust AI by Integration of Knowledge Representation and Machine Learning*"
- Possible réponse conjointe franco-japonaise sur l'appel à projets japonais JSPS qui sera publié au printemps 2022, baptisé RECONNECT, visant à contrecarrer la baisse drastique du nombre de chercheurs japonais partant à l'international et à dynamiser les partenariats internationaux. Cet appel est baptisé RECONNECT et financera des projets d'une durée pouvant aller jusqu'à 7 ans et jusqu'à 500 millions de yens par projet (soit environ 4 millions d'euros) : https://www.jsps.go.jp/j-grantsinaid/35_kokusai/05_sendou/data/sendou_e.pdf

À l'issue de ce projet, Tony Ribeiro a démarré un nouveau projet de recherche appliquée sur l'apprentissage de la dynamique de données pour une entreprise japonaise (SONAR Co. LTD). Il souhaite porter ses efforts sur la mise en lien du monde académique et du monde industriel.