

Annexe 4

RAPPORT D'ACTIVITE

A. IDENTIFICATION DU PROJET

Type d'action financée	Sabbatique Internationale
Titre du projet	Learning Logical Modeling From Dynamic Properties of Biological Systems
Acronyme	LearnDynBio
Porteur	MAGNIN Morgan
Partenaires	National Institute of Informatics (NII), Tôkyô, Japan
Unité de rattachement	LS2N, Ecole Centrale de Nantes

Période du projet	19/01/2020 - 14/07/2020
Période faisant l'objet du rapport d'activité	19/01/2020 - 14/07/2020

B. REALISATION DU PROGRAMME SCIENTIFIQUE

B1. Description des travaux effectués

Le travail mené pendant ce séjour de 6 mois au National Institute of Informatics (NII) s'est concentré sur trois axes majeurs.

(A1) **Sur le plan scientifique**, je me suis concentré sur l'**approfondissement des travaux autour du framework d'apprentissage *Learning From Interpretation Transition (LFIT)*** qui constitue le cœur de notre collaboration actuelle avec l'équipe de Katsumi INOUE. LFIT est un paradigme de programmation logique inductive construisant automatiquement un modèle de la dynamique d'un système à partir de l'observation de ses états-transitions. Dans le cas des systèmes biologiques, cela signifie que nous construisons un modèle logique (compréhensible) à partir de données de séries temporelles (par exemple des données d'expression génétique). Cette méthode souffre de deux contreparties : d'une part, elle aboutit possiblement à un grand nombre de règles logiques, difficiles à lire et à analyser pour un collaborateur biologiste ; d'autre part, elle nécessite de définir a priori la sémantique de la dynamique (synchrone/asynchrone/généralisée/etc.) du modèle cible. Au cours de ces six mois au NII, le travail a visé à surmonter ces limites, en interaction avec Katsumi INOUE et l'un de ses étudiants (Hikaru IKEDA) d'une part, avec Tony RIBEIRO (post-doc avec qui nous avons été lauréats de l'AAP « Challenge Post-doctorants Internationaux » du RFI A2020 et qui était invité de mars à septembre 2020 par Katsumi INOUE) d'autre part.

(A2) **En termes d'animation de la recherche**, je me suis investi dans la **consolidation des échanges** entre l'équipe de Katsumi INOUE et l'équipe MeForBio du Laboratoire des Sciences du Numérique de Nantes (LS2N).

(A3) Enfin, je tenais à profiter de ce séjour pour **contribuer à l'élargissement des échanges entre le National Institute of Informatics et, respectivement, l'école Centrale de Nantes (ECN), le Laboratoire des Sciences du Numérique de Nantes (LS2N) et le RFI Atlanstic 2020 (A2020).**

Enfin, j'ai poursuivi mon apprentissage du japonais en suivant des cours hebdomadaires (1h30 par semaine, sans compter le travail en autonomie) à l'école « *Coto Japanese Academy* ». La

connaissance de la langue me semble en effet nécessaire pour une meilleure compréhension tant de l'environnement que des spécificités du système japonais et permettre d'inscrire mes actions dans la durée. Disposant d'un niveau équivalent B1-B2 en CECRL, je me suis concentré ce séjour sur un vocabulaire plus technique en lien avec mes domaines et les compétences de rédaction par écrit.

B2. Résultats obtenus

(La numérotation reprend les axes évoqués dans la partie précédente)

(A1) Nous avons obtenu de nouveaux résultats sur le framework LFIT.

D'abord, pour répondre à l'enjeu de l'apprentissage à partir de données biologiques, **nous avons proposé une nouvelle variante du framework LFIT**, baptisée LOO-LFIT (Leave-One-Out LFIT). Cette approche s'inspire d'une démarche suivie dans les expériences dans le domaine de la génétique. Partant d'un jeu de données initial, il s'agit d'appliquer plusieurs fois l'algorithme LFIT à des variantes de ce jeu initial, ces variantes étant générées en excluant à chaque fois un gène différent de l'ensemble de données d'origine. L'algorithme se termine en fusionnant les résultats obtenus pour identifier les règles importantes. Nous avons validé cette démarche sur des données d'expression génétique artificielles et réelles issues des challenges de machine learning DREAM. Les modèles ainsi produits sont compréhensibles (en ce qu'ils sont constitués d'un nombre restreint de règles), tout en garantissant un pouvoir prédictif comparable aux outils de référence dans le domaine (notamment dynGENIE3).

Ensuite, pour élargir l'applicabilité de LFIT à nos systèmes d'intérêt (notamment l'étude des mécanismes génétiques à l'échelle d'une cellule), **nous avons étendu le framework LFIT pour qu'il soit en mesure d'apprendre la sémantique du modèle** à mesure qu'il analyse les données qui lui sont fournies en entrée. Jusqu'à maintenant, LFIT exigeait une hypothèse du modélisateur sur la sémantique la plus appropriée en termes de représentation des phénomènes observés. Nos précédentes contributions permettaient ainsi de fournir une méthode d'apprentissage logique pour une sémantique synchrone, asynchrone ou généralisée. Or ces dix dernières années, de nombreux travaux se sont portés sur la sémantique la plus pertinente pour la modélisation de réseaux booléens ou multi-valués (discutant notamment la conservation de certaines propriétés comme les attracteurs cycliques ou les points fixes), sujet encore en cours de discussions. Nous avons conçu un nouvel algorithme qui capture la sémantique du modèle en même temps qu'il apprend les règles logiques régissant le comportement du système.

(A2) Afin d'inscrire les échanges avec l'équipe de Katsumi INOUE dans la durée, je me suis investi dans les deux actions suivantes.

D'une part, **l'obtention pour Katsumi INOUE du statut de « adjunct professor » de l'école Centrale de Nantes**, dossier que j'ai présenté et défendu lors du Conseil Scientifique de l'ECN du 13 février 2020.

Le titre de « *adjunct professor* » a été mis en place à l'ECN en 2018. Il s'agit d'un statut visant à renforcer les liens internationaux au niveau recherche avec certains partenaires privilégiés de haut niveau. Ce statut n'inclut pas de salaire, mais cristallise une volonté d'avancer sur des sujets communs, volonté matérialisée par des actions telles que le co-encadrement de thèses, des séjours annuels sur place, etc. L'analyse de ces dossiers est instruite par le Conseil Scientifique de l'ECN. Le titre de « *adjunct professor* » donne lieu à une convention entre l'établissement et le partenaire international, d'une durée de 5 ans. Formellement, la signature de cette convention sera effectuée lors de la prochaine visite de Katsumi INOUE à Nantes (date pour le moment inconnue en raison des circonstances sanitaires internationales).

D'autre part, **la mise en œuvre pratique de l'accord de co-tutelle de thèse** (« doctoral dual degree program ») signé en novembre 2019 entre l'ECN et SOKENDAI (la *Graduate School* interdisciplinaire à laquelle sont inscrits tous les doctorants en thèse au NII). Avec Katsumi INOUE et le soutien de nos établissements respectifs, nous avons réussi ainsi l'an dernier à écrire un accord cadre de coopération entre nos institutions. Il restait à traiter, à l'échelle globale, des problématiques d'implémentation pratique du programme (par exemple les périodes auxquelles les

candidatures à de telles co-tutelles peuvent être reçues) et, dans le cas spécifique de notre collaboration, à travailler à sa mise en œuvre. Nous avons ainsi obtenu en mai 2020 l'accord de principe de nos établissements qui, chacun, apportera un demi-financement de thèse pour inscrire en 2021 un doctorant sur nos thématiques, en co-encadrement.

(A3) J'ai travaillé avec Emmanuel PLANAS, directeur des relations internationales du NII, à identifier les correspondances développables entre membres du NII et ceux des laboratoires visés par le RFI Atlantic 2020. Nous avons échangé sur ce sujet avec Pierre COINTE, dans la perspective de l'organisation d'un workshop NII-A2020 à l'automne 2020. Évidemment, les circonstances mondiales ont mis en pause l'organisation d'une telle rencontre, mais ce travail préparatoire a suscité l'intérêt des collègues sollicités. Nous pourrions donc le reprendre dès que l'horizon se sera éclairci en termes de voyages internationaux.

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**. Dès la fin février, le gouvernement japonais a appelé à restreindre les déplacements sur le territoire. De fait, les réunions de travail auxquelles j'avais initialement prévu de participer avec les collaborateurs de Katsumi INOUE à l'université de Kyôto, Nagoya, Wakayama et Kôbe ont été **annulées**. Heureusement, j'ai déjà eu l'occasion de rencontrer les chercheurs en question, notamment au cours d'un séjour de trois semaines effectué en août 2019 (financé à la fois par le NII et par un projet PHC Sakura que je porte avec l'Université de Kôbe), de sorte que **les conséquences scientifiques de cette crise mondiale sur mon projet scientifique ne se sont pas fait ressentir de manière cruciale**.

Par contre, d'un point de vue thématique, l'enjeu que représente la compréhension d'un virus émergent a suscité **un intérêt partagé** avec l'équipe de Katsumi INOUE. En 2016, nous avons en effet participé à un challenge international de machine learning sur la susceptibilité génétique à l'infection par des virus respiratoires (Respiratory Viral DREAM Challenge), participation qui avait nourri nos travaux ultérieurs. Nous avons donc mené une veille active sur la question de la modélisation qualitative du COVID19 qui a fédéré une partie de la communauté impliquée dans l'intelligence artificielle et la biologie des systèmes. Notre objectif était de **déterminer si nos méthodes d'apprentissage logique pouvaient avoir un intérêt** pour une meilleure compréhension des mécanismes moléculaires et génétiques en jeu et, si l'intérêt n'était pas immédiat, **d'identifier des pistes de recherche futures pour faciliter une applicabilité de nos approches**.

Le projet international « *COVID-19 Disease Map* » (<https://disease-maps.org>), auto-constitué et mené par Marek Ostaszewski de l'Université du Luxembourg, repose à ce jour sur des connaissances expertes : il s'agit de curation « à la main » des mécanismes moléculaires impliqués dans le fonctionnement du virus. Nos méthodes d'apprentissage ne sont donc pas actuellement adaptées à ce type de problématique, a contrario de certaines méthodes d'analyse de la dynamique. Nous en dégagons deux perspectives : (1) mesurer l'apport possible des approches d'analyses de la dynamique sur les modèles qualitatifs formés (de premiers modèles en SBML-qual ont vu le jour en mai) ; (2) rester en veille par rapport aux données quantitatives possiblement publiées sur ce sujet et sur lesquelles nous pourrions travailler. A compter d'avril 2020, Katsumi INOUE a lancé un étudiant de M2 sur le sujet « Analyse quantitative à plusieurs échelles de la cinétique d'infection par le nouveau coronavirus SARS-CoV-2 », étudiant dont nous suivons ensemble les avancées.

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 : chapitre de livre dont la parution est prévue d'ici la fin de l'année 2020
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 LHOSSAINE. 40 pages. À paraître chez ISTE Sciences, 2020.

Publication en cours de soumission en conférence internationale avec comité de lecture (CMSB*2020) :

H. IKEDA, K. INOUE, M. MAGNIN and T. RIBEIRO. Identification and Inference of Gene Regulatory Networks from Interpretation Transition. 15 pages.

Publication en cours de rédaction, dont l’objectif est une soumission au numéro de la revue « Machine Learning Journal » dont l’échéance de soumission est le 31 août :

T. RIBEIRO, M. FOLSCHETTE, M. MAGNIN and K. INOUE. Learning any semantics from dynamic programs (titre provisoire).

Médiation : Participation aux Journées Portes ouvertes du National Institute of Informatics (12 et 13 juin 2020) qui se sont tenues exceptionnellement en ligne cette année (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/>

D. INDICATEURS DE VISIBILITE

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

- les publications référencées au paragraphe C.
- les futurs séjours respectifs que Katsumi INOUE et moi-même effectueront annuellement dans le laboratoire de l’autre.
- la mise en œuvre d’une co-tutelle de thèse à partir de 2021
- la mise à disposition publique, en licence libre, des implémentations du framework LFIT : <https://github.com/Tony-sama/pylfit>

E. PERSPECTIVES

Katsumi INOUE et moi bénéficions désormais d’un statut quasi-similaire dans nos établissements respectifs : je suis « *visiting professor* » au NII, lui devient « *adjunct professor* » à l’ECN. De la sorte, nous serons amenés à poursuivre nos échanges à travers, au moins, les deux points de synchronisation annuels que représentent le séjour de l’un dans le laboratoire de l’autre. Ces échanges seront précieux pour se coordonner dans la suite de nos travaux :

- le suivi du projet, financé par le RFI Atlanstic 2020, de « challenge post-doc international » intitulé *Learning Adequate Semantic from Transition Observations with Functional Uncertainty and Stochasticity* (LAST OF US). Ce projet a démarré en mars 2020 avec l’accueil au NII pendant 6 mois de Tony RIBEIRO (financement de cette partie par le budget recherche de Katsumi INOUE).
- l’encadrement du doctorant en co-tutelle que nous comptons recruter en 2021.
- la préparation de futures réponses à appels à projets internationaux (par exemple bi-latéraux ANR-CREST).

Le projet a donc permis de consolider la collaboration et de passer à une étape supplémentaire dans la complémentarité entre l’équipe française et l’équipe japonaise.

Ce fonctionnement synergique pourrait sans doute être reproduit par d’autres équipes dans le périmètre du RFI Atlanstic 2020. En la matière, les prochains enjeux sont donc :

- d’alimenter les couplages possibles entre chercheurs du RFI Atlanstic 2020 et du NII. Le workshop initialement prévu entre les deux institutions en serait une bonne occasion.
- de développer les projets de co-tutelles de thèses, déjà possibles entre SOKENDAI et l’ECN, et sans doute à reproduire à l’échelle d’autres établissements du RFI Atlanstic 2020 qui y verraient un intérêt.