

Proposition de stage de Master 2 - 2019

Intégration de connaissances dynamiques préalables dans l'apprentissage de modèles biologiques

Encadrants : Morgan MAGNIN et Olivier ROUX (Équipe MeForBio, ECN, LS2N)

Web site: <https://ls2n.fr/?lang=en>
Email: morgan.magnin@ec-nantes.fr

Mots clés : intelligence artificielle, programmation logique, biologie des systèmes

Contexte

Dans tous les domaines impliquant des interactions dynamiques entre différents composants - et en particulier en biologie, via l'analyse et le traitement des données d'expression des gènes dans le temps -, nous avons constaté que la problématique de construction automatique de modèles chronométriques est cruciale pour la bonne compréhension de nombreux systèmes. Au cours de précédents travaux, nous avons conçu des approches permettant de construire automatiquement des modèles qualitatifs dynamiques (c'est-à-dire le niveau d'expression qualitatif d'un composant - par exemple « présent », « moyennement présent », « absent » - d'un composant en fonction du temps) à partir de données de séries temporelles.

Les solutions que nous développons au sein de l'équipe MeForBio du LS2N ont pour objectif l'apprentissage et l'analyse de la dynamique d'un système complexe à partir de l'observation de ses transitions d'états [RMI+15]. Celles-ci sont formalisées sous la forme de séries temporelles. L'apprentissage se fait conjointement et itérativement sur des représentations discrètes de ces données et par le passage à des modèles continus. L'utilisation de nos algorithmes portant sur la représentation discrète est destinée à comprendre comment fonctionne un système en se basant juste sur l'observation que l'on peut en faire de l'extérieur, et sans avoir accès à sa structure interne. L'intérêt du modèle que nous pouvons ainsi construire, exprimé sous forme de programme logique ou de réseaux d'automates synchronisés, est de pouvoir l'utiliser pour prédire l'évolution du système réel dans d'autres conditions que celles observées (et donc avec une possible application, par des expériences *in-silico*, pour des thérapies géniques).

Objectif du stage

Ce stage aura pour objectif d'adapter et d'améliorer l'algorithme d'apprentissage que nous avons précédemment développé à partir de données de séries temporelles. Le but sera d'intégrer, dans l'algorithme d'apprentissage actuel, les connaissances préalables sur le système que nous avons pour le moment mises de côté (propriété dynamique connue a priori telle une connaissance sur le mécanisme oscillatoire du système, knockdown/knock-out d'un gène, etc.) dans le processus d'apprentissage. Le travail s'appliquera sur l'un des deux jeux de données cibles envisagés ci-après (l'application cible sera définitivement arrêtée au moment où le projet débutera) :

- soit un **jeu** de données « in silico » : le **jeu** de données associé au DREAM4 Challenge, basé sur les systèmes biologiques *E.coli* et *S.cerevisiae* [GMO+10]. Les défis scientifiques baptisés « DREAM Challenge » sont organisés depuis près d'une dizaine d'années au sein de la communauté de la biologie des systèmes et de la médecine. Ils sont proposés par un consortium rassemblant académiques et établissements privés impliqués dans le secteur de l'informatique et de la santé. Ils ont constamment gagné en notoriété, et constituent désormais une source reconnue de données – et de questions scientifiques associées – dans le domaine de la bio-informatique. Le **jeu** de données du DREAM4 a l'avantage de disposer d'un outil permettant de générer autant de données que nécessaire, et de fournir désormais une base incontournable pour la validation *in silico* d'algorithmes d'apprentissages. Si c'est ce **jeu** de données qui est retenu, le projet visera à intégrer l'intégralité des données fournies dans le cadre de ce challenge.

- soit un jeu de données « réelles » portant soit sur l'impact de mécanismes environnementaux (lumière, rythmes alimentaires, etc.) sur la robustesse du cycle circadien (mécanisme oscillatoire présent chez tous les mammifères sur une période de 24h environ), soit sur l'impact de doses de radiothérapie sur le devenir de cellules tumorales.

À travers ce stage de Master, l'étudiant aura l'opportunité de développer les connaissances et compétences suivantes :

- comprendre les problématiques liées à l'apprentissage de modèles en biologie ;
- comprendre les mécanismes liés à l'apprentissage automatique par programmation logique ;
- savoir construire un modèle à partir de données « brutes » ;
- savoir modifier un algorithme existant pour enrichir ses possibilités ;
- savoir analyser la qualité d'un modèle ;

Références

[GMO+10] A. Greenfield, A. Madar, H. Ostrer, and R. Bonneau. Dream4: Combining genetic and dynamic information to identify biological networks and dynamical models. PloS one, 5(10):e13397–e13397, 2010.

[RMI+15] T. Ribeiro, M. Magnin, K. Inoue, and C. Sakama. Learning Delayed Influences of Biological Systems. In Frontiers in Bioengineering and Biotechnology, 2, 81. 2015.