

Proposition de sujet de thèse (AAP AiBy4 2022, axe Santé du futur)

Contact : christine.sinoquet@univ-nantes.fr

1 Informations sur la thèse

- Titre** Modélisation de l'évolution physiologique d'un patient en réaction à un process de soins. Application à la simulation d'un patient au bloc opératoire, dans une nouvelle modalité de formation en anesthésie, assistée par le numérique.
- Mots-clés** **Intelligence artificielle, santé du futur, patient digital, formation assistée par le numérique, bloc opératoire, anesthésie, simulation, modélisation, apprentissage semi-supervisé, apprentissage profond, séries temporelles, traces d'événements**
- Encadrement** La personne recrutée sera encadrée par Christine Sinoquet et Corinne Lejus-Bourdeau.
Christine Sinoquet (<https://christinesinoquet.wixsite.com/christinesinoquet>) est Maître de Conférences en Informatique, habilitée à diriger des recherches (HdR 2014), au **Laboratoire des Sciences du Numérique de Nantes / UMR CNRS 6004**. Depuis 2008 en particulier, ses principales activités de recherche concernent l'apprentissage automatique, pour des applications dans le domaine de la biologie, la santé, et le diagnostic de l'état des machines.
Corinne Lejus-Bourdeau est Professeur des Universités – Praticien Hospitalier, Docteur en Médecine au Service d'Anesthésie-Réanimation chirurgicale du **CHU de Nantes / Hôtel Dieu – Hôpital Femme-Enfant-Adolescent**, Directeur du **Laboratoire Expérimental de Simulation de Médecine Intensive de l'Université de Nantes (LESiMU)**.
Christine Sinoquet et Corinne Lejus-Bourdeau collaborent depuis 2019, dans le cadre du projet régional RFI OIC EXAN (Acquisition et consolidation d'EXpertise en Anesthésie - Développement et comparaison de diverses modalités de formation assistées par le Numérique).
- Partenaires** Les travaux doctoraux bénéficieront également de l'expertise de Raphaël Cinotti et de Nicolas Grillot, tous deux Docteurs en Médecine au Service d'Anesthésie-Réanimation chirurgicale du CHU de Nantes / Hôtel Dieu – Hôpital Femme-Enfant-Adolescent et Praticiens Hospitaliers.
- Financement** Ces travaux de recherche seront financés dans le cadre du projet nantais AiBy4 (AiBy4). AiBy4 est l'un des 22 projets retenus par l'ANR pour son appel "contrats doctoraux en IA" (2020-25).

Candidature à la bourse de thèse

Compétences souhaitées

- Master ou équivalent en **Mathématiques** ou **Mathématiques / Informatique** ou **Informatique** avec spécialisation en science des données ou probabilités / statistique, ainsi qu'en apprentissage automatique (incluant deep learning de préférence)

- Compétences théoriques et expérience requises en probabilités / statistiques, mathématiques appliquées, apprentissage automatique
- Si la personne candidate n'a pas d'expérience dans la modélisation par composition de modèles utilisant des réseaux de neurones profonds, qui est l'un des axes d'investigation de la thèse, elle devra en revanche montrer une forte motivation pour s'investir dans ce domaine
- Intérêt pour l'interdisciplinarité (santé)
- Expérience en programmation et bon niveau de programmation
- Bonnes capacités rédactionnelles
- Aptitude au travail en équipe, capacité à rendre compte de l'avancement de ses travaux

Calendrier de la sélection

Les dossiers de candidature seront **de préférence** envoyés avant le **jeudi 21 avril 2022 12H00**.

Date **limite** d'envoi des candidatures : **mardi 26 avril 2022 12H00**

Les personnes correspondant au profil demandé seront convoquées pour une audition par visio-conférence (au fil de l'eau).

Décision : **mi-mai 2022**

Pièces à fournir

- **CV détaillé**
- **lettre de motivation**
- **relevé de notes de Master 1** (avec rang de classement et effectif de la promotion)
- **notes de Master 2 hors stage** (avec rang de classement et effectif de la promotion)
- **résumé du stage en cours** (entre 2 et 4 pages, références bibliographiques en sus)
- **lettres de recommandation** pour l'année en cours
- **coordonnées de personnes référentes** (prénom, nom, statut, institution (détailler les sigles le cas échéant), ville, adresse mail, numéro de téléphone)

Questions et envoi des dossiers de candidature (archive zip) à christine.sinoquet@univ-nantes.fr

Résumé du sujet

Les travaux de thèse s'inscrivent dans un contexte fortement interdisciplinaire (IA / Santé). Le but de la thèse est de développer un gestionnaire automatisé de scénarios de simulation d'anesthésies sur un patient digital, pour l'entraînement d'internes et d'infirmiers anesthésistes. Ces scénarios doivent être réactifs aux actions de l'interne ou de l'infirmier en formation. La clé du problème consiste à savoir simuler l'évolution des paramètres physiologiques du patient digital en fonction de l'action déclenchée par la personne en formation ou par le reste de l'équipe médicale (qui est virtuelle). Il faut également savoir prédire la prochaine action de l'équipe médicale (hors personne en formation). On peut tirer parti des données enregistrées lors des anesthésies d'une cohorte de patients réels, opérés pour la même chirurgie, pour prédire l'évolution du patient numérique au fil des actions réalisées sur le patient. Pour chaque patient de la cohorte, nous disposons également de la trace des actions médicales réalisées sur le patient, au bloc opératoire. Ces traces sont utilisées pour prédire la prochaine action à réaliser, au fil d'un scénario. Du point de vue IA, nous sommes donc confrontés à un problème d'apprentissage semi-supervisé à partir de séries temporelles multivariées et de traces d'événements interdépendantes, dans un objectif de prédiction de séries temporelles multivariées à court-terme et en temps réel, et de prédiction du prochain événement.

2 Objectifs applicatifs

La simulation est un outil pédagogique moderne dédié à l'apprentissage et au maintien des compétences dans le domaine des sciences de la santé. Il permet aux apprenants de s'entraîner à la prise en charge de situations réalistes dans un environnement sécuritaire, sans risque pour le patient ("jamais la première fois sur le patient"). Le Laboratoire Expérimental de Simulation en Médecine Intensive (LE SiMU) de l'Université de Nantes permet ce type d'apprentissage dans sept spécialités médicales, dont l'anesthésie.

LE SiMU intervient sur simulateurs de patients haute-fidélité (mannequins), avec immersion des agents en formation dans une équipe médicale humaine pleine-échelle interprofessionnelle. Actuellement, les scénarios sont écrits à l'avance, et le formateur fait évoluer les paramètres physiologiques du mannequin "à la main", en réponse aux actions de l'équipe médicale (y compris celles de l'apprenant immergé dans cette équipe). D'une part, il y a peu de variabilité dans les scénarios. D'autre part, un apprenant n'a accès qu'à un nombre limité de scénarios (environ 3 par demi-journée de formation).

Dans le but d'améliorer la sécurité et la qualité des soins peropératoires, les formateurs du SiMU souhaitent varier la diversité des scénarios à proposer aux internes en anesthésie et infirmiers anesthésistes, en formation initiale, ainsi qu'aux praticiens plus expérimentés, en formation continue. Pour varier les scénarios, il est envisagé à terme d'automatiser la génération de scénarios réalistes de simulation, en s'appuyant sur tout ou partie de la base de profils anesthésiques enregistrés par le CHU de Nantes depuis 2004 (500 000 profils anesthésiques). Dans cette modalité assistée par le numérique, la personne qui suit la formation (interne ou infirmier), fait partie de l'équipe médicale. Les autres membres de l'équipe médicale sont simulés très simplement (icônes réalisant des actions et émettant des informations, sur l'écran de l'ordinateur assigné à l'apprenant).

3 Sujet de thèse

3.1 Objectifs opérationnels

Le problème posé consiste à savoir faire évoluer les paramètres physiologiques d'un patient numérique en réponse à des actions médicales successives. Ce problème équivaut à prédire l'évolution d'une série temporelle, à court terme, jusqu'au déclenchement de la prochaine action médicale, par l'apprenant ou par le reste de l'équipe médicale (qui est virtuelle). Excepté si c'est l'apprenant qui déclenche la prochaine action, cette action doit également être prédite.

Nous devons définir un modèle capable de générer une série multivariée à d variables, et évoluant sous l'effet d'une séquence d'événements horodatés dans le temps continu.

Nous proposons d'apprendre ce modèle à partir des profils anesthésiques d'une cohorte de patients. Le profil anesthésique d'un patient est constitué d'une série temporelle multivariée et d'une trace d'événements. La série temporelle décrit l'évolution dynamique de d paramètres physiologiques (e.g., la fréquence cardiaque) enregistrés via des capteurs, à fréquence constante. La trace d'événements est la séquence horodatée des actions asynchrones de l'équipe médicale (e.g., administration d'un anesthésique).

La prise en charge de ce problème est inédite.

3.2 Limites de l'état de l'art et verrous à lever

Les méthodes statistiques classiques capturent efficacement les relations linéaires dans les séries temporelles, et obtiennent de bons résultats lorsque le jeu de données est de petite taille (e.g., modèle linéaire autorégressif, modèles MA (Moving Average) et ARMA (AutoRegressive Moving Average)). Cependant, ces approches statistiques sont peu efficaces lorsqu'elles sont appliquées à des séries temporelles non linéaires complexes à grande échelle [LZG21]. Dans le cas où la non-linéarité est liée à la présence de changements de régime dans l'évolution des systèmes dynamiques, des propositions majeures de la littérature ont été par exemple le modèle autorégressif à seuil (TAR, Threshold AutoRegressive model) [T90] et le modèle autorégressif à changement de régime Markovien (MSAR, Markov-Switching AutoRegressive model) [H90] pour les changements discrets, le modèle à changement de régime continu (STAR, Smooth Transition AutoRegressive model) [LT94].

Cependant, **ces modèles sont inadaptés pour notre objectif**, pour diverses raisons. Par exemple, dans le modèle TAR, les changements de régime sont déterminés par les seules variations de la série temporelle. Tous les modèles évoqués précédemment **sont incapables de prendre en compte l'existence de dépendances à long terme entre événements**, ainsi que la **variabilité de la portée des dépendances entre événements**.

3.3 Modélisation jointe d'une série temporelle et d'une trace d'événements

Des revues de l'état de l'art récentes consacrées à la prédiction du temps de survie ou du temps de récurrence d'une pathologie discutent des diverses façons de représenter les dépendances entre un sous-modèle dédié à la trace d'événements et un sous-modèle dédié à la série temporelle, dans les cadres fréquentiste [HPJ16] et Bayésien [ASG20]. Cependant, ces modèles ne permettent pas la **double prédiction nécessaire** pour résoudre notre problème.

Dans le domaine de la **prédiction de série temporelle multivariée**, les approches les plus efficaces actuellement pour prendre en charge les **aspects non linéaires** et les **corrélations entre variables** reposent sur les **architectures de réseaux de neurones profonds**, éventuellement hybridées entre elles, ou encore hybridées avec un modèle classique [LZ20].

Par ailleurs, la littérature récente indique que les **processus temporels ponctuels** constituent une classe de modèles très utilisée pour la **prédiction du prochain événement** [Z22]. Les processus de Poisson constituent la classe des processus ponctuels les plus populaires et les plus simples, pour modéliser l'occurrence d'événements indépendants les uns des autres. La composition de processus ponctuels simples (e.g., [SJ10]), et des processus ponctuels plus complexes comme les processus de Cox [C55] et de Hawkes [H71], par exemple, permettent de prendre en compte des dépendances entre événements. Un récent état de l'art est consacré à ces processus [JO20].

Cependant, lorsqu'aucune information n'est disponible sur la distribution des temps d'occurrence des événements, des approches non paramétriques peuvent être utilisées pour apprendre des distributions sophistiquées à partir des données, grâce à des **réseaux de neurones**. Les réseaux neuronaux récurrents, et spécialement les LSTM (Long Short Term Memory) sont particulièrement utilisés pour apprendre des dépendances complexes (e.g., [TVR17]).

Hormis pour la prédiction du temps de survie ou de récurrence évoquée précédemment, la modélisation jointe de trace d'événements et de série temporelle a été peu abordée, jusqu'à ce jour. Deux publications seulement ont été recensées, qui se situent dans la même veine que les travaux envisagés pour la thèse.

Choi et co-auteurs informent un processus de Hawkes au moyen d'une série temporelle, afin de prédire l'évolution de la pathologie d'un patient [CDC15]. Xiao et collaborateurs décrivent un modèle joint combinant deux réseaux de neurones récurrents profonds (des LSTM en l'occurrence), dans le but de réaliser la prédiction informée du prochain événement [XYF19]. Dans cette approche, l'un des LSTM capture les dépendances entre les événements, et l'autre modélise l'influence de la série temporelle sur l'occurrence des événements.

Le cadre de travail de la thèse **se démarque des deux approches précédentes** puisque notre objectif est de prédire, non seulement le prochain événement, mais aussi l'évolution de la série temporelle.

Pour répondre aux objectifs de la thèse, nous investiguerons donc dans un premier temps comment réaliser **une double prédiction** en intégrant un **processus ponctuel ou une variante de processus ponctuel** et une architecture de **réseau de neurones** profond dédiée à la modélisation de la série temporelle.

Nous examinerons ensuite comment modéliser un processus d'occurrences d'événements à l'aide d'un **réseau de neurones profond** et comment hybrider ce sous-modèle avec un **réseau de neurones** dédié à la modélisation de la série temporelle.

4 Plan de travail

Le planning prévu pour la thèse est le suivant :

1. Etude de l'état de l'art sur la prédiction de série temporelle multivariée par utilisation d'un réseau de neurones profond ; étude de l'état de l'art sur la prédiction d'événement via un processus temporel ponctuel d'une part, via un réseau de neurones profond d'une part ; étude de l'état de l'art sur la modélisation d'un processus temporel ponctuel à l'aide d'un réseau de neurones profond.
2. Appropriation du simulateur de profils anesthésiques développé dans le cadre du projet EXAN.

3. Choix de l'environnement de programmation (PyTorch, Keras, TensorFlow etc) pour l'apprentissage profond.
4. **Mise au point d'un modèle combinant un modèle neuronal profond** modélisant un **processus temporel ponctuel** pour la prédiction d'événement et un **réseau de neurones** pour la prédiction de série temporelle ; proposition éventuelle de plusieurs architectures ; développement des prototypes correspondants ; examen de la nécessité et des possibilités d'utiliser des modèles pré-entraînés, examen de la nécessité de recourir à l'*augmentation de données*.
5. **Mise au point du protocole de validation** de l'architecture ou des architectures proposées ; mise au point du protocole de validation de l'usage de ces architectures par les utilisateurs de la plateforme de simulation LE SiMU.
6. idem point 4, pour un modèle neuronal profond **hybridant deux réseaux de neurones** dédiés respectivement à la prédiction d'événement et à la prédiction de série
7. idem point 5, appliqué à l'architecture ou aux architectures proposées en point 6.

Références bibliographiques

- [ASG20] M. Alesfri, M. Sudell, M. García-Fiñana and R. Kolamunnage-Dona (2020) Bayesian joint modelling of longitudinal and time to event data: a methodological review. *BMC Medical Research Methodology* 20, 94.
- [C55] D.R. Cox (1955) Some statistical methods connected with series of events. *Journal of the Royal Statistical Society* 17(2), 129-164.
- [CDC15] E. Choi, N. Du, R. Chen et al (2015) Constructing disease network and temporal progression model via context-sensitive Hawkes process. In: *IEEE International Conference on Data Mining*, pp 721-726.
- [H71] A.G. Hawkes (1971) Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 58, 83-90
- [H90] J. D. Hamilton (1990) Analysis of time series subject to changes in regime. *Journal of Econometrics*, 45(1-2), 39-70.
- [HPJ16] G. L. Hickey, P. Philipson, A. Jorgensen and R. Kolamunnage-Dona (2016) Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. *BMC Medical Research Methodology*, 16,117.
- [JO20] J. Jang, and R. Oh (2021). A review on Poisson, Cox, Hawkes, shot-noise Poisson and dynamic contagion process and their compound processes. *Annals of Actuarial Science* ,15(3), 623-644.
- [LT94] C.-F. J. Lin and T. Teräsvirta (1994) Testing the constancy of regression parameters against continuous structural change. *Journal of Econometrics*, 62(2), 211-228.
- [LZ20] B. Lim B, and S. Zohren S (2020) Time series forecasting with deep learning: a survey. <https://arxiv.org/abs/2004.13408>
- [LZG21] Z. Liu, Z. Zhu, J. Gao and C. Xu (2021) Forecast methods for time series data: a survey. *IEEE Access*, 9, 91896-91912.
- [SJ10] A. Simma and M.I. Jordan (2012) Modeling events with cascades of Poisson processes. In: *International Conference on Uncertainty in Artificial Intelligence*, pp. 546-555.
- [T90] H. Tong (1990) Non-linear time series: a dynamical system approach. Oxford University Press.
- [TVR17] N. Tax, I. Verenich, M. L. Rosa, M. Dumas. (2017) Predictive business process monitoring with LSTM neural networks. In : *International Conference on Advanced Information Systems Engineering, CAiSE2017*, Lecture Notes in Computer Science 10253, pp. 477-492.
- [XYF19] S. Xiao, J. Yan, M. Farajtabar et al (2019) Learning time series associated event sequences with recurrent point process networks. *IEEE Transactions on Neural Networks and Learning Systems*, 30(10),3124-3136.
- [Z22] L. Zhao (2022) Event prediction in the big data era : a systematic survey. *ACM Computing Surveys*, 54(5),1-37.