Offre de stage de recherche • Master 2, dernière année d'école d'ingénieur ou équivalent

Apprentissage contrastif de dynamiques latentes identifiables pour la surveillance des équipements de radiothérapie

Mots-clés: IA de confiance • Apprentissage non-supervisé • Identifiabilité • Explicabilité • Séries temporelles • Pronostics and Health Management • Radiothérapie

Informations clés

Période	Durée	Lieu
À partir de février 2027	**5 à 6 mois**	**Toulouse** – Oncopole Claudius Regaud - IUCT-Oncopole (4 j/sem.) et LAAS-CNRS (1 j/sem.)

Encadrement **Kélian Poujade, Louise Travé-Massuyès, Anthony Réveillac et Laure Vieillevigne**

Sujet

1. Contexte d'étude

Les machines complexes produisent de nombreuses mesures de capteurs, séries temporelles et journaux d'événements (*logs*) renseignant sur leur état de fonctionnement et l'apparition de défauts. En *Prognostics and Health Management* (PHM), l'enjeu est notamment d'apprendre à caractériser et suivre l'état de santé de ces systèmes afin de détecter les comportements anormaux, comprendre les défauts et anticiper leur évolution vers la défaillance, alors même que les annotations décrivant précisément les dégradations sont souvent rares.

L'apprentissage contrastif permet d'apprendre des représentations sans supervision explicite en construisant un signal auto-supervisé à partir des relations entre observations [1]. Dans le cas de données temporelles, ces relations peuvent notamment exploiter la proximité dans le temps, la continuité des trajectoires ou les transitions entre états. L'objectif est alors d'apprendre un espace latent dans lequel certaines observations ou transitions partageant des propriétés communes sont rapprochées, tandis que d'autres sont distinguées. Cependant, proximité temporelle et similarité d'état ne coïncident pas nécessairement: deux observations proches dans le temps peuvent correspondre à des dynamiques différentes, tandis que des états comparables peuvent apparaître à des instants éloignés. La définition de relations contrastives pertinentes pour des systèmes dynamiques constitue donc un enjeu essentiel.

2. Problématique scientifique

Les modèles à espace latent cherchent à représenter un système par des variables cachées décrivant, par exemple, son état de fonctionnement ou sa dégradation. Plusieurs représentations différentes pouvant parfois expliquer les mêmes observations, se pose la question de l'**identifiabilité**: peut-on retrouver de manière unique les facteurs latents ayant généré les données, éventuellement à certaines transformations caractérisées près? Cette propriété permet de limiter les ambiguïtés de représentation et constitue une base théorique pour obtenir des variables latentes plus stables et comparables entre différents apprentissages. Lorsque ces variables peuvent ensuite être reliées à des grandeurs ou à des mécanismes du système, cette structuration de l'espace latent peut également faciliter leur interprétation et l'analyse du modèle.

Le stage étudiera une variante contrastive de **CANDELA**, un modèle causal dynamique à espace latent structuré développé par les encadrants. Ce modèle cherche notamment à représenter séparément différentes composantes de l'état du système et les mécanismes gouvernant leur évolution au cours du temps. Ses garanties actuelles d'identifiabilité reposent sur des hypothèses concernant les innovations latentes, c'est-à-dire l'information nouvelle intervenant dans l'évolution du système à chaque instant. La variante proposée cherchera à exploiter directement la structure temporelle des observations afin de remplacer certaines hypothèses distributionnelles explicites sur ces innovations par des hypothèses portant sur les relations entre états ou transitions observés.

La question centrale est alors: **sous quelles hypothèses peut-on identifier, à partir des observations séquentielles et de leur structure temporelle, les facteurs latents ainsi que les mécanismes gouvernant leur évolution ?** Le travail cherchera à caractériser précisément les ambiguïtés résiduelles, à déterminer les classes de transformations sous lesquelles la représentation peut être identifiée, puis à établir des conditions suffisantes d'identification et à en préciser les limites au moyen de contre-exemples.

Les travaux sur les représentations contrastives et l'identification de systèmes dynamiques constitueront les principaux points de départ [2, 3, 4].

3. Application aux machines de radiothérapie

Le cas d'étude concerne notamment les systèmes **TomoTherapy** du constructeur **Accuray Inc.** (Madison, Wisconsin, États-Unis). Mieux surveiller leur fonctionnement vise à limiter les interruptions imprévues et à préserver la continuité des séances de radiothérapie et de la prise en charge des patients.

Les représentations apprises seront ensuite étudiées dans plusieurs tâches de PHM, notamment la **détection** d'anomalies, le **diagnostic** des défauts et le **pronostic** de leur évolution, selon les informations disponibles dans les différents jeux de données. L'évaluation reposera d'abord sur des données synthétiques permettant de contrôler les facteurs latents et les mécanismes générateurs, puis sur des jeux PHM de référence et des données techniques issues des équipements de radiothérapie.

Objectifs du stage

1. **Formalisation du problème d'identifiabilité.** Étudier les résultats pertinents en apprentissage contrastif, en représentation identifiable et en identification de systèmes dynamiques; formaliser les facteurs latents, les mécanismes de transition et les classes d'équivalence pertinentes pour CANDELA.
2. **Contribution théorique principale.** Concevoir un objectif contrastif adapté aux transitions dynamiques et caractériser les représentations compatibles avec cet objectif. Le travail cherchera à déterminer sous quelles hypothèses l'espace latent et, lorsque cela est possible, certains mécanismes de transition sont identifiables à une classe de transformations près. Des garanties théoriques seront établies lorsque cela est possible et complétées par des contre-exemples permettant d'en préciser les limites.
3. **Validation expérimentale.** Implémenter l'approche en Python (PyTorch) et confronter les résultats théoriques à des données synthétiques contrôlées. Selon l'avancement, l'intérêt des représentations apprises sera ensuite étudié sur des jeux PHM de référence et des données de radiothérapie, notamment pour des tâches de détection, de diagnostic ou de pronostic.

Références

[1] K. Rombach et al. *Learning Informative Health Indicators Through Unsupervised Contrastive Learning*. IEEE Transactions on Reliability, 74(1):2408–2420, 2025.
[2] D. Klindt et al. *Towards Nonlinear Disentanglement in Natural Data with Temporal Sparse Coding*. ICLR, 2021.
[3] R. S. Zimmermann et al. *Contrastive Learning Inverts the Data Generating Process*. ICML, 2021.
[4] R. González Laiz, T. Schmidt et S. Schneider. *Self-Supervised Contrastive Learning Performs Non-Linear System Identification*. ICLR, 2025.

Profil recherché

- **Diplôme préparé:** Niveau BAC+5 (Master 2, dernière année d'école d'ingénieur ou équivalent), spécialité science des données, intelligence artificielle ou domaine proche.
- **Compétences techniques:** Goût prononcé pour les aspects mathématiques de l'apprentissage automatique; bonnes bases en probabilités, statistiques et algèbre linéaire; bonnes connaissances en apprentissage automatique supervisé et non supervisé; pratique solide de Python et bonnes bases de PyTorch.
- **Un plus:** Connaissances en séries temporelles, modèles génératifs à espace latent, apprentissage auto-supervisé ou optimisation.
- **Qualités recherchées:** Curiosité scientifique, goût pour la compréhension théorique des méthodes, autonomie, rigueur et capacité à conduire une recherche bibliographique. Capacité à communiquer clairement à l'oral comme à l'écrit et intérêt pour la confrontation des développements théoriques à des problématiques de terrain en santé.

Conditions pratiques

Organisation	Travail partagé entre l'Oncopole Claudius Regaud - IUCT-Oncopole et le LAAS-CNRS, au contact d'équipes spécialisées en intelligence artificielle, diagnostic des systèmes et physique médicale. Les deux sites sont accessibles en transports en commun.
Gratification et avantages	Stage financé par l'Oncopole Claudius Regaud - IUCT-Oncopole. Accès au restaurant du personnel pendant la durée du stage.

Candidature

Envoyer un **CV**, un **court texte de motivation** et un **relevé de notes** de l'année précédente à:

poujade.kelian@iuct-oncopole.fr, **louise.trave-massuyes@laas.fr**, **anthony.reveillac@insa-toulouse.fr** et **vieillevigne.laure@iuct-oncopole.fr**.