

Université de technologie de Compiègne - Proposition de thèse

1^{er} partie : fiche scientifique	
Intitulé de la thèse	Apprentissage de modèles graphiques probabilistes pour des tâches complexes
Type de financement	Allocation MESR
Laboratoire d'accueil	Équipe CID, Heudiasyc UMR 7253
Directeurs de thèse	Vu-Linh NGUYEN (Chaire Professeur Junior) Yves GRANDVALET (Directeur de Recherche CNRS)
Domaine	Informatique
Descriptif	<p>Dans le contexte de l'apprentissage statistique, de nombreux travaux ont été consacrés aux classificateurs probabilistes qui prédisent la distribution de probabilité d'une variable de classe unique. Ces modèles permettent de calculer les prédictions optimales pour différentes fonctions de perte ainsi que des informations sur l'incertitude associée à leurs prédictions.</p> <p>La classification multidimensionnelle est une généralisation de la classification multi-classe dans laquelle plusieurs variables de classe non exclusives sont prédites, dans un modèle qui tient compte des dépendances entre ces variables de classe [3, 4]. Dans ce cadre, les réseaux bayésiens multidimensionnels [3] constituent l'une des rares familles de modèles prédictifs qui peuvent être interprétés, en fournissant une représentation éventuellement compacte des relations entre les différentes variables (caractéristiques et classes). Cependant, les réseaux bayésiens multidimensionnels interprétables et fiables ne parviennent toujours pas à passer à l'échelle pour résoudre les problèmes difficiles du monde réel, dont les descripteurs comprennent des caractéristiques continues et discrètes ainsi que des objets à haute dimension tels que les images.</p> <p>À notre connaissance, le projet en cours « Classification probabiliste multidimensionnelle » (initié par Linh à l'Université de technologie d'Eindhoven) est la première tentative d'aborder ce problème d'une manière générique et rigoureuse. Notre ambition est de mettre en œuvre les méthodes développées dans des applications à fort enjeu, telles que l'analyse des données de santé ou les véhicules autonomes. Nous nous appuyons sur les avancées réalisées sur plusieurs fronts : en apprentissage de modèles graphiques probabilistes [1], en optimisation [2, 8], et en apprentissage profond [5] pour développer des algorithmes d'apprentissage et d'inférence efficaces généralisant ceux qui existent pour la classification multi-label (voir [6, 7] et ses références).</p> <p>De nombreuses questions spécifiques méritent d'être abordées dans le cadre de la classification multidimensionnelle . Nous mentionnons ci-dessous trois manières possibles pour aborder cette thèse. Le choix de l'une d'entre elle pourra être laissé à l'étudiant.e en fonction de ses goûts et de ses compétences.</p> <p><i>Regularisation du problème d'apprentissage.</i> Un classificateur multidimensionnel est défini par la structure du réseau modélisant les dépendances entre les variables et ses paramètres. L'apprentissage peut retourner un classificateur multidimensionnel trop complexe. Dans cette situation, une approche générale consiste à développer des fonctionnelles de régularisation qui favorisent les modèles simples. Nous avons développé une famille de fonctions de vraisemblance conditionnelle régularisées. Leur amélioration et leur implémentation dans un package Python pourrait être une première étape de la thèse.</p> <p><i>Incorporation de règles logiques dans l'inférence.</i> Des règles logiques fiables et interprétables (AND, OR, XOR, ...) peuvent parfois être extraites des données ou fournies par des experts du domaine. Cependant, les techniques d'inférence pour la classification multidimensionnelle ne sont pas capables d'intégrer ces informations pour réaliser leurs prédictions. Nous avons développé de nouvelles techniques d'inférence bayésienne.</p>

	<p><i>Apprentissage actif.</i> La modélisation de la classification multidimensionnelle par un réseau bayésien offre un potentiel pour le développement de protocoles de décision adaptatifs pour recommander l'acquisition des variables non observées les plus influentes et les plus pertinentes (selon un critère de performance défini). Ces approches correspondent à des stratégies d'apprentissage actif nécessitant des algorithmes de recherche efficaces pour calculer les requêtes optimales.</p> <p>[1] J. Cussens, M. Järvisalo, J. H. Korhonen, and M. Bartlett. Bayesian network structure learning with integer programming : Polytopes, facets and complexity. <i>Journal of Artificial Intelligence Research</i>, 58 :185–229, 2017.</p> <p>[2] N. Friedman, D. Geiger, and M. Goldszmidt. Bayesian network classifiers. <i>Machine Learning</i>, 29(2) :131–163, 1997.</p> <p>[3] S. Gil-Begue, C. Bielza, and P. Larrañaga. Multi-dimensional bayesian network classifiers : A survey. <i>Artificial Intelligence Review</i>, 54(1) :519–559, 2021.</p> <p>[4] B.-B. Jia and M.-L. Zhang. Decomposition-based classifier chains for multi-dimensional classification. <i>IEEE Transactions on Artificial Intelligence</i>, 3(2) :176–191, 2021.</p> <p>[5] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. <i>Nature</i>, 521(7553) :436–444, 2015.</p> <p>[6] V.-L. Nguyen and E. Hüllermeier. Multilabel classification with partial abstention : Bayes-optimal prediction under label independence. <i>Journal of Artificial Intelligence Research</i>, 72 :613–665, 2021.</p> <p>[7] J. Read, B. Pfahringer, G. Holmes, and E. Frank. Classifier chains : a review and perspectives. <i>Journal of Artificial Intelligence Research</i>, 70 :683–718, 2021.</p> <p>[8] T. Roos, H. Wettig, P. Grünwald, P. Myllymäki, and H. Tirri. On discriminative bayesian network classifiers and logistic regression. <i>Machine Learning</i>, 59(3) :267–296, 2005.</p>
Mots-clés	Classification multidimensionnelle, apprentissage multimodal, modèles graphiques probabilistes, confiance et interprétabilité des algorithmes
Profil recherché	Master 2 ou ingénieur en informatique, bonnes compétences en programmation (Python, PyTorch, TensorFlow, ...) et/ou une solide formation en mathématiques.
Date de début	01/10/2023
Lieu de travail	Heudiasyc UMR 7253, Université de Technologie de Compiègne

2^e partie : fiche de poste	
Durée	36 mois
Missions complémentaires	Possibilité d'enseigner si souhaité
Laboratoire d'accueil	Heudiasyc UMR 7253, Université de Technologie de Compiègne et CNRS opère dans le domaine des sciences de l'information et du numérique, notamment l'informatique, l'automatique, la robotique et l'intelligence artificielle.
Moyens matériels	Bureau partagé, ordinateur personnel, accès aux serveurs GPU du laboratoire et du supercalculateur Jean Zay de l'IDRIS, ainsi qu'aux plateformes du laboratoire
Moyens humains	Le laboratoire compte 33 enseignant-chercheurs permanents, 6 chercheurs CNRS, 14 techniciens, ingénieurs et administratifs en support de la recherche, 46 doctorants et 5 post-doctorants, 12 post-docs/ingénieurs CDD. Collaborations internes et externes

Modalités de travail	<p>L'équipe d'encadrement propose un projet pédagogique en deux étapes. Au cours de la première étape, la personne recrutée sera guidée dans le choix des résultats, des solutions algorithmiques, des logiciels existants, sur lesquels elle pourra baser ses propres résultats, logiciels, protocoles expérimentaux, et les communiquer par le biais d'articles scientifiques.</p> <p>Une fois les connaissances et les compétences nécessaires acquises, des problèmes plus difficiles pourront être attaqués en collaboration avec les superviseurs et des collaborateurs, afin de développer les compétences de recherche nécessaires pour travailler à la fois de manière indépendante et en collaboration. La thèse sera financée par une Allocation MESR, et un soutien financier est disponible pour les déplacements (conférences, ateliers, écoles d'été, ...).</p>
Projets liés à la thèse	<p>Chaire en IA de confiance, Chaire SAFE AI</p>
Collaboration	<p>Cassio de Campos, UAI team, Eindhoven University of Technology, The Netherlands.</p>
Contact	<p>Les candidatures doivent comprendre une lettre de motivation, un curriculum vitae et les coordonnées d'au moins une personne de référence.</p> <p>Les candidatures et les questions doivent être envoyées à :</p> <ul style="list-style-type: none"> - Vu-Linh Nguyen (vu-linh.nguyen@hds.utc.fr) - Yves Grandvalet (yves.grandvalet@hds.utc.fr)