

Sujet de stage M2

Application des Réseaux Bayésiens Linéaires Gaussiens pour la Gestion de l'État de Santé des Éoliennes

Laboratoires: Laboratoire Computer Science & Artificial Intelligence (LaCSAi) CRIGEN-ENGIE et le Lip6 (Laboratoire d'informatique de Paris 6) de Sorbonne Université.

- **Mots clés** : aide à la décision, réseau bayésien, linéaire Gaussiens, apprentissage automatique
- **Discipline** : Informatique
- **Localisation** : 4 Place Jussieu, 75005 Paris (Lip6) et Site du Landy, 361 Avenue du Président Wilson, 93210 Saint-Denis (LaCSAi)
- **Contacts** :
 - o Ahmed Mabrouk, ahmed.mabrouk@engie.com
 - o Christophe Gonzales, christophe.gonzales@lip6.fr
 - o Philippe Calvez, philippe.calvez1@engie.com.

Contexte de la recherche

Étant fournisseur d'énergie engagé dans la transition énergétique et fortement impliqué dans le développement des énergies renouvelables, ENGIE déploie toute sa capacité en matière de recherche et d'innovation pour relever ces défis et offrir des services de plus en plus respectueux de l'environnement à ses clients. Avec la décarbonisation et la décentralisation, la digitalisation constitue un des piliers essentiels de la transformation d'ENGIE en faveur de la transition énergétique dans le monde. Dans ce contexte, la plateforme digitale de gestion des données Darwin en est la concrétisation. En effet, cette plateforme permet à ENGIE de recueillir, d'analyser et d'exploiter en temps réel 24/24 et 7/7 les données transmises par les capteurs installés sur les éoliennes du parc "La Haute Borne" (Meuse, France). D'autres installations à l'instar des parcs solaires et hydroélectriques sont également connectées à la plateforme Darwin.

Face à ce déluge de données, la première question à se poser est la suivante : comment réutiliser toutes cette masse de données ? En réalité, sur le plan des données acquises, la gestion (acquisition et visualisation) des données issues de différentes activités a considérablement évolué ces dernières années, drainant derrière elle autant de nouvelles problématiques que de résultats. Parmi ces nouvelles problématiques, on peut citer celle de l'analyse et l'extraction des connaissances à partir des données relatives aux fonctionnements des éoliennes, de façon à ce qu'elles soient manipulables et compréhensibles par les utilisateurs. Ces données-là sont en effet assez complexes de part leur taille et leur variabilité. Afin d'optimiser ces tâches laborieuses, il est impératif de se doter d'outils permettant d'exploiter au mieux l'historique des données relatives aux différents phénomènes étudiés et d'interpréter les différents signaux en provenance de plusieurs sources de données. Dans ce contexte, l'utilisation des techniques de machine learning et d'aide à la décision s'avère cruciale.

Le projet du stage vise à mettre en place un outil à base de réseau Bayésien[1] permettant de synthétiser et hiérarchiser l'ensemble des données (principalement quantitatives) relatives aux différents phénomènes étudiés et d'interpréter les différents signaux en provenance de plusieurs sources de données. Pour cela, il réunit un consortium composé de deux laboratoires complémentaires sur les problématiques d'aide à la décision, d'apprentissage automatique et de modèles décisionnels (Lip6) et de l'acteur majeur de l'énergie (LaCsAi-ENGIE).

Objectif

Au sein d'un environnement confus, complexe et aux enjeux multiples, le décideur fait généralement appel à plusieurs bases de données et de connaissances pour l'aider dans son processus de prise de décision. Cependant, quand le problème décisionnel est complexe et de surcroît mal défini, le décideur éprouve de plus en plus de difficultés à comprendre les données qu'il a à sa disposition, ce qui complique par conséquent le processus de prise de décision. Comme c'est déjà le cas pour la plupart des problèmes étudiés chez ENGIE, nous avons ressenti le besoin d'aider le décideur à bien comprendre le problème étudié, et de l'aider à parvenir à une meilleure interprétation des données en provenance de son environnement. Dans ce cadre, l'objectif principal de ce stage consiste à mettre en place une solution logicielle et algorithmique valorisant les données en provenance du parc éolien et permettant ainsi d'explicitier les problèmes décisionnels complexes et mal définis.

L'utilisation des réseaux bayésiens[1] dans ce domaine d'application est particulièrement appropriée, dans la mesure où ils permettent de modéliser d'une manière synthétique et hiérarchique un système complexe, et d'en extraire de l'information (diagnostic, étude de la fiabilité, pronostic) pour la prise de décision. L'un des avantages majeurs de ce type de modèles réside dans le fait que, contrairement à d'autres modèles de type « boîte noire », les réseaux bayésiens sont aisément interprétables par des experts du domaine, ces derniers pouvant éventuellement participer à leur construction. Ce travail de recherche s'appuiera sur des travaux déjà effectués au niveau du CRIGEN portant sur la représentation du fonctionnement des éoliennes « dans un état normal » à l'aide d'un réseau bayésien. Il s'appuiera aussi sur les travaux déjà réalisés au sein du LIP6 dans le département DESIR, notamment l'utilisation d'aGrUM, une librairie qui implémente plusieurs algorithmes destinés à la construction et la manipulation des modèles graphiques probabilistes. Le travail à réaliser dans le cadre du stage doit passer par les étapes suivantes :

Étape 1 : État de l'art

Au cours de cette étape, vous serez amené(a) à conduire une étude bibliographique détaillée sur les réseaux bayésiens ainsi que les principaux algorithmes dédiés à l'apprentissage de structure et paramètres de ce type de modèle.

Étape 2 : Implémentation d'un réseau bayésien

A la lumière de cette étude, le stagiaire aura pour tâche d'implémenter un algorithme de construction d'un réseau bayésien potentiellement dynamique et du type conditionnel Gaussien[2] adapté à la nature des variables d'entrée (temporelles et continues). Pour ce faire, vous serez amené(e) à passer par les deux étapes suivantes :

1. Analyse des données descriptives des fonctionnements des éoliennes dans le but de sélectionner les variables pertinentes à prendre en considération lors de la construction du modèle
2. À la lumière de cette analyse, vous ferez le choix ainsi que l'implémentation de l'algorithme d'apprentissage automatique du réseau bayésien le plus adapté aux données d'entrée

Étape 3 : Expérimentations

Le stagiaire aura également pour tâche de proposer une méthodologie de validation du modèle élaboré à travers la confrontation des résultats expérimentaux avec les simulations du modèle.

Étape 4 : Bilan

À l'issue de cette étape, vous élaborerez un bilan sous la forme :

1. Une présentation détaillée au sein du lab CsAI dans le cadre de nos sessions hebdomadaires
2. La rédaction d'un rapport de stage illustrant les différentes étapes du projet (résultat de l'état de l'art de premier niveau, problématique, hypothèses scientifiques, explication de l'approche considérée, expérimentation, perspectives)
3. Si le temps le permet, il vous sera demandé(e) de rédiger/participer à un article scientifique qui récapitule l'ensemble des travaux réalisés.

Profil recherché

Le candidat doit suivre un Master M2 en informatique/machine learning. Des connaissances en mathématiques, machine learning et en modèles graphiques probabilistes (réseaux bayésiens) sont nécessaires. Des connaissances sur les énergies renouvelables et plus particulièrement les éoliennes seraient notamment appréciées.

Une expérience en programmation avec Python et/ou C++ serait aussi un plus. Enfin, le candidat devra démontrer son intérêt pour le domaine de l'éolien et des réseaux bayésiens.

Le dossier de candidature doit être envoyé le plus tôt possible à Philippe Calvez (philippe.calvez1@engie.com), Ahmed Mabrouk (ahmed.mabrouk@engie.com) et Christophe Gonzales (christophe.gonzales@lip6.fr) avec :

- CV détaillé (ensemble des expériences et technologies maîtrisées)
- une lettre de motivation
- éventuellement des recommandations.

Durée du stage

6 mois – Date de début flexible

Divers

Convention de stage obligatoire.

Le stage pourra en fonction des résultats déboucher sur une thèse CIFRE.

Références :

[1] Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems : Networks of Plausible Inference. Morgan Kaufmann, second edition in 1991. 5, 13, 30, 48, 194

[2] Lauritzen, S. and Jensen, F. (2001) . Stable local computation with conditional gaussian distributions. Statistics and Computing, 11:191–203