

Internship offer (EN¹): Co-constructing a graphical model through interactive visualization

Anastasia.Bezerianos@universite-paris-saclay.fr (Polytech, Univ. Paris-Saclay),
Nadia.Boukhelifa@inrae.fr (INRAE, Univ. Paris-Saclay), and
Melanie.Munch@u-bordeaux.fr (Univ. Bordeaux)

Keywords: interactive visualization, graphical models, comparative visualization, Artificial Intelligence (AI) learning models.

Context

Sustainable food production is one of the main challenges facing our world today. At INRAE (Institut national de recherche pour l'agriculture, l'alimentation et l'environnement) different experts come together to study production methods. These experts can be biologists and agronomists (domain experts) as well as statisticians and mathematicians (modelers) producing models that simulate complex biological and agricultural processes.

The internship work is part of the ANR EVAGRAIN [1] research project which aims to develop, in collaboration with stakeholders in the cereal sector, a tool that predicts the quality of wheat from data collected each year in order to better anticipate the difficulties caused by climate events. To achieve this we have already constructed a knowledge base with existing data generated in the project, through the use of an existing ontology. From this knowledge base we can generate explanatory models (in the form of Bayesian networks, e.g., figure 1) that can be combined with expert knowledge. From this, explanatory models are generated, learned by integrating expert knowledge.

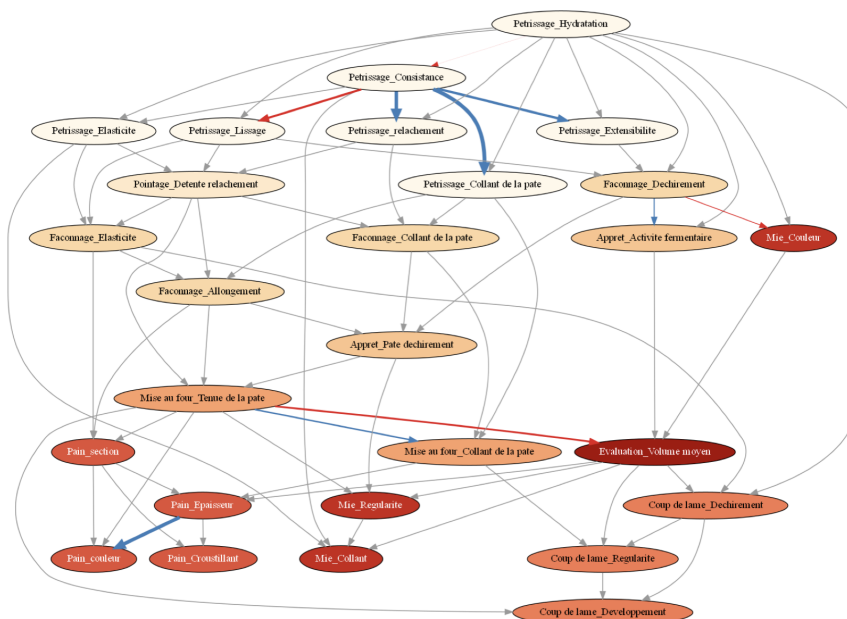


Figure 1 : An example of a learned graphical model (a Bayesian network) generated for the EVAGRAIN project

¹ A french version of the internship offer is available below (page 4/6)

Currently the learning models are developed offline, and then presented to domain experts in a scheduled meeting. Through discussion, experts give feedback with regards to what variables to consider, the correct dependencies, and those that need to be modified. Although experts find it very useful to be involved in the modeling process, there are currently a number of issues that hamper the collaboration between the modelers and domain experts: (a) feedback about the correctness of the model is manual and tedious, and existing model quality indicators can be hard to interpret by the experts. The modeler currently takes notes and then goes back to the code to re-implement the changes; and (b) the process is time consuming and error prone.

Objectives of the internship

The aim of this internship is to facilitate the dialogue between domain experts in agronomy and food technology and the model builders through an online visualization system. The selected student will perform the following tasks:

- a literature review on visualization / explanation methods for Bayesian networks, and expert knowledge elicitation interfaces.
- follow a user-centered design methodology to design and implement an online visualization system that provides:
 - a graphical representation of the learned model (e.g., node-link diagram or an adjacency matrix);
 - a dashboard containing a set of indicators such as entropy, confidence, precision, robustness, *etc*, allowing domain experts to have interpretable information about the quality of the model.
 - user interactions to allow experts to edit the graphical model taking into account their own expertise (e.g., to order or group variables, add constraints, etc); and
 - a history of the different iterations on the model allowing for easy comparisons between the different versions.
- evaluate the developed prototype with domain experts and modelers.

The selected student will build on existing work from the ANR EVAGRAIN project. The following resources are currently available for an immediate start of this internship: data about the different wheat and flour quality tests; a preliminary learning model implementation in Python; video recordings of online exchanges and discussions between modelers and domain experts when confronted with a learning model; and access to domain experts to gather user requirements and feedback.

Existing research that can help guide work for this internship include explainable methods for Bayesian Networks [2,3], BayesPiles [4], and comparative visualization [5].

Required and desirable skills

Required skills include: web development; programming (JavaScript, Python, other); user-centered design; information visualization. Interest in working with real-world data and domain experts. Knowledge in machine learning is not required but experience is a plus.

Work environment

- **Supervisors** : send CV and motivation letter to Anastasia Bezerianos anastasia.bezerianos@universite-paris-saclay.fr (Polytech, Univ. Paris-Saclay), Nadia Boukhelifa nadia.boukhelifa@inrae.fr (INRAE, Univ. Paris-Saclay), and Mélanie Munch melanie.munch@u-bordeaux.fr (Univ. Bordeaux)
- **Collaborators** : Cédric Baudrit (INRAE), Kamal Kansou (INRAE)
- **Internship duration** : 6 months, start in February or March 2023
- **Location** : AgroParistech Saclay Campus, 22 place de l'agronomie 91120 palaiseau
- **Allowance** : around 600 euros a month

References

- [1] ANR project EVAGRAIN <https://anr.fr/Projet-ANR-20-CE21-0008>
- [2] Lacave, C., & Díez, F. J. (2002). A review of explanation methods for Bayesian networks. The Knowledge Engineering Review, 17(2), 107-127. <https://tinyurl.com/ussdjfby>
- [3] <https://lstu.fr/06zWFGHx>
- [4] Vogogias, A., Kennedy, J., Archambault, D., Bach, B., Smith, V. A., & Carrant, H. (2018). BayesPiles: Visualisation support for bayesian network structure learning. ACM Transactions on Intelligent Systems and Technology (TIST), 10(1), 1-23. <https://tinyurl.com/c3rtn2b4>
- [5] Gleicher, M., Albers, D., Walker, R., Jusufi, I., Hansen, C. D., & Roberts, J. C. (2011). Visual comparison for information visualization. Information Visualization, 10(4), 289-309. <https://tinyurl.com/y87ddfbz>

Offre de stage (FR): Co-construire un modèle graphique à travers la visualisation interactive

Anastasia.Bezerianos@universite-paris-saclay.fr (Polytech, Univ. Paris-Saclay),
Nadia.Boukhelifa@inrae.fr (INRAE, Univ. Paris-Saclay), and
Melanie.Munch@u-bordeaux.fr (Univ. Bordeaux)

Mots clés : visualisation interactive, modèles graphiques, visualisation comparative, modèles d'apprentissage (Intelligence Artificielle IA).

Contexte

La production alimentaire durable est l'un des principaux défis auxquels notre monde est confronté aujourd'hui. Différents types d'experts à l'INRAE (Institut national de recherche pour l'agriculture, l'alimentation et l'environnement) étudient cette question. Ces experts peuvent être des biologistes et des agronomes (experts du domaine) ainsi que des statisticiens et des mathématiciens (modélisateurs) produisant des modèles qui simulent des processus biologiques et agricoles complexes.

Ce travail s'inscrit dans le cadre du projet de recherche ANR EVAGRAIN [1], dont l'objectif est de développer en collaboration avec des acteurs de la filière céréalière un outil permettant de prédire la qualité de blé à partir de données collectées chaque année afin de pouvoir mieux anticiper les difficultés dues aux événements climatiques. Pour ce faire, une base de connaissances structurée par une ontologie a été établie, permettant d'intégrer les nouvelles données générées pendant le projet. A partir de celle-ci, des modèles d'apprentissage explicatifs sont générés, appris en intégrant des connaissances expertes (ex., figure 1).

Actuellement, les modèles d'apprentissage sont développés hors ligne, puis présentés à des experts du domaine lors d'une réunion planifiée. Au cours de la discussion, les experts donnent leur avis sur les variables à prendre en compte, les dépendances pertinentes et celles qui doivent être modifiées. Bien que les experts trouvent très utile d'être impliqués dans le processus de modélisation, il existe actuellement un certain nombre de problèmes qui entravent la collaboration entre les modélisateurs et les experts du domaine : (a) la critique du résultat appris est manuelle et fastidieuse, et les indicateurs de qualité des modèles existants sont difficilement interprétables par les experts. Le modélisateur prend actuellement des notes, puis revient au code pour implémenter les modifications ; et (b) le processus prend du temps et est sujet aux erreurs.

Objectif du stage

L'objectif de ce stage est de faciliter le dialogue entre les experts du domaine en agronomie et technologie alimentaire et les modélisateurs en concevant et en mettant en œuvre un système de visualisation en ligne. L'étudiant.e sélectionné.e doit effectuer les tâches suivantes :

- une revue de littérature sur la visualisation / explication des réseaux Bayésien et les interfaces d'élicitation des connaissances expertes.

- suivre une méthodologie de conception centrée sur l'utilisateur pour concevoir et mettre en œuvre un système de visualisation en ligne qui fournit:
 - une représentation graphique du modèle appris (par exemple, un graphe ou une matrice d'adjacence);
 - un tableau de bord contenant un ensemble d'indicateurs tels que l'entropie, la confiance, la précision, la robustesse, *etc.*, permettant aux utilisateurs d'avoir des informations interprétables sur la qualité du modèle.
 - des interactions utilisateurs pour permettre aux experts d'éditer le modèle graphique en tenant compte de leur propre expertise (par exemple, pour ordonner ou regrouper des variables, ajouter des contraintes, *etc.*) ; et
 - un historique des différentes itérations sur le modèle permettant des comparaisons faciles entre les différentes versions.
- évaluer le prototype développé avec des experts du domaine et les modélisateurs.

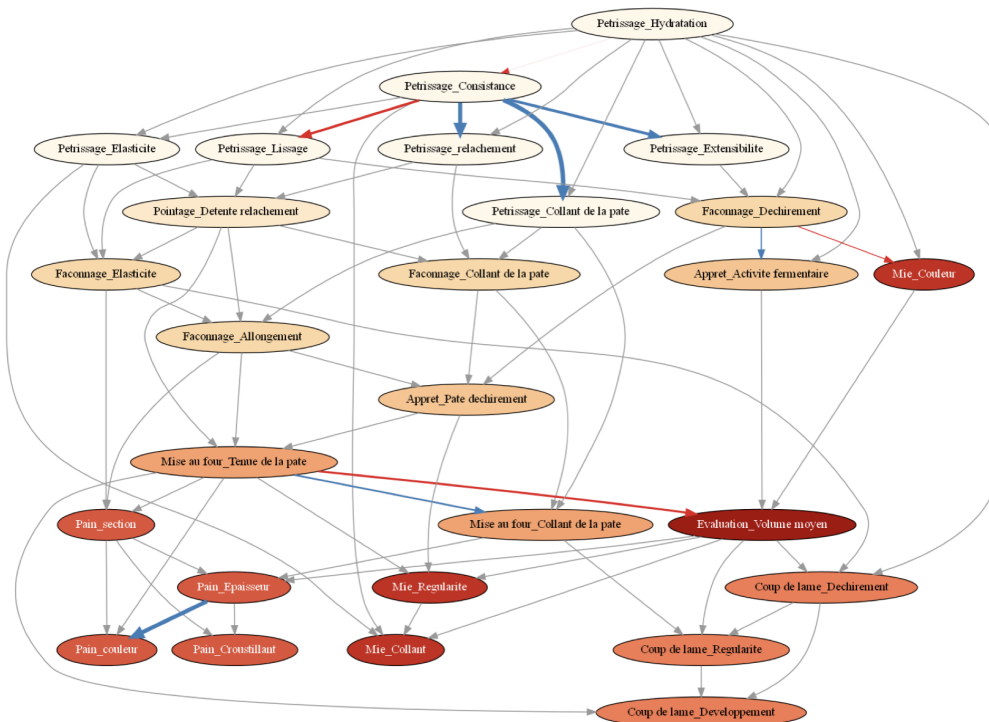


Figure 1 : Un exemple de modèle graphique d'apprentissage (réseau Bayésien) généré dans le cadre du projet EVAGRAIN

L'étudiant.e sélectionné.e s'appuiera sur des travaux existants du projet ANR EVAGRAIN. Les ressources suivantes sont actuellement disponibles pour un démarrage immédiat de ce stage : des données sur les différents tests de qualité du blé et de la farine ; une implémentation préliminaire du modèle d'apprentissage en Python ; des enregistrements vidéo d'échanges et de discussions en ligne entre modélisateurs et experts du domaine lorsqu'ils sont confrontés à un modèle d'apprentissage ; et l'accès à des experts du domaine pour recueillir les besoins et les commentaires des utilisateurs.

Les recherches existantes qui peuvent guider le travail pour ce stage comprennent des méthodes d'explicabilité pour les réseaux Bayésiens [2,3], BayesPiles [4] et la visualisation comparative [5].

Profil souhaité

Les compétences requises incluent : développement web ; programmation (JavaScript, Python, autre); conception centrée sur l'utilisateur ; visualisation d'informations. Intérêt pour travailler avec des données et des experts du domaine. Des connaissances en machine learning ne sont pas requises mais un plus.

Environnement scientifique

- **Encadrants** : envoyer CV et lettre de motivation à Anastasia Bezerianos anastasia.bezerianos@universite-paris-saclay.fr (Polytech, Uni. Paris-Saclay) et Nadia Boukhelifa nadia.boukhelifa@inrae.fr (INRAE, Uni. Paris-Saclay), Mélanie Munch melanie.munch@u-bordeaux.fr (Univ. Bordeaux).
- **Collaborateurs** : Cédric Baudrit (INRAE), Kamal Kansou (INRAE).
- **Durée du stage** : 6 mois, début Février ou Mars 2023
- **Lieu** : AgroParistech Campus de Saclay, 22 place de l'agronomie 91120 Palaiseau.
- **Allocation** : environ 600 euros par mois.

Références

[1] ANR project EVAGRAIN <https://anr.fr/Projet-ANR-20-CE21-0008>

[2] Lacave, C., & Díez, F. J. (2002). A review of explanation methods for Bayesian networks. The Knowledge Engineering Review, 17(2), 107-127. <https://tinyurl.com/ussdjfby>

[3] <https://lstu.fr/06zWFGHx>

[4] Vogogias, A., Kennedy, J., Archambault, D., Bach, B., Smith, V. A., & Carrant, H. (2018). BayesPiles: Visualisation support for bayesian network structure learning. ACM Transactions on Intelligent Systems and Technology (TIST), 10(1), 1-23. <https://tinyurl.com/c3rtn2b4>

[5] Gleicher, M., Albers, D., Walker, R., Jusufi, I., Hansen, C. D., & Roberts, J. C. (2011). Visual comparison for information visualization. Information Visualization, 10(4), 289-309. <https://tinyurl.com/y87ddfbz>