Sustainable Food Production and AI+Visualization Research Internship

Co-constructing a graphical model through interactive visualization

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, mathematicians or ML experts (modelers) producing models that simulate complex biological and agricultural processes.

The internship work is part of the ANR EVAGRAIN project [1] which aims to model *bread-making*, in order to create decision support tools for millers and bakers to classify optimal French wheat, balancing bread quality and societal demands for healthy, sustainable wheat production [2]. The bread-making process is modeled using a Bayesian Network (BN), a Machine Learning (ML) probabilistic approach that generates scalable models under uncertainty [3]. The model was trained on a large bread-making dataset and input from domain experts [2] (see Figure 1). In BNs, nodes represent random variables, directed arrows represent probabilistic dependencies, and relationships between nodes are quantified through conditional probability tables at each node.

These models are so far 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. While Bayesian Networks and other ML methods have powerful modeling capabilities for complex data, they often lack *explainability* for domain experts: expressing the logic behind the network so that experts can understand predictions and significant flows of information. Furthermore, 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.
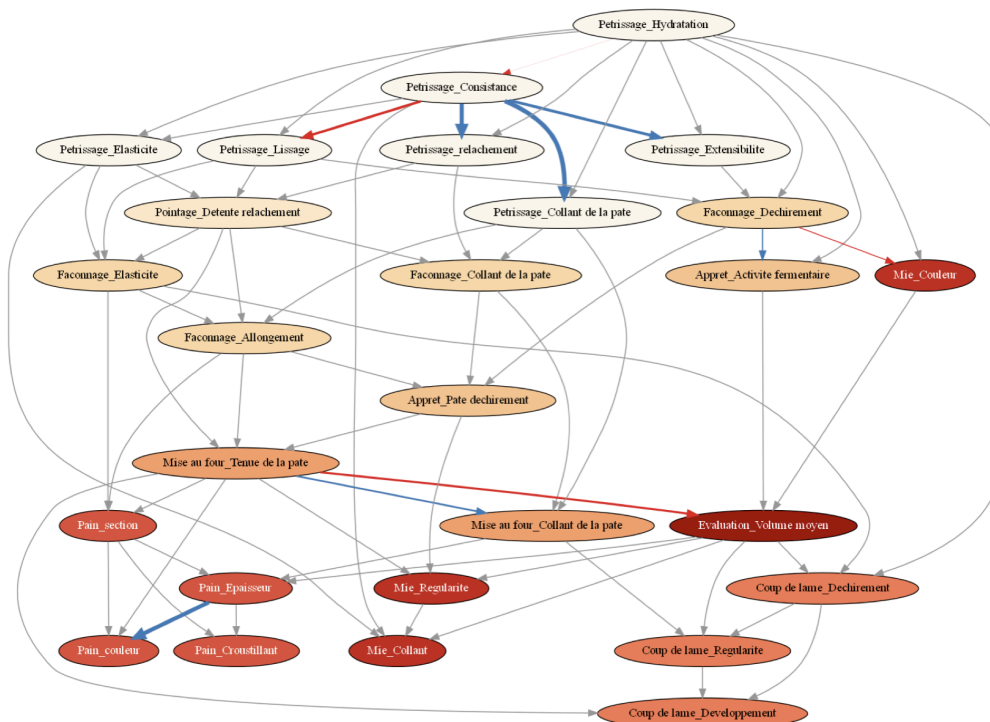
Figure 1 : An example of a learned graphical model (a Bayesian network) generated for the EVAGRAIN project, which models bread-making steps for use in wheat quality assessment.

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 build on existing work (namely an existing Bayesian model and an early prototype visualization tool) to perform the following tasks:
- get up to speed on literature on visualization / explanation methods for Bayesian networks, and expert knowledge elicitation and refinement interfaces.
- follow a user-centered design methodology to design and implement interactive visualizations that provide:
  - visualization of a set of indicators such as entropy, confidence, precision, robustness, *etc,* to be shown in conjunction with the main visualization of the model, 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; a preliminary web-based visualization platform in JavaScript, video recordings of online exchanges and discussions between modelers and domain experts when confronted

with a learning model and the visualization; 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 [4,5], BayesPiles [6], and comparative visualization [7].

Required and desirable skills

Required skills include: web development; programming (JavaScript/D3.js, Python, other). Interest in working with real-world data and domain experts. Knowledge in machine learning is not required but experience is a plus. Experience in user-centered design and information visualization are a plus.

Work environment

- **Supervisors** : send CV, motivation letter and unofficial transcripts (marks) to Nadia Boukhelifa nadia.boukhelifa@inrae.fr (INRAE, Univ. Paris-Saclay); and anastasia.bezerianos@universite-paris-saclay.fr (LISN, Univ. Paris-Saclay)
- **Collaborators (INRAe)**: Mélanie Münch, Cédric Baudrit, Kamal Kansou
- **Internship duration** : 6 months, ideally starting in February or March 2025
- **Location** : AgroParistech Saclay Campus, 22 place de l'agronomie 91120 palaiseau
- **Allowance** : around 650 euros a month

References

[1] ANR project EVAGRAIN https://anr.fr/Projet-ANR-20-CE21-0008
[2] Münch, M., Baudrit, C., Kansou, K., and Fernandez, C. 2023. Conception d'un Outil de Diagnostic: Application à l'Essai de Panification en Industrie Boulangère. In JFRB 2023: 11èmes Journées Francophones sur les Réseaux Bayésiens et les Modèles Graphiques Probabilistes. https://hal.science/hal- 04190423/ Munch, Mélanie, et al. "Diagnosis based on sensory data: Application to wheat grading quality." Innovative Food Science & Emerging Technologies 96 (2024): 103771.
[3] Pearl, J., and Russell, S. 2000. Bayesian Networks.
ftp://cobase.cs.ucla.edu/pub/stat_ser/R277.pdf
[4] 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
[5] https://lstu.fr/06zWFGHx
[6] Vogogias, A., Kennedy, J., Archambault, D., Bach, B., Smith, V. A., & Currant, 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
[7] 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