[CoCo-IML] Collaborative interactive machine learning : Co-constructing trustworthy predictive models to improve wheat quality assessments

Abstract

In interactive machine learning, humans and machine learning algorithms collaborate to achieve tasks mediated by interactive visual interfaces. In traditional interactive machine learning, models are often developed with single users in mind. However, real world applications often require the collaboration of different types and levels of expertise to help reliably develop and assess the results of machine learning models. The goal of this PhD is to develop a *collaborative interactive machine learning* (CoCo-IML) framework, and a prototype system that implements it, in order to: (i) on the one hand support the dialogue between different types of domain experts and modelers, and (ii) on the other hand, support the dialogue between the experts and the machine learning algorithm. The proposed work will be demonstrated through a real world use-case application from the agronomy and food technology domains, following a participatory design methodology. The evaluation of the framework and prototype system will take into account various criteria including the interpretability and user trust of the model, as well as the quality of collective decision making under uncertainty.

Keywords: Interactive Machine Learning, explainability, Human-Computer Interaction, Visualization, Collaboration, agronomy, sustainability.

Description of the project

1. Context and challenges

Interactive machine learning (IML) is a promising emerging research area within visual analytics, a discipline focused on the use of interactive visual interfaces to enhance analytical reasoning (Thomas, 2006). It involves a close integration of human expertise with machines in the learning process via interactive interfaces and visualization (Boukhelifa, 2020). The typical IML workflow consists of a two-step process within each iteration (Jiang, 2019). *First*, intermediate results of the model and the underlying data are visually displayed to the user, who then engages with these visualizations, extracts insights about both the data and the model, and subsequently provides feedback to the model. *Second*, the model is iteratively refined by incorporating the user's input. Thus, IML enables human-driven interactive steering over machine learning models, offering more advantages for **tasks that require human expertise** in the analytical process (Jiang, 2019). For example, an IML system can help users more effectively sort and cluster large numbers of documents where the user adds items to clusters according to their prior knowledge, and the system learns from those interactions to then recommend new items and clusters (Drucker, 2011). In EvoGraphDice (Boukhelifa, 2013), the user ranks scatterplots according to their preference, and the system learns from their interactions and proposes new plots more pertinent to their search. In all those systems, feedback from the human helps steer the machine learning (ML) algorithm, but also the ML feedback informs user tasks.

Much of the traditional Interactive Machine Learning (IML) focuses on designing systems for **a single user**. In some applications however, **multiple types of experts** need to come together to be able to interact with and interpret the results of IML (Boukhelifa, 2019). Robert et al. (2016) define collaborative interactive machine learning as a set of methods where at least one human is integrated into the ML algorithm using a specific user interface that facilitates the manipulation of both the algorithm and its intermediate stages, to find an optimal solution. Specifically, users can manipulate the parameters of the algorithm during its execution. Indeed, collaborative IML can go further than single-person interaction with ML to take into account larger teams, richer interactions and multiple types of users (e.g., model builders, domain experts, decision makers) having different levels of expertise (novices and experts). These teams can work together in a synchronous or asynchronous fashion (Isenberg, 2011).

The interactive and collaborative context within which this PhD work will take place brings numerous challenges. In this PhD, the primary emphasis is on tackling two key issues related to human-factors, with a specific focus on communicating uncertainty: **(1) human-computer interaction and collaboration challenges:** how to create collaborative human-machine interfaces that assist the multi-user interaction needed to co-construct these IML models. These user interfaces need to take into account the various types of knowledge and expertise of the team involved in the process, the different types of machine learning algorithms used and their inherent uncertainty, the various styles of collaborations (synchronous or asynchronous), and aim to build common ground (Gergle, 2013); and **(2) visualization and explainability challenges**: how to represent and explain the constructed models and their uncertainty to multiple types of users (that may have different levels of data and visualization expertise), track and visualize the intermediate results of the machine learning process and the associated quality measures. Particular attention will be paid to the different sources of uncertainty, from the data, the learning process, and the human-human collaboration (Boukhelifa 2012, Boukhelifa 2017).

The goal of this PhD is to develop a general collaborative interactive machine learning framework (CoCo-IML) to help domain experts and modelers effectively work together to build reliable and trustworthy machine learning models. We will demonstrate the framework through a real-world case study on wheat quality assessment.

2. Use Case : Predictive models for wheat quality assessment using Bayesian networks

Assessing the quality of wheat is a technological challenge that many countries, including France, must be able to meet in order to enhance its wheat production. This assessment relies on tests, such as for moisture and protein content, which overall aim to evaluate the ability of a grain to reach the specifications defined for its use (e.g., for bread or biscuit making). The tests that are currently available are often limited, imprecise and time-consuming, making it difficult to establish robust wheat quality criteria (nutritional, sanitary, technological). To address this problem, researchers at INRAe are working in collaboration with industrial partners to develop a more robust wheat quality assessment system that meets the requirements of a wide range of users, and is able to predict wheat quality variations such as from extreme weather.

This PhD will rely on findings from an ongoing funded project (ANR EVAGRAIN[1]) which aims to develop, in collaboration with stakeholders in the cereal sector, a Machine Learning (ML) model that predicts the quality of wheat from data collected each year in order to better anticipate the difficulties caused by climate events. Specifically, the model encodes the dependencies between the behavioral properties of wheat (e.g., capacity of aggregation of proteins) and the quality criteria (e.g., viscosity of the dough, volume of the bread, etc), and integrates knowledge and data from different sources: data from the project, various databases, literature, existing models, and expert knowledge. The results are shown as a graphical model (a **Bayesian Network,** Figure 1) which shows key variables in the modeled system (the nodes), relationships between the variables (links), and the nature of these relationships (positive/negative correlation in color).

Currently the learning model is developed offline, and then presented to domain experts in a scheduled (often online) meeting. Through discussion, experts give feedback to the modeling team with regards to what variables to consider, the correct dependencies, and those that need to be modified. Although experts find it 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. The modeling team currently takes notes and then goes back to the code to re-implement the changes; **(b)** Existing model quality (uncertainty) indicators which can be complicated to interpret by the experts, are not currently communicated; and **(c)** The collaborative process is time consuming and can be error prone.

Objectives of the PhD : The goal of this PhD is to facilitate the dialogue between domain experts in agronomy & food technology, and model builders through a collaborative and interactive machine learning platform. This will undoubtedly provide a valuable contribution to the use-case domain (wheat quality assessment) with the creation of more robust models and tools to access them. But it will also contribute to the research domains of human-computer interaction and interactive machine learning.

This PhD project specifically tackles challenges provided by a concrete real world use-case, but the ultimate goal is to develop a general framework that supports collaborative steering of machine learning models for a variety of collaborative use cases and types of models.

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

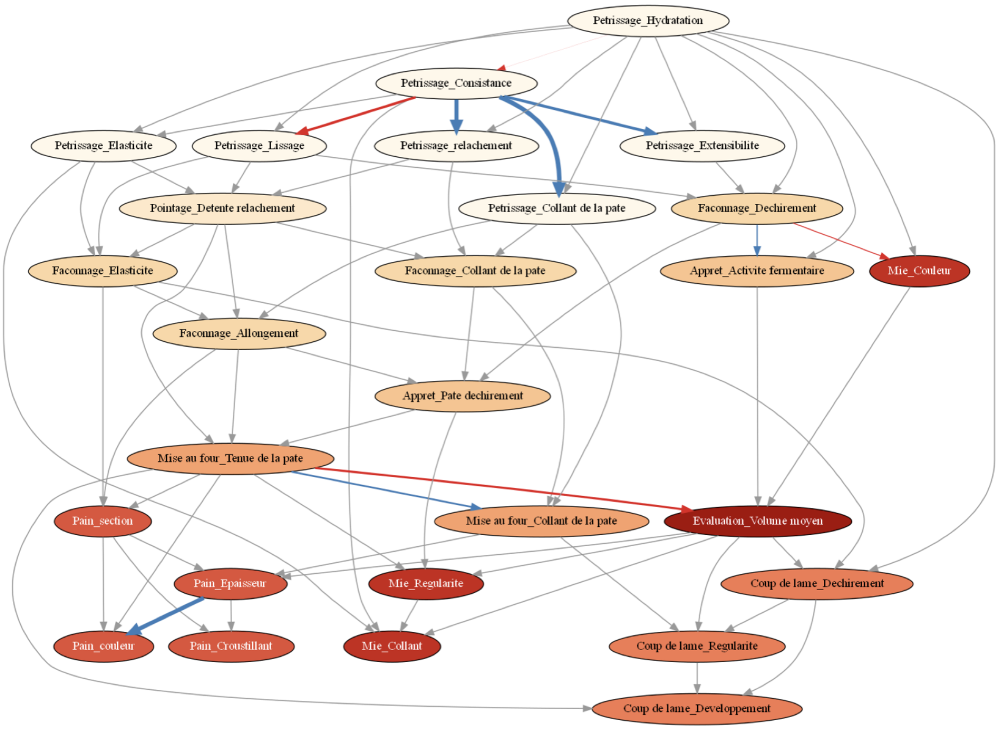

Figure 1 : An example of a (simplified) Bayesian network generated for the EVAGRAIN project

3. Related work

Collaborative approaches for interactive ML systems Robert et al., (2016) define collaborative interactive machine learning as a set of methods where at least one human is integrated into the ML algorithm using a specific user interface that facilitates the manipulation of both the algorithm and its intermediate stages, to find an optimal solution. Specifically, users can manipulate the parameters of the algorithm during its execution. Much of the traditional Interactive Machine Learning (IML) focuses on designing systems for a single user (Boukhelifa, 2013). In the use case under study for this PhD, a group of domain experts (typically 3-5) need to work together to assess the quality of the ML model. These experts need to consult each other, as each might have expertise that only covers part of the biological domain (Boukhelifa, 2019).
Previously, collaboration between domain experts (and the modeling team) was carried out through informal meetings (synchronous distributed collaboration as described by Isenberg et al., 2009 and Isenberg et al., 2011). The goal of this PhD project is to provide a consistent platform for a small to medium group of domain experts and modelers to work together (human-to-human collaboration), and to collaborate with the ML algorithm (human-machine collaboration) to construct a learning model able to assess reliably the quality of wheat.
To develop this collaborative framework, we will build on existing work from the field of collaborative visualization, such by Heer et al., (2008) to draw an initial set of requirements (e.g., support for division and allocation of work, improve shared context and awareness, decide on discussion models, etc). In comparison to traditional IML systems (e.g,. by Stark et al., 2012; Micheils et al., 2021) which focus on human-machine collaborative data mining, **the novelty** here is in integrating techniques from collaborative visualization (i.e. human-human collaboration) into the design of interactive machine learning systems; and in explaining the resulting models and their quality (uncertainty) to different types of user groups.

Visualization and explanation methods for Bayesian networks A Bayesian Network (BN) is a probabilistic graphical model that represents a set of variables and their probabilistic relationships. It is used to encode and represent probabilistic knowledge and to make probabilistic inferences about the relationships between variables. Despite the widespread use of BNs across many domains, the inference and visualization of these networks can be challenging for users without programming skills (Chen, 2019). Moreover, Lacave et al. (2002) point out that because the reasoning process of BNs follows a normative approach, it is more challenging to explain their inference results

compared to methods that try to imitate human reasoning. As a result, Lacave et al. argue that the importance of explanation methods is even greater in BNs than in heuristic expert systems.
According to Lacave et al., the properties of explanations are categorized into three groups pertaining to *what* is explained (e.g., evidence, model, reasoning), *how* it is explained (e.g., text, graphics, multimedia), and to *whom* it is explained (e.g., adapt to user's knowledge about the domain or the reasoning method). This PhD project will investigate which explanation method/properties are more suitable for the collaborative IML context described in the wheat assessment use-case. **The novelty** of the approach will be in adapting explanations to different types of expertise, and conducting formal user studies to investigate the effect of a type of explanation on user comprehension and trust.
Moreover, this way of visualizing datasets helps to better understand their strength and limitations (e.g., by bringing to light **biases** not taken into account during their conception). Munch et al. (2021) gives an example of the effectiveness of this approach, by learning a BN from a dataset on bio-based products transformation processes combined with expert knowledge. The final results showed some intriguing relations between the variables, which led to the design of new experiments to validate them. More generally, experts could use such information to formulate hypotheses, leading to a new way of comprehending knowledge discovery, as well as new criteria to define data quality in regards to ML algorithms.

Expert knowledge elicitation interfaces While there has been a growing interest in exploring the interaction between machine learning and artificial intelligence systems and human knowledge and beliefs, the process of eliciting knowledge from domain experts in the development of ML models has received limited attention. Based on a review of 28 papers, Kerrigan et al., (2021) highlighted four areas where incorporating expert knowledge into machine learning can prove beneficial: problem specification, feature engineering, model development, and model evaluation. In terms of medium of elicitation, they found a number of forms of "conversations" between the domain expert(s) and the ML researcher or ML system, including custom interfaces made specifically for the given task, written communication, meetings or interviews, and shadowing.
We will focus on model development and evaluation as areas of ML, in order to design and implement expert knowledge elicitation methods via dedicated user interface components. For example, domain experts can use their knowledge to validate an existing link (making a binary decision), or provide richer feedback to the ML (e.g., probabilistic information). **The novelty** will be in conducting formal user studies to investigate the impact of how knowledge is elicited, on the usefulness of the information for the modeling process.

4. Method

The work proposed in the PhD focuses on facilitating human-to-human collaboration as well as human-machine collaboration in the context of Bayesian modeling for a use case application in agronomy & food technology (wheat quality assessment). The aim is to investigate the following research questions:
RQ1. what type of collaboration support is needed to assist model development and evaluation, in a (initially) synchronous distributed context?
RQ2. does visualization improve the group's understanding of the model and its quality or uncertainty?
RQ3. are experts able to express their knowledge accurately using the visualization interface?
RQ4. does the co-construction process lead to more user trust in model predictions ?
RQ5. can the ML model help the domain expert reach better decisions and design new experiments?

A collection of methods will be used to answer these research questions: **(i)** Human-Computer Interaction & User-centred design methods: to capture user requirements; iterative evaluation of the developed prototypes; HCI methodologies to evaluate interactive systems; mixed evaluation methods (quantitative and qualitative); **(ii)** Interactive & collaborative visualization techniques: new interaction techniques to allow users to edit graph-based models and express their expertise. Comparative visualization techniques (Vogogias, 2018) will be used to highlight similarities, differences and the evolution between different versions of the model (for a single user, and/or across different users). The Marcelle toolkit (Françoise, 2021) for composing interactive machine learning workflows and interfaces may be used to create early prototypes; and **(iii)** Explainable AI: to help domain experts understand what factors are influencing model predictions and their associated uncertainties, and why certain outcomes were generated.

Note that data about the different wheat and flour quality tests and a **preliminary learning model** implementation in Python are already available (from the EVAGRAIN project).

5. Expected results

The selected student will perform the following tasks. The following steps are given for guidance and may be re-defined with the selected candidate and according to findings from meetings with stakeholders:

- a literature review on collaborative visualization for AI/ML, explanation methods for Bayesian networks, and expert knowledge elicitation interfaces.
- follow a user-centered design approach to develop an online collaborative system that provides:
  - a graphical representation of the learned model (node-link diagram, an adjacency matrix,…) that avoids clutter;
  - a dashboard containing a set of indicators such as entropy, confidence, precision, robustness, *etc,* allowing users 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 easy comparisons between the different versions, in order to help experts reach consensus with regards to the final chosen model.
- evaluate the developed prototype with experts from agronomy & food technology domains.

In order to attain the objectives of CoCo-IML listed above, the project relies upon three intermediate milestones: **(a) Collaborative model co-construction framework:** design a collaborative visualization framework for model editing and exploration, and an open software prototype that implements it. The challenge is to determine what aspects of the AI can or should be co-designed, and to provide an online framework that caters for multiple types of experts, including non-technical user groups; **(b) Visualization of model uncertainty and evolution:** leverage the history of collective and individual model building and explorations (e.g., through log files). The challenge will be to improve awareness in collaborative work by providing a concise visualization of past models, their quality and evolution/differences; **(c) Validation Studies**: evaluate the developed prototypes. For instance, we could compare the outcome (decisions) of the co-construction process with and without online collaboration. The challenge will be to design controlled experiments that take into account the different components of the interactive ML system.

6. PhD Calendar

[0-6 months] Literature review on collaborative visualization for AI/ML, explanation methods for Bayesian networks, and expert knowledge elicitation interfaces.
[6-9 months] Participatory design workshops with end-users to gather use case scenarios and design requirements.
[9-12 months] Design of an online collaborative system taking into account user requirements and findings from related work. A research publication is expected given the novel architecture.
[12-18 months] Implementation of a prototype system including the knowledge elicitation user interface, and evaluation in smaller controlled studies. A research publication is expected given the novel insights the evaluation will provide.
[18-24 months] Design & Implementation of model comparison widgets.
[24-30 months] Framework evaluation with expert users to assess if the tool meets their needs. Note that intermediate evaluations will also take place in the prototype development phase (stage 3--5) to make sure basic functionalities meet user needs; and to get early feedback. A research publication is expected given the novel insights of combining human and machine intelligence to solve real world problems.
[30-36 months] Writing of thesis manuscript.

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 PhD: 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.

7. Candidate

Computer science background (Masters) or related disciplines. Required skills include: web development; programming (JavaScript, Python, other); user-centered design; information visualization. Interest in working with real-world data and applications. Knowledge in machine learning is not required but experience is a plus.

This PhD topic is participating to the Université Paris-Saclay EU COFUND DeMythif.AI program : https://www.dataia.eu/actualites/cofund-demythifai-appel-sujets-de-these. It is reserved to international students who have spent less than 12 months in France in the last 3 years. The candidates will be evaluated by a jury who will select 15 PhD to start in fall 2024. The successful candidates will be fully funded for 3 years, have access to specific scientific and non-scientific training, and be fully part of the Université Paris-Saclay AI community.

To apply, please send your CV and a motivation letter to nadia.boukhelifa@inrae.fr and anastasia.bezerianos@universite-paris-saclay.fr

8. Work environment

Internal Collaborators: The supervising team is in the larger Paris-Saclay area, divided between three labs INRAe/MIA-Paris-Saclay; LISN/ILDA and UMR Sayfood; **INRAe/MIA-Paris-Saclay**: Nadia Boukhelifa; **LISN**: Anastasia Bezerianos (HDR) from the ILDA team, **INRAe/Sayfood**: Sophie Berland.
External collaborators: **INRAe/STLO:** Mélanie Münch; **INRAe/BIA**: Kamal Kansou; **INRAe/Univ. Bordeaux**: Cédric Baudrit.

References

- **Boukhelifa**, N., **Bezerianos**, A., Isenberg, T., & Fekete, J. D. (2012). Evaluating sketchiness as a visual variable for the depiction of qualitative uncertainty. IEEE Transactions on Visualization and Computer Graphics, 18(12), 2769-2778.
- **Boukhelifa**, N., Cancino, W., **Bezerianos**, A., & Lutton, E. (2013). Evolutionary visual exploration: evaluation with expert users. In Computer Graphics Forum (Vol. 32, No. 3pt1, pp. 31-40). Oxford, UK: Blackwell Publishing Ltd.
- **Boukhelifa**, N., Perrin, M. E., Huron, S., & Eagan, J. (2017, May). How data workers cope with uncertainty: A task characterisation study. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (pp. 3645-3656).
- **Boukhelifa**, N., **Bezerianos**, et al., (2019). An exploratory study on visual exploration of model simulations by multiple types of experts. In Proc. of the 2019 CHI Conf. on Human Factors in Computing Systems (pp. 1-14).
- **Boukhelifa**, N., **Bezerianos**, A., Chang, R., Collins, C., Drucker, S., Endert, A., ... & Sedlmair, M. (2020). Challenges in evaluating interactive visual machine learning systems. IEEE Computer Graphics and Applications, 40(6), 88-96.
- Chakhchoukh,M., **Boukhelifa, N.**, **Bezerianos, A.** (2022). Understanding how In-Visualization Provenance can support Trade-off analysis. In IEEE Transactions on Visualization and Computer Graphics (TVCG), 2022, 16 pages.
- Chen, J., Zhang, R., Dong, X., Lin, L., Zhu, Y., He, J., ... & Chen, F. (2019). shinyBN: an online application for interactive Bayesian network inference and visualization. BMC bioinformatics, 20(1), 1-5.
- **Françoise**, J., **Caramiaux**, B., & Sanchez, T. (2021). Marcelle: composing interactive machine learning workflows and interfaces. In The 34th Annual ACM Symposium on User Interface Software and Technology (pp. 39-53).
- Gergle, D., Kraut, R. E., & Fussell, S. R. (2013). Using visual information for grounding and awareness in collaborative tasks. Human–Computer Interaction, 28(1), 1-39.
- Heer, J., & Agrawala, M. (2008). Design considerations for collaborative visual analytics. Information visualization, 7(1).
- Isenberg, P., Carpendale, S., **Bezerianos**, A., Henry, N., & Fekete, J. D. (2009). Coconuttrix: Collaborative retrofitting for information visualization. IEEE Computer Graphics and Applications, 29(5), 44-57.
- Isenberg, P., Elmqvist, N., Scholtz, J., Cernea, D., Ma, K. L., & Hagen, H. (2011). Collaborative visualization: Definition, challenges, and research agenda. Information Visualization, 10(4), 310-326.
- Jiang, L., Liu, S., & Chen, C. (2019). Recent research advances on interactive machine learning. Journal of Visualization, 22, 401-417.
- Kerrigan, D., Hullman, J., & Bertini, E. (2021). A survey of domain knowledge elicitation in applied machine learning. Multimodal Technologies and Interaction..
- Lacave, C., & Díez, F. J. (2002). A review of explanation methods for Bayesian networks. The Knowledge Engineering Review, 17(2), 107-127.
- Michiels, M., Larranaga, P., & Bielza, C. (2021). BayeSuites: An open web framework for massive Bayesian networks focused on neuroscience. Neurocomputing..
- **Munch**, M., Buche, P., Dervaux, S., Dibie, J., Ibanescu, L., Manfredotti, C., Wuillemin, P-H., Angellier-Coussy, H. (2022). Combining ontology and probabilistic models for the design of bio-based product transformation processes. Expert Systems with Applications, 203, 117406.
- Robert, S., et al., (2016). Reasoning under uncertainty: Towards collaborative interactive machine learning. Machine Learning for Health Informatics: State-of-the-Art and Future Challenges, 357-376.
- Stark, et al., (2012). Mixed-initiative data mining with Bayesian networks. In 2012 IEEE Int. Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (pp. 107-110). IEEE.
- Thomas, J. J. (2005). Illuminating the path:[the research and development agenda for visual analytics]. IEEE Computer Society.
- Vogogias, A., et al., (2018). BayesPiles: Visualisation support for bayesian network structure learning. ACM Transactions on Intelligent Systems and Technology.