Vázquez-Ingelmo, A., García-Holgado, A., García-Peñalvo, F. J., Andrés-Fraile, E., Pérez-Sánchez, P., Antúnez-Muiños, P., Sánchez-Puente, A., Vicente-Palacios, V., Dorado-Díaz, P. I., Cruz-González, I., & Sánchez, P. L. (2023). Flexible Heuristics for Supporting RecommendationsWithin an AI Platform Aimed at Non-expert Users. In C. S. González-González, B. Fernández-Manjón, F. Li, F. J. García-Peñalvo, F. Sciarrone, M. Spaniol, A. García-Holgado, M. Area-Moreira, M. Hemmje, & T. Hao (Eds.), Learning Technologies and Systems. 21st International Conference on Web-Based Learning, ICWL 2022 and 7th International Symposium on Emerging Technologies for Education, SETE 2022. Tenerife, Spain, November 21–23, 2022. Revised Selected Papers (pp. 333–338). Springer Nature. https://doi.org/10.1007/978-3-031-33023-0_30

Flexible Heuristics for Supporting Recommendations Within an AI Platform Aimed at Non-expert Users

Andrea Vázquez-Ingelmo[1](✉), Alicia García-Holgado[1], Francisco José García-Peñalvo[1], Esther Andrés-Fraile[1], Pablo Pérez-Sánchez[2], Pablo Antúnez-Muiños[3], Antonio Sánchez-Puente[4], Víctor Vicente-Palacios[5], Pedro Ignacio Dorado-Díaz[2], Ignacio Cruz-González[3], and Pedro Luis Sánchez[3]

[1] GRIAL Research Group, Computer Science Department, Universidad de Salamanca, Salamanca, Spain
{andreavazquez,aliciagh,esther.andres}@usal.es

[2] Biomedical Research Institute of Salamanca (IBSAL), Salamanca, Spain

[3] University Hospital of Salamanca, CIBERCV and Biomedical Research Institute of Salamanca (IBSAL), Salamanca, Spain
pedrolsanchez@me.com

[4] University Hospital of Salamanca and CIBERCV, Salamanca, Spain

[5] Philips Ibérica, Madrid, Spain

Abstract. The use of Machine Learning (ML) to resolve complex tasks has become popular in several contexts. While these approaches are very effective and have many related benefits, they are still very tricky for the general audience. In this sense, expert knowledge is crucial to apply ML algorithms properly and to avoid potential issues. However, in some situations, it is not possible to rely on experts to guide the development of ML pipelines. To tackle this issue, we present an approach to provide customized heuristics and recommendations through a graphical platform to build ML pipelines, namely KoopaML, focused on the medical domain. With this approach, we aim not only at providing an easy way to apply ML for non-expert users, but also at providing a learning experience for them to understand how these methods work.

Keywords: Information system · Medical data management · Medical imaging management · Artificial Intelligence · Health platform · HCI

1 Introduction

Machine Learning (ML) has become a powerful method to tackle complex problems in different contexts. These algorithms ease the analysis of great quantities of data to discover hidden patterns, reach new insights, and even predict events.

Some data-intensive contexts, like the health domain, benefit from developing ML pipelines for their data, to support time- and resource- consuming tasks, such as diagnoses, disease detection, segmentation, assessment of organ functions, etc. [1–3].

However, applying these algorithms is not straightforward. There are some algorithms that work better under some specific circumstances, or with datasets that have certain characteristics [4–8]. Otherwise, if ML algorithms are applied without understanding the process, the outputs of the trained models could lead to wrong conclusions, discrimination, and other hazardous issues [9–11].

While health professionals have a deep understanding of the input data, they could lack skills related to theoretical and practical foundations of ML. In this context, it is necessary to provide novice, lay, and non-expert users with tools that alleviate the learning curve of ML. This way, non-expert users can benefit from the application of ML without risking the quality of their models.

In this work, we present an approach to customize heuristics and recommendations for assisting novice users in the development of ML pipelines. This approach is integrated in a graphical platform (KoopaML) [12] that allow the instantiation and design of ML pipelines through visual means.

The goal of integrating the management of heuristics in KoopaML is to assist users in the development of their pipelines while providing a learning experience related to the suitability of certain algorithms given the input data, or the problem to solve.

The rest of this paper is structured as follows. Section 2 provides an overview of KoopaML's architecture. Section 3 describes the heuristics management module. Finally, Sect. 4 discusses the approach and presents the conclusions of this work.

2 Architecture

Due to the constant improvement and evolution of ML approaches, it is necessary to rely on a flexible architecture. For this reason, KoopaML is based on different modules that communicate with each other through data streams [12].

These modules include the user management, the pipelines management, the tasks management, and the heuristics management.

Particularly, the heuristics management module is in charge of providing an interface for designing heuristics in the form of decision trees. These heuristics sets are stored persistently and can be interchanged to apply different sets of heuristics depending on the problem.

The defined heuristics are then used by the pipelines management module to yield recommendations given the current state of the workspace, as will be detailed in the next section.

3 Heuristics Management

As introduced in the previous section, the heuristics management module of KoopaML provides functionality to create, modify, delete, and apply different rules to obtain useful recommendations for novice users while designing ML pipelines.

Figure 1 shows the main interface for this module, where a list of the available heuristics of the system is displayed. Each available set of heuristics can be modified and deleted at any time, as well as marked as default (so it is applied to every new pipeline).

Fig. 1. Available heuristics interface. Contents in Spanish.

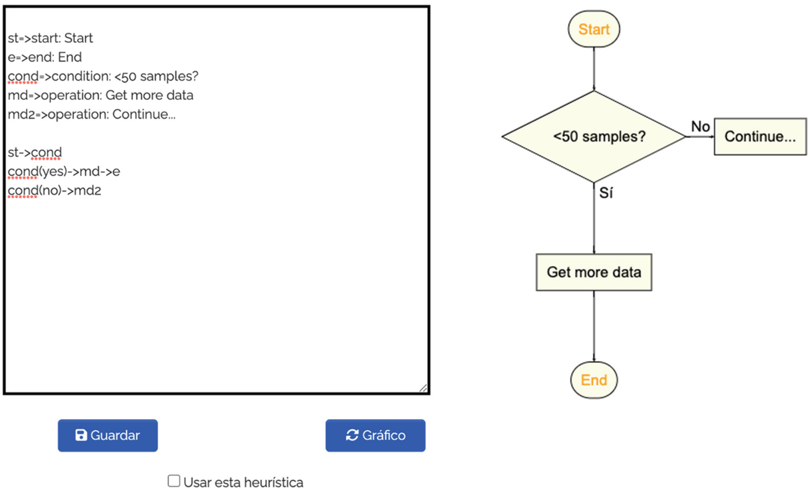

Fig. 2. Definition of a new heuristic. The interface is composed by a text field in which the heuristics are codified (left) and a canvas that shows the generated tree from the DSL (right).

The Domain Specific Language (DSL) provided by the flowchart.js (https://github.com/adrai/flowchart.js) library is then employed to define the rules and conditions of the heuristic's decision tree (Fig. 2). This library allows textual and graphical representation of flow charts, which provides a fine-grained manipulation of heuristics and rule-based recommendation.

The heuristics are then processed in the backend to restructure them as a nested dictionary that can be easily traversed programmatically. This dictionary is stored and finally employed by the workspace to yield the recommendations.

After defining the heuristics and starting a new project, the process carried out in the workspace is as follows:

1. Every time the workspace is initialized or modified, a request is sent to the backend containing the current state of the workspace (number and type of the nodes, dataset characteristics, etc.)
2. The backend retrieves the default heuristic and goes through the dictionary using the current state. In this step, each condition is read and applied using the available data from the workspace.
3. Whenever a condition reaches a leaf node, the process is stopped, and the recommendation is sent back to the client.
4. The interface shows the text obtained after applying the rules of the heuristic.

This process is displayed in Figs. 4 and 5. In Fig. 4, the workspace is empty, so a non-expert user could have some difficulties knowing where to start. In this case, the default heuristic guides the user and recommends them to insert a "Data import" node to upload their data.

Once the requested node has been included, the process described above is fired again, but, as the state of the workspace has changed and now it includes a "Data import" node, the recommendation text differs from Fig. 3. At this point, the system recommends the user to add a node to deal with missing data before training any model (Fig. 4).

Fig. 3. Recommendation yielded in an empty workspace: "Insert the first node «Data import» and select the data file that you want to use". Contents in Spanish.

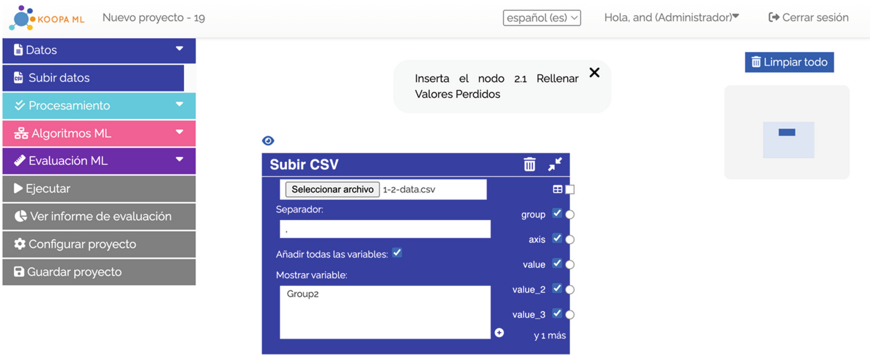

Fig. 4. Recommendation yielded after including a "Data import" node: "Insert the node «Fill missing values»". Contents in Spanish.

4 Discussion and Conclusions

This work presents a flexible approach based on heuristics for guiding non-expert users in the development of ML pipelines without the necessity of having programming skills.

The heuristics can be dynamically defined through a graphical interface using a DSL, which enables the possibility of evolving the rules based on new evidence, or even applying different sets of rules depending on the type of user.

Using tangible heuristics has been pointed out in previous works. Some authors, for example, have noticed that heuristics can lead to greater insights and greater engagement of the users in the process of applying ML [13, 14]. In addition, some works remark the importance of using human-readable rules when selecting AI algorithms in certain contexts [15].

Another potential benefit of using these kinds of recommendations in KoopaML is the educational experience offered to the users, which can learn from the heuristics and trace the explanations related to the recommendations yielded by the system, providing more transparency to the recommendation process.

Future research will involve the evaluation of the heuristics management module and their usefulness with non-expert users.

Acknowledgements. This research was partially funded by Ministry of Science and Innovation through the AVisSA project grant number (PID2020-118345RB-I00). This work was also supported by national (PI14/00695, PIE14/00066, PI17/00145, DTS19/00098, PI19/00658, PI19/00656 Institute of Health Carlos III, Spanish Ministry of Economy and Competitiveness and co-funded by ERDF/ESF, "Investing in your future") and community (GRS 2033/A/19, GRS 2030/A/19, GRS 2031/A/19, GRS 2032/A/19, SACYL, Junta Castilla y León) competitive grants.

References

1. Litjens, G., et al.: A survey on deep learning in medical image analysis. Med. Image Anal. **42**, 60–88 (2017)
2. González Izard, S., Sánchez Torres, R., Alonso Plaza, Ó., Juanes Méndez, J.A., García-Peñalvo, F.J.: Nextmed: Automatic Imaging Segmentation, 3D Reconstruction, and 3D Model Visualization Platform Using Augmented and Virtual Reality. Sensors (Basel) **20**, 2962 (2020)
3. Izard, S.G., Juanes, J.A., García Peñalvo, F.J., Estella, J.M.G., Ledesma, M.J.S., Ruisoto, P.: Virtual reality as an educational and training tool for medicine. J. Med. Syst. **42**(3), 1–5 (2018). https://doi.org/10.1007/s10916-018-0900-2
4. Rivolli, A., Garcia, L.P.F., Soares, C., Vanschoren, J., de Carvalho, A.C.P.L.F.: Meta-features for meta-learning. Knowledge-Based Systems **240**, 108101 (2022)
5. Vanschoren, J.: Meta-learning. Automated machine learning, pp. 35–61. Springer, Cham (2019)
6. Taratukhin, O., Muravyov, S.: Meta-learning based feature selection for clustering. In: IDEAL 2021: Intelligent Data Engineering and Automated Learning – IDEAL 2021, pp. 548-559. Springer International Publishing (Year)
7. Nayak, A., Božić, B., Longo, L.: An ontological approach for recommending a feature selection algorithm. In: ICWE 2022: Web Engineering, pp. 300–314. Springer International Publishing (Year)
8. Vilalta, R., Drissi, Y.: A perspective view and survey of meta-learning. Artif. Intell. Rev. **18**, 77–95 (2002)
9. Weyerer, J.C., Langer, P.F.: Garbage in, garbage out: The vicious cycle of ai-based discrimination in the public sector. In: Proceedings of the 20th Annual International Conference on Digital Government Research, pp. 509–511 (Year)
10. Ferrer, X., van Nuenen, T., Such, J.M., Coté, M., Criado, N.: Bias and discrimination in AI: a cross-disciplinary perspective. IEEE Technol. Soc. Mag. **40**, 72–80 (2021)
11. Hoffman, S.: The emerging hazard of ai-related health care discrimination. Hastings Cent. Rep. **51**, 8–9 (2021)
12. García-Peñalvo, F.J., et al.: KoopaML: a graphical platform for building machine learning pipelines adapted to health professionals. Int. J. Interactive Multimedia and Artificial Intelligence (In Press)
13. Reilly, D., Taylor, M., Fergus, P., Chalmers, C., Thompson, S.: The categorical data conundrum: heuristics for classification problems—a case study on domestic fire injuries. IEEE Access **10**, 70113–70125 (2022)
14. Lu, Z., Yin, M.: Human reliance on machine learning models when performance feedback is limited: heuristics and risks. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, pp. Article 78. Association for Computing Machinery, Yokohama, Japan (2021)
15. Golshanrad, P., Rahmani, H., Karimian, B., Karimkhani, F., Weiss, G.: MEGA: Predicting the best classifier combination using meta-learning and a genetic algorithm. Intelligent Data Analysis **25**, 1547–1563 (2021)