Use of a i*extension for Machine Learning: a real case study

Jose M. Barrera[1,2], A. Reina-Reina[1,2], Sandra García-Ponsoda[1,2] and Juan Trujillo[1,2]

[1]*Lucentia dpt. of Software and Computing Systems, University of Alicante, Carretera San Vicente del Raspeig s/n, Alicante 03690, Spain*

[2]*Lucentia Lab, Av. Pintor Pérez Gil, 16, 03540 Alicante*

Abstract

Capturing requirements in machine learning projects is a challenging task. It requires domain knowledge as well as experience in the machine learning field. The i* framework is a popular high abstraction-layer requirements capturing tool. However, the use of i* directly in the machine learning field (ML) is unfeasible due to it cannot capture all the restrictions and relationships of ML elements. In previous works we have extended i* to better capture machine learning requirements. In this paper, we apply the i* for machine learning extension to a real machine learning case study, in the context of a project focused on the diagnosis and treatment of Attention-Deficit/Hyperactivity Disorder (ADHD). The results show that the use of the i* for machine learning extension provides insights about the correct path to follow, aiding in the definition and selection of machine learning solutions that better fulfill the project requirements. Moreover, it facilitates faster development of the machine learning solution in a more structured way, avoiding errors and making the application of i* an effective tool for managing machine learning requirements.

Keywords

machine learning, iStar, requirements engineering, conceptual modelling, methodology

1. Introduction

Requirements engineering (RE) in machine learning (ML) field is still in development. Although some academic proposals exist, in practice machine learning projects are being carried out without a clear approach to manage the requirements phase. This implies that capturing requirements is still based on two main elements: (i) the domain-knowledge of the requirements engineer; (ii) the expertise or know-how of the ML engineer, who has to provide the suitable initial steps to start with the project.

Since there is no methodology established, only generic requirements tools can be used. In this sense, the options are either general software engineering requirements approaches, which are inadequate since they focus on different constructs such as classes and layers, or high-level abstraction frameworks such as the i* framework. The i* framework is high-abstraction tool for capturing requirements that can be applied to any field. However, the flexibility comes

iStar'22: The 15th International i Workshop, October 17th, 2022, Hyderabad, India*

jm.barrera@ua.es (J. M. Barrera); alejandro.reina@ua.es (A. Reina-Reina); sandra.gp@ua.es (Sandra García-Ponsoda); jtrujillo@dlsi.ua.es (J. Trujillo)

0000-0002-3359-2685 (J. M. Barrera); 0000-0001-9949-2735 (A. Reina-Reina); 0000-0002-0739-6680 (Sandra García-Ponsoda); 0000-0003-0139-6724 (J. Trujillo)

at a cost: i* cannot capture all the relationships and rules of a specific field. Thus, it must be tailored for each specific field, under the form of i* extensions, with the aim of capturing the aforementioned rules.

Our extension of i* has been developed following the steps provided in PRISE [1] and a set of guidelines provided in [2].

In our previous work [3], we proposed a i* extension for ML. Moreover, we presented the 8 questions that guide the ML RE process, and we provided the metamodel that captures the interactions of ML field. In our previous work we provided the theoretical metamodel and methodology. Now, in this paper, we apply our proposal to a real case study, showing its application and impact on the results obtained. More specifically, in this paper we (i) have refined our model by adding a new question (question 9) associated with quality aspects of ML model; (ii) provide an example of use of our ML extension in a real case study focused on the research of Attention-Deficit/Hyperactivity Disorder (ADHD); (iii) show how the use of the i* ML extension aids in the selection of machine learning solutions (configurations) that better fulfill the project requirements.

2. Related work

Goal-oriented requirements engineering (GORE) has gained great interest from the community in recent years due to its ability to capture different aspects to analyze requirements, conceptualization to elicit, model or capturing alternatives and conflicts [4]. In this sense, i* framework allows GORE methodology providing a high abstraction requirement capturing tool that allows to the users focusing in goals. Various/multiples versions from i* has been released [5] to refine the metamodel until the actual version 2.0 [6]. More extensions of i* have been developed through years, as we can see in [7].

In a recent survey, authors highlight the requirements in ML development inherent requirements elicitation, iterative process of machine learning or uncertainty of data[8]. Although other solutions has been proposed in literature, there few aspects has not been covered yet. For instance, authors in [9][10] focus their efforts in non-functional requirements (NFR), while our proposal covers both non-functional requirements and functional requirements (FR). In [11] authors proposed GORE-MLops for ML requirements capturing. However, this tools is still in a too high-abstraction layer, which makes more difficult to capture ML constraints. Our proposal is based in a lower abstraction-level, and it captures the relationships and restrictions of ML elements. Finally, in [12], authors proposed patterns for solving the ML requirements capturing. However, their proposal can lead to errors when planning goals (for example, a f-measure can be used as a clustering metric). Our proposal is less prone to produce errors, due to we have established more specific relationships between more specific elements (for example, a classification goal is related only with classification metrics and a classification tasks). Consequently, it do not rely so heavily in ML knowledge.

To sum up, in this paper we provide a real use case using a more-directed less-free ML i* extension, where we demonstrate how it can helps to capturing requirements and selecting the algorithms that satisfy the organization requirements.

3. Attention-deficit/hyperactivity disorder: a case study

ADHD is a chronic neuropsychiatric disorder that appears during childhood and is characterised by inattention, social impairment, age-inappropriate activity levels and/or hyperactivity and impulsivity [13]. Due to there is lack of a standard test to detect this disorder objectively, each physician must rely on his or her professional experience, which means that the same patient may or may not be diagnosed with ADHD, depending on the physician he or she sees [14]. Consequently, less subjective techniques such as Electroencephalography are used to help diagnose in an even-handed manner. Thus, we are trying corroborate with empirical data the diagnosis of the patients. That could help to reduce the subjectivity in ADHD diagnosis.

The electroencephalogram (EEG) contains as many signals (temporal series) as electrodes we have placed on the scalp. As brain signals have low voltage, the small electrodes have to be powerful enough to gather this brain activity. By contrast, they also collect unwanted information, called noise, Therefore, a pre-processing work of the signal is very important to obtain only the necessary information.

The EEG that is being used in this research provides 128 signals/second in 19 channels. The data for this experiment can be found here [1].

4. Use of i* ML extension of real case study

The key point of our proposal is the relation between the specifications of *MLGoal—MLTask—Indicator*. Our proposal filters non-valid configurations: the degree of compliance of a classification task can be only measured with a classification metric, which is pursuing only classification goals. Moreover, the *MLQualityAspects* are gathered from the business need. Consequently, the *MLTasks* that can help to achieve the *MLGoals* are detected from the very beginning, and the *MLTasks* that are not suitable (according to business need) are discarded as possible algorithms for the ML model. This ensures only valid ML algorithms for the task at hand are selected, while margin for mistakes is greatly reduced. Furthermore, the use of enumerated classes ensures that only a correct indicator to the ML task at hand can be selected.

In our proposal, 9 questions guide the process of RE [3]. Through the answers of these 9 questions, we can capture the ML requirements. These questions are:

1. **Which problem must be solved?:** The goal of the project is to discriminate if a patient has ADHD or not. Moreover, this goal can be divided into two different classification goals: to classify if the patient has Attention Disorder (AD) or not; and to classify if the patient has Hyperactivity Disorder (HD) or not. The answer to this question provides both *ClassificationGoals* (*Detection of AD* and *Detection of HD*) in our specific i* goal scheme.
2. **In which time frame should we have the answer?:** Due to the nature of the disorder, the prediction is not subject to deadlines. It must be done as soon as possible, to provide the suitable treatment, but there is no a time frame associated to it. Although the answer to this question could have an impact over *Dataset* element, in our case study it has no effect, due to data there is no specification of time frame to our i* goal scheme.

[1]https://ieee-dataport.org/open-access/eeg-data-adhd-control-children

3. **Which data do you think is important for the model?:** With the aim of providing a empirical-based prediction, only the data provided by the EEG device will be used. Although the researchers have considered to use more data related with other aspects (age, social background, etc), it is outside the scope of this project and it will be developed in future projects. Data provided by the EEG device is only numerical. The answer to this question will affect the *Dataset* element (*EEG signals*).
4. **Which granularity of data is available to tackle the problem?:** The EEG device works with a frequency of 128 Hz. That implies 128 samples/second. The granularity has been considered enough for the problem. The answer to this question establishes the value of parameter *temporalResolution* of the *Dataset* element (*EEG signals*).
5. **Which metrics and hit rate would be valid to consider the project as a success?:** Due to the nature of the research field (e-health), the ML model must provide four metrics: accuracy, sensitivity, recall and f1-score. The organization has not established threshold metrics associated to the project success. The answer to this question establishes the *ClassificationIndicators* elements (*Accuracy, Precision, Recall, F1-Score*) for both *ClassificationTasks* (*Detection of AD* and *Detection of HD*)
6. **Is the explainability of the model necessary?:** Explainability is absolutely paramount. The presented ML models must provide the explainability and interpretability [15]. The answer to this question specifies one *MLQualityAspects* element (*Explainability*). Moreover, this *MLQualityAspects* will filter the *ClassificationTasks*, with the aim of focusing on ML algorithms that fulfil the desired quality aspects.
7. **Is it likely that the data distribution will change?:** No, data distribution of AHDH patients is not bound to change. Consequently, no drift detection is required. The answer to this questions would specify if *DriftDetection* (another *MLQualityAspects*) is required. However, it has no effect in our use case, due to it is not.
8. **Is there any bias in the data? Is data fair from an equity point of view?:** There is a strong bias in data. Although no special task must be done for dealing with that bias, other metrics than accuracy must be used (precision, recall and f1-score). The answer to this questions would specify if *Equity* (another *MLQualityAspects*) is required. However, it has no effect in our use case, due to it is not.
9. **Is another ML quality aspect required?:** It would be advisable to build models with scalabity MLQualityAspect. However, scalability is a secondary quality aspect while explainability is mandatory. The answer to this question specifies one *MLQualityAspects* element (*Scalability*). Moreover, this *MLQualityAspects* will filter the *ClassificationTasks*, with the aim of focusing on ML algorithms that fulfil the desired quality aspects. Due to *Scalability* is a secondary *MLQualityAspect*, we have focused on *ClassificationTasks* that provides *Explainability*, despite the effects on *Scalability*.

As we can see in figure 1, we have a main high-abstraction goal focused in ADHD detection. Moreover, this goal can be divided into two specific classification goals: a classification of AD disorder, and a classification of a HD disorder. Furthermore, we will focus each classification in four specific classification metrics: accuracy, precision, recall and f1-score.

Due to the organization requirements, explainability is the foremost $MLQualityAspect$ that must be taken into account. On the one hand, we have selected as potential initial candidates all

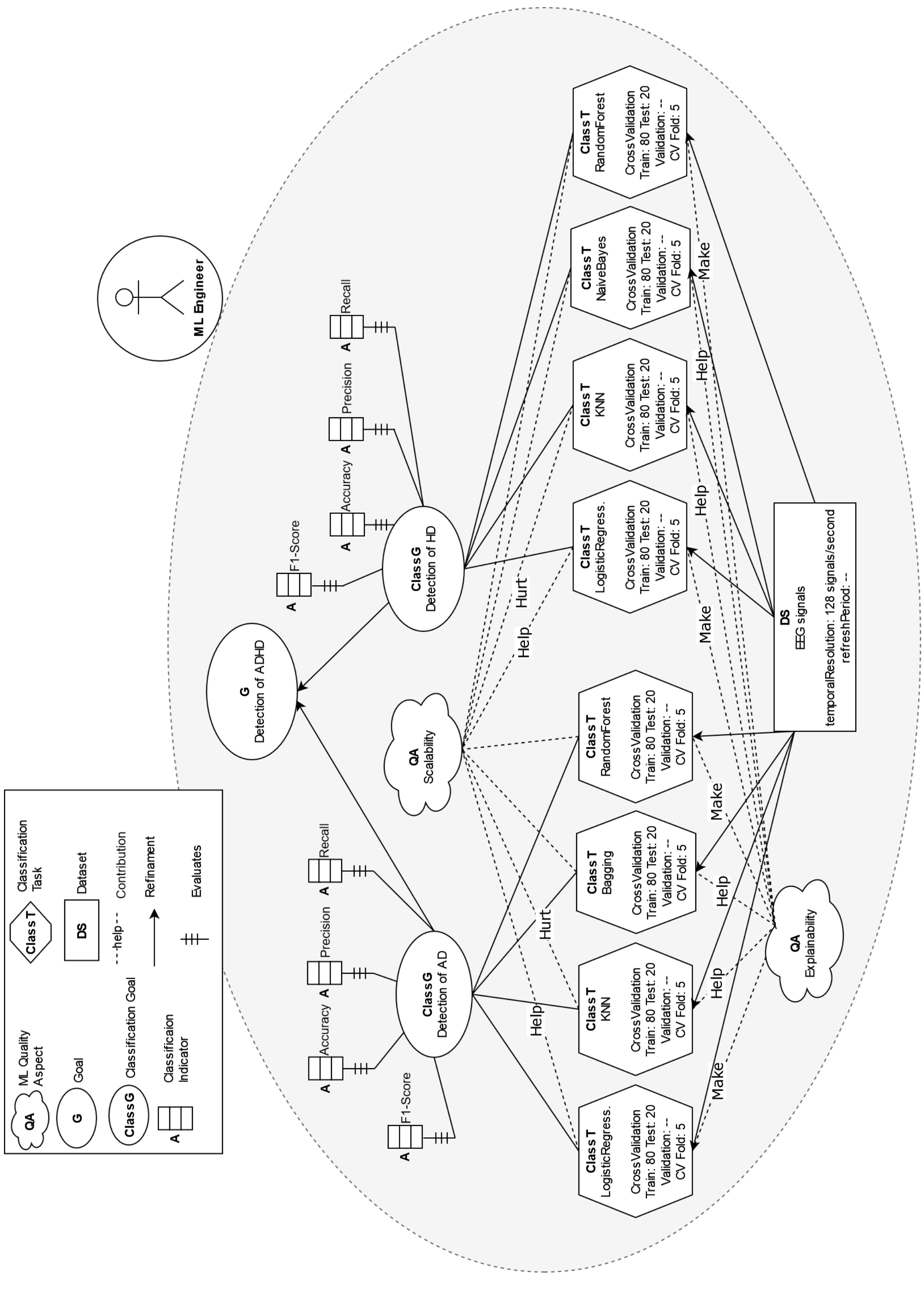

Figure 1: Application of ML i* extension to case study.

algorithms that have a contribution of ”make” and ”help” with *Explainability*, according to their characteristics in [16] and [17]. Thus, these four algorithms will conform the *ClassificationTasks* for both goals: LogisticRegression, KNN, Bagging and RandomForest.

On the other hand, we have taken into account Scalability as another quality aspect required by organization. However, due it has been considered secondary. As a result of that, we have represented the relationship with the ”explainable” algorithms. As we can see in figure 1, LogisticRegression is an algorithm that provides scalability (”helps” scalability), while KNN has a worse performance (”hurts” scalability) when data is increased.

We must highlight that other algorithms such as SVM, Adaboost or XGBoost will not fulfil organization requirements. Thus, it would be a futile work to train those ML models in pursue of good values of the forementioned metrics. Thus, following this i* extension we are focusing our efforts in ML models that solve our task. Finally, after following our proposal, the ML engineers have decided to put efforts in creating a new algorithm focused in explainability, since they have detected a lack of explainable algorithms for similar real cases.

5. Conclusions

In this paper, we have presented an application of i* for ML extension in the context of a project focused on the diagnosis and treatment of Attention-Deficit/Hyperactivity Disorder (ADHD). Moreover, we have refined our previous work by adding a new question focused on the quality aspects of ML projects.

Compared to common practice, by following the our i* for ML extension we guide the user through a series of questions that allow us to capture in i* models the requirements of ML projects, providing a more systematic way to tackle the requirements phase. As a result, our approach helps to cover the gaps between requirements engineering (RE) and ML. Our approach helps the ML developer to translate high-level organization requirements into technical ML goals, which consider both the functional and non-functional requirements (NFR’s) of the project. Therefore, we ensure that the main aspects of the projects are covered, avoid invalid ML configurations, and aid in the selection of better ML solutions from the project requirements perspective.

As a part of our future work, we plan to carry out an empirical evaluation through a controlled experiment with two different groups. This will allow us to test and estimate the exact impact of the our approach in terms of time, errors, and cost compared to the current practice for ML projects.

References

[1] E. Gonçalves, J. Araujo, J. Castro, PRISE: A process to support iStar extensions, Journal of Systems and Software 168 (2020) 110649. doi:10.1016/J.JSS.2020.110649.

[2] E. Gonçalves, M. A. de Oliveira, I. Monteiro, J. Castro, J. Araújo, Understanding what is important in istar extension proposals: the viewpoint of researchers, Requirements Engineering 24 (2019) 55–84. doi:10.1007/s00766-018-0302-5.

[3] J. M. Barrera, A. Reina, A. Maté, J. C. Trujillo, Applying i* in conceptual modelling in machine learning, volume 2983, 2021, pp. 56–62. URL: http://ceur-ws.org.

[4] J. Horkoff, F. B. Aydemir, E. Cardoso, T. Li, A. Maté, E. Paja, M. Salnitri, L. Piras, J. Mylopoulos, P. Giorgini, Goal-oriented requirements engineering: an extended systematic mapping study, Requirements Engineering 24 (2019) 133–160. URL: https://link.springer.com/article/10.1007/s00766-017-0280-z. doi:10.1007/S00766-017-0280-Z/TABLES/8.

[5] E. Gonçalves, T. Heineck, J. Araújo, J. Castro, Catie: A catalogue of istar extensions, 2018. URL: http://istarextensions.cin.ufpe.br/catalogue/.

[6] F. Dalpiaz, X. Franch, J. Horkoff, istar 2.0 language guide, CoRR abs/1605.07767 (2016). URL: http://arxiv.org/abs/1605.07767. arXiv:1605.07767.

[7] E. Gonçalves, J. Castro, J. Araújo, T. Heineck, A systematic literature review of istar extensions, Journal of Systems and Software 137 (2018) 1–33. doi:10.1016/j.jss.2017.11.023.

[8] Z. Wan, X. Xia, D. Lo, G. C. Murphy, How does machine learning change software development practices?, IEEE Transactions on Software Engineering (2020) 1–1. doi:10.1109/TSE.2019.2937083.

[9] J. C. L.M., do Prado Leite, Enterprise, Business-Process and Information Systems Modeling, volume 387, Springer International Publishing, 2020. URL: http://link.springer.com/10.1007/978-3-030-49418-6. doi:10.1007/978-3-030-49418-6.

[10] J. Horkoff, Non-functional requirements for machine learning: Challenges and new directions, IEEE, 2019, pp. 386–391. doi:10.1109/RE.2019.00050.

[11] F. Ishikawa, Y. Matsuno, Evidence-driven requirements engineering for uncertainty of machine learning-based systems, IEEE, 2020, pp. 346–351. doi:10.1109/RE48521.2020.00046.

[12] S. Nalchigar, E. Yu, Y. Obeidi, S. Carbajales, J. Green, A. Chan, Solution patterns for machine learning, volume 11483 LNCS, Springer Verlag, 2019, pp. 627–642. URL: https://doi.org/10.1007/978-3-030-21290-2_39. doi:10.1007/978-3-030-21290-2_39.

[13] E. Cormier, Attention deficit/hyperactivity disorder: A review and update, Journal of Pediatric Nursing 23 (2008) 345–357. doi:10.1016/j.pedn.2008.01.003.

[14] S. P. Hinshaw, Attention deficit hyperactivity disorder (adhd): controversy, developmental mechanisms, and multiple levels of analysis, Annual review of clinical psychology 14 (2018) 291–316.

[15] R. Marcinkevičs, J. E. Vogt, Interpretability and explainability: A machine learning zoo mini-tour (2020). URL: http://arxiv.org/abs/2012.01805.

[16] M. Fatima, M. Pasha, Survey of machine learning algorithms for disease diagnostic, Journal of Intelligent Learning Systems and Applications 09 (2017) 1–16. doi:10.4236/jilsa.2017.91001.

[17] G. Stiglic, P. Kocbek, N. Fijacko, M. Zitnik, K. Verbert, L. Cilar, Interpretability of machine learning-based prediction models in healthcare, WIREs Data Mining and Knowledge Discovery 10 (2020) e1379. URL: https://onlinelibrary.wiley.com/doi/10.1002/widm.1379. doi:10.1002/widm.1379.