ICARE: the principles of Explainable AI in a Context-aware Recommendation APP *

Anna Dalla Vecchia[1,*,†], Barbara Oliboni[1,†] and Elisa Quintarelli[1,†]

[1]Dept. of Computer Science, University of Verona, Verona, Italy

Abstract

In this paper, we summarize our previous contribution to the research area of Explainable Recommender Systems in the healthcare domain, called ICARE (Intuitive Context-Aware Recommender with Explanations), which is a framework based on data-mining algorithms that can provide personalized recommendations with contextual and intuitive explanations. In particular, we consider the scenario related to physical activities to improve sleep quality, and we now describe how the system satisfies the four principles of Explainable Artificial Intelligence.

Keywords

Health Recommendation System (HRS), Context-Aware Recommendation Systems (CARS), Explainable Artificial Intelligence

1. Introduction

Collecting data on people's behaviors and habits has become easier since sensors and mobile devices are widespread and low-cost elements of modern daily life. Recommendation Systems are commonly used to exploit data, often collected through IoT devices, and support users in decision-making processes that cover different domains. The proposed suggestions are based on the users' past habits or their profiles.

Among these systems, Health Recommendation Systems (HRS) provide recommendations in the healthcare context [1]. Possible applications range from healthy lifestyle or balanced nutrition plans to healthcare information and medical treatment suggestions. Collecting data related to people's behaviors and well-being has also become easier thanks to wearable devices; indeed, many people own them and can monitor with simple apps their movements, record their sleep quality and heartbeats while also being surrounded by IoT devices embedded in common appliances in homes, offices and means of transport. This situation ensures a constant stream of new data that can be integrated with external data sources. It offers increasing opportunities for data analytics in the healthcare domain to produce useful insights. Moreover, collecting data can also help people gain more awareness of their physical health and improve their lifestyle. In addition, since wearable devices do not require that people manually provide data, the participants of health-related studies never stop supplying needed data [2].

A very important aspect to consider for improving the personalization of suggestions is contextual information, i.e., information about the specific scenario in which users are acting. The context can include weather conditions, social and spatio-temporal information. For example, the recommendation related to the physical activity to perform could be different based on weather conditions: it could be a run in the park on a sunny day and a workout in the gym on a rainy day. Context-Aware Recommendation Systems (CARS) provide recommendations taking into account also the contextual dimension [3].

The temporal aspect of data collected from sensors, as wearable devices, is very important and can be considered another contextual dimension for obtaining useful knowledge. Sensors collect information about events happening in succession, and thus (stream) data are temporal. When considering the temporal aspect, detecting events' periodicity or temporal correlations is possible and allows the prediction of future situations by improving the provided suggestion [4].

To be sure that users will accept the provided recommendations, an HRS should not only focus on efficiency and accuracy because clear explanations accompanying suggestions can help users understand and trust the system [5]. Providing clear, timely recommendations and intuitive and simple explanations is very important, especially in healthcare.

Explainable recommendations [6] are usually accompanied by understandable motivations that guarantee transparency and interoperability.

Explainability refers to both interpretability and fidelity to guarantee that explanations make sense for

Published in the Proceedings of the Workshops of the EDBT/ICDT 2024 Joint Conference (March 25-28, 2024), Paestum, Italy

*Summary of the bookchapter ICARE: An Intuitive Context-Aware Recommender with Explanations, in Advances in AI-Enhanced Paradigms and Applications in Healthcare, Springer, 2024.

*Corresponding author.
†These authors contributed equally.
anna.dallavecchia@univr.it (A. Dalla Vecchia); barbara.oliboni@univr.it (B. Oliboni); elisa.quintarelli@univr.it (E. Quintarelli)
0000-0001-7026-5205 (A. Dalla Vecchia); 0000-0003-4395-0893 (B. Oliboni); 0000-0001-6092-6831 (E. Quintarelli)

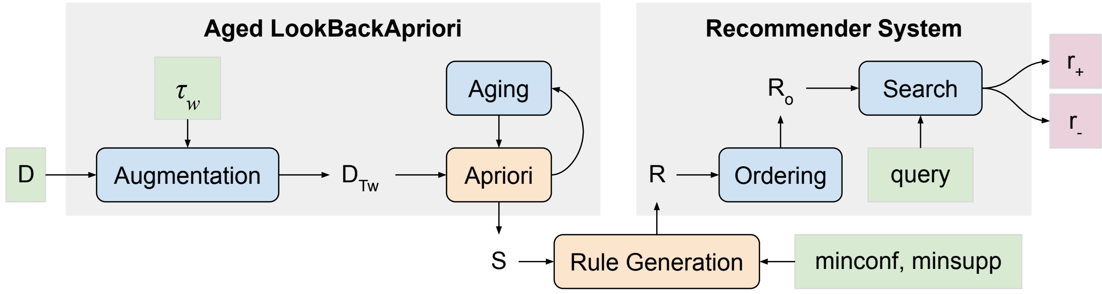

Figure 1: The ICARE Workflow

users and describe how the recommendation has been produced. [7].

In the Artificial Intelligence community, four principles that must be considered to guarantee high quality in explaining an intelligent system output have been proposed [8].

We consider the scenario related to physical activities to carry out to reach and follow a healthy lifestyle and improve sleep quality. To support the user in the choice of the next activities to perform, we propose the ICARE (Intuitive Context-Aware Recommender with Explanations) framework based on data mining algorithms for personalized recommendations. The supplied recommendations, in the form of totally ordered sequential rules, include the context and perform well in providing explanations [9].

This paper briefly summarizes ICARE and shows the system's behavior concerning the four principles of Explainable Artificial Intelligence.

2. ICARE architecture

In this Section, we describe the overall approach of ICARE, which is represented in Figure 1 and the ALBA (Aged LookBackApriori) algorithm used to infer totally ordered sequential rules that are then used to provide contextual recommendations [10].

ICARE needs as input a temporal dataset D (in our scenario, the log of physical activity levels and sleep scores collected by Fitbit), enriched with contextual information. We leverage the temporal dimension to capture the user's habits on some specific days, i.e., we distinguish whether each day is a weekday or part of the weekend. Whenever it is possible, we also integrate weather conditions to establish if the user's preferences are related to specific situations. The dataset D is fed into ALBA, which then considers a temporal window τ_w to construct an augmented data set D_{τ_w}. This augmented dataset is then fed to a version of the Apriori algorithm [11] that calls an aging mechanism at each iteration to calculate the support of items. The aging mechanism enhances the importance of recent patterns and decreases the importance of no more recent data. This process outputs a set of frequent sequences S, used to generate a set of totally ordered sequential rules R, formalized as implications $X \rightarrow Y$, where X and Y are two sets of ordered data items, such that $X \cap Y = \emptyset$, according to specific thresholds for confidence and support. Support is the frequency of the set $X \cup Y$ in the dataset, while confidence is the conditional probability of finding Y, having found X and is given by $sup(X \cup Y)/sup(X)$. Based on the confidence and completeness of rules, the recommender system orders R to produce a set of totally ordered sequential rules that can be queried to extract a positive recommendation R^+ and a negative one R^-.

Two possible rules mined by the ALBA Algorithm are the following:

$$r_1 : \{HA : 3, LA : 2\}_{-1} \wedge \{HA : 3, LA : 2\}_0 \rightarrow \{SL : 1\}_0$$

$$r_2 : \{Rainy, HA : 3\}_{-1} \wedge \{Good, HA : 1\}_0 \rightarrow \{SL : 1\}_0$$

r_1 states that if yesterday the user performed a high level of heavy physical activity ($HA : 3$) and a medium level of light activity ($LA : 2$), and today they perform the same activities, the resulting sleep score will have a low value ($SL : 1$). r_2 contains context information as well; yesterday it was a rainy day, and the user performed only a high level of heavy physical activity ($HA : 3$), today the weather conditions are good, and the user performed a low level of heavy physical activity ($HA : 3$).

3. The four principles of Explainable AI in ICARE

The increasing complexity of decision-making processes and intelligent systems has made them more challenging to understand. This has motivated the research community's attention regarding AI systems' transparency, ethics, and accountability. Explainable AI focuses on creating and deploying AI systems that explicitly explain

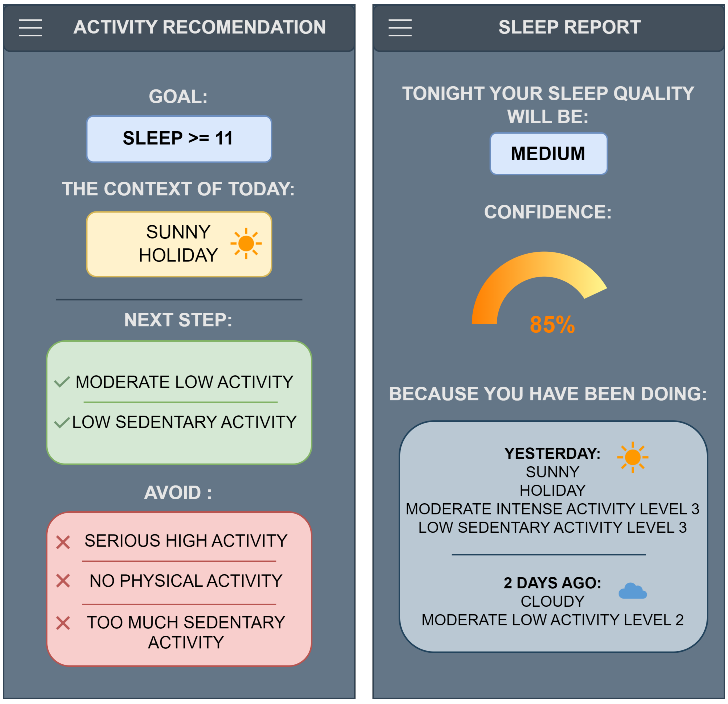

Figure 2: ICARE Positive and Negative Recommendations and ICARE Prediction and Explanation screenshots

their decision-making processes and the outcomes generated by machine learning algorithms.

Four principles for explainable AI have been identified [8] and are:

- *Explanation*: AI systems should be able to provide clear explanations for their actions
- *Meaningful*: explanations offered by AI systems must be comprehensible and relevant to humans, particularly those who are not experts in the field.
- *Explanation accuracy*: explanations need to be precise and accurate. Thus, it is possible to identify the most important variables/features in a decision-making process.
- *Knowledge limits*: AI systems should recognize their limitations and uncertainties, operating exclusively within their designed conditions.

Now, we will show how these principles are naturally integrated into ICARE, since the recommendation process is based on mined sequential rules that contain their explanation.

Explanation This principle asserts that a system can explain, provide evidence, and explicitly support every decision made. Our choice to use ALBA, an algorithm based on Apriori, offers the possibility of fulfilling this requirement because the form of mined rules directly explains the suggestion (i.e., the consequent of mined rules) in their antecedent. For example, consider the following two rules:

$$r_1 : \{Cold, Good, WD, HA : 1, LA : 2\}_0 \rightarrow \{SL : 1\}_0$$

$$r_2 : \{HA : 3, LA : 2\}_{-1} \wedge \{HA : 3, LA : 2\}_0 \rightarrow \{SL : 1\}_0$$

r_1 states that on a cold day with good weather conditions during the week (WD), if the user performs a low level of heavy physical activity ($HA : 1$) and a medium level of light activity ($LA : 2$), the resulting sleep score will have a low value ($SL : 1$).

r_2 states that independently from the context, if the user yesterday performed a high level of heavy physical activity ($HA : 3$) and a medium level of light activity ($LA : 2$), and today performs the same level of physical activity, the resulting sleep score will have a low value ($SL : 1$).

In Fig. 2, we show two screenshots of the ICARE mobile application with information provided to the user. ICARE is an Android app with a Python backend. The left screenshot shows the predicted sleep score and the best positive (green) and negative (red) recommendations. The right screenshot shows the contextual explanation (the antecedents of mined rules) of the sleep score, reported in the lower part *"Because you have been doing"*.

Meaningful The meaningful principle is satisfied when the final user understands the explanations provided by the system, which should be generated together with each recommendation. The development of the ICARE app has been performed by considering this principle; indeed, both positive and negative suggestions should help the user better understand what influences his/her sleep quality. In Fig. 2, the left-hand side screenshot reports positive and negative suggestions and the sleep quality explanation. In this example, the current context is *Sunny Holiday*, and the goal is a score greater than or equal to 11 for sleep quality. Recommendations for activities to perform are represented in part *"Next step"* on a green background, and recommendations for activities to avoid in part *"Avoid"* on a red background.

Explanation Accuracy The Explanation and Meaningful principles require a system to generate explanations that are comprehensible to humans. These principles do not impose that a system provides precise explanations. The Explanation Accuracy principle introduces a challenge in motivating accuracy for a system's explanations through metrics.

During the implementation of ICARE, we have collected data from 4 users, specifically for this study. For our approach's experimental evaluation, we verified that relevant features influencing sleep quality may differ for users. Thus, after a training phase, we have considered only relevant features for each user. We have performed interviews for our evaluation to understand if the mined insights are accurate.

For example, one user has sleep quality that is highly correlated to weather conditions, as shown by the following rules:

$$User1_1 : \{Rainy, WD, LA : 2, MA : 3\}_0 \rightarrow \{SL : 3\}_0$$

$$User1_2 : \{Sunny, Cold, WD, LA : 2, HA : 1\}_0 \rightarrow \{SL : 1\}_0$$

Another user has sleep quality correlated to the part of the week (i.e., weekend or weekday), as shown by the following rules:

$$User2_1 : \{WE, HA : 1, MA : 1\}_0 \rightarrow \{SL : 3\}$$

$$User2_1 : \{WD, HA : 1, LA : 3\}_0 \rightarrow \{SL : 1\}$$

Knowledge limits This principle requires identifying and pointing out cases the system could not manage. ICARE provides suggestions based on collected data, and in selecting rules to consider for producing recommendations, it considers the thresholds set for confidence and support. This means that the ALBA algorithm may not produce suggestions when it does not mine suitable rules, but it cannot be found in a situation out of its scope. When users do not wear their Fitbit constantly, the precision of rules decreases. In the future, we should add a module to delete outliers from data.

4. Related Work

Explainable Artificial Intelligence (AI) contributes to creating trustworthy in suggestions proposed by automated systems, and thus, it characterizes trust in AI systems. Trust is fundamental, especially in the medical domain.

A method to define desired properties is needed to characterize a good explanation from an AI system. The four principles of Explainable Artificial Intelligence have been proposed in [8] for defining the factors to consider for producing a suitable explanation and possibly evaluating its quality. The proposed principles cover different perspectives and consider the importance of explaining the outcome of the process, but also the style, the aim, and the recipient of the explanation itself. In [12], the four principles proposed in [8] have been applied to Biometrics and Facial Forensic Algorithms in order to consider *trust* and societal norm of AI systems and provide a foundation of explainability.

In [13], the authors consider problems related to explaining black-box algorithms and classify them concerning the notion of explanation. Starting from the description of a problem definition, a type of black box, and a preferred explanation, the survey should support the researcher in finding the proposals that best fit their requirements.

In [14], the authors provide an analytical review of the papers in the literature focusing on the explainability of artificial intelligence in the context of machine learning and deep learning. The work briefly describes the different terms used to indicate the understandability of a system and maps out the main challenges to deal with to fulfill the explainability issues.

In [15], the notion of explainable AI has been explored in the biomedical contexts for proposing a functional definition and a conceptual framework that can be used when considering explainable AI.

In this work, we show that data mining algorithms still represent an alternative that fits well for producing explainable recommendations.

5. Conclusions and future work

In this paper, we have described an Intuitive Context-Aware Recommender with Explanations (ICARE), which is a framework for collecting and enriching wearable device data with contextual information to produce relevant insights. Indeed, sequential rules correlating sequences of past events with a specified future goal are discovered by analyzing temporal logs. Finally, ICARE provides explainable recommendations using an intuitive application.

As for future work, we plan to extend the ICARE app to collect the user's feedback on the received predictions and recommendations. Moreover, we are applying the framework in other scenarios, particularly for maintaining glucose levels in the normal range, by suggesting how to organize physical activity and meals. Another possible domain is a recommendation system that suggests balanced meals based on what they recently ate, also considering where meals are consumed (e.g., home, school canteen, ...).

References

[1] A. Calero Valdez, M. Ziefle, K. Verbert, A. Felfernig, A. Holzinger, Recommender Systems for Health Informatics: State-of-the-Art and Future Perspectives, Springer International Publishing, Cham, 2016, pp. 391–414.

[2] W. A. Wendt T, Knaup-Gregori P, Decision support in medicine: a survey of problems of user acceptance, Stud Health Technol Inform. 77 (2020) 852–6.

[3] G. Adomavicius, B. Mobasher, F. Ricci, A. Tuzhilin, Context-aware recommender systems, AI Mag. 32 (2011) 67–80.

[4] A. Dalla Vecchia, N. Marastoni, S. Migliorini, B. Oliboni, E. Quintarelli, Mining totally ordered sequential rules to provide timely recommendations, in: Proceedings of the 27th European Conference on Advances in Databases and Information Systems, ADBIS 2023, volume 1850 CCIS, 2023, p. 197 – 207. doi:10.1007/978-3-031-42941-5_18.

[5] K. Balog, F. Radlinski, S. Arakelyan, Transparent, scrutable and explainable user models for personalized recommendation, in: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR'19, Association for Computing Machinery, New York, NY, USA, 2019, p. 265–274. doi:10.1145/3331184.3331211.

[6] Y. Zhang, G. Lai, M. Zhang, Y. Zhang, Y. Liu, S. Ma, Explicit factor models for explainable recommendation based on phrase-level sentiment analysis, in: Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR '14, Association for Computing Machinery, New York, NY, USA, 2014, p. 83–92. doi:10.1145/2600428.2609579.

[7] A. F. Markus, J. A. Kors, P. R. Rijnbeek, The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies, Journal of Biomedical Informatics 113 (2021) 103655. doi:https://doi.org/10.1016/j.jbi.2020.103655.

[8] P. Phillips, C. Hahn, P. Fontana, D. Broniatowski, M. Przybocki, Four principles of explainable artificial intelligence (draft) (2020).

[9] B. Oliboni, A. Dalla Vecchia, N. Marastoni, E. Quintarelli, ICARE: An Intuitive Context-Aware Recommender with Explanations, Intelligent Systems Reference Library 244 (2023) 65 – 86. doi:10.1007/978-3-031-37306-0_4.

[10] A. Dalla Vecchia, N. Marastoni, B. Oliboni, E. Quintarelli, The synergies of context and data aging in recommendations, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), volume 14148 LNCS, 2023, p. 80 – 87. doi:10.1007/978-3-031-39831-5_7.

[11] R. Agrawal, R. Srikant, Mining sequential patterns, in: P. S. Yu, A. L. P. Chen (Eds.), Proceedings of the Eleventh International Conference on Data Engineering, March 6-10, 1995, Taipei, Taiwan, IEEE Computer Society, 1995, pp. 3–14.

[12] P. J. Phillips, M. Przybocki, Four principles of explainable ai as applied to biometrics and facial forensic algorithms, 2020. arXiv:2002.01014.

[13] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, D. Pedreschi, A survey of methods for explaining black box models, ACM Comput. Surv. 51 (2018). doi:10.1145/3236009.

[14] P. Angelov, E. Almeida Soares, R. Jiang, N. Arnold, P. Atkinson, Explainable artificial intelligence: an analytical review, WIREs Data Mining and Knowledge Discovery 11 (2021). doi:10.1002/widm.1424.

[15] C. Combi, B. Amico, R. Bellazzi, A. Holzinger, J. H. Moore, M. Zitnik, J. H. Holmes, A manifesto on explainability for artificial intelligence in medicine, Artificial Intelligence in Medicine 133 (2022) 102423. doi:https://doi.org/10.1016/j.artmed.2022.102423.