Air Handling Unit Explainability Using Contextual Importance and Utility

Avleen Malhi[1,2(✉)], Manik Madhikermi[2,3(✉)], Matti Huotari[2], and Kary Främling[2,3]

[1] Department of Computing and Informatics, Bournemouth University, Poole, UK
amalhi@bournemouth.ac.uk
[2] Department of Computer Science, Aalto University, Espoo, Finland
{avleen.malhi,manik.madhikermi,matti.huotari}@aalto.fi,
manik.madhikermi@protonmail.com
[3] Department of Computer Science, Umeå University, Umeå, Sweden
kary.framling@cs.umu.se

Abstract. Artificial intelligence has acted as an essential driver of emerging technologies by employing many sophisticated Machine Learning (ML) models, while lack of model transparency and results explanation limits its effectiveness in real decision-making. The eXplainable AI (XAI) has bridged this gap by providing the explanation of outcomes made by these complex ML model. In this paper, we classify the functioning of an air handling unit (AHU) using the neural network and utilise contextual importance and contextual utility (CIU) as an XAI module for explaining outcome of the neural Network. Here, we prove that CIU (XAI module) can generate transparent and human-understandable explanations, which the end-user can therefore utilize for making decisions proving the overall applicability of the method in a novel use-case. Visual and textual explanations for the causes of an individual prediction have been derived from the CIU that are numeric values calculated from the machine learning module results. We also have provided contrasting explanations against some causes that were not involved in the decision. We provide both in our proposed approach.

Keywords: Explainable artificial intelligence · Contextual importance · Contextual utility · Air handling unit

1 Introduction

The artificial intelligence (AI) has advanced increasingly from mature machine learning techniques in the last years. As an effect, a variety of use cases for these technologies has found their way into the everyday tlives of a multitude of users. The absence of transparency in AI decisions, might have an adverse impact on the system trustworthiness. This lack of trustworthiness can also decrease the overall user-experience [5,11]. Even though the research on understandable and transparent AI systems is flourishing, but these explanations are mostly targeted for technical users and are ignored for the end-users in the realistic artificial intelligence

T. Hara and H. Yamaguchi (Eds.): MobiQuitous 2021, LNICST 419, pp. 513–519, 2022.
https://doi.org/10.1007/978-3-030-94822-1_32

systems. Overall, these concerns towards innovative technologies are considered as a critical matter for AI researchers to resolve [2,10]. Hence, the promotion of eXplainable Artificial Intelligence (XAI) is vital for enabling excellent exploitation and establishment of innovative machine learning algorithms in AI systems [1]. Many research studies recommend to model the explanations based on the relevant practical concepts which means providing complete and contrastive explanations to end-users for producing human understandable explanations [8]. Complete explanations provide the justification of an individual instance whereas contrastive explanations explains why a particular prediction was made contrary to other one.

In this paper, for our case-study we use a reference model that has a machine learning module for classifying the air handling unit and a module for making the explanation for the reasons of the classification made. The fact that detecting the failure cases from the normal behavior is a rather laborious task because of high number of dimensions and huge data. Further, because the reasons leading to a particular working state (e.g. fault situation) are often unknown and unique, the reasoning about the event-chain leading to a state is particularly burdensome; therefore Air Handling Unit (AHU) is a particularly good case-study for explainable artificial intelligence [7]. A proven AI method has been applied for the classifying module. For the explanation module we have used the Contextual Importance and Utility (CIU) method for explaining the classification results in more human understandable way [4]. CIU provides the explanations in the form of natural language and visual representations for explaining the test instances [3]. Our method provides both complete and contrastive explanations for the fault detection in air handling unit. The rest of the paper is organized as follows: Sect. 2 studies the related work in the explainable artificial intelligence, Sect. 2 presents the proposed approach based on contextual importance and utility (CIU) for an air handling unit. Section 3 discusses the results in the form of visual and textual explanations and, finally Sect. 4 concludes the paper.

2 Proposed Approach for Explaining the Events in an Air Handling Unit

Generally, it is a human tendency to ask for explanations for making a particular prediction instead of other which implies that what will be outcome if the input is changed. Hence, the explanations also play a major role in explaining the prediction results that something has happened instead of another. Complete explanations provide the justification of an individual instance [9]. The approach of contextual importance and utility can be used for both linear and non-linear models and it explains the predictions of the model for an individual test instance by calculating the contextual importance and utility for each feature. The contextual importance (CI) of an input has been defined by the current values of the inputs and the input range, and the contextual utility (CU) is defined by the output range and current output value. The contextual importance of an input on an output is then defined as the ratio:

$$CI(C,\{i\},j) = \frac{Cmax(C,\{i\},j) - Cmin(C,\{i\},j)}{absmax_j - absmin_j} \tag{1}$$

where $absmax_j$ is the maximal possible value for output j, $absmin_j$ is the minimal possible value for output j, $Cmax(C,\{i\},j)$ is the maximal value of output j observed when modifying the values on inputs $\{i\}$ and keeping the values of the other inputs at those specified by C. Correspondingly, $Cmin(C,\{i\},j)$ is the minimal value of output j observed. The definition of CU is then

$$CU(C,\{i\},j) = \frac{out_{C,j} - Cmin(C,\{i\},j)}{Cmax(C,\{i\},j) - Cmin(C,\{i\},j)} \tag{2}$$

where $out_{C,j}$ is the value of the output j for the context C.

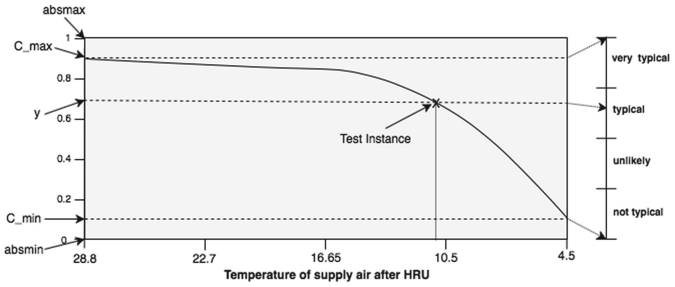

Fig. 1. Contextual importance and contextual utility illustration for case study of air handling unit

CI refers to the ratio of output range obtained by varying the input values for a certain feature x_1 through its whole range from minimum to maximum. The output range lies between the lowest possible output value (C_{min}) by varying feature values and the highest possible output value (C_{max}). CU refers to the position of the predicted output y_i for the selected test instance with context to the output range calculated by CI i.e. if the y_i is close to (C_{max}), it is having high utility and if it is close to (C_{min}), it is having low utility. This has been very clearly illustrated in the Fig. 1 which lists the C_{min}, C_{max}, CI and CU values for a particular test instance of air handling unit. The values are depicted for an input *feature temperature of supply air after HRU* which is an important feature in air handling unit. The results are explained for the normal functioning of air handling unit.

In this paper for our case-study we have used a method for decision making for a black-box model. We are explaining the results derived by neural network method (i.e. a continuous, non-linear method) [6,7]. The method takes into account the selection criteria and the importance of the criteria in a context for the air handling unit in question for making explanations. For our case-study of an air handling unit, the output is "working status" that has two values normal (0) and abnormal (1).

3 Performance Analysis

In this section we introduce the results of diagnosis of the air handling unit's working state (including but not solely faults), which are often onerous to detect based on individual input events of the AHU so that the end-user understands the rationale of the system. The result of explanations for a air handling unit are presented having two output classes, its normal and abnormal functioning. The dataset used in this example is real time data collected from *Enervent* company with 26700 instances consisting 18882 normal and 7818 abnormal instances. Normal state indicates the normal functioning of the air handling unit and Abnormal means there is no heat recovery. We calculated the CI and CU values for the normal and abnormal test instances based on which we provide the human understandable explanations in the form of visual and text based representation. The *absmin* and *absmax* are 0 and 1 respectively and value of C_{min} and C_{max} lies within the range of 0 and 1.

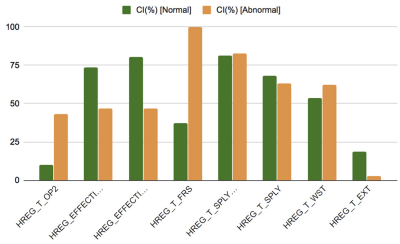

Fig. 2. The comparison of contextual importance values for normal and broken case of AHU

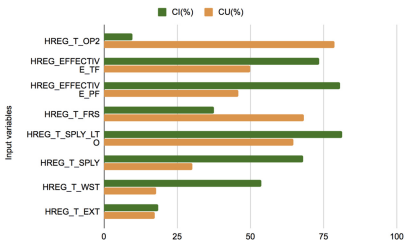

Fig. 3. The contextual importance and utility for all input variables in a normal AHU

The contextual importance and utility values are visually represented in Fig. 3 to depict the importance and utility for each input variable in a normal instance of air handling unit. It can be analyzed that current exhaust speed (EFFECTIVE_PF) and temperature of supply air after HRU (T_SPLY_LTO) are the most important features for the normal AHU working and the utility of T_SPLY_LTO is typical which means it is the considered feature in detection of normal functioning of AHU. The least considered characteristic is the temperature of the extracted air (T_EXT) which has the lowest utility. Again, the temperature of supply air (T_SPLY) which is an important feature is least considered for the normal state of AHU. The contextual importance (CI) is compared for the normal and abnormal instance of air handling unit as depicted in Fig. 2. When the feature's importance is compared for the normal and abnormal

state of AHU, it is seen that the temperature of fresh incoming air (T_FRS) was most important feature in deciding abnormal state of AHU whereas current exhaust speed (EFFECTIVE_PF) and temperature of supply air after HRU (T_SPLY_LTO) were the most important ones for deciding normal AHU state. Considering the abnormal state, the most considered features with high utility values are temperature at operator panel (T_OP2) and current supply fan speed (EFFECTIVE_TF) and current exhaust fan speed (EFFECTIVE_PF). It implies that temperature of fresh incoming air (T_FRS) was most important feature which was considered and even though temperature at operator panel (T_OP2), current supply fan speed (EFFECTIVE_TF) were not very important feature but they were considered highly in deciding the abnormal functioning of air handling unit. Although temperature at operator panel (T_OP2) and temperature of fresh incoming air (T_FRS) are also most considered features with high utility in deciding abnormal functioning of AHU but their importance is less. The temperature of extracted air (T_EXT) is least important with less utility value. The temperature of fresh incoming air (T_FRS) is having least utility value explaining the abnormality of the system as shown in Table 1. In contrast, temperature of supply air after HRU (T_SPLY_LTO) is the most important considered feature for normal functioning of the AHU. The other important feature is current exhaust speed (EFFECTIVE_PF) but having comparatively less utility value. The other features which are most considered in deciding the normal functioning are temperature at operator panel (T_OP2), temperature of supply air (T_SPLY) which have high utility values but their feature importance is really less. The CI and CU values are compared for the normal and abnormal instances of air handling unit in Table 1.

Table 1. Comparison of CI and CU values for normal and abnormal case of AHU

		T_OP2	EFFECTIVE_TF	EFFECTIVE_PF	T_FRS	T_SPLY_LTO	T_SPLY	T_WST	T_EXT
Normal	CI (%)	9.81	73.46	80.63	37.46	81.22	67.98	53.65	18.59
	CU (%)	78.73	49.84	45.83	68.17	64.7	30.2	17.76	17.22
Abnormal	CI (%)	42.96	46.71	46.63	100	82.7	63.11	62.22	2.57
	CU (%)	97.45	89.63	89.79	41.86	50.63	66.34	67.28	49.68

The textual based explanations are shown in Fig. 4. It gives the complete and contrastive explanations for the normal and abnormal state of the air handling unit for justification of prediction of the particular class label (normal or abnormal). The contrastive explanations are also produced to discuss the possible contrasting explanations.

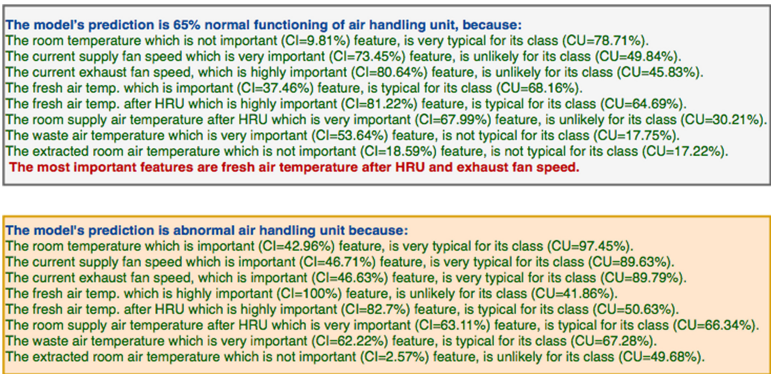

Fig. 4. The normal and abnormal states of AHU explainable by complete and contrastive explanations

4 Conclusion

The explanation method discussed provides adaptability for explaining any "black-box" model. The explanation method used in the study comprises the of getting the contextual importance and contextual utility for the individual instances to provide the complete visual and text-based explanations as well as contrastive explanations. We used case study of air handling unit to explain the fault detection scenarios. We provided the explanations for the normal as well abnormal working conditions of the air handling unit in the human-understandable manner. In the case-study, we used multiple criteria for decision making and transferred the results into understandable verbal format that including certain degree of uncertainty in the explanations as the events are not black-and-white classification situations. Future work includes testing of the approach for more complex building automation data comprising of multiple sensors and multi-class outputs.

References

1. (DARPA). Broad agency announcement, explainable artificial intelligence (XAI) (2016). https://www.darpa.mil/attachments/DARPA-BAA-16-53.pdf
2. Došilović, F.K., Brčić, M., Hlupić, N.: Explainable artificial intelligence: a survey. In: 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 0210–0215. IEEE (2018)
3. Främling, K.: Explaining results of neural networks by contextual importance and utility (1996)
4. Främling, K., Graillot, D.: Extracting explanations from neural networks. In: Proceedings of the ICANN, vol. 95, pp. 163–168. Citeseer (1995)
5. Linegang, M.P., et al.: Human-automation collaboration in dynamic mission planning: a challenge requiring an ecological approach. In: Proceedings of the human factors and ergonomics society annual meeting, vol. 50(23), pp. 2482–2486. SAGE Publications, Los Angeles (2006)

6. Madhikermi, M., Yousefnezhad, N., Främling, K.: Heat recovery unit failure detection in air handling unit. In: Moon, I., Lee, G.M., Park, J., Kiritsis, D., von Cieminski, G. (eds.) APMS 2018, Part II. IAICT, vol. 536, pp. 343–350. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-99707-0_43
7. Madhikermi, M., Malhi, A.K., Främling, K.: Explainable artificial intelligence based heat recycler fault detection in air handling unit. In: Calvaresi, D., Najjar, A., Schumacher, M., Främling, K. (eds.) EXTRAAMAS 2019. LNCS (LNAI), vol. 11763, pp. 110–125. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-30391-4_7
8. Miller, T.: Explanation in artificial intelligence: insights from the social sciences. Artif. Intell. **267**, 1–38 (2019)
9. Molnar, C.: Interpretable machine learning. a guide for making black box models explainable (2018)
10. Shahriari, K., Shahriari, M.: IEEE standard review-ethically aligned design: a vision for prioritizing human wellbeing with artificial intelligence and autonomous systems. In: 2017 IEEE Canada International Humanitarian Technology Conference (IHTC), pp. 197–201. IEEE (2017)
11. Stubbs, K., Hinds, P.J., Wettergreen, D.: Autonomy and common ground in human-robot interaction: a field study. IEEE Intell. Syst. **22**(2), 42–50 (2007)