



HAL
open science

An occupant-centered approach to improve both his comfort and the energy efficiency of the building

Fateh Boulmaiz, Patrick Reignier, Stephane Ploix

► To cite this version:

Fateh Boulmaiz, Patrick Reignier, Stephane Ploix. An occupant-centered approach to improve both his comfort and the energy efficiency of the building. Knowledge-Based Systems, 2022, pp.108970. 10.1016/j.knosys.2022.108970 . hal-03667103

HAL Id: hal-03667103

<https://hal.science/hal-03667103>

Submitted on 22 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

An occupant-centered approach to improve both his comfort and the energy efficiency of the building

Fateh Boulmaiz^{a,*}, 1, Patrick Reignier^a and Stephane Ploix^b

^aUniv. Grenoble Alpes, Grenoble INP, LIG, CNRS, Grenoble, 38000, France

^bUniv. Grenoble Alpes, Grenoble INP, G-SCOP, CNRS, Grenoble, 38000, France

ARTICLE INFO

Keywords:

Case-based reasoning
Bayesian network
Clustering
Data models
Energy Management Systems
Explainable machine learning
Genetic algorithms

ABSTRACT

The accelerating depletion of fossil fuel reserves and the growing awareness about climate issues have put forth a plethora of interesting approaches that attempt to tackle the crucial problem of energy saving, specifically in buildings, known as a major energy consumer. Existing approaches tackle the problem of energy efficiency in buildings by proposing model-based approaches such as knowledge models (thermal, CO₂, cost, etc.) and regressive models. However, different factors make the building of knowledge models particularly challenging, including the very complicated interaction between several heterogeneous phenomena that can impact the use of energy in buildings like the buildings envelope characteristics, their positions, the weather conditions, but also the occupant's behavior is a critical issue in the process. More Recently, techniques from machine learning (ML) to support energy saving in buildings gained increased interest. They learn from collected historical data a model that forecasts the future energy behavior of the building. Although occupant's behavior to save energy in the building is far from trivial, has received less attention from these studies. This paper takes on this challenge and proposes an energy management system based on historical data thanks to case-based reasoning approach. We guide the occupant by proposing an action plan (opening/closing of doors/windows, etc.) to help him in the process of improving his indoor comfort (thermal, air quality, luminosity, etc.) without using more energy if not using less. To encourage the occupant to trust the inference mechanism learnt and cooperate with the energy management system, this approach generates explanations arguing the proposed action plan. We assess the performance of our energy management technique on real-world data collected from a research building at the University of Grenoble, France.

1. Introduction

Global electricity use is projected to increase by 50 % by 2045, says a report published by the BNEF¹ organization [66]. To meet this demand, global electricity production is growing steadily, reaching 1012 watt\hour in 2013, three times more than in 1973. In turn, this production is responsible for about 43% (a yearly increase of 1.1% during the past decade) of the world's carbon dioxide CO₂ emissions, making it by far the largest contributor to the generation of greenhouse gases.

Buildings are the most energy-intensive end-use sector, far ahead of transport, industry, and agriculture. It represents 46% and 92% of final energy consumption in France [69] and Hong Kong [92] respectively, making this sector in France the source of more than 25% of greenhouse gas emissions. With a building park that includes more than 37 million dwellings in France, it is obvious that the measures to be taken as a priority to improve energy consumption focus on buildings.


Numerous studies have investigated the factors that influence the building energy efficiency. Based on a study conducted by the International Energy Agency [103], six parameters can affect the building energy behavior: climate,

building envelope characteristics, energy management system (EMS) characteristics, operation and maintenance of buildings, occupant behavior, and quality of inside design.

One of the main challenges to saving energy in buildings is to reconcile the energy sobriety with the comfort of their occupants. On the policy front, various measures have been adopted to tackle this issue. For instance, governments around the world have set in motion significant reforms involving different measures such as setting energy standards for new buildings or proposing sales contracts with attractive electricity sales rates for off-peak consumption. The purpose is to meet the occupant's needs by providing the best possible thermal comfort, air quality (CO₂ concentration), and cost while cutting down on the energy consumed in the buildings.

The Majority of the above initiatives are primarily concerned with addressing the physical envelope of the building and rarely take into account the role of the occupant in the energy-saving process, albeit an essential one, especially since, paradoxically, the occupants often remain inside these buildings about 90% of the time [94]. The behavior of the occupant and its influence on the energy efficiency of the building are extensively studied in the literature (e.g., see [79, 57, 18, 77]) from different aspects: psychological, cultural, economic, physiological, and time. These studies are unanimously in agreement that the behavior of occupants has a considerable bearing on the buildings energy use. Therefore, guiding the occupant by intelligently utilizing his actions can bring an enhancement of his thermal and air quality comfort whilst consuming equivalent energy if not

*Corresponding author

 fateh.boulmaiz@univ-grenoble-alpes.fr (F. Boulmaiz)

ORCID(s):

¹Bloomberg New Energy Finance

lesser. It is, therefore, now necessary to turn to strategies involving the behavior of the occupant in the improvement of the energy saving of a building.

As buildings are erratic consumers of energy [104], control-based approach is a key strategy to attend occupants to elaborate an optimized energy consumption plan while respecting their comfort. Traditionally, smart thermostats are used. The latter only might show the actual temperature and may provide basic energy use recommendations. Moreover, the cooperation and interaction between occupants and these systems are limited to define only the setpoints and some states.

Another control-based strategy is the rule-based approach [48, 4] which considers "If-Then" rules to guide the occupant's actions to reduce energy consumption. The interaction between the system and the occupant is both more advanced and more flexible than with intelligent thermostats since it allows to set various data on the environment status and customize some action plans. However, these systems do not allow to simulate the environment or to assess its evolution. For example, "If-Then" systems cannot predict how changes in weather conditions might impact the comfort of the occupant.

To reach a global optimization of the energy performance of buildings, the building control process must be optimized. It is required to look for new methods to estimate the weather, the occupancy behavior and to improve the energy consumption without jeopardizing the occupant's comfort. An approach deeply investigated in the research literature and largely applied in the professional sector (e.g., [8, 9, 44, 88]), involves developing a knowledge model of the considered building and applying optimization methods to identify a combination of actions (windows/doors opening, heating setpoints, etc.) leading to the desired comfort level, whilst minimizing the required energy costs. Alternative approaches frequently adopted in the modeling of buildings use techniques from artificial intelligence such as genetic algorithms, Bayesian networks, and Artificial Neural Network [59, 19, 61]. Many considerations affect the relevance of a modeling approach to a specific setting, for instance the aim of the modeling, the data availability, and the level and nature of the user's interaction with the system. While Knowledge modeling approaches certainly provide benefits in precision, are challenging to be specified, and remain an open scientific issue [9]. They require a considerable time to design, their implementation requests a significant effort, and have difficulty addressing the wide range of causal relationships influencing energy behavior in buildings. Knowledge models must be fitted to each energy area, and be readjusted or rebuilt whenever the environment changes. For example, installing an additional cooling system or heater would necessitate changing or even redesigning the model. The coming paragraph details the problems that hamper the deployment of those approaches.

1.1. Knowledge model challenges

Modeling a knowledge model is a challenging process given the character of the data (volume, variety, velocity, etc.) as well as the causal connections that may exist between these data.

Data properties

A building is a complex system composed of a variety of interlocking phenomena which influence the energy behavior of the building. For formalizing the model of the system, each phenomenon is modeled using variable that can be qualitative or quantitative (continuous, discrete).

Based on the literature [26, 98], the overall building energy efficiency can be impacted by different factors which can be grouped into four categories including building conception, building materials and techniques of construction, environmental features, and occupant behavior. Doors and windows positions, orientation of the building, and aperture-to-wall ratio are examples of the design parameters. Quality of materials used to insulate walls and ceilings, and glazing thermal performance are mentioned as the second category. As environment parameters, weather characteristics, and position of the building about the site surroundings, can be cited. Occupant behavior is defined by the presence and the interaction of the occupant regarding his environment by performing (or not) actions to change the indoor environment. As actions, windows/doors opening/closing and adjusting the HVAC system settings can be mentioned.

Data interactions

The process of determining the links, specifically the causal ones, between the features is a key stage in the development of a knowledge system, as the efficiency of the latter is closely tied to the accuracy of this stage. Although considerable research exists on causality (cf. [39, 93] for an overview), this matter is only at its beginning because existing approaches are challenged by the ongoing exponential complexity of the systems to be modeled and, consequently, are restricted to fairly simple applications using a limited number of variables. An extensive comparative study of different approaches to inferring causality is presented in [55], which highlighted that neither of these techniques is meaningfully more efficient than the others.

Cognitive aspect

In the theory control literature, the qualifier *Human-in-the-loop* is commonly employed to depict the implication of the human in systems² control. A human could take an active role in controlling the system when the latter is fully or partially controlled by the human. The human is referred to be passive in the loop when he is supervised, if not outright controlled, by the system.

A common finding from psychological studies on theory control is that humans need to feel in control over the system [56, 89]. In EMS, the desire of the occupant to control should

²The term system is used in the broad sense to include physical systems such as machines or software applications.

be considered as a component of his comfort [12]. Indeed, it has been established in the literature that people have a higher tolerance to variations in indoor thermal conditions if they can exercise some control over them (for example, occupants like to activate blinds by themselves and not by actuators) [16]. But also, for the same sensation of discomfort, we can observe different reactions [17].

However, modelling occupant behavior is not at all straightforward and still a very challenging task. The existing approaches still fall short, they are either very general or very specific for a particular scenario. Inadequacies due to the stochastic nature and randomness of people, humans are diverse in their knowledge and abilities, and finally, besides the rationality of decision making, a person's behavior has emotional aspects as well [51].

Interaction hurdle

The occupant is incapable of correctly interpreting and understating the functioning of his environment, owing to the complex nature of the various phenomena and the different dynamic variables that occur in the building. The occupant is aware of his intents and comfort needs, but he lacks the knowledge to achieve them. Whilst the system has the expertise to optimize the various aims, it is not often conscious of the occupant's desires. A major difficulty of the EMS is its opaqueness, i.e., its lack of transparency in providing clear reasoning and persuasive justification for its advice to occupants. The study reported in [34] concluded that unless there is confidence in the recommendations, the actions and the outcomes generated by the EMS as an intelligent advisor, its adoption and therefore the expected results will be seriously restricted.

The lack of interaction between the EMS and occupant could lead to a conflict situation since the occupant fails to interpret the behavior of the system and even seems to be battling against his house to meet his comfort. A representative scenario is when an occupant shuts the shades to dim the lights to take a nap. Consequently, the EMS turns on the lights automatically as the space is not well lit, which can be uncomfortable. In addition, shortly afterward, the EMS can pull back the shades to warm the indoor of the house as the sun appears on a day of winter.

It is not possible for the occupant nor for the EMS to reach alone and independently of each other the best compromise between comfort and energy efficiency. Both the EMS and occupant have a part of the total knowledge needed to meet the objectives. Consequently, close cooperation between the EMS and occupant is essential as their knowledge is complementary. A well-designed EMS should provide a mechanism to encourage the occupant to cooperate with it.

1.2. Contribution

As outlined in the studies [72, 9], identifying the set of variables that model the phenomena characterizing a complex system is a daunting and expensive task. Therefore, this paper develops an alternative methodology for the design of

an EMS using a data-based model instead of a knowledge-based model. The proposed framework does not require any expert knowledge particular to each building but draws its knowledge and expertise from the past data describing the various phenomena (energy behavior) specific to each building as well as from the actions of its occupants. Our approach is based on assumption that historical data holds hidden information about the environment variables as well as about the occupant's behavior and provides information on their interactions with each other. Thus, that inferred information could be utilized to develop a model to forecast building behavior from environment features and occupant's behavior thanks to the case-based reasoning (CBR) approach.

Trust depends on the perception of the user of the competence of the system [70]. Several studies [102, 41, 33] have established that the capacity of the system to explain recommendations is an influential factor in user's confidence and acceptance. As such, an EMS that can motivate its actions with explanations seems competent and thereby enhance the confidence of the occupant in its accuracy. To foster cooperation between EMS and occupants, a new explanation method combining case-based reasoning and Bayesian networks is developed.

In summary, our main contributions included the following:

- we investigate the use of a data-driven approach applying the CBR method to improve the comfort of building occupants while optimizing energy expenditures.
- we propose new methods for features weighting and similarity estimation.
- we present a simple and effective approach to the adaptation process.
- We design three algorithms to explain the reasoning process adopted by the system to advise the occupant.
- we gauge the effectiveness of the CBR framework through an extensive experimental evaluation using real data.

The remainder of this paper is structured as following. Section 2, briefly provides the relevant background for understanding the proposed approach. Related works are reviewed in the Section 3. Afterwards, a description of our contribution is given in Sections 4 and 5. Section 6, reports and comments on experiments with realistic data. In closing, Section 7 concludes and suggests some future research guidelines.

2. Review of related literature

We survey the pertinent literature for two main aspects of our research. First, we give a review of related notions in the area of CBR. Second, we review the literature on the explanation of intelligence systems.

2.1. Case-based reasoning

Reasoning from cases is a reasoning mode that consists of solving a new problem by matching one or more similar problems already solved whilst fitting their solutions if any [54]. It can be used either for problem-solving or decision-explanation [53].

The interest in CBR is pushing the emergence of its application for the diagnostic, maintenance, conception, engineering, etc. in various domains such as law [82], medicine [96], banking and insurance [60], etc.

Broadly, the existing literature adopts the original model first proposed in [2] that is, the 4R model: Recover, Reuse, Revise, and Retain. We argue that the standard 4Rs scheme has to be completed with a 5th step, the elaboration of the problem, and assume a holistic perspective for the development of the CBR system. Indeed, considering the different CBR steps independently of each other is usually not very effective, as they are complementary steps to achieve a common objective.

Elaboration

The purpose of this step is to query the system environment to collect the information required to structure the problem in a format exploitable by the other stages. The case structure is freely defined and is heavily dependent on the application domain. However, whatever the domain, certain information is compulsory. A classical composition consists minimally of two elements:

- the description of the problem which defines the current state of the environment and describes the request to be solved.
- the description of the retained solution, if any, at the end of the problem-solving process.

Retrieve

In this stage one seeks out one or more past solved cases having problems estimated to be similar to the problem to be solved. The objective is to identify first and foremost similar cases, but above all cases that can be easily adapted [91]. The major issue at this stage is to identify the similarity. A detailed overview of the existing approaches to similarity estimation is given in [52, 63].

The different features of the system do not have the same influence on the solution of the problem, to model their importance to each other, the features are weighted in the similarity estimation. Indeed, in [101], it has been shown that assigning weights to features may significantly enhance the precision of conventional techniques such as kNN and SVM.

Reuse

Ideally, in the reuse step, cases with problems matching the problem to be solved are found and consequently, their solutions are applied unmodified for solving the target problem. Unfortunately, the new problem seldom perfectly fit the existing ones. It is, therefore, necessary to adapt the solutions of the latter to consider their differences to the target problem. While many approaches are reported

to adapt, these can be categorized into two broad classes according to the component of the retrieved cases targeted by the modification. Whereas in *structural adaptation* one acts directly on the solutions of the retrieved cases, *derivational adaptation* takes up the different phases of reasoning leading to the solutions of the similar cases..

Revise

During the revision stage, the proposed solution is revised to check its relevance to resolve of the target case. The revision process is a way to learn new knowledge following the passing or failing of the reuse process. The review process is performed either by simulation or by soliciting directly domain expert/user to judge the solution.

Retain:

The retention step is the opportunity to enrich the case base with the new solving experience. Besides the knowledge of the novel case, the knowledge of the other steps and notably that of adaptation and similarity could be improved.

2.2. Explanations

Whereas research on explanations is not recent (the first explanation systems date back to rule-based systems in the 1970s), in last few years, the role of explanations in the design of machine learning models has received surge interest both from technical and ethical perspectives [3].

It is necessary to point out that the notion of explanation is always confused with that of justification. The former is a form of reasoning that aims at providing an integer report on the process that led to the recommendation to make the user understand the causes that allowed to generate the suggestion, whereas justification is the system's capacity to assist the user in the comprehension of reasons for recommending a particular action or item[71]. That is to say, justification consists of giving reasons why a recommendation is deemed to be appropriate, without attempting to describe the actual process of reasoning that led to this decision.

The principal objective of the explanation is to motivate the behavior of a model in a human-friendly way [99, 49]. There are several possible intents that a system can formulate in providing explanatory information for a recommendation. Table 1 states the results of an in-depth analysis study [97] on explanation objectives. It is interesting to point out that is very challenging to find a trade-off between all these objectives as certain of these aims are conflicting. For instance, persuasiveness could be antagonistic to the effectiveness. Therefore, few works in the literature consider more than one of these goals, and the bulk of the existing works seek to optimize a single objective. Based on a broad-based literature review [73], the most prominent explanation objective is transparency, which is considered a pivotal factor in earning the trust of users [1].

Several taxonomies have been suggested in the literature to categorize explanation techniques. They may vary considerably according to the criteria used and techniques' characteristics. A detailed overview of these taxonomies is

Table 1
Explanation objectives [97].

Objective	Definition
Trust	Boost user's trust in the system
Satisfaction	Improve ease of use or enjoyment
Efficiency	Assist users with decision-making faster
Effectiveness	Assist users in making a good decisions
Scrutability	Give users the ability to report to the system that it is wrong
Transparency	Clarify the workings of the system
Persuasiveness	Persuade users to test or purchase

provided in [15]. Based on how recommendations are communicated to the recipient, explanations can be grouped into three categories as:

- attribution (feature)-based explanations: explain the model in terms of the relevance of the input features by calculating the explanatory power of the input feature on each output feature.
- example-based explanations: refer to historical situations in which the recommendation turned out to be true or false. An illustrative example of such explanation is: the outside temperature is currently 21°, and it is advised to open the window for 45 minutes at 2 pm. Before yesterday, the situation was similar, it was 22° and the same advice found to be correct.
- rule-based explanations: these methods target to describe the reasoning of a model in simplified or human-understandable rules, such as "If-Then-Else" rules or logical rules. In practice, this is the most used technique as far as it is easier to understand by humans [27], as they tend to analyze events based on a cause-effect approach[78].

While several techniques have been explored to generate explanations, the criteria for choosing the appropriate approach of explanation for a specific domain in a given situation are still quite unclear. Research in the assessment of explanation approaches is very active, however to date, there is no consensus on evaluation metrics. Authors in [105], present an excellent overview of existing efforts in the literature to evaluate machine learning explanations' quality.

3. Related work

To position our research, this section sheds light on existing frameworks. For convenience, it follows the same outline as the previous section. First, we discuss some case-based EMS. Then, we present frameworks considered as references in the field of systems explanation.

3.1. Machine learning-based EMS

The conception of energy management systems is extensively investigated in the literature. The goal here is far

from being an exhaustive review of the existing literature. Attempts in this regard are given in [65, 67]. Authors in [8] propose a set of criteria to be fulfilled to assist in the effortless design of energy monitoring and control systems in buildings. They evaluate a few EMS against these criteria. The study [25] gives a structured review of the existing research on energy management systems according to a machine learning outlook by classifying them into two categories. Firstly, the occupant-oriented approach consists of occupancy identification, analysis of activities, and prediction of preferences and behavior. Secondly, the appliance- and energy-oriented approach involves the profiling of devices, detection of anomalies, and monitoring of energy.

Artificial Neural Networks (ANN) with various network architectures are broadly applied to forecast building energy consumption [62, 95], heating and cooling loads [75, 47, 90], control of HVAC Systems [85, 46], energy efficiency [14] and comfort enhancing [24, 76].

Undoubtedly, ANNs provide a valuable tool for energy behavior modeling in the building. However, the performance of ANN models is highly conditioned on the quantity and the quality of the data on which ANNs are trained. Indeed, results from different researchers [40, 13, 20, 6] show that ANNs require an appropriate quantity of input samples during the training phase to achieve a suitable precision. Otherwise, ANNs require an appropriate design of the structure of the network and a careful tuning of their different hyper-parameters for training. In the current state of research, they are done in an ad-hoc fashion by testing several types of architecture and different parameters to find the most efficient configuration. Finally, a major drawback of these approaches is their unawareness of physical laws. It is quite possible to generate results that are in contradiction with their physical natures, for example, a negative CO₂ concentration or a humidity level that exceeds 100%.

An alternative approach widely adopted in the literature consists of using CBR for predicting the energy behavior of building. Thereafter some existing material is discussed that we found relevant to the current research, i.e. using CBR to enhance the energy performance of buildings while respecting the comfort of the occupants.

The application of CBR in the study of the energy behavior of buildings was mainly focused on the forecast of electricity use. In alignment with existing studies [86, 80, 45], The authors in [21] and [28] describe two closely approaches to predicting electricity consumption in buildings using the CBR. They propose to apply the Analytical Hierarchy Processing (AHP) method to structure the problem to be solved into a hierarchy of objectives and then a pairwise influence comparison matrix is performed by the expert opinions to determine the weight of each variable on the global objective. The similarity between the cases is estimated by computing a similarity distance based on the cosine metric. The two algorithms Kmeans and KNN are combined to retrieve similar cases. The adaptation process is assumed to be an optimization problem which is addressed by the particle swarm optimization (PSO) method [50]. The

revision process involves the Support Vector Regression model combined with a threshold to improve the solution proposed in the adaptation step. However, as discussed in the introduction, even with high expertise knowledge, it is not straightforward to determine the features modelling a building energy behavior, and harder still to draw a comparison between these features.

Other research, such as [36], tackles the issue of energy performance by better monitoring the HVAC system avoiding abrupt changes, as this would drastically increase the energy consumption. To remedy the inefficient piloting of HVAC systems, the authors propose a framework relying on a multi-agent system which learns the social behavior of the occupant and the building's energy consumption thanks to a CBR approach. The latter is used to estimate the social activities of each employee in the Building's offices (arrival time, lunchtime, leaving time). To minimize the harsh shifts in temperature (temperature change by several degrees in a short period), an optimization algorithm is used to adjust the indoor temperature and turn on/off the air conditioning system. The algorithm is based on the employees' schedules and different variables that can affect the indoor temperature. However, this category of approaches considers only a very limited set of variables that can influence the energy behavior of buildings. Furthermore, it does not take into account the weight of each variable nor the impact of each of them compared to other variables. Finally, this approach provides no mechanism for capturing the occupant's feedback. The only possible interaction between the occupant and the system is limited to the occupant's preferred temperature setting.

Another direction of research is focused on the combination of several approaches to increase the efficiency of the proposed approaches. This is the case of the study [43] where the authors proposed a CBR-based model to control the electricity consumption in school buildings. The precision of the model is enhanced by combining ANN, genetic algorithm and multiple regression analysis (MRA). Particularly, a genetic algorithm is used to look for similar cases. MRA and ANN techniques are used to bound the range of possible values for the predicted solution.

3.2. Explanation

Recently, explanation research has experienced a surge of interest, after slacking off after the explosion of explanation research in expert-based systems during the past decades. This resurgence is driven by the craze for artificial intelligence methods and fostered by the rapid development of machine learning techniques and the capabilities of computing machines.

Although the wealth of literature on the matter, the EMS explanation variant of the intelligent systems explanation problem has attracted very limited consideration so far. Failing this type of research and for the sake of brevity, we review in this section only frameworks that are model agnostic, i.e., techniques that have no prior knowledge of the model structure (consider the model to explain as a black

box) and thus they are generic enough to be applied to any machine learning model.

LIME (Local Interpretable Model-Agnostic Explanations) framework [81] provides a model-independent method for explaining machine learning models. The overall concept is to generate a local substitution model driven by interpretable data. The local substitution model is locally equivalent to the model to be explained, which is often complicated. Local modeling is supported by easy-to-interpret models such as SVM and decision trees. The local model is trained on a sample of data randomly chosen from the training dataset and weighted based on the distance from the data to be explained. The data features are filtered to retain only the most significant ones.

In [29], the authors draw on LIME to introduce a data-driven model to explain how the EMS make decisions in forecasting the coefficient of performance (COP) of HVAC system in buildings. Specifically, they use five approaches to locally model the data, namely: generalized linear models, SVM, ANN, extreme gradient boosting trees, and random forests. The former is applied to grasp the linear relations among the local data. The remaining techniques are used to model more complex and non-linear ones. Each item in the surrogate model is assigned a weight that corresponds to its similarity with the item to be explained. The more similar an item is to the item to be explained, the more interesting it is to generate local explanations. The similarity is calculated using the Gower's metric [38]. Although the philosophy of providing a framework independent of the model to be explained is very practical since it facilitates the task of the designer of machine learning algorithms, the LIME framework suffers from an important shortcoming, namely, its inconsistency, i.e., the same forecasted result could have various explanations. This is the result of the random selection of the neighborhood when generating the local model. This lack of robustness is highlighted in the study presented in [5] and confirmed by another study of uncertainty in [37].

Other plot-based methods such as Individual Conditional Expectations (ICE) [35] and Partial Dependence Plot (PDP) [32], allow inspecting the relationship between an explanatory variable (or even two) and the phenomenon to be explained for which a prediction is calculated. The procedure for ICE consists in plot the curves of the prediction corresponding to the variation of the value of the explanatory variable to be inspected (so that it takes all possible values) for each item in the database. The PDP plot is simply the average of the items plots for all values of the explanatory variable. These methods are easy to implement and offer an intuitive and causal interpretation of the model (if the explanatory variable changes for this value, the impact will be visible on the graph). However, correlations between variables can lead to the creation of incoherent situations (e.g., in the case where the explanatory variables are height and weight of a person, a situation with height = 2 m and weight = 5 kg is very likely). As such, these methods are

quite far from the requirements of an explanatory model of an EMS (several highly correlated explanatory variables).

Recently, authors in [83] proposed the Shapley Additive exPlanations (SHAP) method which slightly extends the LIME model by borrowing the notion of Shapley values [87] from game theory. For a given item, the Shapley value of an explanatory variable (or of several variables) represents its contribution to the difference between the value predicted by the model and the average of the predictions of all the items. However, like the ICE and PDP methods, SHAP could lead to inconsistent situations because of the correlations between the variables. Moreover, the contributions' calculation, if exhaustive, can be very time-consuming.

4. Proposed solution

This research seeks to investigate quantitatively and qualitatively the feasibility of using CBR approach to enhance occupant comfort while minimizing energy consumption in buildings. To meet this goal, the following desiderata outline the properties deemed essential for an effective system.

Desideratum 1 (data model). *Proposed approach should be based on a "data model" which is a conceptual representation of the existing relations among the data extracted from the sensors measurements and does not need the use of a physical building model.*

Desideratum 2 (flexibility). *The system should be easily adaptable to the needs and comfort requirements of the occupant, e.g., the occupant is looking for thermal comfort, air quality, sound comfort, or a combination of some or all of these criteria.*

Desideratum 3 (cooperation). *Cooperating of the occupant with the EMS by accepting the proposed action plans to change his behavior to more responsible ones.*

4.1. Architecture overview

Our working assumption that will drive our approach states that:

Hypothesis 1. *Performing the same actions on two days with the same context generates the same effects, and therefore the occupant attains a similar level of satisfaction.*

Another hypothesis that we consider in this research, especially to satisfy Desideratum 3, suggests that:

Hypothesis 2. *Giving the occupant transparent information about the reasoning process of the EMS through an explanation mechanism could improve the occupant's trust in the recommendations provided by the EMS.*

Looking from an overall perspective, we look for among the historical days having the same context as the day to predict, the day having the best performance (this notion is developed in Section 4.6) and propose its actions to the occupant.

To estimate the weight of the variables, historical days are initially classified into similar days, i.e. those which have similar contexts and identical actions and therefore have the same effects constitute a unique group. For instance, turning on the heating with the window open does not produce the same effect with the window closed. Among the action plans of days similar to the day to be predicted the action plan that earlier yielded the best effects is advised to the occupant. These actions are also exploited to learn a Bayesian network describing the causal relations in the neighboring context to generate causal explanations of the system's reasoning.

More concretely, we have developed a five-stage approach as depicted in Figure 1:

1. gathering homogeneous variables (having the same type and the same physical magnitude) and standardize³ them.
2. a genetic algorithm approach is applied to calculating the weights of the variables.
3. finding days with similar context variables as the day to be predicted.
4. extracting the best schedule of actions to present to the occupant.
5. generating and presenting explanations that correspond to the recommended action plan.

4.2. Case structure

This research presents a finer classification of variables involved in building modeling. The different variables are grouped into three categories, based on the causal relations among the phenomena they model:

- *context variables:* describe the uncontrollable phenomena of the environment. They are the causing of other phenomena. For instance, the number of occupants and the meteorological conditions (humidity, precipitation, sunlight, wind direction and speed, etc.).
- *action variables:* represent phenomena as causes of other phenomena but which can be monitored by the occupant to regulate the energy consumption or the comfort. For instance, the door opening and the switching of the air conditioning system on/off.
- *effect variables:* model the phenomena representing the consequences of the application of the actions to the contexts. For instance, indoor temperature, noise level, and the recommended actions number.

Each historical day D corresponds to one case C which is described by a triplet (p^D, s^D, st^D) , where p^D is the problem part, s^D is the solution part, and st^D the environment state.

- the problem part describes non-controllable environmental features for day D , it is defined by a set of k descriptors representing the k variables \mathcal{E}_i^D of the context of day D . So, $p^D = \{\mathcal{E}_i^D\}_{i \in \{1, \dots, k\}}$.

³The terms "normalization" and "standardization" are used in the same sense.

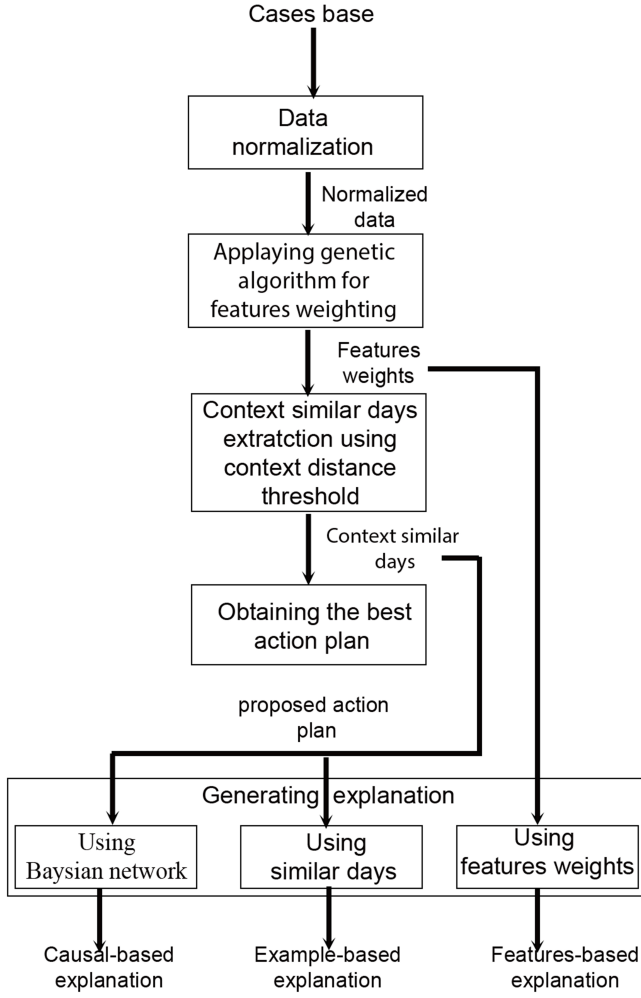


Figure 1: An overview of the EMS architecture

- the solution part is represented by the set of l descriptors describing the l variables \mathcal{A}_i^D of the actions of the day D . Thus, $s^D = \{\mathcal{A}_i^D\}_{i \in \{1, \dots, l\}}$.
- the environment state post the running of the solution is described by the part st^D comprising m elements corresponding to the variables \mathcal{E}_i^D of the effects of the day D , $st^D = \{\mathcal{E}_i^D\}_{i \in \{1, \dots, m\}}$.

Note that the descriptors describing the features of the context, action, and effect are time-series data since the data collected are sequences of values evolving over time. The time frame chosen to sample the data does not impact the applicability of the approach.

Before discussing our methodology for similar days extracting, let stop our attention on the notion of the comparison between days.

Definition 1 (Days comparison). *The comparison between two days consists of comparing all or a type of variables that model days. In context-based comparison, for instance, the comparison consists of computing the distance between vectors representing each context variable.*

4.3. Normalizing data

Incompatibility of measuring units and magnitudes between variables may bias the results, it is, therefore, necessary to adjust them according to transformation functions to have a comparable impact.

The proposed approach is in two phases. the first one is a discretization phase whose goal is to eliminate the noise due to the measurement sensors. The second one is to rescale features values to the same range to correctly interpret their effects.

This study defines a sensitive threshold assigned to each variable according to expert knowledge as a tool to discretize the features.

Definition 2 (Sensitive distance). *It reflects the sensitivity level of human perception to the changes in values of a variable.*

For instance, according to physical findings, any change in CO₂ concentration below 100 ppm is not perceptible to the occupant. Indeed, fresh air contains 400 ppm of CO₂ on average and the occupant generates 4000 ppm of CO₂ when exhaling. Meanwhile, a temperature variation of less than 2 °c is a change that is not felt by the occupant.

The discretization formula is described as follows:

$$New(V) = \frac{V}{s_V}, \text{ with } s_V - \text{sensitive distance of feature } V.$$

According to the expert knowledge, the sensitive distance for some variables is given by:

- CO₂ concentration: 400 ppm,
- tempreature: 2 °c,
- wind speed: 2 m/s.

Thereafter, a linear transformation on the data is carried out thanks to Min-Max normalization method. Suppose, V_{min} and V_{max} are the minimal and maximal values of variable V , the normalization function matches a value V_D^t of V to $V_D'^t$ in the range $[0, 1]$ by evaluating the formula (1).

$$V_D'^t = \frac{V_D^t - V_{min}}{V_{max} - V_{min}} \quad (1)$$

With V_D^t - the value of variable V of day D at time t .

4.4. Features weighting

The weighting of context and action features is done to reflect the extent to which these features are important in generating effects. The overall goal is to approach the optimal level of influence of each feature. To do this, if an action or context feature is more significant for effects, it would be assigned a relatively high weight compared to other features, knowing that while a couple of days are within a great context-action distance $Dist_{CA}$ (cf. Equation 2), they should be within a great effect distance $Dist_E$ (cf. Equation 3) and conversely. It is imperative to introduce the context-action distance in the calculation of the weights

of context and action variables because the impact of the latter is not completely isolated from each other. There are situations where the influence of a variable is correlated to other variables. For example, the sunshine in a room does not only depend on the state of the weather (absence of storms) but also on the position of the blinds (open or not). That is why one has to consider this distance to estimate the impact of the features between each other.

More specifically, we provide an approach using a genetic algorithm to approximate features weights and fulfill the identified constraints. The pseudocode corresponding is given in Algorithm 1.

1. create an initial random number of vectors of weight w_i for context and action variables. Each context and action weight variable is represented by a unique value in the vector w_i , such that the size of the vector is equal to the total number of context and action variables.
2. cluster the days, which are close, using a k-means clustering approach based on the weighted context-action distance defined in equation (2).
3. the optimization problem is to solve the objective function, which corresponds to the optimization of two criteria: minimizing the number of clusters to ensure the variety of actions in each cluster (to be able to propose at the end a pertinent set of actions to the user for the current day) and reducing the average of the distance defined in equation (4) within the effects in each cluster to meet our starting hypothesis (see Hypothesis 1).
4. generation of new individuals w_{i+1} from w_i by applying crossover and mutation operators and back to step 2 until the convergence criteria are met. The algorithm terminates either if the total of the generations in the population reaches a given value without fitness enhancement, or when a maximum number of generations has been attained,

$$Dist_{CA}(D', D'') = \left(\sum_{i=1}^{n_C} w_{V_C} * (V_C^{D'} - V_C^{D''})^2 + \sum_{i=1}^{n_A} w_{V_A} * (V_A^{D'} - V_A^{D''})^2 \right)^{\frac{1}{2}} \quad (2)$$

With w_{V_C} , w_{V_A} - the weight of context $V_C^{D'}$ and action $V_A^{D'}$ features respectively of day D' , n_C - number of context variables, n_A - number of action variables.

$$Dist_E(D', D'') = \left(\sum_{i=1}^n (V_E^{D'} - V_E^{D''})^2 \right)^{\frac{1}{2}} \quad (3)$$

With $V_E^{D'}$ - the effect variable of day D' .

$$Dist_{w_C, w_A} = \frac{1}{n} \left(\sum_{C_i \in C_{w_C, w_A}} \frac{\sum_{D', D'' \in C_i} Dist_E(D', D'')}{|C_i|} \right) \quad (4)$$

With n - number of clusters, D' , D'' - cluster members, $|C_i|$ cardinality of cluster C_i .

Algorithm 1: Pseudocode of genetic algorithm for features weighting

```

// Get All variables from the base of cases
Input: AllVariables = { $\zeta, \Lambda, \xi$ }
Output:  $w$ , Weights of context and action variables
// Generate an initial Population of  $n$  vectors
 $w_0$  of size  $|\zeta| + |\Lambda|$ 
1 Initialize( $w_0$ )
2  $w \leftarrow w_0$ 
3 Function Reproduce( $w$ ):
4    $SelW \leftarrow Select(n)$  // Selection of  $n$ 
   chromosomes for reproduction into New
   population using Roulette wheel
5   while  $|NewW| < |w|$  do
6     // From  $SelW$ 
7     Select 2 chromosomes,  $w1$  and  $w2$ 
8      $NewW \leftarrow CrossOver(w1, w2)$ 
9   end while
10   $w \leftarrow NewW$ 
11 return  $w$ 
12 Function Evaluation( $w, labels$ ):
13   // Calculate the fitness score
14    $Fitness \leftarrow \frac{1}{Dist_{w_C, w_A} + 0.1}$  //  $Dist_{w_C, w_A}$ 
   defined in equation (4)
15 return  $Fitness, |labels|$ 
16 while not termination () do
17   for  $i \leftarrow 1$  to  $n$  do
18     Calculate  $Dist_{CA}(w_i, \zeta, \Lambda)$ 
19      $Labels \leftarrow kmeans(Dist_{CA})$ 
20      $score[w_i] \leftarrow Evaluate(w_i, Labels)$ 
21   end for
22    $w \leftarrow Reproduce(w)$ 
23    $w \leftarrow Mutate(w)$  // Choose randomly 2 position
    $x1, x2$  on each  $w_i$ : swap( $x1, x2$ )
24 end while
25 return  $w$ 
    
```

4.5. Retrieving similar days

Once the weights of action and context variables are determined, days having a similar context to the one forecasted can be extracted. The similarity estimation is done using the context variables exclusively as the action variables are not available (that is what we want to determine) and, consequently, effect variables are either unknown. The unique data at disposal are the context features (based on weather predictions and planned occupancies) and these are used to compare context variables between two days by calculating context distance through the formula (5).

$$Dist_C(D', D'') = \left(\sum_{i=1}^n (w_{V_C} * V_C^{D'} - w_{V_C} * V_C^{D''})^2 \right)^{\frac{1}{2}} \quad (5)$$

With w_{V_C} - the weight of context variable V_C , $V_C^{D'}$ - the context variable V_C of day D' .

To determine the days with similar context to the day to be predicted, the kNN algorithm is used to evaluate the similarity according to the distance defined in the equation (5). The challenge at this stage is how to define the criterion for determining the distance beyond which the days are not considered neighbors. Indeed, if the context distance is far too high, it could result in some selected days which are not similar. On the other hand, if context distance is too low, there might be some similar days which are not selected, and therefore, the variety of the actions set might not be sufficient to recommend actions to the occupants.

To address this issue, the notion of *sensitive distance* introduced in Section 4.3 is used again. This is a value defined by experts in the domain (physical rule), above which the difference between the effect variable values will not be perceived by the occupant.

Definition 3. *Two days are similar based on their effects if the difference between the values of each effect variable does not exceed the corresponding sensible distance.*

For instance, in the case of the effect variables temperature and CO₂ concentration, according to definition 3, two days are similar based on their effects if:

- the maximum difference in indoor CO₂ concentration should be less than 400 ppm,
- the maximum difference in indoor temperature should be less than 2 °c.

The target is to identify the clusters of days with the same context while respecting the sensitive distance for the effect variables. From the context distance and the physical rules, it is, therefore, possible to set a metric to determine the threshold context distance T_C below which the days are considered similar.

$$Dist_C(D', D'') \leq T_c \Rightarrow \text{similarity of } D', D'' \text{ according to sensitive distance.} \quad (6)$$

However, once again, it is not possible to exactly define the T_C which satisfies the previous condition. The reason is that the number of neighboring days, and thus potentially the context distance, that verify the similarity based on physical rules (Equation (6)) varies from day to day in the base of historical days. So, if T_C is set too low, likely, some similar days meeting the condition will not be chosen (recall). On the other hand, when T_C is set too high, the result may include irrelevant days (precision). Figure 2 illustrates this challenge.

To find the best trade-off of precision and maximizing the number of similar days based on physical rules, a method of classification using the F1-measure metric is applied. To calculate the F1-measure defined in Equation (7), each pair of days is labeled according to the T_C and the physical rules. Each element of the set of labels is defined by the equation (8).

The threshold T_C retained is the one that maximises the F1-measure on the set of labeled days as shown in formula (9).

$$F1\text{-measure}(T_C^i, \ell^i) = \frac{2}{\text{Precision}^{-1} + \text{Recall}^{-1}} \quad (7)$$

Whith ℓ^i - the set of labeled days corresponding to T_C^i .

$$T_C = \underset{T_C^i}{\operatorname{argmax}}(F1\text{-measure}(T_C^i, \ell^i)) \quad (9)$$

4.6. Getting the best action schedule

Assuming our starting hypothesis, the principle of the search for the best action plan to improve the occupant's comfort consists in identifying, among all the days having the same context as the day to be predicted, the day D^{best} that provides the best performance. The actions Λ^{best} performed on this day are extracted and recommended to the occupant.

Comparing the performances of two days D' and D'' consists in comparing the effects ξ' of applying the actions Λ' to the context ζ' of the day D' with the ξ'' obtained by applying the actions Λ'' to the context ζ'' of the day D'' . However, it does not make much sense to compare a pair of effect values. For example, the comparison between two indoor temperatures at a given time does not give any indication of the occupant's satisfaction level following this temperature change. To address the issue, we introduce the notion of occupant's satisfaction.

Definition 4 (Satisfaction). *Satisfaction refers to the evaluative judgment that the occupant makes about the effects of the application of an action plan according to the adequacy perceived between the desired effects and the real effects obtained.*

Different criteria can be considered to model the satisfaction such as indoor temperature, air quality, and noise level. The acceptability levels of each criterion are determined by a domain expert.

In the following, two criteria are introduced to exemplar the measure of the occupant's satisfaction Y , thermal and air quality comfort. These indices are calculated based on the recommended values cited in the standards [30, 11] for thermal comfort level and in [10, 31] for indoor air quality. More precisely:

- satisfaction based on temperature: a thermal comfort level is judged as acceptable for temperatures between 18°c and 26 °c. The occupant is considered fully satisfied for temperatures ranging from 21 °c to 23 °c. To simplify the calculation, we compute dissatisfaction instead of satisfaction.

The formula to calculate the thermal dissatisfaction $\bar{Y}_i^j(t)$ at the i -th hour is given in equation (10) that

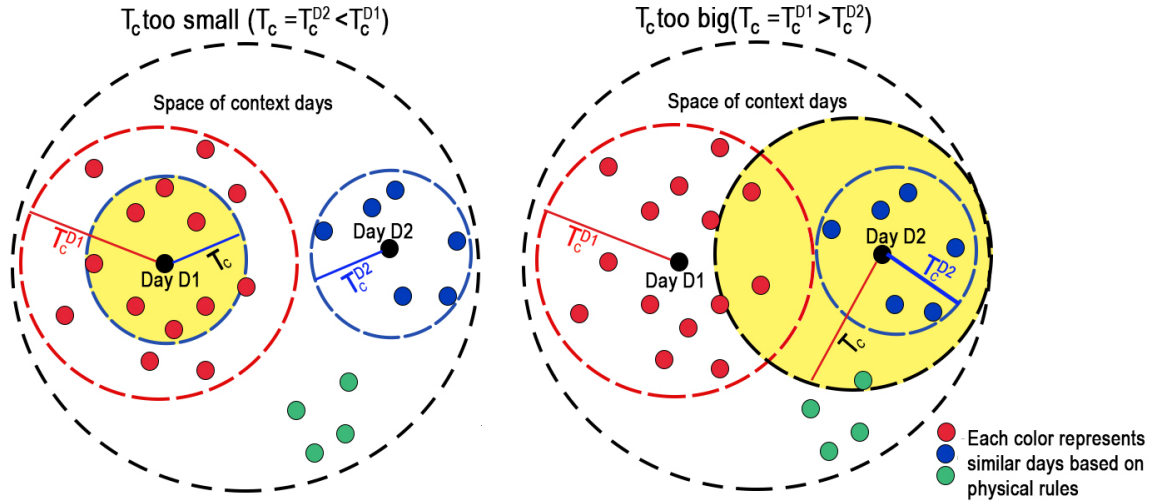


Figure 2: Challenge of T_c distance

$$\forall \ell' \in \mathcal{L}, \forall D, D' \in CB, \ell'(D, D') = \begin{cases} 1 & \text{if } T_{in}^D - T_{in}^{D'} < 2 \wedge CO_2^D - CO_2^{D'} < 400 \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

With \mathcal{L} - the set of labels, CB - the case base, T_{in}^D - the indoor temperature of day D , CO_2^D - the indoor CO_2 concentration of day D .

is represented graphically in the Figure 3.

$$\bar{Y}_t^i = \begin{cases} \frac{21-t}{21-18} & \text{if } t < 21 \\ \frac{t-23}{26-23} & \text{if } t > 23 \\ 0 & \text{if } t \in [21, 23] \end{cases} \quad (10)$$

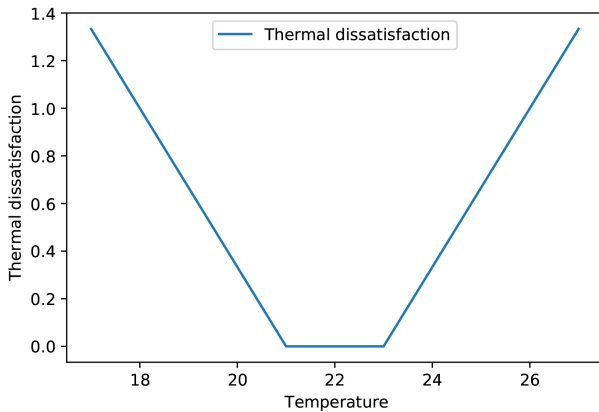


Figure 3: Thermal dissatisfaction

- satisfaction based on air quality: the air quality is deemed to be excellent when the CO_2 concentration is less than 500 ppm. It is considered to be satisfactory

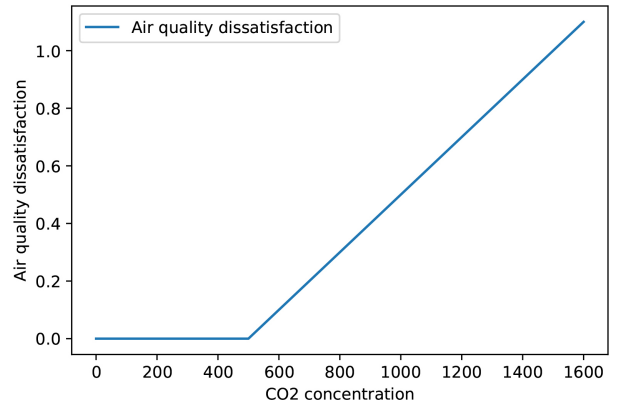


Figure 4: Air quality dissatisfaction

between 500 ppm and 1500 ppm. Still, for simplification matters, we compute dissatisfaction air quality instead of satisfaction.

The air quality dissatisfaction $\bar{Y}_c^i(C)$ at the i -th hour is calculated by the formula (11). The variation of the air quality as a function of the CO_2 concentration is plotted in the Figure 4 which follows from equation (11).

$$\bar{Y}_c^i(c) = \begin{cases} \frac{c-500}{1500-1000} & \text{if } c > 500 \\ 0 & \text{if } c \leq 500 \end{cases} \quad (11)$$

Where c - the CO_2 concentration.

The global dissatisfaction \bar{Y}^i at the i^{th} hour is formulated by Equation (12).

$$\bar{Y}^i = \sum_{V \in V_C \cup V_A} \Omega^V * \bar{Y}_V^i \quad (12)$$

with $\sum \Omega^V = 1$ - the relative weight of variable V over the other action and context variables.

Global satisfaction Υ is obtained through the equation (13).

$$\Upsilon = \frac{1}{\bar{Y}} \quad (13)$$

The best action plan Λ^{best} to recommend to the occupant is the one of the day with the highest satisfaction. More formally:

$$\Lambda^{best} = \underset{\Lambda^D}{\operatorname{argmax}}(\Upsilon(D)), D \in \{\text{similar context days}\} \quad (14)$$

Where $\Upsilon(D)$ - the satisfaction of occupant for day D ,

5. Generation of explanations

Although machine learning systems give interesting results, they can occasionally be wrong in ways that humans would not be aware. This overuse of machine learning systems results in misuse (overtrust phenomenon [23]). At times, these systems are misunderstood, which leads to their being overlooked or underused by people (so we speak of under-trust [23]). To avoid falling into such situations, we propose an approach that allows presenting the previously generated action plan in a way to improve the calibrated trust [58] of the occupant in the EMS. The notion of calibrated trust consists in (not) following the (in)correct advice. Its principle is to give the occupant enough explanatory information to assess the quality of the proposed action plan and so provide him the ability to lean on the recommendations of the system when deemed plausible, if not to manage otherwise. By assuming this notion, the contribution presented hereafter contrasts with the existing studies that only seek to strengthen the blind confidence of the occupants. The aim is to encourage the occupant to explore further action plans whenever he is not convinced by the explanations provided by the system to generate the action plan, which would be a wonderful opportunity for the system to learn better.

We hypothesize that an approach that supports all three types of explanations namely features-based explanations, examples-based explanations, and rules-based explanations could improve the calibrated confidence of occupants in the explanations provided by the EMS. In this chapter, we contribute to the construction of such model by leveraging the CBR approach and the Bayesian network, taking into account findings from multidisciplinary studies on explanations from the perspective of the user. The synthesis of these studies led to the formulation of three key principles that guided the design of the explanations framework:

- *explanations are chosen*: people seldom, if ever, look for explanations that capture the overall causes of an occurrence. They prefer, albeit incomplete, explanations that convey the most important information that is relevant to the decision [68].
- *explanations are contrastive*: explanations do not clarify why a specific prediction was made, but rather motivate why that prediction was given instead of another, to meet people's mental need to seek out why event P (referred as fact [64]) occurred instead of some event Q (called foil [64]) rather than to justify why event P occurred. The foil could be a real event or an imagined one.
- *explanations are contextual*: the explanation is a form of social interplay involving the explicator and the explanation receiver[42]. It is therefore important to look at the context in which the explanation occurs, i.e. define the receiver's profile to determine the content of the explanation and the way to transmit it. An explanation can be conveyed to the user in a variety of ways, and choosing the appropriate one is crucial to the overall success of an explanation approach.

Now, we will take advantage of the calculations made in the previous steps to generate the different explanations.

5.1. Features-based explanation generation

In this type of explanation, the model forecast is explained by indicating the overall importance of each feature in impacting the predictions of the model. We adopted a method that attributes scores of importance to features according to their contribution to the forecasted value of an item. The scores are visualized using a graph. Specifically, we used the results of the method presented in the Section 4.4 to estimate the importance of each feature. The scores are ordered and the first n features with the scores are presented to the occupant as a bar graph.

5.2. Example-based explanation generation

Example-based explanations provide reasons for a decision-making process by supplying instances of similar situations which resulted in the same outcome earlier. For instance, today, the weather conditions are very similar to the one of 2013/04/13, the same proposed recommendations were satisfactory in this situation.

Several declensions were reported in the literature to implement this approach. We focus on the explanation in its counterfactual variant. Explanations based on counterfactual [100] produce examples that yield a distinct model outcome given minimal shifts in input features.

The main challenge of counterfactual-based explanation is twofold. Firstly, is to select the relevant examples that motivate the decision-making process. Secondly, how to identify the features to be adjusted and how to measure the impact of this change on the behavior of the model, especially since in our case we have no possibility to simulate the results.

Our approach to overcoming the former issue is both intuitive and simple. Intuitive regarding the global approach adopted to answer the research problem of this paper. i.e., finding example cases to solve the case to be predicted. Simple since all the required computations to extract the example cases have already been carried out in the previous steps. Specifically, the cluster of neighbors (similar days) of the day to be predicted is already computed in the retrieving similar days' step (see Section 4.5). From the set of days in the cluster of neighbors, the appropriate day to be presented as an example of explanation is the one that has the second-best satisfaction since the best one is the day recommended to the occupant.

The generation of explanations consists in examining the difference between the scenarios of the two days to generate an indicator of the influence of the change of action variables on the effects. Since one of the objectives of our approach is to make the occupant aware of the influence of his actions on the energy behavior of the building, it is therefore interesting to show him the effect of modifying his actions on the building's energy performance. The comparison between the two scenarios consists thus of the calculation of two values for each time slice:

- the difference on each action variable for each corresponding time slot of scenario S_c^{best} (corresponding to the recommended action plan) and scenario $\tilde{S}c$ (corresponding to the explanatory example). It corresponds to Equation (15).

$$\forall \Lambda^{best} \in \{S_c^{best}\}, \forall \tilde{\Lambda} \in \{\tilde{S}c\}, \Lambda_k^{best} - \tilde{\Lambda}_k = \Delta \Lambda_k \quad (15)$$

Where Λ^{best} , $\tilde{\Lambda}$ is the action plan of the recommended scenario S_c^{best} , explanatory scenario $\tilde{S}c$ respectively.

- the difference on each effect variable (indoor temperature, air quality, etc.) for each corresponding time slot of scenario S_c^{best} and scenario $\tilde{S}c$. This corresponds to Equation (16).

$$\forall \xi^{best} \in \{S_c^{best}\}, \forall \tilde{\xi} \in \{\tilde{S}c\}, \xi_k^{best} - \tilde{\xi}_k = \Delta \xi_k \quad (16)$$

Where ξ^{best} , $\tilde{\xi}$ is the effect variables of the scenario S_c^{best} , $\tilde{S}c$ respectively.

The comparison between the scenarios provides quantitative information about the difference between the actions of the two scenarios and between the associated effects. Generally, people are reluctant to accept numerical knowledge and prefer knowledge that draws on a qualitative aspect to describe a phenomenon [22]. For instance, it is desirable for an occupant to have a recommendation like opening the door at 4:00 pm could create an airflow that will improve the air quality than to say, opening the door at 4:00 pm could increase the airflow by 25% that would allow decreasing the CO₂ concentration by 700 ppm and the humidity by 10%.

For the sake of acceptability, it is, therefore, worthwhile to turn the obtained numerical values into qualitative values. This transformation is obtained by dividing the continuous domain of each action/effect variable into a limited number of intervals (discretization process), resulting in a non-overlapping split of the continuous domain. Each interval is then mapped to a discrete value (label). For a granularity k , the number of intervals associated with each variable is defined by a symmetric partition of $(2k + 1)$ classes with an interval having only one value 0. Each class is labeled by the upper limit of its interval. So, the set of associated labels is denoted by an ordered set B of $(2k + 1)$ values as shown by the equation (17).

$$B = \{-a_{k-1}, \dots, -a_0, 0, a_0, \dots, a_{k-1}\} \quad (17)$$

Discretization values are defined by a domain expert based on their impact on the occupant perception. The transformation process of the numerical values into its qualitative form consists of using 1 to k arrows to represent visually the change magnitude corresponding to each discretization label and the associated sign of variation (i.e., arrows up for positive values and down for negative values). The qualitative conversion process is described by the pseudocode presented in Algorithm 3, whereas the pseudocode representing the generation of counterfactual explanations is shown in the Algorithm 2.

5.3. Causal-based explanation generation

The two types of explanations suggested earlier are intuitive and easy to understand, they are therefore intended for occupants who have no background in the domain or who are not looking for a deep explanation of the reasoning that led to the proposed action plan. For the wise user, the explanations should be causal to be consistent with the user's expectations of the system's reasoning process.

Causal explanations are conditioned on the existence of causal information. This work provides a probabilistic method for generating causal information and for estimating the magnitude of the impact of each causal variable (action or context variable) on the effect variables based on Bayesian networks.

Bayesian networks are a category of probabilistic graphical methods consisting of a set of random variables, the eventual dependencies between variables are depicted by directed acyclic graphs. The starting node of an edge is a causal event, referred to as the parent node, and the target node is a consequence of the cause, referred to as the child node. Each node is assigned a conditional probability table from information provided by either domain experts or statistical correlations that reflect the influences of the parent on its child nodes. The use of Bayesian networks allows obtaining a compact representation of the sets of dependencies between variables thanks to the graph structure (see section 5.3.1) and the notion of conditional probability tables (see section 5.3.2), from which it will be easier to reason.

The benefit of Bayesian networks in explanations generation is the ability to directly exploit the causal links (arcs)

Algorithm 2: Tabulating counterfactual explanations

Input: Best scenario plan

$$S_c^{best} = \{\Lambda^{best}, \xi^{best}, \zeta\}$$

// Get the best scenario from Section 4.6

Output: Γ : table for counterfactual explanations

 1 Obtain action plan from S_c^{best} : $\rho^{best} = \{\Lambda^{best}\}$

 2 Obtain effects vector from S_c^{best} : $V_E^* = \{\xi^{best}\}$

3 Obtain the scenario of a similar context day

$$\tilde{S}_c = \{\tilde{\Lambda}, \tilde{\xi}, \zeta\}$$

 4 Obtain action plan from \tilde{S}_c : $\tilde{\rho} = \{\tilde{\Lambda}\}$

 5 Obtain effects vector from \tilde{S}_c : $V_{\tilde{E}} = \{\tilde{\xi}\}$

 6 **for** $t = t_{start}$ **to** t_{end} **do**

 7 $row \leftarrow t - t_{start} + 1$

 8 $\Gamma_{row,1} \leftarrow t$

 9 Get Λ_t^{best} from ρ^{best} // Different actions at instance t

 10 Get $\tilde{\Lambda}_t$ from $\tilde{\rho}$

 11 Compute $\Delta\Lambda_t = \Lambda_t^{best} - \tilde{\Lambda}_t$

 12 Get ξ_t^{best} from ξ^{best} // Different effects at instance t

 13 Get $\tilde{\xi}_t$ from $\tilde{\xi}$

 14 Compute Υ_t^{best} from ξ^{best} // Get user satisfaction at instance t

 15 Compute $\tilde{\Upsilon}_t$ from $\tilde{\xi}_t$

 16 Compute $\Delta\Upsilon_t = \Upsilon_t^{best} - \tilde{\Upsilon}_t$

 17 $\Gamma_{row,2} \leftarrow$ Qualitative conversion ($\Delta\Lambda_t$)

 18 $\Gamma_{row,3} \leftarrow$ Qualitative conversion ($\Delta\Upsilon_t$)

 19 **end for**

 20 $\Gamma_{(row+1),3} \leftarrow$ Qualitative conversion ($mean(\Delta\Upsilon_t)$)

Algorithm 3: Qualitative conversion

Input: q : quantitative value

$$B = \{-a_{k-1}, \dots, -a_0, 0, a_0, \dots, a_{k-1}\}$$

 // Ordered list of discretization labels associated to q
Output: Qualitative explanation

 1 **if** $q \neq 0$ **then**

 2 $d \leftarrow$ discretization(q, K)

 3 $ind = index(d)$

 4 $n \leftarrow ind \bmod k$

 5 **if** $d > 0$ **then**

 6 insert n up arrows

 7 **else**

 8 insert $(k - n)$ down arrows

 9 **end if**

 10 **else**

11 insert '='

 12 **end if**

in the graphical representation to give causal explanations to the occupant. A direct causality between two variables is materialized by the existence of a direct arc between two nodes representing the two variables, however, indirect causality is established if a path (composed of at least two arcs) exists between the two nodes. Furthermore, the cause and effect relations between the variables are not deterministic but probabilistic. Thus, the observation of a cause or several causes does not systematically lead to the effects that depend on them but only modifies the probability of observing them. This allows classifying the causes from the most likely to the least one.

The construction of a Bayesian graph includes two tasks, learning the structure of the graph and learning some associated parameters (conditional probability tables).

5.3.1. Learning the graph structure

Learning the graph structure consists in determining both the set of nodes representing the random variables and the arcs representing the causal links between these variables. This step aims to find the globally optimal structure that fits the given data. Two approaches are used in the literature to define the structure of a graph. The principle of the constraints-based approach is to look for all the relations of conditional independence between variables to obtain the Markov coverage of each variable and thus the structure of the graph. The Markov blanket of node N is the only knowledge needed to predict the behavior of this node, it is composed of the parents of N , its children, and the parents of its children. The process of searching for the best structure is generally very expensive, so the second approach is often preferred. Score maximization-based approaches consist in maximizing a given function that evaluates the scores on a combinatorial space of the different graphs generated from the data set. So, learning the graph structure necessitates a structure definition mechanism and a scoring function to evaluate it. Many heuristics are used in the literature to define the structure of a graph, we used the Hill-Climbing algorithm to create different combinations of nodes and arcs to generate graphs. This choice is motivated by the fact that this algorithm allows finding the optimal balance between the computing requirement and the accuracy of the obtained graph.

The algorithm is initialized with a random graph where arcs randomly created between vertex representing context, action, and effect variables. This initial graph is attributed a score calculated thanks to a function f_{score} . The algorithm then performs a random change in the graph (remove, invert, or add arrows). If this modification leads to no increase in the score, the algorithm goes back to the previous state and makes another modification. However, if the score increases, the new state is kept and a new modification is performed. The algorithm terminates when no modification increases the score. There are several score functions whose purpose is to facilitate the selection of the statistical model representing reality as closely as possible. We applied the Bayesian Information Criterion (BIC) algorithm [84] to estimate the

score of the generated structures. It allows searching for a compromise between over-learning and under-learning. The chosen structure is the one that minimizes the BIC score. The corresponding pseudocode is shown in Algorithm 4.

Algorithm 4: Causal explication graph building

Input: Db : Database of context, action, and effect variables
 V : Set of vertex representing context, action, and effect variables
 f_{score} : A score function
Output: A graph maximizing f_{score}

```

1  $G \leftarrow$  empty graph
2  $G_{optimal} \leftarrow$  random graph
3  $Score_{best} \leftarrow f_{score}(G_{optimal})$ 
4 while  $G_{optimal} \neq G$  do
5    $G \leftarrow G_{optimal}$ 
6   foreach  $v_i \in V$  do
7     foreach  $v_j \notin parents(v_i)$  do
8       if  $f_{score}(G + \widehat{v_j v_i}) > Score_{best}$  and
           $\widehat{v_j v_i}$  does not form a loop in  $G$ 
          then
9          $G_{optimal} \leftarrow (G + \widehat{v_j v_i})$ 
10         $Score_{best} \leftarrow f_{score}(G + \widehat{v_j v_i})$ 
11      end if
12    end foreach
13    foreach  $v_j \in parents(v_i)$  do
14      if  $f_{score}(G - \widehat{v_j v_i}) > Score_{best}$  then
15         $G_{optimal} \leftarrow (G - \widehat{v_j v_i})$ 
16         $Score_{best} \leftarrow f_{score}(G - \widehat{v_j v_i})$ 
17      end if
18    end foreach
19    foreach  $v_j \in parents(v_i)$  do
20      if
           $f_{score}(G + \widehat{v_i v_j} - \widehat{v_j v_i}) > Score_{best}$ 
          then
21         $G_{optimal} \leftarrow (G + \widehat{v_i v_j} - \widehat{v_j v_i})$ 
22      end if
23    end foreach
24  end foreach
25 end while
26 return  $G_{optimal}$ 

```

5.3.2. learning the parameters

In this section, we assume that we already know the structure of the Bayesian network to be learned thanks to the previous step and that we only need to evaluate the conditional probability tables (parameters of the Bayesian network).

Definition 5. A parameter ϑ_i is a table consisting of the set of probabilities for each of the possible values of the variable V_i given each of the possible configuration or combination values of its parents $\wp(V_i)$.

Let V_1, \dots, V_n be n random variables, $V = \{\zeta, \Lambda, \xi\}$, and DB a complete database of examples. Let also denote $\{\vartheta_1, \dots, \vartheta_n\}$ the parameters of the model to be estimated. The variable V_i can take its values in $\{v_{i1}, \dots, v_{ik}\}$. Let N_{ij} be the number of times the set of parent variables of V_i in G , $\wp(V_i)$, takes its j^{th} configuration in D , for all $j \in \{1, \dots, qi\}$ and let N_{ijk} be the number of times where $V_i = k$ and $\wp(V_i)$ takes its j^{th} value in the database D , for all $k \in \{1, \dots, r\}$.

In the where all variables are observed, the simplest and most widely used method is the statistical estimation which consists in estimating the probability of an event by the frequency of occurrence of this event in the database. This approach, called maximum likelihood, provides an estimate of the parameter ϑ_{ijk} by applying the Formula (18).

$$\vartheta_{ijk} = \frac{N_{ijk}}{\sum_{k=1}^r N_{ijk}} \quad (18)$$

6. Experimental analysis

The two main points we look to verify in our experimental study are the following: (1) Validate our working hypothesis (see Hypothesis 1). (2) Validate that the action plan proposed to the occupant improves his comfort. To that end, the experimental process consists of the following steps:

1. description of the data used for the validation.
2. evaluation of the genetic algorithm's performance to estimate the features' weighting.
3. discussion on the threshold of the distance T_C calculated by our approach to retrieve similar days.
4. checking the validity of our working hypothesis.
5. assessment of the proposed action schedule's performance.
6. presentation of the results of the implementation of the different explanation strategies.

6.1. Setup of testbed

To evaluate the performance of the proposed approach, an experiment was carried out on realistic data obtained by deploying several sensors in a university office building occupied by several researchers. The sensors collect mainly information on the indoor environment conditions as humidity, temperature, occupancy, window opening, etc. These data are enriched by information about the outdoor environment conditions using weather forecasts provided by a service provider. Table 2, presents the recorded features indicating their units and the relevant category (context, action, effect) to which they were affected.

While the values recorded in the base of cases represent the average of the values obtained from the sensors and the weather forecast service during a time step (in

this experiment, a time step of one hour is adopted), an exception is made concerning the *Window_opening* and the *Door_opening* variables, since their values are the percentage of opening in regards to the time step. For instance, $Door_opening = 0.2$ in the k^{th} hour of the day, indicates that the door was open for 12 minutes (60×0.2) during that hour.

Data were registered for a continuous period beginning in 2015-04 and extending through 2016-10. Data are filtered to eliminate periods with missing values, weekends, and public holidays (since the office is unoccupied on these days and therefore there are no recorded actions). After screening, the retained periods are: 2015-05-01/2015-05-28, 2015-06-01/2015-07-23, 2016-05-11/2016-05-31, 2016-06-02/2016-06-19 and 2016-06-22/2016-07-30. Altogether, 98 days were retained. The base of cases is randomly split into disjointed training and testing data (75% and 25% respectively). The final train data consists of 74 days and the test data contains 24 days. Knowing that the periods retained were during the season of ambient temperature see high, the heating was switched off.

For results reproducibility, a knowledge model of the office in which the sensors have been installed is used. The model is presented in [7], it describes mathematically, the physical interactions between the office characteristics. Specifically, it is an implementation of the thermal and CO₂ models of the office. These models will allow reproducing the effects ξ^{best} of the proposed action plan Λ^{best} in the context ζ . To ensure that the results are consistent, any real effect features in the recorded data are overwritten with the simulated ones using the office knowledge model. Therefore, the objective of applying the knowledge model is twofold:

$$\left\{ \begin{array}{l} \zeta, \Lambda^{best} \xrightarrow{\text{production}} \xi^{best} // \text{use KM model to produce results} \\ \zeta, \Lambda \xrightarrow{\text{simulation}} \xi // \text{use KM to ensure results consistency} \end{array} \right.$$

Where Λ^{best} - recommended actions, Λ - real actions, KM - knowledge model.

Experiments were conducted on an Intel® Core™i5-2,7 GHz processor, 16 GB RAM machine, running Python 3.8.

6.2. Variables weighting

The proposed genetic algorithm was applied to the dataset described above to generate weights for different variables. The optimisation procedure is ended after a maximum of 500 iterations. However, to save time, if the minimum distance of a solution from the best solution changes by less than 10^{-5} over 15 iterations, the optimisation is interrupted.

The genetic algorithm was able to compute the weights reported in Table 3 corresponding to the context and action variables described in Table 2. The algorithm has converged in 4784 seconds after 11 iterations.

It is observed from results reported in Table 3 that in the case study, the variable that models the temperature of the corridor has the highest impact on effect variables

($W_{T_{corridor}} = 0.21136$) whereas the variable models outside nebulosity has the lowest influence on effect variables compared to the other context and action variables ($W_{Nebulosity} = 0.00391$).

6.3. Threshold of context distance

This section discusses through the case study how the F1-measure is used to learn optimal context distance for nearest similar days.

The result of application of the proposed approach to determine the context distance threshold is schematized in the Figure 5. We observe that if the threshold distance T_C rises, the number of similar days (recall curve) that would be detected also rises, unlike the accuracy (precision curve) that evolves inversely. While the precision of prediction is very high (99%) when $T_C = 0.4$, the percentage of similar days (recall) is low (2%). Inversely, if $T_C = 1.5$, the recall is very higher (96%) but the accuracy is the only 52%. For the case study, the F1-measure reaches its maximum (72%) at $T_C = 1.2$ when precision is 64% and recall is 81%.

6.4. Hypothesis proving

The objective of this step is to check that the hypothesis about dependencies between the action, context and effect variables is valid (Hypothesis 1). In concrete terms, this amounts to verifying that the effects ξ^{best} of a day resulting from the application of the best plan Λ^{best} extracted from a historical day D^{hist} should be similar to the effects ξ^{hist} of the latter-day.

For each day D from the 24 test days, assuming $\Delta\xi = |\xi^{best} - \xi^{hist}|$ is the difference between the effects of applying the best action plan to day D and the reel effects of the best historical day D^{hist} . For each effect variable (temperature, CO₂ concentration), the difference is given by the difference between the values for each hour of the two days. Table 4 reports the obtained results. Specifically, the maximum difference of temperature ΔT does not exceed 2 °c and the maximum difference of CO₂ concentration ΔCO_2 is less than 400 ppm. This means that the days to be predicted and the days with the best action plan are similarly based on the expert rules (see Section 4.5). This illustrates the correctness of our starting hypothesis.

As an illustration, Figure 6 shows the predicted temperature and CO₂ concentration for the days 2016-07-09 and 2016-07-10 as well as the same values for the day 2015-07-21, identified as the best historical day. The results show that their effects are similar, since for 2016-07-09, the maximum temperature difference ΔT , which is obtained at 17:00, does not exceed 1.7 °c, and the same value for CO₂ concentration ΔCO_2 is recorded at 18:00 and does not go beyond 140 ppm. For 2016-07-10, the maximum ΔT does not exceed 0.9 °c (at 19:00) and the maximum ΔCO_2 is 150 ppm (at 18:00).

6.5. Evaluating the recommended action plan

Table 2
Features in building.

Feature	Unit	Description	Type
Toffice_reference	°c	Indoor temperature	Effect
Office_CO2_concentration	ppm	Indoor CO ₂ concentration	Effect
Tcorridor	°c	Corridor temperature	Context
Illuminance (new, old)	lux	Luminosite in office	Context
Solar radiation	w/m ²	Solar radiation in office	Context
Wind speed	m/s	Outside wind's speed	Context
Corridor_CO2_concentration	ppm	Corridor CO ₂ concentration	Context
Power (block east, west, total)	w	Electricity consumption	Context
Theat	°c	Heater temperature	Context
Occupancy		Number of occupancy in office	Context
Nebulosity	%	Outside nebulosity	Context
Tout	°c	Outside temperature	Context
Windows_opening	%	The duration of opening window	Action
Door_opening	%	The duration of opening door	Action

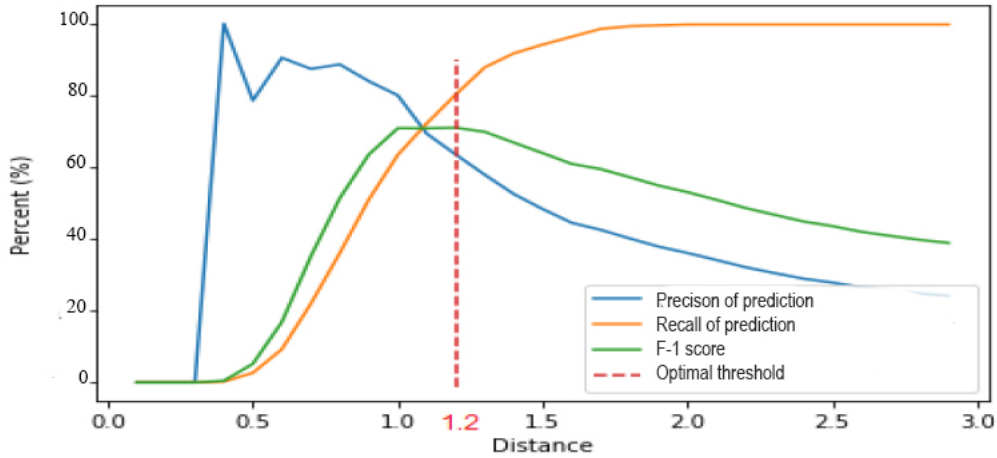


Figure 5: Threshold of context distance.

The present part aims to gauge the effectiveness of the approach by assessing the improvement (or not) of the occupant satisfaction following the carrying out of the suggested actions. The evaluation principle consists in comparing, for each day D from the test set, the effect ξ^{best} of the application of recommended actions Λ^{best} with the real result ξ obtained by applying the real actions Λ (actions performed by the occupant without support from EMS) on the context of the same day.

The improvement evaluation of the satisfaction Υ^{best} achieved by the effect ξ^{best} following the carrying out of the suggested actions Λ^{best} , against the satisfaction Υ obtained from applying the real actions Λ for the day D , is calculated by the equation (19).

$$H_D(\Upsilon^{best}, \Upsilon) = \frac{\Upsilon^{best} - \Upsilon}{abs(\Upsilon)} \quad (19)$$

Where $abs(\Upsilon)$ - absolute value of Υ .

$H_D(\Upsilon^{best}, \Upsilon) > 0$, implies that the advised actions have enhanced the occupant satisfaction regarding the real actions. The accuracy of the proposed method is assessed according to the function (20).

$$Accuracy = \frac{|N_D^{best}|}{|N_D|} \quad (20)$$

With - $N_D = \{D(\Upsilon^{best}, \Upsilon)\}$, the set of test days.

- $N_D^{best} = \{D(\Upsilon^{best}, \Upsilon) \in N_D, H_D(\Upsilon^{best}, \Upsilon) > 0\}$

This experiment considers two criteria to rate the occupant's satisfaction, thermal and air quality comfort. The

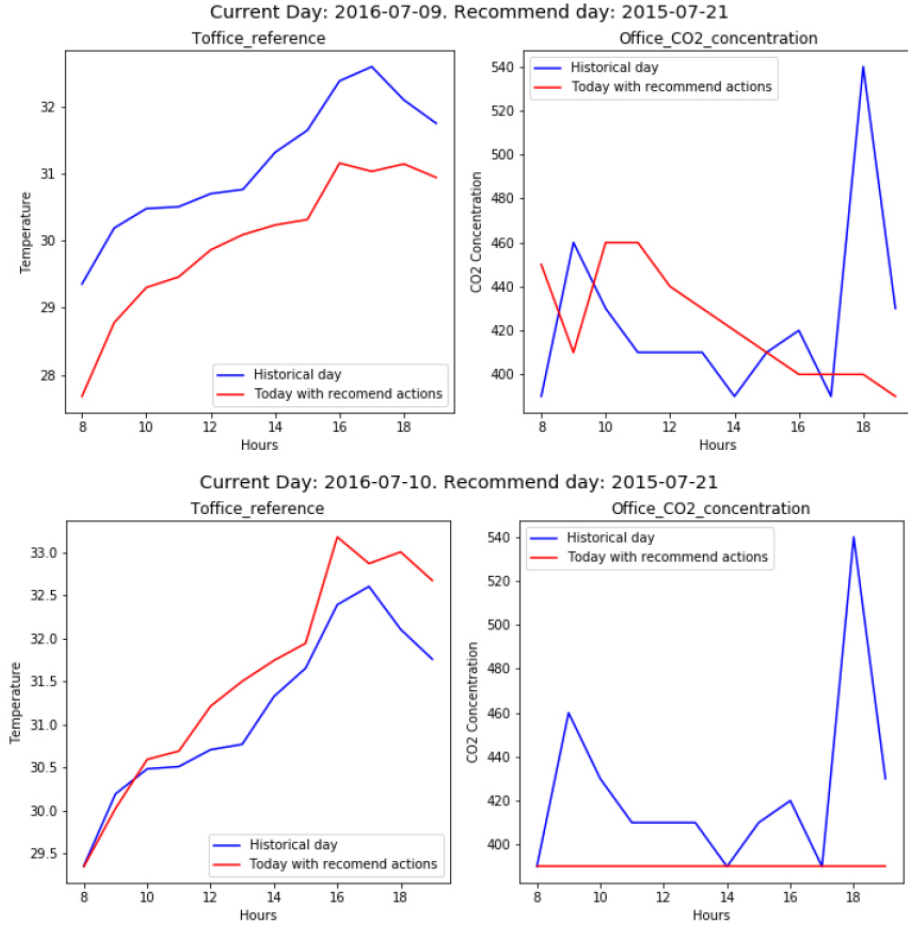


Figure 6: Example applying recommended actions.

importance of the temperature dissatisfaction is equivalent to that of the air quality (i.e., $\Omega^T = \Omega^{CO_2} = 0.5$). In this case, the global dissatisfaction at the i^{th} hour defined by the Equation (12) becomes:

$$\bar{Y}^i = 0.5 * \bar{Y}^T + 0.5 * \bar{Y}^{CO_2} \quad (21)$$

With a context distance threshold $\tau_C = 1.2$ and owing to the limitation of the data, the approach misses to find similar days for 11 days among the 24 test days, and therefore, we can suggest to the occupant actions for 13 days. ($|N_D| = 13$). To be consistent with the results, when the process of extracting similar days fails, the recorded actions of the test day are considered for the simulation. The obtained results of the comparison between the satisfaction resulting from the applying of the recommended action plan and the satisfaction resulting from the applying of the real actions are shown in Figure 7, demonstrating that over the 13 days, the advisable actions improve performance by 10 days ($|N_D^{best}| = 10$), showing the effectiveness of the proposed method in 77% of cases. We can observe that the improvement is in the range of 8% to 18%.

This is confirmed by plots in Figure 8 which illustrate two examples of test days and their effect variables as results

of the execution of the real and recommended actions. We can easily observe for both days that the overall performance of our approach is superior to that of the real actions performed by the occupant without assistance. Indeed, it is worthwhile observing that our approach succeeded in reaching the optimal air quality comfort by limiting the CO_2 concentration at values significantly lower than the upper limit (500 ppm) of the optimal comfort range. For the day of 2016-07-14, the proposed action plan improves the overall comfort by almost 18%, since it allows to bring the indoor temperature even closer to the optimal thermal comfort interval. Note that our approach allows to reduce the temperature by $1^\circ C$ on average and the maximum temperature difference is obtained at 08H00 with a reduction of $1.39^\circ C$. Analogously, for the day of 2016-07-16, the temperature is improved by $0.57^\circ C$ on average with a maximum reduction of $0.77^\circ C$ at 19H00, thus improving the comfort by 11.4%.

6.6. Explanation

Features-based explanations

Features' weights generated in the Section 6.2 thanks to the genetic algorithm we proposed in Section 4.4 are used to explain to the occupant the importance of different features in the prediction of an action plan. Figure 9 shows the

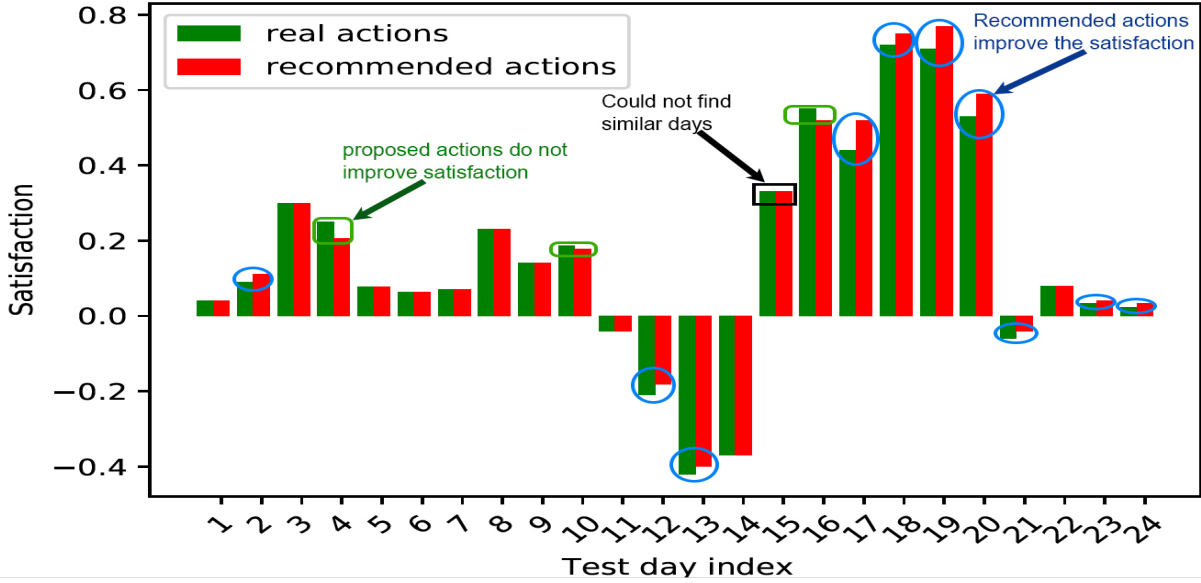


Figure 7: Performance evaluation regarding satisfaction improvement.

Table 3
Generated weights.

Variable	Weight
Tcorridor	0.21136
Illuminance (new)	0.10247
Illuminance (old)	0.08193
Solar radiation	0.02583
Wind speed	0.04188
Corridor_CO ₂ _concentration	0.04069
Power block east	0.08788
Power block west	0.05852
Power block total	0.05186
Theat	0.10789
Occupancy	0.03950
Nebulosity	0.00391
Tout	0.02946
Windows_opening	0.01576
Door_opening	0.09380

importance of the five features having the highest influence on the behavior of the prediction system. It can be easily remarked that the feature models the temperature of the adjacent corridor to the office has a remarkably high impact compared to the other features.

Example-based explanations

In this experiment, the generation of example-based explanations consists of comparing the recommended plan of actions for the day to be predicted with the actions plan for the day with the second-best satisfaction in the neighborhood

of that day and their effects. In the Figure 10, the system explains to the occupant the choice of the action plan of the day of 2015-06-17 to recommend it for the day of 2016-07-25, by comparing it to the day having the second-best satisfaction which is the day of 2015-05-26. The last row of this table shows that the overall satisfaction in terms of temperature and air quality would be better if the proposed plan is adopted.

Causal-based explanations

To increase the accuracy of the Bayesian network in generating causal explanations, the learning of the Bayesian network is performed on the contextual neighbors of the day to be predicted. This is motivated by the fact that the learning of a global network from the overall base of cases is unlikely to succeed since the discovery of the causal relations between the variables of the days with distant context distances can generate conflicts and thus to the failure of the process. For example, the algorithm could infer the relation between heating and indoor temperature on days when the heating is on, while this explanation is not consistent for a summer day when the heating is off.

The graph depicted in Figure 11 represents the causal relations between the different features on the day of 2016-07-25. The corresponding Bayesian network is learned from the data observed in a neighborhood composed of eight days within a context distance of $T_C = 1.2$ (see Section 6.3). Each type of feature is depicted by a different color, i.e. all orange nodes are effect variables, etc. The arc thickness is proportional to the importance of the dependence probability between a node and its parent. We can see that occupancy is a factor having a stronger influence on the features *Window_opening* and *Door_opening* than on the variable

Table 4
Summary of the validation results of the working hypothesis.

	ΔT ($^{\circ}C$)		ΔCO_2 (ppm)	
	Max value	Mean value	Max value	Mean value
Max difference	1.26	0.59	139.0	55.0
Mean difference	2.0	1.56	400.0	199.0

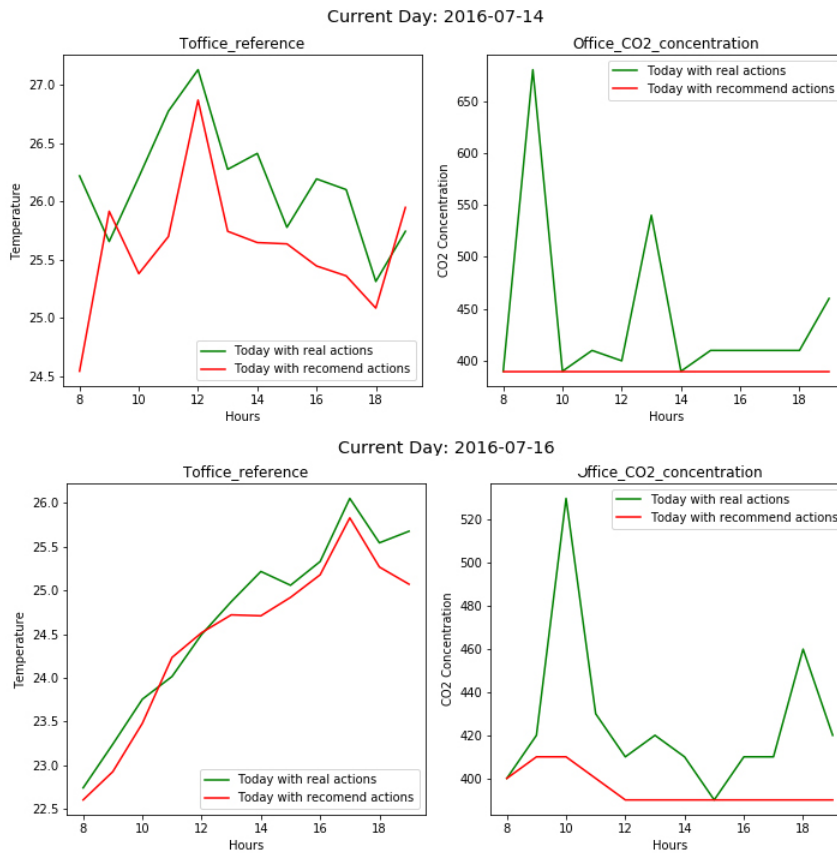


Figure 8: Comparison between recommended and recorded actions effects for 2016/07/14 and 2016/07/16.

total_electric_power. We observe also that the indoor temperature is impacted in the same proportion by the factors *Tcorridor*, *Door_opening*, and *total_electric_power*.

Discussion

The different types of explanations provided to the occupant aim at helping him to better understand the reasoning of the system, which in turn seeks to make the occupant aware of the impact of his behavior on energy use and help him to modify his behavior to improve this impact. We are aware that, reading a complex table such as the one generated to present the example-based explanations in Section 5.2 or observing a graph to understand the causal reasoning of the system such as the Bayesian network presented in Section 5.3, could be a rather embarrassing experience even for an experienced user.

To capture the occupant’s intention, it is desirable that the generated explanatory data be transformed into a natural

language. Indeed, prior work [74] has shown that verbalizing an explanation in a natural human language is preferred if it is destined to lay user. Although translating the graphical data into human language is not the focus of this study, we consider it a particularly important orientation for upcoming work.

7. Conclusion

This article lays the groundwork for a practical framework for enhancing building occupant comfort while optimizing energy expenditure. Specifically, a data-driven approach is used to calculate and recommend an action plan to the occupant to make the occupant aware of the impact of his behavior on increasing the energy efficiency of the building without scarifying his comfort. The approach presented tries on two schemes. Firstly, CBR is used to calculate the best action plan to propose to the occupant. It tries

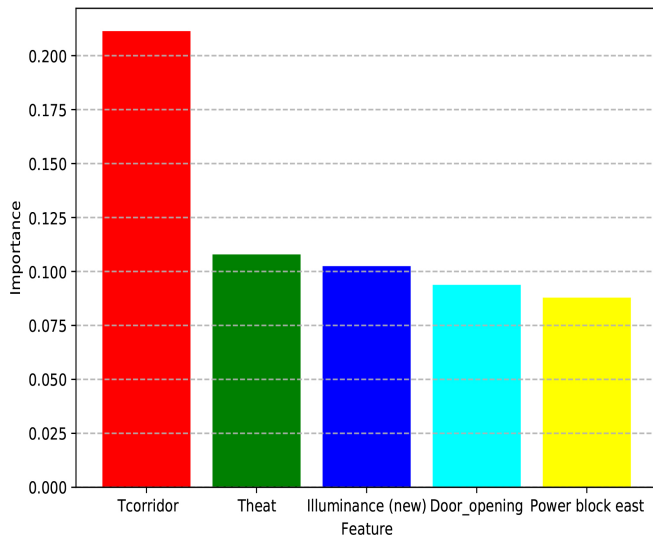


Figure 9: Features-based explanation

on an innovative method based on a genetic algorithm for features weighting and a straightforward but surprisingly efficient method to retrieve the set of actions to propose. Secondly, an explanation-oriented approach is used to foster the occupant's trust in the recommendations proposed by the system.

This technique is easily deployed in any type of environment without the need for domain expert knowledge. The experimentation performed with real data shows the feasibility of the technique. The test results of the CBR method indicate that in 77% of the cases the satisfaction of the occupant is improved up to 18%. To further enhance this work, some guidelines are planned for the future. We aim to investigate a more sophisticated approach to compute the best action plan using a local optimization method to generate an action plan from a model learned on the neighborhood of the day to be predicted. Also, we intend to integrate further criteria in the estimation of the occupant's comfort as the energy cost, the outdoor noise, and the natural lighting level. Furthermore, Like any data-based approach, the performance of the proposed approach may be significantly influenced by the quality and quantity of the data used, one future perspective is to explore how to extend the proposed method to handle cases without similar data. Finally, to avoid overloading the occupant, we plan to enhance the presentation interface of the recommendations and explanations by carrying out large-scale experiments.

References

- [1] Felfernig A and Gula B. Consumer behavior in the interaction with knowledge-based recommender applications. In *In ECAI 2006 Workshop on Recommender Systems*, page 37–41, 2006.
- [2] A. Aamodt and E. Plaza. Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Commun.*, 7:39–59, 1994.
- [3] Ashraf Abdul, Jo Vermeulen, Danding Wang, Brian Y. Lim, and Mohan Kankanhalli. Trends and trajectories for explainable, accountable and intelligible systems. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, apr 2018.
- [4] Maytham S. Ahmed, Hussein Shareef, Azah Mohamad, Jamal Abd Ali, and Ammar Hussein Mutlag. Rule base home energy management system considering residential demand response application. *Applied Mechanics and Materials*, 785:526–531, aug 2015.
- [5] D. Alvarez-Melis and T. Jaakkola. On the robustness of interpretability methods. *ICML Workshop on Human Interpretability in Machine Learning (WHI 2018)*, Stockholm, Sweden., 2018.
- [6] Ahmad Alwosheel, Sander van Cranenburgh, and Caspar G. Chorus. Is your dataset big enough? sample size requirements when using artificial neural networks for discrete choice analysis. *Journal of Choice Modelling*, 28:167–182, 2018.
- [7] Amr Alzouhri Alyafi, Monalisa Pal, Stephane Ploix, Patrick Reignier, and Sanghamitra Bandyopadhyay. Differential explanations for energy management in buildings. In *2017 Computing Conference*, pages 507–516, 2017.
- [8] S. Aman, Y. Simmhan, and V. K. Prasanna. Energy management systems: state of the art and emerging trends. *IEEE Communications Magazine*, jan 2013.
- [9] Fatima Amara and al. Comparison and simulation of building thermal models for effective energy management. *Smart Grid and Renewable Energy*, 2015.
- [10] Inc American Society of Heating, Refrigerating and Air-Conditioning Engineers. *Ansi/ashrae standard 62: Ventilation for acceptable indoor air quality*, 2019.
- [11] Refrigerating American Society of Heating and Atlanta. Air-Conditioning Engineers. *Ashrae standard 55-2010. thermal environmental conditions for human occupancy*, 2010.
- [12] Pascal Amphoux. Vers une théorie des trois confort. *Annuaire 90*, pages pp. 27–30, 1990.
- [13] Martin Anthony and Peter L. Bartlett. *Neural Network Learning: Theoretical Foundations*. Cambridge University Press, 1999.
- [14] Fabrizio Ascione, Nicola Bianco, Claudio De Stasio, Gerardo Mauro, and G.P. Vanoli. Artificial neural networks to predict energy performance and retrofit scenarios for any member of a building category: A novel approach. *Energy*, 118, 11 2016.
- [15] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bannetot, Siham Tabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, and Francisco Herrera. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information Fusion*, 58:82–115, 2020.
- [16] Gail Brager, Gwelen Paliaga, Richard de Dear, Bjarne Olesen, J. Wen, F. Nicol, and M. Humphreys. Operable windows, personal control, and occupant comfort. *ASHRAE Transactions*, 110:17–35, 01 2004.
- [17] Gaëtan Briseperrière. Pratiques de consommation d'énergie dans les bâtiments performants : consommations théoriques et consommations réelles. Technical report, ADEME - Agence de la transition écologique, July 2012.
- [18] Jiayu Chen, John E. Taylor, and Hsi-Hsien Wei. Modeling building occupant network energy consumption decision-making: The interplay between network structure and conservation. *Energy and Buildings*, 47:515–524, apr 2012.
- [19] Liang Chen. Agent-based modeling in urban and architectural research: A brief literature review. *Frontiers of Architectural Research*, 1(2):166–177, 2012.
- [20] Dan C. Cireşan, Ueli Meier, and Jürgen Schmidhuber. Transfer learning for latin and chinese characters with deep neural networks. In *The 2012 International Joint Conference on Neural Networks (IJCNN)*, pages 1–6, 2012.
- [21] Naomi Dassi and al. A case based reasoning coupling multi-criteria decision making with learning and optimization intelligences: Application to energy consumption. *EAI Endorsed Transactions on*

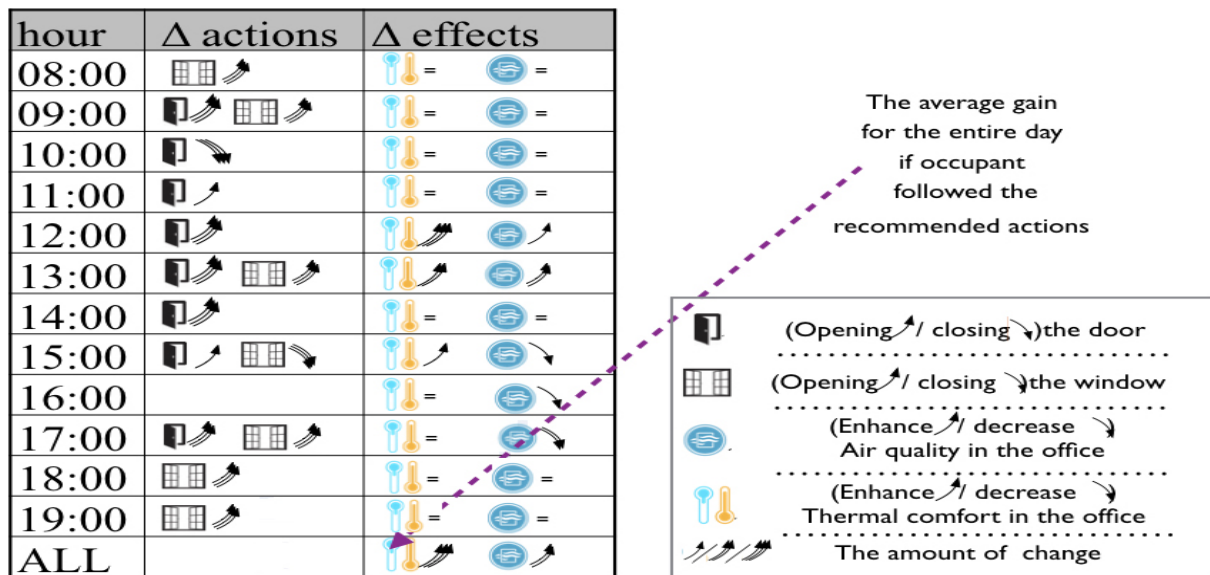


Figure 10: Example-based explanation

Smart Cities, 2019.

[22] Johan de Kleer. Qualitative and quantitative knowledge in classical mechanics. Technical report, USA, 1975.

[23] Ewart de Visser, Marvin Cohen, Amos Freedy, and Raja Parasuraman. A design methodology for trust cue calibration in cognitive agents. volume 8525, pages 251–262, 06 2014.

[24] Zhipeng Deng and Qingyan Chen. Artificial neural network models using thermal sensations and occupants’ behavior for predicting thermal comfort. *Energy and Buildings*, 174:587–602, sep 2018.

[25] Djamel Djenouri and al. Machine learning for smart building applications: Review and taxonomy. *ACM Comput. Surv.*, 52(2), March 2019.

[26] S. Dornbush, A. Joshi, Z. Segall, and T. Oates. A human activity aware learning mobile music player. In *AITAmI@IJCAI*, 2007.

[27] Marek Druzdzel. Verbal uncertainty expressions: Literature review. Technical report, University, Pittsburgh., 10 1998.

[28] Ricardo Faia and al. Case based reasoning with expert system and swarm intelligence to determine energy reduction in buildings energy management. *Energy and Buildings*, 155:269–281, 2017.

[29] Cheng Fan, Fu Xiao, Chengchu Yan, Chengliang Liu, Zhengdao Li, and Jiayuan Wang. A novel methodology to explain and evaluate data-driven building energy performance models based on interpretable machine learning. *Applied Energy*, 235:1551–1560, feb 2019.

[30] International Organization for Standardization. Ergonomics of the thermal environment—analytical determination and interpretation of thermal comfort using calculation of the pmv and ppd indices and local thermal comfort criteria, 2005.

[31] International Organization for Standardization. Building environment design - indoor air quality - methods of expressing the quality of indoor air for human occupancy, 2008.

[32] Jerome Friedman. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29:1189–1232, 10 2001.

[33] Fatih Gedikli, Dietmar Jannach, and Mouzhi Ge. How should i explain? a comparison of different explanation types for recommender systems. *International Journal of Human-Computer Studies*, 72(4):367–382, apr 2014.

[34] Alyssa Glass, Deborah L. McGuinness, and Michael Wolverton. Toward establishing trust in adaptive agents. In *Proceedings of the 13th international conference on Intelligent user interfaces - IUI '08*. ACM Press, 2008.

[35] Alex Goldstein, Adam Kapelner, Justin Bleich, and Emil Pitkin. Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. *Journal of Computational and Graphical Statistics*, 24, 09 2013.

[36] Alfonso González-Briones, Javier Prieto, Fernando De La Prieta, Enrique Herrera-Viedma, and Juan Corchado. Energy optimization using a case-based reasoning strategy. *Sensors*, 18(3):865, mar 2018.

[37] Alicja Gosiewska and P. Biecek. ibreakdown: Uncertainty of model explanations for non-additive predictive models. *ArXiv*, abs/1903.11420, 2019.

[38] J. C. Gower. A general coefficient of similarity and some of its properties. *Biometrics*, 27(4):857–871, 1971.

[39] Ruocheng Guo and al. A survey of learning causality with data. *ACM Computing Surveys*, 53(4):1–37, sep 2020.

[40] David Haussler. Decision theoretic generalizations of the PAC model for neural net and other learning applications. *Information and Computation*, 100(1):78–150, sep 1992.

[41] Jonathan L. Herlocker, Joseph A. Konstan, and John Riedl. Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work - CSCW '00*. ACM Press, 2000.

[42] Denis J. Hilton. Conversational processes and causal explanation. *Psychological Bulletin*, 107(1):65–81, 1990.

[43] Taehoon Hong, Choongwan Koo, and Kwangbok Jeong. A decision support model for reducing electric energy consumption in elementary school facilities. *Applied Energy*, 95:253–266, jul 2012.

[44] Luis Hurtado, Phuong Nguyen, W.L. Kling, and Wim Zeiler. Building energy management systems — optimization of comfort and energy use. pages 1–6, 09 2013.

[45] Norman Ihle. Case representation and adaptation for short-term load forecast at a container terminal., *24th International Conference on Case-Based Reasoning (ICCBR)*, 2016.

[46] Raghavan Chandran Ilambirai, Parthasarathy Sivasankari, Sankaramurthy Padmini, and Harish Chowdary. Efficient self-learning artificial neural network controller for critical heating, ventilation and air conditioning systems. In *THE 11TH NATIONAL CONFERENCE ON MATHEMATICAL TECHNIQUES AND APPLICATIONS*. AIP Publishing, 2019.

[47] Soteris Kalogirou, Constantinos Neocleousp, and Christos Schizas. Building heating load estimation using artificial neural networks. *Proceedings of the International Conference CLIMA 2000*, 01 1997.

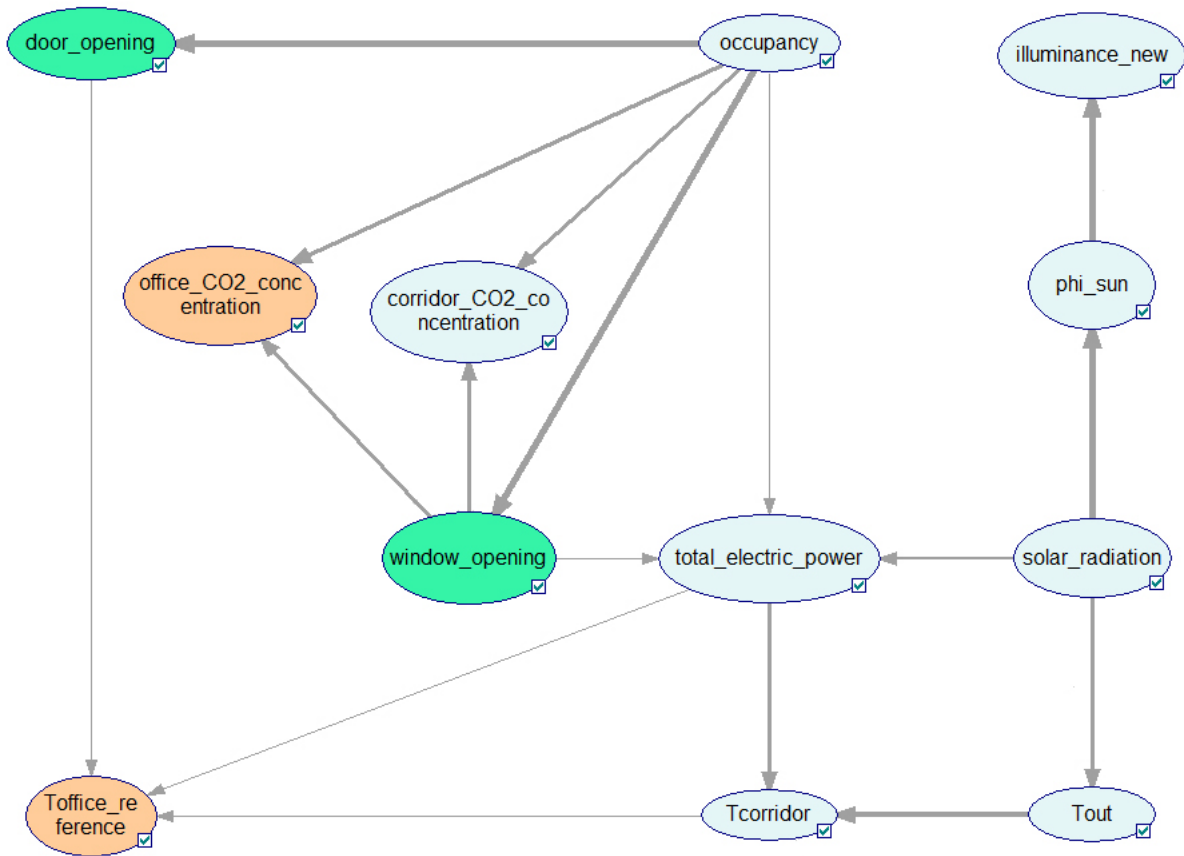


Figure 11: Causal explanations

- [48] Tomoya Kawakami, Naotaka Fujita, Tomoki Yoshihisa, and Masahiko Tsukamoto. An evaluation and implementation of rule-based home energy management system using the rete algorithm. *The Scientific World Journal*, 2014:1–8, 2014.
- [49] F. Keil. Explanation and understanding. *Annual review of psychology*, 57:227–54, 2006.
- [50] J. Kennedy and R. Eberhart. Particle swarm optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks*, volume 4, 1995.
- [51] William Kennedy. *Modelling Human Behaviour in Agent-Based Models*, pages 167–179. 01 2012.
- [52] Halime Khojamli and Jafar Razmara. Survey of similarity functions on neighborhood-based collaborative filtering. *Expert Systems with Applications*, 185:115482, 2021.
- [53] Janet Kolodner. An introduction to case-based reasoning. *Artificial Intelligence Review*, 6:3–34, 03 1992.
- [54] Janet Kolodner. *Case-Based Reasoning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.
- [55] Anna Krakovská and al. Comparison of six methods for the detection of causality in a bivariate time series. *Physical Review E*, 97, 04 2018.
- [56] Klara Królewski. Need for power. In *Encyclopedia of Personality and Individual Differences*, pages 1–3. Springer International Publishing, 2017.
- [57] Y Laaroussi, M Bahrar, M Elmankibi, A Draoui, and A Si-Larbi. Occupant behaviour: a major issue for building energy performance. *IOP Conference Series: Materials Science and Engineering*, 609:072050, oct 2019.
- [58] John D. Lee and Katrina A. See. Trust in automation: Designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 46(1):50–80, 2004.
- [59] Rebecca A. Kelly (Letcher) and al. Selecting among five common modelling approaches for integrated environmental assessment and management. *Environmental Modelling & Software*, 47:159–181, sep 2013.
- [60] Hui Li and Jie Sun. Ranking-order case-based reasoning for financial distress prediction. *Knowledge-Based Systems*, 21(8):868–878, 2008.
- [61] Zhengwei Li, Yanmin Han, and Peng Xu. Methods for benchmarking building energy consumption against its past or intended performance: An overview. *Applied Energy*, 124:325–334, 2014.
- [62] Ziwei Li, Jian Dai, Hongzhong Chen, and Borong Lin. An ANN-based fast building energy consumption prediction method for complex architectural form at the early design stage. *Building Simulation*, 12(4):665–681, may 2019.
- [63] T. W. Liao, Zhiming Zhang, and C. Mount. Similarity measures for retrieval in case-based reasoning systems. *Appl. Artif. Intell.*, 12:267–288, 1998.
- [64] Peter Lipton. Contrastive explanation. *Royal Institute of Philosophy Supplement*, 27:247–266, mar 1990.
- [65] N. Loganathan, P. Mayurappriyan, and K. Lakshmi. Smart energy management systems: a literature review. *MATEC Web of Conferences*, 225:01016, 01 2018.
- [66] Bloomberg Finance L.P. New energy finance new energy outlook, 2018.
- [67] D. Mariano-Hernández, L. Hernández-Callejo, A. Zorita-Lamadrid, O. Duque-Pérez, and F. Santos García. A review of strategies for building energy management system: Model predictive control, demand side management, optimization, and fault detect & diagnosis. *Journal of Building Engineering*, 33:101692, 2021.

- [68] Tim Miller. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267:1–38, feb 2019.
- [69] France Ministère de la transition écologique. Chiffres clés de l'énergie- édition 2019, 2019.
- [70] BONNIE M. MUIR. Trust in automation: Part i. theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics*, 37(11):1905–1922, nov 1994.
- [71] Cataldo Musto, Pasquale Lops, Marco de Gemmis, and Giovanni Semeraro. Justifying recommendations through aspect-based sentiment analysis of users reviews. In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*. ACM, jun 2019.
- [72] H. Q. Nguyen and al. Meta-optimization and scattering parameters analysis for improving on site building model identification for optimal operation. *Building Simulation*, 2016.
- [73] Ingrid Nunes and Dietmar Jannach. A systematic review and taxonomy of explanations in decision support and recommender systems. *User Modeling and User-Adapted Interaction*, 27(3-5):393–444, oct 2017.
- [74] Ingrid Nunes and Dietmar Jannach. A systematic review and taxonomy of explanations in decision support and recommender systems. *User Modeling and User-Adapted Interaction*, 27, 12 2017.
- [75] Nnamdi I. Nwulu. An artificial neural network model for predicting building heating and cooling loads. In *2017 International Artificial Intelligence and Data Processing Symposium (IDAP)*. IEEE, sep 2017.
- [76] Domenico Palladino, Iole Nardi, and Cinzia Buratti. Artificial neural network for the thermal comfort index prediction: Development of a new simplified algorithm. *Energies*, 13(17):4500, sep 2020.
- [77] Antonio Paone and Jean-Philippe Bacher. The impact of building occupant behavior on energy efficiency and methods to influence it: A review of the state of the art. *Energies*, 11(4):953, apr 2018.
- [78] N. Pennington and R. Hastie. Explanation-based decision making: effects of memory structure on judgment, 1988.
- [79] Gabriel Peschiera and al. Response–relapse patterns of building occupant electricity consumption following exposure to personal, contextualized and occupant peer network utilization data. *Energy and Buildings*, 42(8):1329–1336, 2010.
- [80] Radu Platon, Jacques Martel, and Kaiser Zoghalmi. Cbr model for predicting a building's electricity use: On-line implementation in the absence of historical data. volume 9343, 09 2015.
- [81] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "why should i trust you?". In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, aug 2016.
- [82] EDWINA L. RISSLAND, KEVIN D. ASHLEY, and L. KARL BRANTING. Case-based reasoning and law. *The Knowledge Engineering Review*, 20(3):293–298, 2005.
- [83] Raquel Rodríguez-Pérez and J. Bajorath. Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions. *Journal of Computer-Aided Molecular Design*, 34:1013 – 1026, 2020.
- [84] Gideon Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2), mar 1978.
- [85] Byeongmo Seo, Yeo Beom Yoon, Jung Hyun Mun, and Soolyeon Cho. Application of artificial neural network for the optimum control of HVAC systems in double-skinned office buildings. *Energies*, 12(24):4754, dec 2019.
- [86] Aulon Shabani, Adil Paul, Radu Platon, and Eyke Hüllermeier. Predicting the electricity consumption of buildings: An improved cbr approach. pages 356–369, 10 2016.
- [87] L.S. Shapley. A value for n-person games. In *Annals of Mathematics Studies*, number 28, pages 307–317. 1953.
- [88] Isha Sharma, Jin Dong, Andreas Malikopoulos, Michael Street, Jim Ostrowski, Teja Kuruganti, and Roderick Jackson. A modeling framework for optimal energy management of a residential building. *Energy and Buildings*, 130, 08 2016.
- [89] Tatyana A. Shkurko. Socio-psychological analysis of controlling personality. *Procedia - Social and Behavioral Sciences*, 86:629–634, oct 2013.
- [90] S. Sholahudin, Azimil Alam, Chang Baek, and Hwataik Han. Prediction and analysis of building energy efficiency using artificial neural network and design of experiments. *Applied Mechanics and Materials*, 819:541–545, 01 2016.
- [91] Barry Smyth and Mark T. Keane. Adaptation-guided retrieval: questioning the similarity assumption in reasoning. *Artificial Intelligence*, 102(2):249–293, 1998.
- [92] Constantinos Spandagos and Tze Ling Ng. Equivalent full-load hours for assessing climate change impact on building cooling and heating energy consumption in large asian cities. *Applied Energy*, 189:352–368, mar 2017.
- [93] G. Sugihara, R. May, H. Ye, C. h. Hsieh, E. Deyle, M. Fogarty, and S. Munch. Detecting causality in complex ecosystems. *Science*, 338(6106):496–500, sep 2012.
- [94] YouGov survey. Modern indoor living can be bad for your health. Technical report, 15 mai 2018.
- [95] Adnan Sözen, M. Ali Akçayol, and Erol Arcaklioğlu. Forecasting net energy consumption using artificial neural network. *Energy Sources, Part B: Economics, Planning, and Policy*, 1(2):147–155, jul 2006.
- [96] S.L. Ting, S.K. Kwok, Albert H.C. Tsang, and W.B. Lee. A hybrid knowledge-based approach to supporting the medical prescription for general practitioners: Real case in a hong kong medical center. *Knowledge-Based Systems*, 24(3):444–456, 2011.
- [97] Nava Tintarev. Explaining recommendations. In *User Modeling 2007*, pages 470–474. Springer Berlin Heidelberg.
- [98] A. Tóptsis and Alexander Dubitski. Iterative k-line meshing via non-linear least squares interpolation of affectively decorated media repositories. *The Open Artificial Intelligence Journal*, 2:46–61, 2008.
- [99] Jon Vadillo, Roberto Santana, and J. A. Lozano. When and how to fool explainable models (and humans) with adversarial examples. *ArXiv*, abs/2107.01943, 2021.
- [100] Sandra Wachter, Brent Mittelstadt, and Chris Russell. Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *SSRN Electronic Journal*, 2017.
- [101] Duch Włodzisław and Karol Grudziński. Weighting and selection of features. 01 1999.
- [102] L. Richard Ye and Paul E. Johnson. The impact of explanation facilities on user acceptance of expert systems advice. *MIS Quarterly*, 19(2):157, jun 1995.
- [103] Hiroshi Yoshino, Tianzhen Hong, and Natasa Nord. Iea ebc annex 53: Total energy use in buildings – analysis and evaluation methods. *Energy and Buildings*, 152, 07 2017.
- [104] Victor M. Zavala. Real-time optimization strategies for building systems. *Industrial & Engineering Chemistry Research*, 52(9):3137–3150, sep 2012.
- [105] Jianlong Zhou, Amir H. Gandomi, Fang Chen, and Andreas Holzinger. Evaluating the quality of machine learning explanations: A survey on methods and metrics. *Electronics*, 10(5):593, mar 2021.