Introduce structural equation modelling to machine learning problems for building an explainable and persuasive model

Li, Jiarui; Sawaragi, Tetsuo; Horiguchi, Yukio

JCMSI
SICE Journal of Control, Measurement, and System Integration (SICE JCMSI)
SICE

SICE Journal of Control, Measurement, and System Integration

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tmsi20

Introduce structural equation modelling to machine learning problems for building an explainable and persuasive model

Jiarui Li, Tetsuo Sawaragi & Yukio Horiguchi

To cite this article: Jiarui Li, Tetsuo Sawaragi & Yukio Horiguchi (2021) Introduce structural equation modelling to machine learning problems for building an explainable and persuasive model, SICE Journal of Control, Measurement, and System Integration, 14:2, 67-79, DOI: 10.1080/18824889.2021.1894040

To link to this article: https://doi.org/10.1080/18824889.2021.1894040

Published online: 19 Mar 2021.

SICE JOURNAL OF CONTROL, MEASUREMENT, AND SYSTEM INTEGRATION
2021, VOL. 14, NO. 2, 67–79
https://doi.org/10.1080/18824889.2021.1894040

Introduce structural equation modelling to machine learning problems for building an explainable and persuasive model

Jiarui Li [a], Tetsuo Sawaragi [a] and Yukio Horiguchi [b]

[a] Department of Mechanical Engineering and Science, Kyoto University, Kyoto, Japan; [b] Faculty of Informatics, Kansai University, Osaka, Japan

ABSTRACT

With the development of artificial intelligence technologies, the high accuracy of machine learning methods has become a non-unique standard. People are beginning to be more concerned about the understandability between humans and machines. The interference procedure of the machines is hoped to accord with human thinking as much as possible, which has spawned the recent and ongoing demands for developing explainable models. The present study proposes a new explainable and persuasive model for machine learning problems by introducing Structural Equation Modelling into the picture. Six parts make up the model, from data collection to model evaluation. The model can be used for data analysis, machine learning, and causal analysis. The proposed model is also transparent and can be interpreted from design to application. A practical experiment shows its effectiveness in a healthcare problem.

ARTICLE HISTORY
Received 30 October 2020
Revised 12 January 2021
Accepted 26 January 2021

KEYWORDS
Machine learning; structural equation modelling; explainable model; causal analysis; healthcare

1. Introduction

Machine Learning (ML), which is an application of Artificial Intelligence (AI) technology, is widely used nowadays to enable systems to learn from human experiences automatically. For rapidly and correctly making decisions, high accuracy is always regarded as a golden assessment index for an ML model [1]. However, the primary aim of ML is to teach machines to collaborate with a human user or replace human work. Thus, a machine should be highly praised if it can imitate humans as closely as possible. Besides, we hope machines can improve from the existing data and can cope with the changes such that they can become able to think and perform like a human to anticipate what might happen in the future. The targets mentioned above can only be achieved by explaining the mechanics of the decision-making procedure, which has spawned recent and ongoing demands for ML technology for developing explainable ML models.

Samek et al. [2] explained why an explainable ML model is necessary. First, the structure of the explainable model should be transparent so that domain experts can verify it. In the fields related to safety and security of people's lives and property, such as healthcare, law, and regulations, an ML model that does not conform to common sense is invalid. Second, explainability makes the models easy to optimize. If we thoroughly learn the decision procedure of a model, its weaknesses will be easily found at the same time. The explainable models tell us the basis of the machine's thinking, which supplies the channel for judging whether it is right or wrong. The most important function of an explainable model is the coping-with-change ability. This is the key to whether the model can predict the future. In a 2018 paper, Turing Award winner Judea Pearl [3] discussed the limitations of current ML theories. Because ML systems operate almost entirely in statistics or blind models, they cannot be used for strong AI. Only by making certain of the causal relationship in the ML system can the model react correctly to changes in the data. For a causal analysis, the structure of the model must be transparent and explainable.

Roscher and his team [4] illustrated that the readability of the model should have three levels: transparency, interpretability, and explainability. First, the most basic level is transparency. The transparent models can clearly show the data partitioning mode.

Furthermore, the level of interpretability demands a higher requirement for ML models. According to Roscher et al. [4], the interpretability-level ML models should interpret the specific structure of input and output so that humans can understand them. Decision Trees and their ensemble methods are the most easily interpreted ML models. Thus, multiple tree-based methods have been proposed to make the models interpretable [5,6]. The explainability is the level of involving the human aspects. In other words, explainability refers to the domain knowledge from human experts and pursues AI that performs more like human beings. For designing the explainable data structures, Bayesian Networks (BNT) and Structural Equation Modelling

CONTACT Jiarui Li ljr10225008@gmail.com Department of Mechanical Engineering and Science, Kyoto University, Kyoto 615-8510, Japan

(SEM) are two common tools. Many researchers have extended BNT technology to create explainable models. Constantinou et al. [7] developed a rigorous and repeatable method for building effective BNT models for medical decision support from complex and unstructured data. However, the whole procedure described in this paper needs the support from domain experts, which incurs much time and effort. Other similar works [8,9] also combine BNT with other ML methods, such as Neural Networks (NNs). BNT is a probability model, which cannot explain the correlations or causality among training data. In contrast, SEM is a well-known data modelling method expressed by a series of regression functions, which can intuitively describe the relationship among data features.

We [10] previously introduced SEM into ML problems for simplifying the training data dimensions. In this paper, we extend the previous work and propose a new explainable prediction model transformed by the analysis model provided by SEM. The proposed model can be easily implanted into other existing ML models and shows a competitive accuracy. Also, by an intervention procedure, the model can do causal analysis as well.

The remainder of this paper is organized as follows. In Section 2, the background knowledge of the SEM is reviewed. Section 3 details the specific procedures of the proposed model. In Section 4, the model is applied to a healthcare problem, and Section 5 discusses the results. Finally, concluding remarks are given in Section 6.

2. Structural equation modelling (SEM)

Structural Equation Modelling (SEM) is usually a two-step procedure. One is Exploratory Factor Analysis (EFA). The other is Confirmatory Factor Analysis (CFA).

EFA reliably classifies data items into corresponding factors without a specific hypothesis, which aims at identifying latent factors on the basis of the observed variables [11]. For a research topic, the result of EFA may not be unique. Researchers must balance the number of extracted factors avoiding both parsimony and plausibility. Hence, a repeated operation is necessary for EFA to obtain an excellent fitting model in the follow-up CFA procedure. A total explaining variance over 60% and a Kaiser-Meyer-Olkin (KMO) test result higher than 0.5 are the reference points of EFA.

In contrast to EFA, the hypothesis is necessary for the CFA procedure.Figure 1 shows a conceptual model of CFA. The measurement model and structural model make up the hypothesis for CFA to test. As mentioned above, EFA offers the results of extracted factors and their inclusive manifest variables, which builds up the measurement part. The structural part specifies the logic paths among factors. After constructing the model, the factor loadings between manifest items and latent factors and between every two factors are estimated in accordance with the covariance matrix of the manifest items. For example, the model shown as Figure 1 can be expressed as

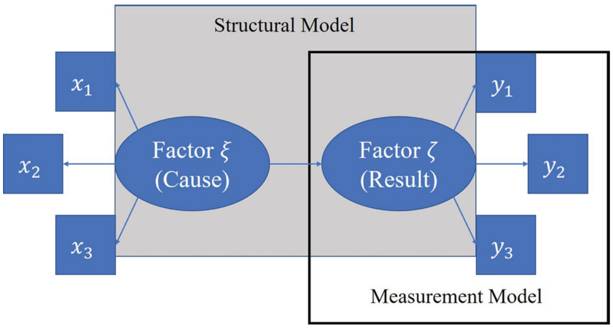

Figure 1. Conceptual model of CFA.

$$X = \Lambda_x \xi + \delta_x \tag{1}$$

$$Y = \Lambda_y \zeta + \delta_y \tag{2}$$

$$\zeta = \Gamma \xi + \varepsilon \tag{3}$$

where X and Y are 3-dimensions manifest variables. ξ and ζ are common factors measured by X and Y respectively. δ and ε are error terms. Using estimation methods, such as maximum likelihood estimation, the loading matrixes Λ_x and Λ_y are easy to calculate, which presents the factor loadings for each manifest variable to its latent factor. Moreover, Γ, the regression weight between two factors, can be estimated as well. The mark of a successful model is obtaining goodness of fit, proving that the hypothesis can express the structure of the data.

3. An explainable and persuasive machine learning model

3.1. Overall structure

The procedure of the proposed method contains six steps: data preparation, data management, structure learning, parameter learning, model utilization, and model validation. The overall structure is shown in Figure 2.

3.2. Data preparation

The starting point of the method is the preparation of data, before which the purpose of the model should be determined. Comprehensively considering all the possible related factors can save many resources for subsequent steps, such as the application fields, users' needs, and the quality of existing datasets. The necessary data should be collected corresponding to the experts' knowledge.

For easy illustration, in the following sections, we assume N-dimensions data have been collected for ML problem A.

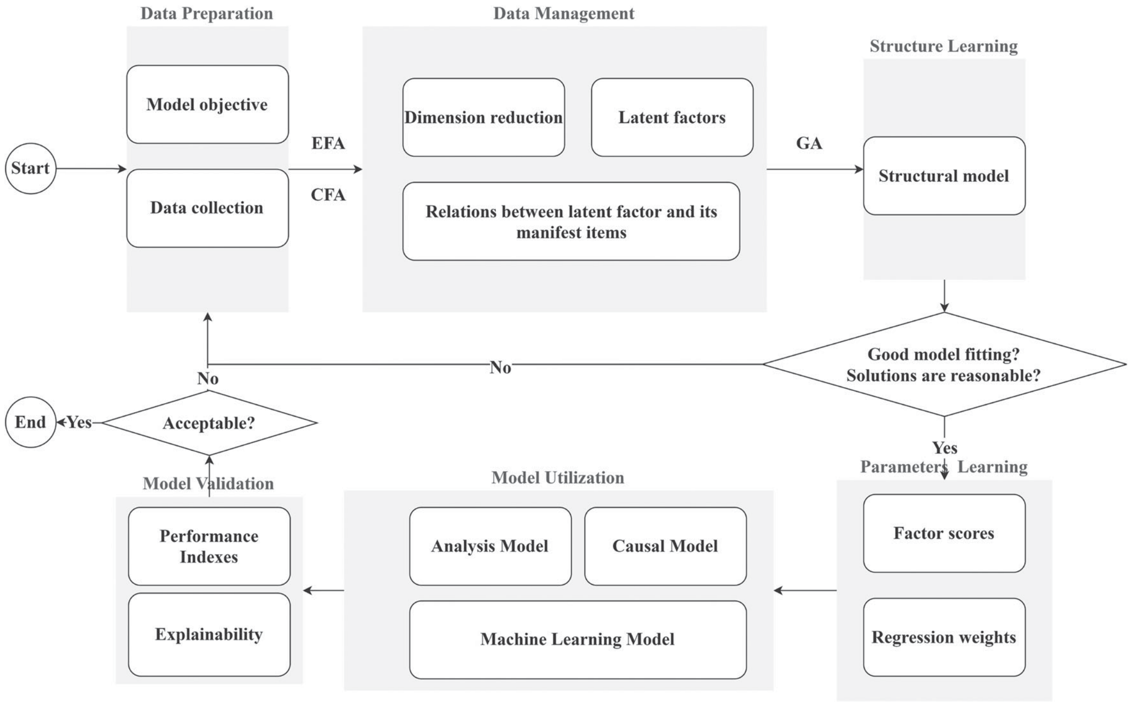

Figure 2. Overall structure of proposed method.

3.3. Data management

Data management aims at simplifying data dimensions, extracting latent factors, and verifying correlations between the latent factors and their manifest items. In the first step, the proposed method collects a large number of data features that relate to the learning target. However, the superfluous data dimensions inevitably cause a computational burden. Usually, not all collected characteristics contribute to the prediction goal. Thus, a filtering and dimensionality reduction process is necessary to extract the feature values closely related to the prediction goal and is sufficient to solve the ML problem.

The proposed method assumes that each dimension of the collected data is a manifest item in SEM, which is the input for data management. Moreover, data dimensions are reduced through EFA and CFA.

For data management, EFA is used to simplify the observed variable and extract latent factors. CFA is used for further reducing items that have low factor loadings to the corresponding latent factors. The initial dataset contains N-dimensions data. EFA gets rid of the variables and extracts a suitable number of factors. Through a factor rotation process, the calculated factor loadings evaluate the variables' ability to explain each common factor. A factor loading over a threshold (> 0.3 in the presented paper) presents the variable belonging to the corresponding factor. Factors that contain fewer than two items are inadvisable, and the final results are more convincing if every observed variable belongs to only one factor. Also, for different research purposes, researchers can reserve or remove factors in accordance with their experience. The final model should reach the reference points mentioned in Section 2.1. Let us assume that for problem A, EFA extracts 15 items belonging to 5 factors.

Next, CFA is used for further confirming the factor loadings. In this step, the emphasis is to verify whether the extracted manifest items are suitable to explain the corresponding factor, and the complexity of the relations among latent factors is not considered here. Also, the differences in connections among latent factors do not affect the factor loadings between the manifest items and its corresponding factor. Thus, the hypothesis model is made with all factors correlated with each other in this step. In EFA, all the factors are compulsively assumed to be mutually independent. However, a structural model used in CFA considers the regressions or correlations among the factors. As a result, the factor loadings obtained from CFA are usually lower than those obtained from EFA. That is why CFA contributes to further reduce data dimensions in this step.

In the example of problem A, the CFA result shows that the factor loading of item 7 is 0.2, which is not suitable for measuring factor 3, so item 7 is removed from the dataset. Finally, the data management procedure extracts 14 items and 5 factors, shown as Figure 3.

3.4. Structure learning

The structure learning procedure aims at specifying the relations between every two latent factors and finding out the best model fitting on the given data. When there is enough domain knowledge, the structure can be given by the experts. Nevertheless, a more automatic way is to use the heuristic method. In the proposed

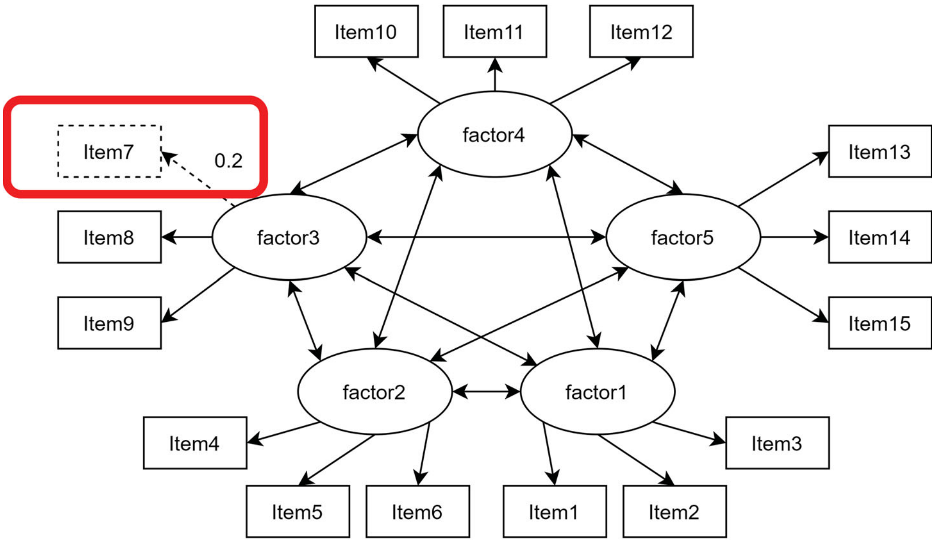

Figure 3. Data management result.

method, we use Genetic Algorithm (GA) to conduct the structure learning procedure, and the steps of applying GA in SEM are as follows.

Step1. Determine the fitness indicators;

Step2. Code the chromosomes and set evolution parameters;

Step3. Generate the initial population and perform pre-evolution iterations for finding "suggestions";

Step4. Add the "suggestions" to the initial population and conduct the evolution steps.

3.4.1. Fitness indicators

The goodness of fit indicators are the criteria for assessing whether SEM models stand or fall. The basic purpose of the indicators is to measure whether the theoretical model constructed by researchers reasonably explains actual observed data. In the proposed study, for obtaining a simple and clear explainable model, the complexity of the model is also noteworthy. As a result, apart from the commonly reported evaluation indexes, the Goodness of Fit Index (GFI), Chi-square (χ^2), and Comparative Fit Index (CFI), the indexes measuring the Degree of Freedom (DoF) are also considered by the proposed method, which are the Root Mean Square Error of Approximation (RMSEA) and the Adjusted Goodness of Fit Index (AGFI) [12]. When the number of factors is fixed, the higher the DoF, the simpler the model. The organized and used indicators in this research are illustrated as follows. The different index evaluates the goodness of fit of a model from different aspects. Only choosing one index as the GA fitness function is not all-inclusive, so we combine all five indexes and define a Comprehensive Evaluation Index (CEI).

$$CEI = GFI + AGFI + CFI + \left(\frac{1}{\chi^2}\right) + \left(\frac{1}{RMSEA}\right) \tag{4}$$

Also, every singular index is checked simultaneously as *CEI* changes to avoid the situation that a certain indicator does not meet the fitting requirements.

3.4.2. Chromosomes encoding and parameters setting

The corresponding GA terms to their meaning in SEM are shown in Table 1.

In the proposed method, each gene indicates one path from one factor to another. The gene will be coded as "1" if the relation is true and "0" if false. What should be paid attention to here is that the path has the direction, and the difference between the directions affects the results of model fitting. Thus, when "1" is given to the gene of factor A pointing to factor B, "0" should be given to the gene of factor B pointing to factor A at the same time. Also, a factor cannot point to itself. One chromosome contains $n * (n - 1)$ genes if n factors are used in the model.

Additionally, the double arrows connection in an SEM model means two factors are correlated, but the causal relationship remains unclear. One function of the proposed method is to do causal analysis, so a double-direction arrow and the circle structure are not permitted in the model. The population number is set in accordance with the number of factors, which should be higher when there are more latent factors in the model.

Because the gene in the proposed method is simply encoded in binary, it is not very strict in the choice of crossover, mutation, and selection methods. If there is no domain knowledge, the probability of the crossover

Table 1. GA-SEM terminology.

GA term	Meaning in SEM
Gene	Hypothesis path among factors
Chromosome	Hypothesis model
Population	Group of chromosomes
Fitness function	CEI

rate is recommended to be set as 0.8. However, the mutation rate should be set 0.3–0.5, which is higher than the commonly recommended mutation rate in many applications of GA. SEM cannot calculate all solutions of GA. When there are unreasonable relationships in the model, SEM will return an error message indicating that the model cannot be calculated. We think these solutions are invalid. On this occasion, we order GA to return to the minimum value. As a result, a relatively higher mutation rate is set to enhance the calculation effectiveness.

3.4.3. Initial population generation and pre-evolution for finding out suggestions

This step is conducted to avoid GA being caught in a local extremum. The procedure of structure learning is conducted after EFA and CFA. The factors extracted by EFA and CFA accord with the correlations of the manifest items. As long as SEM can calculate the model, it will not obtain a very low value in fitting indexes, such as GFI of almost all solutions ranging between 0.8 and 1. The changing range of CEI is small, causing GA to be caught in the local extremum if no pre-processing is operated. However, if the extracted factor is confirmed, the strong or weak relations among the factors will be determined. Besides, the stronger relations that are established, the higher the fitness value. Thus, we create random initial populations and conduct multiple but fewer iterations to extract these strong relations. Here, the factor loading higher than 0.3 is thought as a strong relationship between two factors. Then we give suggestions to the algorithm.

For a suggestion, the genes presenting the strong relations are coded as “1,” and other genes as “0.” The suggestions should be inherited as the dominant population. The crossover and mutation in the dominant population help GA escape from the local extremum. It is not necessary to pour all possible solutions with strong relations into the initial population, and the final solution is not always the same as one or several of the suggestions. If there is no domain knowledge, three suggestions are enough.

3.4.4. Evolution steps

After finding the suggestions, a new initial population containing the suggestions is given to GA. The evolution procedures will stop when CEI is not improved after several evolutions, or the program meets a set maximum iteration criterion. The solution (or solutions) is decoded as the path between factors, and every fitness index should be checked.

If the goodness of fit is acceptable, the next step of parameter learning will begin. Alternatively, if the collected data is not sufficient for building a model, the procedure should go back to data collection. For problem A, one of the results of structure learning is as Figure 4.

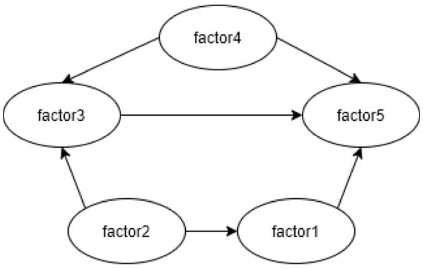

Figure 4. Structure learning result.

For a particular problem, there may be multi-solutions obtained from GA because CEI turns out to be the best fitness value of all these models. We call these possible solutions the candidate models. All the candidate models should be retained for the following steps.

3.5. Parameters learning

The parameter learning of the proposed model contains two parts. One is the structure simplification according to the factor loadings between factors. The other is a regression procedure for separating the learning target from the training data.

There are many methods for SEM to estimate the factor loadings, such as maximum likelihood estimation, general least squares, and asymptotically distribution-free methods. Different methods apply to different data distributions. For example, maximum likelihood estimation requires the data to approximate a normal distribution, whereas the general least squares method does not. The asymptotically distribution-free method can deal with missing data. Thus, before conducting SEM, a priori analysis of the normality of data is necessary. A suitable method should be selected accordingly. The same estimation method is used in the EFA procedure, structural learning procedure, and parameter learning procedure for maintaining consistency.

The factor loadings can be calculated using the estimation method, which represents the strong or weak relations among factors. The calculation is conducted using functions (1)–(3). In the proposed method, we define a factor loading ≥ 0.3 as showing two factors that have a relatively strong relationship. The factors that have factor loadings < 0.3 with all the other factors are thought to have no efficacy for constructing the model. Furthermore, these factors and their contained items should be removed from the model. We call this procedure a structure simplification.

For example, in problem A, the factor loadings of factor 2 are lower than 0.3 regardless of other factors, so factor 2 and items 4, 5, and 6 ought to be removed from the model. We call this procedure as the Structure arrangement shown as Figure 5

As mentioned in Section 3.4, there may be multi-solutions obtained from the structure learning procedure.

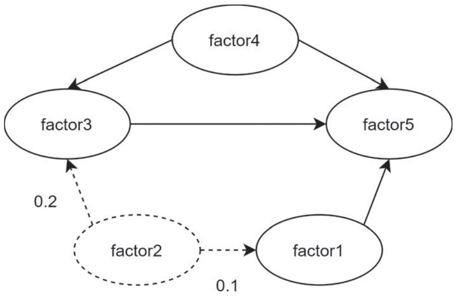

Figure 5. Structure arrangement.

In this situation, the factor(s) in all the candidate models that have factor loadings < 0.3 should be removed.

After the structure simplification, the selected estimation method is used once more for calculating the factor loadings, which can be used for analysing the relations between every two factors. However, for an ML problem, the purpose of the model is classification or prediction. The classification or prediction target is used as one of the manifest items in the built SEM model. Thus, a further step needs to be taken to extract the classification or prediction target and use other manifest items to estimate the target. For example, as shown in Figure 6, for problem A, item 15 is our classification target. It is one of the measuring items for factor 5 in the SEM model.

The estimation methods described above calculate the regression relations between factors and their contained items, which measures the measuring ability of each factor to its items. In contrast, SEM can also estimate the factor scores of each factor using the manifest items. In the shown example, the following function estimates the factor scores of the ith factor.

$$FS\_i = \beta_i + \omega_{i\_1} * item_1 + \cdots + \omega_{i\_j} * item_j + \cdots + \omega_{i\_15} * item_{15} \quad (5)$$

In function (5), β_i is the constant term, and $\omega_{i_j}$is the regression weight of $item_j$ for Factor i. Maximum likelihood estimation is usually used here for estimating factor scores. As mentioned above, many candidate models may be obtained by the structure learning procedure. However, the models with the same CEI value turn out the same factor score calculation results. Thus, the parameter learning shows the same results of all the candidate models. Function (5) shows that for each factor score, the classification target, $item_{15}$, is used as one of the evaluation items for calculating factor scores. As a result, the SEM model cannot be used directly for a classification or prediction model. For using other items (training items) to learn the target item (predicting item), the proposed method conducts a multiple linear regression procedure using the training items on the factor scores. Then, in the presented example, the New estimated Factor Scores (NFS) are obtained, as shown in function (6).

$$NFS\_i = N\beta_i + N\omega_{i\_1} * item_1 + \cdots + N\omega_{i\_j} * item_j + \cdots + 0 * item_{15} \quad (6)$$

Function (6) shows that only the training items estimate the NFSs. The target $item_{15}$ is released from all the factors. Also, the parameters, $N\beta_i$, the constant item for *Estimated Factor score i*, and $N\omega_{i\_j}$ the regression weight for $item_j$ of Estimated Factor score i can be obtained at the same time. Moreover, the final model can be built as shown in Figure 7.

3.6. Model utilization

The model can be applied to different purposes, such as data analysis, machine learning, and causal analysis. A practical example showing the specific utilization of the proposed model will be presented in Section 4.

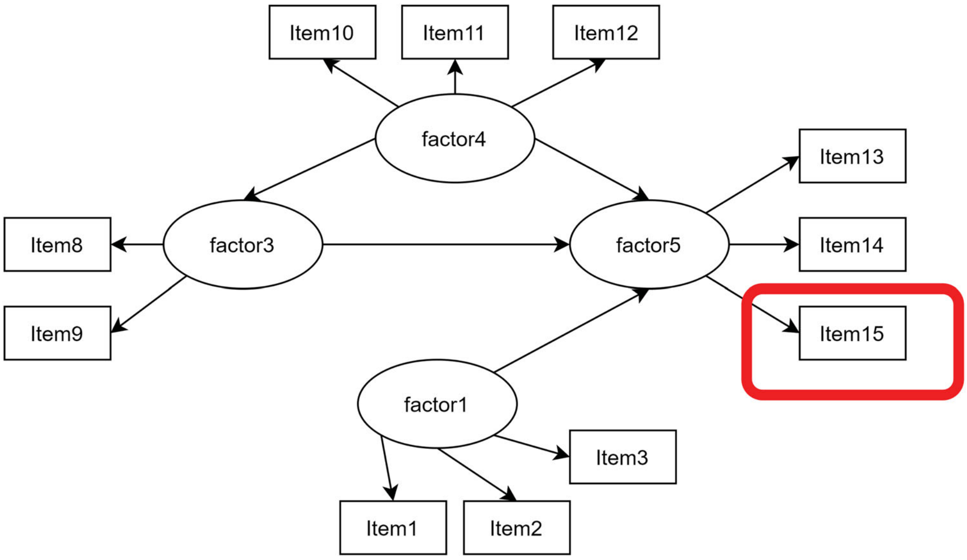

Figure 6. Item 15 is the classification target.

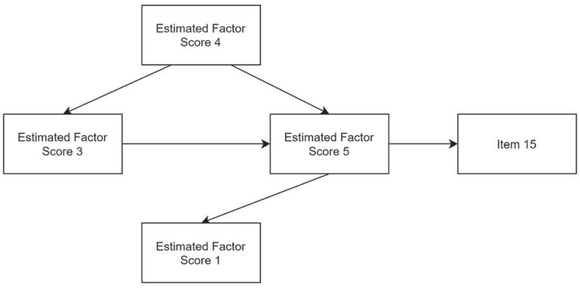

Figure 7. Final model.

For different application purposes, the model should be validated from different aspects. For example, the goodness of fit is the most important evaluation index for the analysis model. The accuracy is the focal point for the ML model. The effectiveness of the intervention is the key to causal models. Besides, for an explainable and persuasive model, the model structure should be simple and easily understood by humans. Also, domain experts should accept its rationality. If the model cannot meet the mentioned requirements, data will need to be repeatedly collected.

4. A practical experiment for a healthcare problem

This section describes a practical application of the proposed method to data analysis, ML, and causal analysis on a common sleep disorder disease, Obstructive Sleep Apnea (OSA).

For testing OSA, the most precise device is Polysomnography (PSG) with a peripheral capillary oxygen saturation (SpO2) test. However, it is expensive and hard for people to use at home. Instead of professional devices, questionnaires are better choices to diagnose OSA in primary care and are self-diagnostic. There are many kinds of questionnaires containing enormous amounts of questions about these three aspects, such as the Quality of Life (QoL) questionnaire, Epworth sleepiness scale, and Stop-Bang questionnaire. Much data is available, but it is impossible and not necessary to use all of these questionnaires at the same time.

On the other hand, the rationality of the model used by a healthcare problem must be recognized by the doctors. Thus, explainable models are necessary. A comprehensible model that can be easily understood by humans also enhances the ease of communication between doctors and patients. Considering the demands mentioned above, we explain how to apply the proposed method to provide a simple and useful analysing, predicting, and causal analysing model for the OSA problem.

4.1. Data preparation

Before collecting data, we review the factors relating to OSA. According to the recently published literature [13–17], OSA relates closely with the following aspects: age, gender, body mass index (BMI), sleep quality including daytime tiredness, snore, health status, and underlying diseases. Thus, we collected questionnaire data considering these factors – the data used for analysis comes from the Sleep Heart Health Study (SHHS) database [18,19]. Apnea-Hypopnea Index (AHI) data can be made on the basis of PSG collection. Among all 5408 participants, 3931 subjects completed all data collection and had no history of OSA diagnosis. AHI ≥ 5 is an indicator of suffering from OSA. A total of 70% of subjects had an AHI ≥ 5 in our study (3931 in total, 1863 males, 2068 females, age 63.7 ± 11.3).

Additionally, there are 66 items collected from the self-rated questionnaires, including Anthropometrics (6 items), Health interview (11 items), Sleep habits and quality (41 items), and SF_36 questionnaires (8 calculated items). Besides, the AHI ≥ 5 treated as undiagnosed OSA is the 67th item input to EFA explained by the next section.

4.2. Data management

EFA and CFA were conducted on the collected items. Table 2 shows the EFA results.

The meaning of the abbreviations in Table 2 are as follows: Sn: Snore, SC: Sleep Complaint, He: Health, HBN, Hard Breath at Night, UD: Underlying Disease, UO: Undiagnosed OSA, Ge: Gender, HoS: Snore Frequency, HLD: Loudness of the Snore, CS: Changes in the severity of the Snore over time, TFA: Frequency of having trouble falling asleep, WN: Frequency of Wake up at Night, WE: Frequency of Wake up Early and cannot go back to sleep, RP: Role-Physical index, VT: Vitality index, RE: Role-Emotion index, WC: Frequency of Woken by Cough, CP: Frequency of Waken by Chest Pain, SoB: Frequency of Woken by Short of Breath, Hy: Hypertension, and Nu: Nocturia.

The 18 items express a total variance of 62.33%, and the KMO test of 0.72. From Table 2, the EFA results

Table 2. EFA results.

	Sn	SC	He	HBN	UD	UO
Ge	**−0.438**	0.172	−0.145	0.044	−0.199	−0.237
HoS	**0.866**	0.031	0.007	0.042	−0.099	0.119
HLD	**0.873**	0.012	0.024	0.056	−0.093	0.079
CS	**0.793**	−0.035	0.020	−0.007	0.000	−0.058
TFA	−0.098	**0.752**	−0.090	0.097	−0.012	0.032
WN	−0.014	**0.880**	−0.072	0.098	0.076	−0.024
WE	0.012	**0.817**	−0.030	0.058	0.064	0.000
RP	0.045	−0.025	**0.802**	−0.128	−0.206	−0.025
VT	−0.011	−0.190	**0.752**	−0.169	−0.012	−0.110
RE	0.083	−0.004	**0.787**	−0.058	−0.022	0.030
WC	0.043	0.097	−0.076	**0.699**	−0.007	0.041
CP	−0.019	0.077	−0.113	**0.811**	0.041	−0.036
SoB	0.025	0.070	−0.131	**0.825**	0.055	0.036
Age	−0.155	−0.040	−0.077	−0.066	**−0.811**	−0.021
Hy	−0.046	0.008	−0.111	0.082	**0.605**	0.212
Nu	0.134	0.262	−0.032	0.080	**0.454**	−0.092
BMI	0.073	0.026	−0.122	0.058	0.139	**0.812**
AHI	0.141	−0.012	0.032	−0.015	0.297	**0.709**

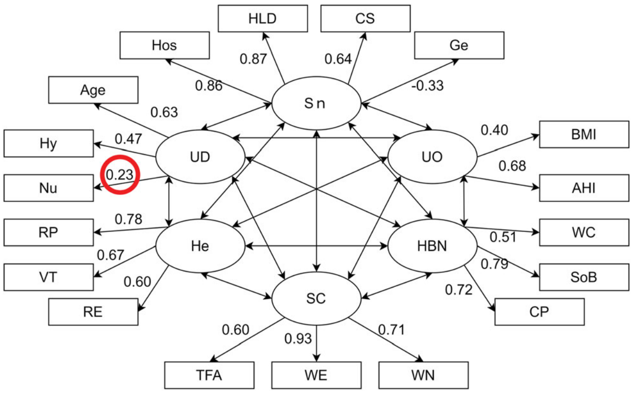

Figure 8. CFA model.

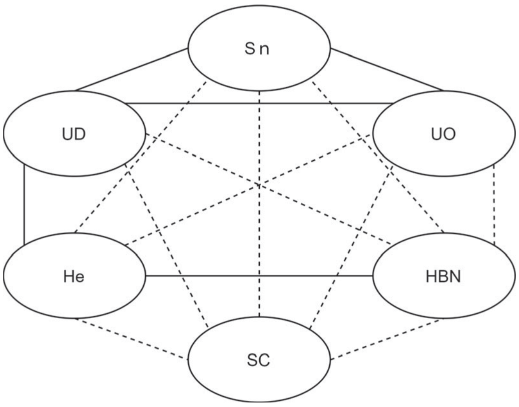

Figure 9. Suggestions.

show that 18 items are classified into 6 factors, and all variables have factor loadings higher than 0.3 to only one factor. Furthermore, we draw a hypothesis model using the extracted 18 items–6 factors and further evaluate the factor loadings using the CFA model, as Figure 8 shows.

As shown in Figure 8, the factor loading of Nocturia to the underlying disease is lower than 0.3, which is not favourable. After removing the Nocturia variable from the model, Table 3 shows the final factor loadings.

The abbreviations in Table 3 have the same meanings as in Table 2.

4.3. Structure learning

The extracted six factors are used for structure learning. The GA procedure specifies the structural model. There are 6 factors, so every chromosome contains 30 genes encoded by “0” or “1.” The crossover, mutation, and selection methods are chosen as Single-Point crossover, Uniform Mutation, and Linear Ranking Selection. Because there are only a few genes in each chromosome, the Single-Point crossover method is selected. For a binary encoding GA, there are not many kinds of mutation methods from which to choose, and Uniform Mutation is the most commonly used. Ranking Selection is mostly used when the individuals in the population have very close fitness values. In the presented application, the CEI is used as the fitness function, which usually changes in a small range at the end of the run. Thus, Ranking Selection leads GA to better select parents in this situation.

Table 3. Factor loadings.

Measured variable	←	Factor	Factor loadings
Ge	←	Sn	−0.328
HoS	←	Sn	0.861
HLD	←	Sn	0.867
CS	←	Sn	0.636
TFA	←	SC	0.600
WN	←	SC	0.931
WE	←	SC	0.708
RP	←	He	0.780
VT	←	He	0.667
RE	←	He	0.603
WC	←	HBN	0.511
CP	←	HBN	0.718
SoB	←	HBN	0.788
Age	←	UD	0.638
Hy	←	UD	0.459
BMI	←	UO	0.382
AHI	←	UO	0.708

Table 4. Parameters for the final evaluation.

Population size	Crossover rate	Mutation rate	Maxi	Max_Run
70	0.8	0.4	2000	300

After choosing the crossover, mutation, and selection methods, the pre-evolution is conducted for finding out suggestions. The result is shown in Figure 9 From Figure 9, three suggestions are chosen randomly with the full line parts coded as “1” and imaginary line coded as “0.” Adding the suggestions to the initial populations with the parameters shown as Table 4 is given to GA.

GA is conducted 10 times, and three answers with the same CEI value, 23.925, are obtained. Figure 10 shows the answers.

As shown in Figure 10, the architectures of the models are the same, but parts of the arrow directions differ among the three candidates. GA finds the best answer to CEI in the 560 generations. At the same time, AGFI and RMSEA also reach the extremum. The values of GFI, CFI, and Chi-square are the second-best ones, which is acceptable. As mentioned in Section 3.4.1, the goodness of fit is not the only target for structure learning in the proposed method, and we also hope a simpler structure can be obtained. AGFI and RMSEA consider the freedom degree of the model, and the better the two indexes are, the simpler the model will be. Thus, the results of GA in the presented example prove that utilizing CEI as the fitness function is effective. The value of the goodness of fitting is shown in Table 5.

As shown in Table 5, all the indexes show that the three candidate models fit well.

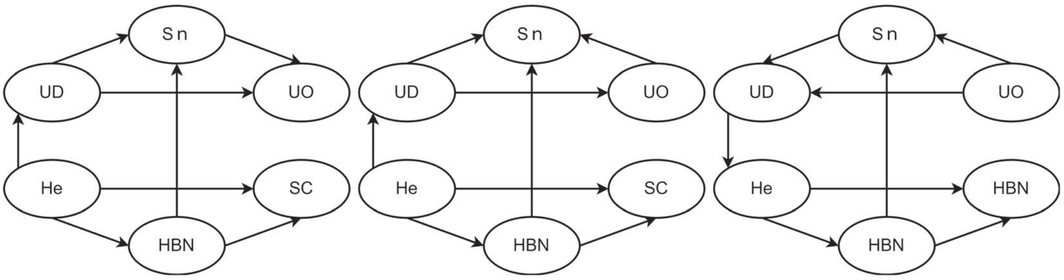

Figure 10. Three candidate solutions.

Table 5. Goodness of fitting.

GFI($>$ 0.90)	CFI($>$ 0.90)	χ^2	AGFI($>$ 0.90)	RMSEA($<$ 0.06)
0.967	0.941	1095	0.954	0.048

4.4. Parameters learning

First, factor loadings are calculated to verify if any factors do not have strong enough relationships with others. The results are shown in Figure 11.

As shown in Figure 11, the relations in the red circles of all three candidates are lower than 0.3, which presents Sleep Complaint (SC) does not have strong relations with any other factors. As a result, SC and its contained manifest items are removed from the dataset. The remaining 14 items and their corresponding factors are shown in Table 6.

As shown in Table 6, the item intended to be analysed or predicted is AHI, which is one of the manifest items of Undiagnosed OSA (UO). Thus, in the next step, a regression procedure is conducted using the other 13 items with their corresponding factor scores calculated by the candidate models. As mentioned above, all the candidate models have the same fitting results, so their parameter learning results are the same as well. By using the learned regression weights and the 13 items (items are shown in Table 6 except AHI), the estimated factor scores can be calculated. Furthermore, the final models made up by the estimated factors and AHI are shown in Figure 12.

As shown in Figure 12, three final candidate models are obtained. The validation of the fitting indexes are shown in Table 7.

4.5. Model utilization and validation

4.5.1. Data analysis

By using maximum likelihood estimation, the standard regression weights between every two factors are calculated, and results are shown in Figure 13.

First, Snore and Underlying Diseases directly affect OSA, and Health and Hard Breath at Night affect OSA

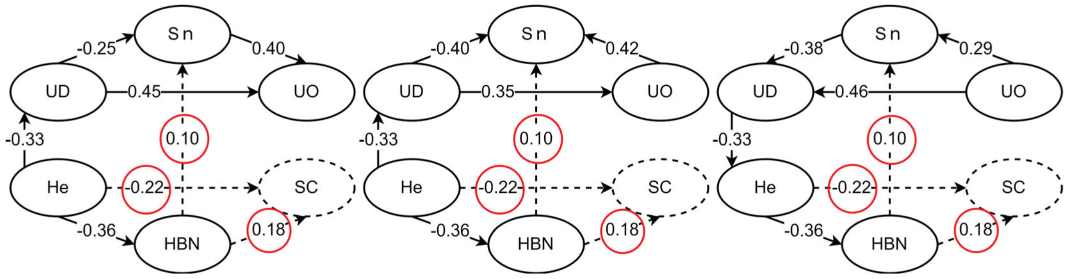

Figure 11. Factor loading verification.

Table 6. Retained items and their corresponding factors.

Items	Content	Factor
Age	–	Underlying disease
Gender	1: Men; 2: Women	Snore
BMI	Calculated by height and weight	Undiagnosed OSA
Snore frequency	Snore frequency	Snore
Loudness of snore	Snore loudness	Snore
Change in snore	Snore becoming stronger or weaker	Snore
Woken by cough	Frequency of waking up due to a cough	Hard breath at night
Woken by chest pain	Frequency of waking up due to chest pain	Hard breath at night
Woken by short of breath	Frequency of waking up due to shortness of breath	Hard breath at night
Hypertension	Hypertension is present or undertreated by hypertension medicine	Underlying disease
Role-physical	The role-physical score calculated from the SF_36 questionnaire	Health
Role-emotion	The role-emotion score calculated by the SF_36 questionnaire	Health
Vitality	Vitality score calculated by SF_36 questionnaire	Health
AHI $\geqslant$ 5?	Apnea-Hypopnea Indexes calculated from PSG	Undiagnosed OSA

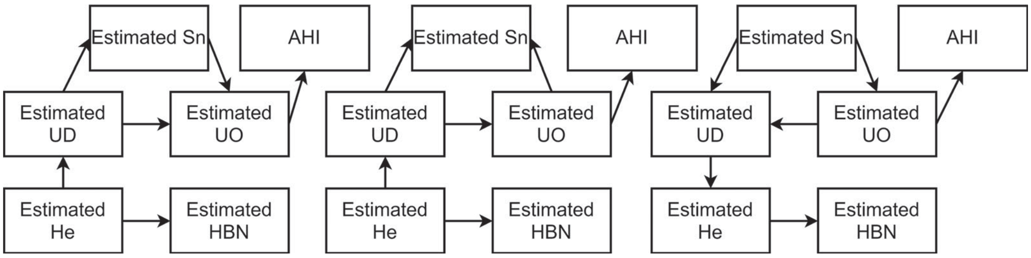

Figure 12. Final models.

Table 7. GFIs of final models.

GFI(> 0.90)	CFI(> 0.90)	χ^2	AGFI(> 0.90)	RMSEA(< 0.06)
0.993	0.990	82	0.984	0.0453

indirectly. Additionally, the factor loadings of Health factors with the other factors are negative, which indicates that health status indirectly reflects the probability of having OSA. The worse one's health, the higher the probability of suffering from OSA.

Considering the analysis described above, a new screening tool to evaluate the risk of having OSA has been created by our team.

4.5.2. Machine learning model

The purpose of this application is to predict whether AHI $\geqslant$ 5. In the previous steps, 13 items were extracted, which can be used to estimate the factor scores. The proposed method uses the estimated factor scores to predict AHI. We validate the model from two aspects: prediction ability and structure effectiveness.

(1) Prediction ability An effective model with high prediction ability requires the model to extract useful features from the dataset accurately and classify the target with high accuracy. Decision Trees and its variances are commonly used methods that can simplify data dimensions and extract useful features. They also provide transparent models. In this part, we use three kinds of Decision Trees and its variants (the ordinary Decision Tree (DT), Bag-ensembled Random Forest (BRF), and AdaBoost-ensembled Random Forest (ARF)) to make classification models for AHI and compare them with the proposed model.

As shown in Figure 12, no matter which candidate model is used, Undiagnosed OSA is the only factor measuring AHI. We classify the estimated Undiagnosed OSA score to predict AHI. The unsupervised classification method, CSCDFCM, proposed by our team in previous research, is used here [20]. Simultaneously, we conducted Decision Trees to extract 13 items with the highest importance of the 66 items. The extraction results are different from those of the proposed method. Table 8 shows the extraction results of the three Decision Tree methods.

Moreover, Table 9 shows the accuracy, F1_score, and the sensitivity of the positive of the three Decision Tree

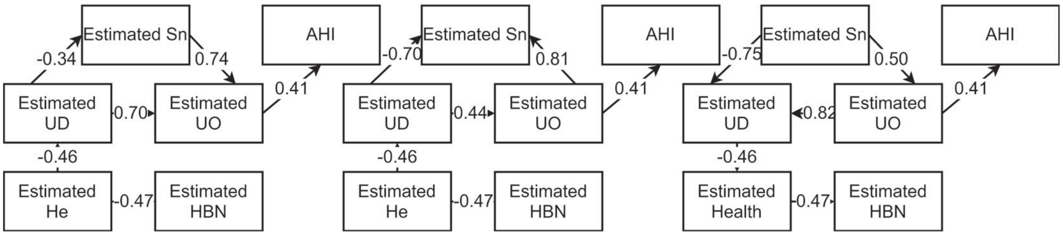

Figure 13. Analysis models.

Table 8. Items extracted by Decision Trees.

DT	BRF	ARF
Age	Age	Age
Height	Height	Height
Weight	Weight	Weight
BMI	BMI	BMI
Fall asleep while watching TV	Physical function	Physical function
Cups of coffee drunk every day	Mental health	Mental health
General health	General health	General health
Vitality	Vitality	Vitality
Minutes to fall sleep	Minutes fall into sleep	Minutes fall into sleep
Time wake up on weekdays	Time wake up on weekdays	Time wake up at weekday
Time wake up at the weekend	Time wake up at the weekend	Time wake up at the weekend
Snore Frequency	Snore Frequency	Snore Frequency
Neck circumference $\geqslant$ 40 cm	Neck circumference $\geqslant$ 40 cm	Neck circumference $\geqslant$ 40 cm

Table 9. Comparison of the accuracy.

Method	Accuracy	F1_score	Sensitivity
DT	67.7%	0.614 [0.46, 0.77]	77.7%
ADT	72.8%	0.657 [0.47, 0.82]	86.4%
BDT	74.1%	0.668 [0.47, 0.83]	89.4%
Proposed model	74.5%	0.672 [0.48, 0.83]	90.0%

models and the classification result of the proposed model. All models conducted 5-fold cross validation.

As shown in Table 9, the proposed model obtained the best accuracy and F1_score, which proves it is more effective as an ML model than the similar explainable model, Decision Trees. Additionally, for a healthcare problem, doctors care about the sensitivity of the positive rate, and the proposed method reaches 90%, which is ideal.

(2) Structure effectiveness

This experiment aims to test the structure effectiveness of the proposed method. As shown in Figure 12, three candidate models are built. Applying the candidate models to BNTs, six factors are the nodes, and the arrows are arcs building up the network. As the estimated factor scores are continuous numbers, CSCD-FCM is conducted to the factor scores for discretizing the data. Furthermore, the estimated factor scores are the evidence used for interfering AHI.

There are three candidate models obtained from the proposed method. The above sections discussed that the estimated scores of the factors are the same in different candidates. Also, the structures of the three candidates are the same, and only a few directions of the arrows are different from each other, which does not affect the interference result of BNT. Thus, when applying the candidate models to BNT, the same result of prediction is obtained.

Besides, K2 is a commonly used method to train structures for BNTs. However, K2 requires domain knowledge to offer the order of nodes to the algorithm. Let us number the nodes of the factors as Hard Breath at Night: 1, Health: 2, Snore: 3, Underlying Disease: 4, Undiagnosed OSA: 5, and AHI: 6. We randomly put them in two orders, [1,6,3,2,4,5] and [6,1,4,2,5,3]. The structures trained by K2 are shown in Figure 14.

Figure 14 shows that the structures trained by BNT under different orders of nodes are different from each other, and Table 10 compares the interference accuracy of AHI on the BNT trained structures and the proposed model structures.

The proposed model structure has the highest accuracy among the three. The results also show that the structures trained by BNT models only present the probability dependency of the nodes, but there is no way to train a reasonable BNT model without domain knowledge. For example, according to the analysis by the proposed model, there are no direct relations between Health and Snore (factor loading between them is lower than 0.3). However, there is a strong relationship between Health and Underlying Disease, and Health affects Snore indirectly through Underlying Disease. However, in the two models trained by BNT, wrong information is transferred by the structure.

This experiment shows that the proposed method can easily apply a simple, reasonable, and effective model structure to BNT networks automatically. There is no need for human experts to participate in the procedure of constructing the model, so much time and labour can be saved.

4.5.3. Causal models

Another function of the proposed model is to analyse the causal relationships among factors. Although statistical dependency between factors can be obtained from the models shown in Figure 13, they cannot reflect the actual causal relationships for which model surgery is necessary. Introducing do(calculus) to the three candidate modes, the intervention models can be obtained. We use one of the candidate models to illustrate the model surgery procedure. The other two are similar.

Figure 15 conducts do (Undiagnosed OSA) for the OSA factor, so the connections between OSA and Snore and the Underlying Disease should be removed. Furthermore, the process of human intervention is conducted to OSA, such as medical treatment. If the causal

Table 10. Comparison of proposed method and BNT.

Method	Accuracy	F1_score
BNT structure 1	73.7%	0.614 [0.46, 0.77]
BNT structure 2	74.2%	0.657 [0.47, 0.82]
Proposed model	74.6%	0.675 [0.49, 0.83]

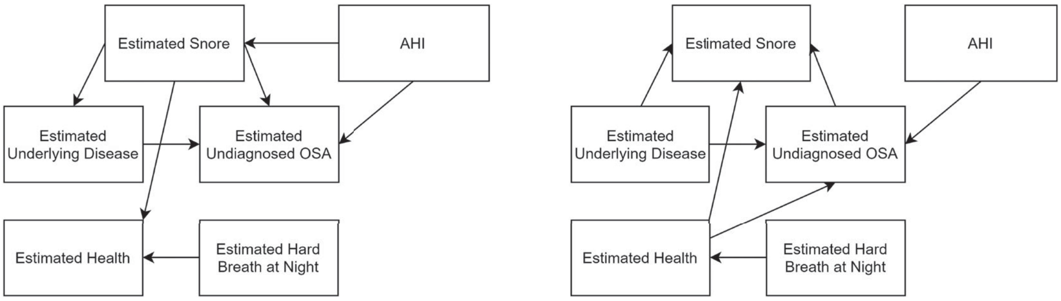
Estimated Snore
AHI
Estimated Underlying Disease
Estimated Undiagnosed OSA
Estimated Health
Estimated Hard Breath at Night
Estimated Snore
AHI
Estimated Underlying Disease
Estimated Undiagnosed OSA
Estimated Health
Estimated Hard Breath at Night

Figure 14. Structures trained by BNTs.

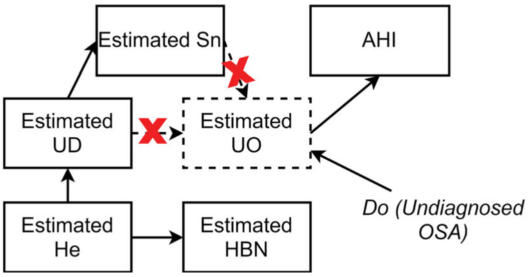

Figure 15. Invention model of Do (Undiagnosed OSA).

relations in this model are true, no change will happen in Snore or Underlying Disease. Similarly, conducting do (Snore) and do (Underlying Disease) for the other two candidate models leads to different conclusions.

By analysing the causal relationships, doctors can determine the most suitable treatment plan for patients, especially when the existing data is insufficient. The presented example shows three kinds of possible causal models. All three factors (Underlying Disease, Snore, and Undiagnosed OSA) can be reasons or results. However, fewer or more candidate models may be obtained from the other applications.

5. Discussions

With the development of ML technology, in addition for the accuracy of learning, the understandability between humans and machines is being paid more attention. Machines are hoped to imitate human behaviour as closely as possible so that humans and machines can collaborate better or even mutually improve. For achieving human-machine understandability, the structures of the learning procedure have to be shown in front of the human eyes. In other words, the degree of explainability of a model is the premise for mutual understanding between humans and machines.

Several existing ML technologies were developed with the explainability, such as Decision Tree methods, BNTs, and their variants. However, some defects of these methods limit their application in practical cases. For example, Tree-type methods judge the necessity of the data features used for prediction by comparing the importance weight of the training data. The Trees cannot express the dependency relationship among the chosen data features, so the reasonability of the inference has no way to be estimated. The partial explainability makes the accuracy of the Tree-type methods dissatisfactory. In the other category, the BNTs methods, although the inference structures are clearly shown, the construction of the structure relies on the domain experts' knowledge. As shown in this paper's medical case, BNTs are incapable of creating the correct structure without prior knowledge. The inference structure's validity affects learning accuracy and relates to the further application, the causal analysis. In the field related to people's life and property, such as medicine and economy, causal analysis is an indispensable means to predict the future. The proposed method supplies an explainable ML model from design to application. The structure is transparent, and rationality can be guaranteed, which endows the model with multi-functions with high quality, including data analysis, machine learning, and causal analysis.

6. Conclusions

The presented paper proposed an explainable machine learning (ML) model by introducing Structural Equation Modelling to the problems. The model is transparent and interpretable from design to application. The human user can recognize the rationality of the model structure so that credible data analysis, ML, and causal analysis can be conducted simultaneously. An application example in the healthcare field shows the practice effectiveness of the model. Future work will be to apply the causal model analysis function of the proposed model in other fields.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Jiarui Li received her BS, and MS degrees from Beijing Jiaotong University, China, in 2014 and 2017, respectively. She is currently a PhD student at Kyoto University. Her research interests include Fuzzy Logic and explainable machine learning technologies.

Tetsuo Sawaragi received his BS, MS and PhD degrees from Kyoto University, Japan. He is currently a Professor of the Department of Mechanical Engineering. His research interests include System Engineering, Human-Machine System, Human-Machine Interference and Cognitive Engineering.

Yukio Horiguchi received his BS, MS and PhD degrees from Kyoto University, Japan. He is currently a Professor of the Faculty of Informatics, Kansai University. His research interests include human factors, human-machine systems, and interface designs.

ORCID

Jiarui Li http://orcid.org/0000-0003-3306-7810

References

[1] Mitchell TM. The discipline of machine learning. Pittsburgh: Carnegie Mellon University, School of Computer Science, Machine Learning Department; 2006.

[2] Samek W, Wiegand T, Müller K-R. Explainable artificial intelligence: understanding, visualizing and interpreting deep learning models. arXiv, preprint arXiv:1708.08296; 2017.

[3] Pearl J. Theoretical impediments to machine learning with seven sparks from the causal revolution. arXiv, preprint arXiv:1801.04016; 2018.

[4] Roscher R, Bohn B, Duarte MF, et al. Explainable machine learning for scientific insights and discoveries. IEEE Access. 2020;8:42200–42216.

[5] Zhang Q, Yang Y, Ma H, et al. Interpreting CNNs via decision trees. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2019. p. 6261–6270.

[6] Frosst N, Hinton G. Distilling a neural network into a soft decision tree. arXiv, preprint arXiv:1711.09784; 2017.

[7] Constantinou AC, Fenton N, Marsh W, et al. From complex questionnaires and interviewing data to intelligent Bayesian network models for medical decision support. Artif Intell Med. 2016;67:75–93.

[8] Le F, Srivatsa M, Reddy KK, et al. Using graphical models as explanations in deep neural networks. 2019 IEEE 16th International Conference on Mobile Ad Hoc and Sensor Systems (MASS); 2019. p. 283–289.

[9] Keppens J. Explainable Bayesian network query results via natural language generation systems. Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law; 2019. p. 42–51.

[10] Li J, Horiguchi Y, Sawaragi T. Data dimensionality reduction by introducing structural equation modeling to machine learning problems. The Society of Instrument and Control Engineer (SICE) Annual Conference 2020; 2020. p. 826–831.

[11] Ullman JB, Bentler PM. Structural equation modeling. In: Handbook of psychology. John Wiley & Sons, Inc.; 2003. Available from: http://dx.doi.org/10.1002/0471264385.wei0224

[12] SchermellehEngel K, Moosbrugger H, Müller H. Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. Methods Psychol Res Online. 2003;8(2):23–74.

[13] Senaratna CV, Perret JL, Lodge CJ, et al. Prevalence of obstructive sleep apnea in the general population: a systematic review. Sleep Med Rev. 2017;34:70–81.

[14] Mansukhani MP, Kolla BP, Somers VK. Hypertension and cognitive decline: implications of obstructive sleep apnea. Front Cardiovasc Med. 2019;6(96). doi:10.3389/fcvm.2019.00096

[15] Mendelson M, Bailly S, Marillier M, et al. Obstructive sleep apnea syndrome, objectively measured physical activity and exercise training interventions: a systematic review and meta-analysis. Front Neurol. 2018;9(73). doi:10.3389/fneur.2018.00073

[16] Chang ET, Baik G, Torre C, et al. The relationship of the uvula with snoring and obstructive sleep apnea: a systematic review. Sleep Breath. 2018;22(4):955–961.

[17] Quan SF, Budhiraja R, Kushida CA. Associations between sleep quality, sleep architecture and sleep disordered breathing and memory after continuous positive airway pressure in patients with obstructive sleep apnea in the apnea positive pressure long-term efficacy study (APPLES). Sleep Sci. 2018;11(4):23.

[18] Zhang GQ, Cui L, Mueller R, et al. The national sleep research resource: towards a sleep data commons. J Am Med Inform Assoc. 2018;25(10):1351–1358.

[19] Quan SF, Howard BV, Iber C, et al. The sleep heart health study: design, rationale, and methods. Sleep. 1997;20(12):1077–1085.

[20] Li J, Horiguchi Y, Sawaragi T. Cluster size-constrained fuzzy C-means with density center searching. Int J Fuzzy Log Intell Syst. 2020;20(4):346–357.