Proceedings of the 2025 SICE Festival with Annual Conference
September 9-12, 2025, Chiang Mai, Thailand

For future thinking of Data-driven sensitivity analysis: Designing human life and behavior.

Fumie Ogawa[1,2†]

[1]Japan Woman's University, Tokyo, Japan
[2]Mazda Motor Corporation, Yokohama, Japan
(Tel : +81-70-7571-9419; E-mail: ogawa.f@mazda.co.jp)

Abstract:

As a proposition of this study, I am thinking to investigate how the work environment and information can improve employee well-being. It will then be used for causal analysis with a view to be addressing for MaaS and EMCs.

These include invisible communication relationships. These efforts were made to visualize and activate classified questionnaires and other surveys. An approach was developed to visualize the degree of activation using structural equation modelling (e.g., SEM), it was based on a Bayesian approach to the data of the choices. In addition, validation was carried out to develop an approach to visualize the degree of activation from the data; this visualization of means between variables based on SEM and multi-group analysis were also considered. The paper shows how important pre-processing is in SEM prediction. Also, as the previous report was able to report the main results obtained when SEM was carried out on questionnaire data extracted according to Bayesian methods, the present paper proposes whether the study is also feasible in terms of spatial construction.

Keywords: Well-Being, SEM, Maas, EMC

1. INTRODUCTION

One recent societal demand has been for work environments, interpersonal relationships and behavioral guidelines that enhance employees' well-being to be promoted. One example is work on visualizing sample responses to questionnaires. In a previous related study, Kevin (2019) argued that presenting both the variance value illustration of the data type and the SEM model more effectively presents the results of the analysis. This paper uses data visualization to describe the strengths of combining scatter plots and SEM in terms of the model's path coefficients. This shows the relationship between the strengths and weaknesses of the variables. Fife et al. (2023) argue that looking at the full distribution of the underlying data can detect unexpected anomalies and provide insights into unmodelled parts of the data [2]. This detects such problems in the shape and non-linear relationships in the overall data. In both cases, the emphasis is on how data visualization can display the accuracy and correctness of the model. Expanding the perspective to include recent machine learning methods, Nishio et al. (2024) analyzed large amounts of educational data (learning history data) to provide learners with 'adaptive learning' feedback, which enables an understanding of their level of proficiency and understanding in the learning process [3]. A method is introduced that enables matching to be performed alongside simulations of conventional methods of estimating learner ability and predicting reactions. The study was conducted from two perspectives: (a) the background of the study (the data used should be normalized) and (b) the accuracy of the model when the variance σ is varied. Ideally, this method should include a large amount of input data, with more than 2,000 pieces of data. In the system designed this time, the hypothesis of deriving 'well-being' was created and set as a hypothetical SEM (path diagram) [4]. The analytical model was obtained by automatically defining indicators of psychological change and control factors from each questionnaire item in relation to 'well-being', using Bayesian networks and factor analysis. Non-arbitrary pre-processing of the alternatives was proposed (Fig. 1,2) In the previous study, the objective variable was happiness. The results depend on the data examined, and the quality of the input data.

***No personally identifiable information is included in this data.**

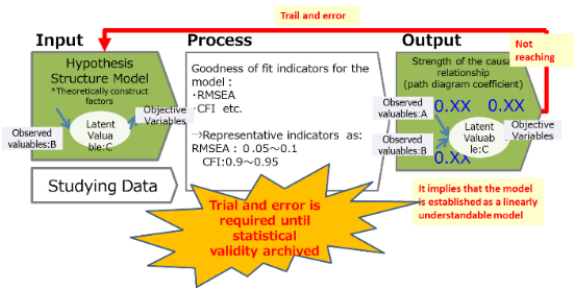

Fig. 1 One scene of research data

2. METHOD

In this study, the SEM was used to introduce a latent variable that is data-driven and not directly observable,

and to examine the assumed causal relationship between this latent variable and the observed variable. The square figures in the model describe the observed variables and the oval figures describe the latent variables (also called factors). These latent variables can measure directly by questionnaire. The strength of the SEM has by questionnaire for multiple regression analysis. The author works in the automotive industry and being student of design and architecture. A recently announced mobility city project aims to create a “mobility test track”. The aim is to create a place where new values can be generated. It is including solutions to mobility-related problems concerning people, goods, information, and energy, they are as well as the development of technologies and services that contribute to global sustainability. There is much discussion about the need to test not only hardware, but also infrastructure.

(1) Test case 1: A web-based survey was conducted with 4,295 participants from a specific region. Survey item: We measured subjective feelings of well-being using questionnaires to assess product usability and sensitivity. Psychological questionnaires were used to measure responses. The 27 items related to self-motivation and satisfaction with life, for example, and were answered using a five-point scale.

(2) Test case 2: A web-based survey was conducted with 10,736 employees of a manufacturing company. The survey items were: The internal working environment (e.g., relationships with superiors), psychological factors (e.g., self-efficacy) and self-motivation at work were measured. Responses to 43 items relating to self-motivation were sought using a five-point scale.

(3) Test case 3: The ALSE calibration field is known for its change in calibration method. And description of the measurement method notes for the new uniformity provisions of the field-measured data. This time I analysis by CAE for ALSE calibration as sample data.

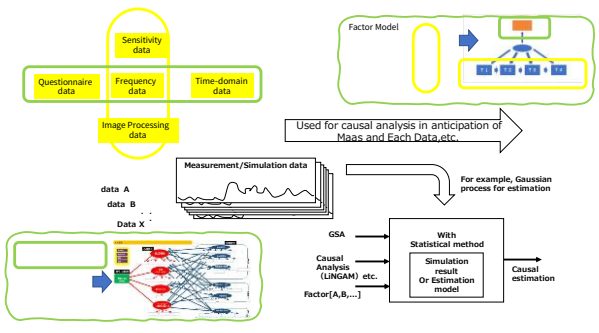

Fig. 2 Data Mapping for different industries Modeling

3. ANALYSIS

In this chapter, we describe the data preprocessing involved in the two approaches to creating an SEM model. The SEM structure means that the survey questions are prepared in advance. These questions are called 'observed variables'. SEM also requires factors, which are also called latent variables, which cannot be measured directly.

3.1 About Test case1

Test Case 1 refers to the results of the causal analysis of the questionnaire responses from users, regardless of prefecture, numbered 4295. Conventional approaches have been combined with LiNGAM [5] (NEC causal analysis software).

Modelling:

The situations in which a questionnaire is created, and the extracted results are used to calculate whether they fit by SEM (using AMOS/R software). The results are shown in Fig.4. It shows Grp. 2: Well-being; Grp. 3: Satisfaction with driving ability; and Grp. 4: Circumstances. These groups came from questionnaires in each situation.

Results:

With the system designed this time, the hypothesis of deriving 'well-being' was created, with the model: SEM (path diagram), it observed variables: data obtained from the questionnaire, latent variables: This variable names from each item of the questionnaire to 'well-being' were assumed from the words of the questions, as the hypothesis can arbitrarily decide the variable names.

3.2 About Test case2

Test case2’s process of the Bayesian network is based on the network score 'BDeu' (Bayesian Dirichlet equivalence uniform), which consists of a Dirichlet prior distribution with hyperparameters that satisfy 'likelihood equivalence'[6].

Modelling:

(i) BDeu is maximized under the unique use of tabu search, one of the metaheuristic search algorithms. Implemented by employing a function to learn the graph structure.

(ii) In the post-processing, when the target variable (child node) is specified from the completed graph structure itself, the substructure (hierarchical structure formed by parent nodes) that leads to it is automatically extracted without excess or deficiency as necessary, as a process that is separate from Bayesian theory (Using software by R). These conference systems are in Fig.5.

Results:

Causal estimation in GSA theory was carried out based on an arbitrary selection of data obtained from the responses to questionnaires in the workplace. The results are shown in Fig.6. Here, original functions are newly formulated and implemented, and all structure is

visualized while changing the format and background. The arrows indicate causality based on conditional probability in Bayes.

3.3 About Test case3

This data has X and Y-axes, these sets are based on frequency data to time domain data. It assumes that noise is occurring in area, for example. In advance, we must prepare for simulation for total modeling by Moment method.

Modelling:
- Creating electric field probes for reception
- Creation of scattering field probes
- Creation of masts
- Creation of horn antenna
- Creation of the basic model (only the Field Probe for reception is placed, which is defined as the basic model)
- Creation of the Scattering Model (Scatterer based on the Basic Model)
-Create each of the constituent parts separately.

Results:
It assumes that noise is occurring in areas where the data are scattered. In radio irradiation tests, other information is often acquired at the same time. However, measurements are usually taken at the same time. The first task is to acquire data that can be digitally demonstrated and to align formats. These issues are to store the measurement data in a multifaceted way, to promote digitalization. And to record what happened at what time as time domain data. It was found that the records were not sufficient.

4. DISCUSSION/CONCLUSION

In the aim of this study, we considered an approach for inferring and utilizing dependencies between variables based on GSA, and we also discussed multi-group analysis The placement invariance of the multi-group analysis's path was carried out and the fit of the model in each population was checked after the model was constructed. As a result of the study, it was found that adding pre-processing using the theory of GSA to the SEM visualisation has a strong degree of use as a proposed scenario for data-driven modelling. The focus should be on people and mobility at demonstration test sites, but there are many examples in architecture where the environment is prioritized over the car.

It assumes that noise is occurring in areas where the data are scattered. In radio irradiation tests,other information is often acquired at the same time.However, measurements are usually taken at the same time dy time domain timing.The first task is to acquire data that can be digitally demonstrated and to align formats. In the previous study, the objective variable was target to happiness. The results depend on the data examined, and the quality of the input data. Digging deeper, it overlaps with the idea of EUTHENICS, which Ellen Richards once advocated. Richards thinks that the goal of life is to find joy in work and to take a shortcut to living a fun life. Regarding the meaning of goals in life, we all need to decide on clear goals at an early stage in life.

The aim was to use causal analysis with a view to deal with other topics, but it was found that it was difficult to digitize the data before the analysis, as it remained as images. I fell that necessary to confirm in advance what we want to do at the preprocessing stage, even our busy daily experimental work.

INDEX

CFI(Comparative Fit Index:Bentler): We compared the analytical model with the independent model (It is different by the index of the researcher) and confirmed it with an in dex to evaluate how good it was. I take a value from 0 to 1 here and an interpreted as a good model so that 1 is near. Generally, 0.90 or more are used as a criterion of the excellent conformity (they vary according to the indexes of the researcher).

RMSEA(Root Mean Square Error of Approximation) is an index of thebadness of fit of analytical model.

I confirmed the index to measure the badness of the analytical model. Generally, 0.05 or less is used as a criterion of the excellent conformity.

AGFI is an index to evaluate the precision of the analysis from the point of view of the explanation rate of the model for the dispersion of the observated variables. Generally, 0.95 or more is used as a criterion of the excellent conformity.

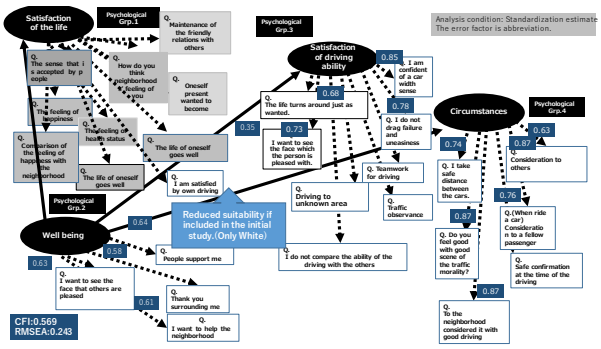

Fig.4 The result of LiNGAM

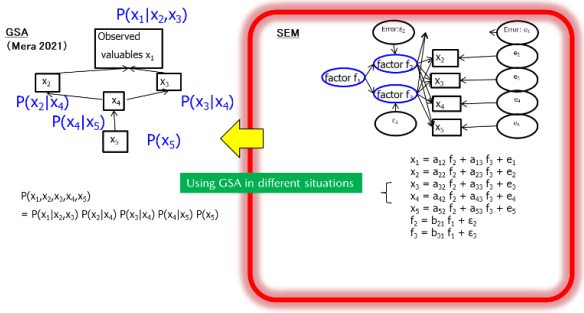

Fig.5 The flowchart of approach of GSA and SEM

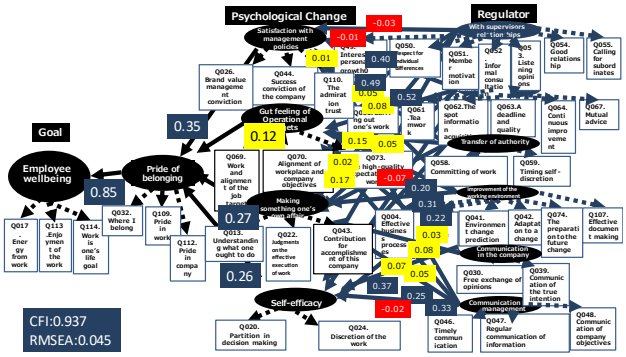

Fig.6 Mapping of SEM

REFERENCES

[1] K.A.Hallgren,C.J.McCabe, K.M.King ,D.C.Atkins, Beyond Path Diagrams: Enhancing Applied Structural Equation Modeling Research through Data Visualization, Addictive behaviors,Volume 94, pp74-82, July,2019.

[2] Fife, D.A.Brunwasser, S.M.Merkle, E.C., Seeing the impossible: Visualizing latent variable models with flexplavaan. Psychological Approachs, *28*(6), 1456–1477,2023.

[3] T.Nishio, E.Tsutsumi and M.Ueno, Deep-IRT with a temporal Convolutional Networkto Convolutionalto Convolut student's Past Ability Change,IEICE D,Vol.J107-D,No.3,pp.98-10,2024.

[4] H.Toyota, Structural Equation Modeling AMOS Edition,Tokyo-Tosho,2014.

[5] S.Shimizu, P.O Hoyer et al.,A Linear Non Gaussian Acyclic Model for causal Discovery, Journal of Machine Learning research 7,2003-2030,2006.

[6] M.Ueno, Bayesian Network, Corona Publishing Co.,ltd. ,2013.

[7] M.Mera, Y.Nakamura et al., Study on Approach of Utilization of Bayesian Networks for Analyzing Time Series Data, Journal of The Japan Society for Industrial and Applied Mathematics, Vol.31, No.2, pp.76-104, 2021.

[8] Leo Breiman,Statistical Modeling: The Two Cultures, Statistical Science 2001,Vol.16,No.3,pp.199-231

[9] E.H.Richards,The art of Right Living,Whitcomb & Barows,Boston,1904.

[10] Seisho Yoshiyama,relation of work to adoptations of the Natural Environment in E.H.Richards,the fourth R idea, Yokkaichi university environmental information collection, pp.17-27, 2008.