Prospects of Continual Causality for Industrial Applications

Anonymous submission

Abstract

We have been working on causal analysis of industrial plants' process data and its applications, such as material quantity optimization using intervention effects. However, process data often has some problems such as non-stationary characteristics including distribution shifts, which make such applications difficult. Combined with the idea of continual learning, causal models may be able to solve these problems. We will present the potential and prospects for industrial applications of continual causality, showing previous work. We also introduce the brief idea of a specific new causal discovery method using a continual framework.

Our Position and Purpose

We have been working on researches and business applications about industrial plants where predictive models are created from data and in addition the results are used for later actions to achieve specific objectives especially. We have found that the concepts of continual learning and causality are important to achieve these goals. In terms of the combination of continual learning and causality (**continual causality**), we will present the challenges we have faced so far and the prospects for their solutions. In particular, we are currently considering a new causal discover method to deal with non-stationarity and non-linearity by continual learning, and its brief idea will also be discussed later in this paper.

Discussion

Causality in Industrial Applications

In many industrial applications of AI, the purpose of its predictions is often to stabilize or maximize a specific variable using the predicted value, e.g. optimizing the output product for material input in plants. Since a mere prediction model may not capture the data generation process, it may not be possible to estimate intervention effects, such as how much the production rate will increase when the material input is increased. Therefore, causal analysis are important for such applications. Causality is also useful from the standpoint of interpretability, since plants have high risk of accidents and its damage, and it is important to understand the basis and reasons for various types of predictions.

While causality is useful, the complete picture of causal relationship is rarely available in plant process data. This is also because plant processes often include feedback loops and material reuse, and there may be time-delayed effects among processes, so its causal relationships and directions are often nontrivial. Therefore, it is important to identify unknown causal relationships as well as to estimate intervention effects. However, it is difficult to conduct experiments such as Randomized Controlled Trials (RCTs) to identify causal relationships in non-operating conditions, because of the risk of accidents, its damage and business factors in plants. Therefore, the framework of **causal discovery**, which finds causal relationships and directions only from data, is important.

Potential of Continual Learning

Modeling plant process data also shows difficulties in terms of maintenance over time. For example, there are many non-stationary characteristics such as instability at start-up, distribution shifts due to changes in the quantity or type of products, trends by equipments aging, and seasonality caused by outdoor temperature changes (Kadlec, Gabrys, and Strandt 2009). By using the concept of continual learning, the system can be expected to continuously train models and adapt to changes in system conditions. We have working on this as a new method JIT-LiNGAM (Anonymous submitted to CLeaR 2023), which is mentioned bellow.

Moreover, it may also be possible to reconsider the aforementioned tasks of stabilizing and maximizing plant process variable not only in the context of system control or causality, but also in the context of reinforcement learning and continual learning (future work).

Past Efforts and Future Prospects

In this paper, we discuss the following challenges and describe the efforts of researchers including us so far and future prospects.

Causal Discovery

Causal discovery is a framework identifying unknown causal relationships and directions only from data. As mentioned above, this framework is important because causal relationships are often unknown in plant process data. Discov-

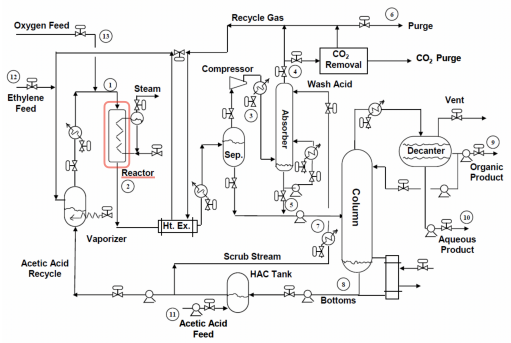

Figure 1: Flows of the vinyl acetate production plant simulator (Luyben and Tyréus 1998).

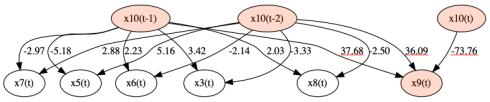

Figure 2: Results of applying VAR-LiNGAM to plant simulator data. Edges represent linear causality coefficients. The node x7∼x10 means a certain process variable, and for example x7(t-1) means the value of x7 one step before at time t.

ered causal relationships are used for later intervention effect estimation and optimization, as well as for variable selection and model interpretation. LiNGAM (Shimizu et al. 2006) is a representative linear causal discovery method. There are also its extensions with partial prior knowledge (Shimizu et al. 2011) and with latent variables (Hoyer et al. 2008). Several non-linear methods are also known (Peters et al. 2014; Zheng et al. 2020; Uemura et al. 2022).

We have applied these methods to actual plant process data, but we have faced the common problem that the "true causal relationships" are unknown and so it is difficult to evaluate the results. However, there is a possibility that the causal model can be continuously evaluated indirectly based on the results of later interventional actions and be updated based on this. This should be considered future work of continual causality.

Time-series Extension

It is necessary to introduce time-series models to account for time-lagged variables in causal discovery. Furthermore, it enables us to construct causal models without contradiction by expanding feedback loops along the time direction. Specific methods include VAR-LiNGAM (Hyvärinen et al. 2010). We have conducted numerical experiments applying VAR-LiNGAM to simulation data of vinyl acetate plants (Luyben and Tyréus 1998). The results of these experiments are briefly presented in Figures 1 and 2.

Optimal Intervention

After getting a complete causal model, the optimal amount of intervention to an operable variable can be calculated

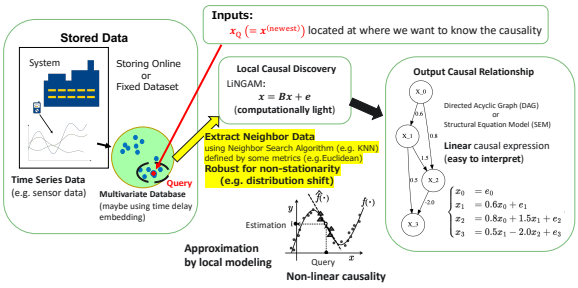

Figure 3: The idea of JIT-LiNGAM

backward such that a certain variable has a specific value (Pearl, Glymour, and Jewell 2016). There are some extensions of this approach, for example, a method including predictive models (Blöbaum and Shimizu 2017) and a method that estimates the optimal individual-level intervention (Kiritoshi et al. 2021).

Continual Causal Discovery : JIT-LiNGAM

We proposed a causal discovery method for non-stationarity such as distribution shift and non-linear causal relationships, JIT-LiNGAM (Anonymous submitted to CLeaR 2023) where LiNGAM is combined with Just-In-Time Modeling (JIT) (Stenman, Gustafsson, and Ljung 1996; Bontempi, Birattari, and Bersini 1999). JIT is a method conventionally used for soft sensors (pseudo-sensors in plants for difficult-to-measure locations using regression models, etc.) where local linear models are trained continually by extracting neighboring samples of the current input sample from a database. Based on Taylor ' s theorem, non-linear phenomena in plants can be approximated by local linear models. And by using neighboring samples for the modeling, it can follow continual changes in plants. Moreover, the database can be updated by adding samples online, but due to limitations of memory and computational complexity, efficient use of data is necessary. For example, the way using data from the most recent several years, and other developments, such as the use of influence functions, or utilizing continual learning methods and reinforcement learning, may be considered (future work).

Extensions to time-delayed causality as described above and development to optimal intervention are also possible. In addition, since it can capture snapshots of non-linear, non-stationary and dynamically changing causal relationships, it may be able to even deal with cases where causal directions are being reversed. This can be a solution to the plant feedback loop problem described above.

Conclusion

We presented our positions in the causal analysis research area relevant to continual learning problems. Now we are working on each introduced theme independently, but in the future we need to integrate them. Particularly, JIT-LiNGAM is expected to be extended in various ways in the future. Continual causality is still unexplored areas, and much further research will be conducted in the future.

Appendix:JIT-LiNGAM Algorithm

We will show the details of JIT-LiNGAM in Algorithm1. More details are in our submitted paper (Anonymous submitted to CLeaR 2023).

Algorithm 1: JIT Algorithm for Time-Series Causal Discovery (JIT-LiNGAM)

Inputs:

stored data $\mathcal{D} = \left\{ \boldsymbol{x}^{(t)} \mid t = 1, \ldots, T-1 \right\}$,

query point $\boldsymbol{x}_{\mathrm{Q}} = \boldsymbol{x}^{(T)}$, distance function $d(\cdot, \cdot)$, number of neighbors K.

Outputs:

weighted adjacency matrix $\boldsymbol{J}(\boldsymbol{x}^{(T)})$: representing the causality defined in the neighborhood for query point $\boldsymbol{x}^{(T)}$.

Procedure 1

Extract K-data of $\boldsymbol{x}^{(t)}$ from $\mathcal{D}$, based on $d(\boldsymbol{x}^{(t)}, \boldsymbol{x}_{\mathrm{Q}})$, which is the distance from the query point $\boldsymbol{x}_{\mathrm{Q}}$. (The details of how to extract K-data are described in the paper (Anonymous submitted to CLeaR 2023)) The resulting K-data subset $\Omega(\boldsymbol{x}_{\mathrm{Q}}; d, K)$ is:

$$\Omega(\boldsymbol{x}_{\mathrm{Q}}; d, K) = \left\{ \boldsymbol{x}^{(\sigma(k))} \mid k = 1, \ldots, K \right\},$$

where $\sigma(k)$ is a function that returns the k-th nearest time index t in $\Omega(\boldsymbol{x}_{\mathrm{Q}}; d, K)$.

Procedure 2

Centralize $\Omega(\boldsymbol{x}_{\mathrm{Q}}; d, K)$ and get $\tilde{\Omega}(\boldsymbol{x}_{\mathrm{Q}}; d, K)$, where mean is subtracted from each element of $\Omega(\boldsymbol{x}_{\mathrm{Q}}; d, K)$ along each dimension of $\boldsymbol{x}$.

Procedure 3

Train LiNGAM using $\tilde{\Omega}(\boldsymbol{x}_{\mathrm{Q}}; d, K)$, and get resulting weighted adjacency matrix $\boldsymbol{J}(\boldsymbol{x}^{(T)})$.

References

Anonymous. submitted to CLeaR 2023. Causal Discovery for Non-stationary Non-linear Time-series Data Using Just-In-Time Modeling.

Blöbaum, P.; and Shimizu, S. 2017. Estimation of interventional effects of features on prediction. In *2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP)*, 1–6. IEEE.

Bontempi, G.; Birattari, M.; and Bersini, H. 1999. Lazy learning for local modelling and control design. *International Journal of Control*, 72(7-8): 643–658.

Hoyer, P. O.; Shimizu, S.; Kerminen, A. J.; and Palviainen, M. 2008. Estimation of causal effects using linear non-Gaussian causal models with hidden variables. *International Journal of Approximate Reasoning*, 49(2): 362–378.

Hyvärinen, A.; Zhang, K.; Shimizu, S.; and Hoyer, P. O. 2010. Estimation of a structural vector autoregression model using non-gaussianity. *Journal of Machine Learning Research*, 11(5).

Kadlec, P.; Gabrys, B.; and Strandt, S. 2009. Data-driven soft sensors in the process industry. *Computers & chemical engineering*, 33(4): 795–814.

Kiritoshi, K.; Izumitani, T.; Koyama, K.; Okawachi, T.; Asahara, K.; and Shimizu, S. 2021. Estimating individual-level optimal causal interventions combining causal models and machine learning models. In *The KDD'21 Workshop on Causal Discovery*, 55–77. PMLR.

Luyben, M. L.; and Tyréus, B. D. 1998. An industrial design/control study for the vinyl acetate monomer process. *Computers & Chemical Engineering*, 22(7-8): 867–877.

Pearl, J.; Glymour, M.; and Jewell, N. P. 2016. *Causal Inference in Statistics: A Primer*. John Wiley & Sons.

Peters, J.; Mooij, J. M.; Janzing, D.; and Schölkopf, B. 2014. Causal discovery with continuous additive noise models. *Journal of Machine Learning Research*, 15: 2009–2053.

Shimizu, S.; Hoyer, P. O.; Hyvärinen, A.; Kerminen, A.; and Jordan, M. 2006. A linear non-Gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7(10).

Shimizu, S.; Inazumi, T.; Sogawa, Y.; Hyvärinen, A.; Kawahara, Y.; Washio, T.; Hoyer, P. O.; and Bollen, K. 2011. DirectLiNGAM: A direct method for learning a linear non-Gaussian structural equation model. *The Journal of Machine Learning Research*, 12: 1225–1248.

Stenman, A.; Gustafsson, F.; and Ljung, L. 1996. Just in time models for dynamical systems. In *Proceedings of 35th IEEE Conference on Decision and Control*, volume 1, 1115–1120. IEEE.

Uemura, K.; Takagi, T.; Takayuki, K.; Yoshida, H.; and Shimizu, S. 2022. A multivariate causal discovery based on post-nonlinear model. In *Conference on Causal Learning and Reasoning*, 826–839. PMLR.

Zheng, X.; Dan, C.; Aragam, B.; Ravikumar, P.; and Xing, E. 2020. Learning sparse nonparametric dags. In *International Conference on Artificial Intelligence and Statistics*, 3414–3425. PMLR.