

Feature based causality analysis and its applications in soft sensor modeling [★]

Feng Yu^{*} Liang Cao^{**} Weiyang Li^{*} Fan Yang^{*} Chao Shang^{*}

^{} Department of Automation; Beijing National Research Center for Information Science and Technology (BNRist), Tsinghua University, Beijing 100084, China (e-mail: yf18@mails.tsinghua.edu.cn; liwy18@mails.tsinghua.edu.cn; yangfan@tsinghua.edu.cn; c-shang@tsinghua.edu.cn)*

*^{**} Department of Chemical and Biological Engineering, University of British Columbia, Vancouver, BC V6T 1Z3, Canada (e-mail: clubc19@mail.ubc.ca)*

Abstract: In industrial processes, causality analysis plays an important role in fault detection and topology building. Aiming to attenuate the influence of common correlation and noise, a feature based causality analysis method is proposed. By using the orthogonality and de-noising in feature analysis, it can capture more efficient causal factors. Moreover, better causal factors can make better predictions. Soft sensors based on least-squares regression and two neural networks are tested to compare the performance when using different causal factors and not using causal factors. The results show that the causal feature based soft sensors obtain the best performance and causal factors are crucial to prediction performance. Hence, it has great application potential owing to its strong interpretability and good accuracy.

Keywords: Feature learning; causality analysis; soft sensor

1. INTRODUCTION

For large-scale industrial processes, the system behavior is influenced by complex inter-relationship of different variables. To identify such inter-relationship, causality analysis, which aims to find a causal network of measured variables, plays an important role in analyzing influence mechanisms and understanding dynamic behavior. Moreover, with an accurate topology capturing causality, it is easy to perform fault detection when determining root cause and fault propagation.

Nowadays, several data-based causality analysis methods have been proposed like Granger causality analysis (GCA) and transfer entropy (TE). Considering the bivariate situation, if the augmentation can get better prediction accuracy when the regression of a variable on its lagged values is compared to the augmented regression on lagged values of the other variable, then it is said this variable is Granger-caused by the other variable (Yang et al., 2014). Since it is easy to understand and implement (Duan et al., 2014), GCA becomes a popular method in causality analysis. Transfer entropy (Schreiber, 2000) is another popular approach which can be applied to nonlinear situation. It has been shown GCA and TE are equivalent in case of Gaussian distributed variables (Barnett et al., 2009). Similarly, the basic idea of transfer entropy is to which extent the uncertainty is reduced when the past observations of the other variable is included compared to the situation where information of one variable is obtained using the

past values of itself alone. The complexity of calculating probability density function and too many parameters limit the application of TE.

In addition, the influence of noise makes all data-based causality analysis method difficult to obtain an accurate result. Nalatore et al. (2007) and Overbey & Todd (2009) pointed out the influence of noise in GCA and TE respectively. To attenuate the noise influence, feature based causality analysis is proposed since extracted features contain less information about noise and reflect the essential rule of process itself.

Principal component analysis (PCA, Valle et al., 1999) and slow feature analysis (SFA, Shang et al., 2015) stand out among different feature learning methods in process monitoring. PCA aims to extract latent variables carrying most variance information while SFA aims to extract latent variables with slow changes. Both methods are seen as good approaches to extracting the essential rule of processes and reducing the influence of noise. Hence, substituting features for the original variables to perform causality analysis is a better choice.

In industrial processes, soft sensors aim to predict those hard-to-measure primary variables online with those easy-to-measure variables which have high correlations with primary variables (Wang et al., 2019). Since correlation does not mean causality, selected variables using correlation coefficients will have redundant information about primary variables. Meanwhile, the original variables contain noise, which reduces the prediction accuracy. As a result, causal feature based soft sensor is proposed in this paper.

[★] This work was supported by the National Natural Science Foundation of China (Nos. 61873142 and 61433001).

It should be noted that although some studies have been made to combine causality analysis and feature learning methods (Yuan, 2013; Zhou et al., 2009), they only value the convenience of computation and the preservation of information in extracted features since the dimensions of the original variables are rather high. This work, however, pays more attention to whether combinations of variables could exert more causal influence on the primary variables than the original variables and applies the feature based causality analysis to the soft sensor modeling.

The rest of the paper is arranged as follows. In section 2, two feature learning methods are reviewed. In section 3, Granger causality analysis is introduced. In section 4, feature based causality analysis and causal feature based soft sensor is proposed. Performance of proposed method using Tennessee Eastman benchmark process is evaluated in section 5. Finally, some conclusions are made in section 6.

2. FEATURE LEARNING

2.1 Principal component analysis

Principal component analysis (PCA) is a classical feature-extracting method aiming to find those latent variables who carry the most variance information from the original data. For a given standardized data matrix where n denotes the number of observations and m denotes the number of variables, PCA tries to decompose it as the following equation:

$$\mathbf{X} = \widehat{\mathbf{X}} + \mathbf{E} = \mathbf{T}\mathbf{P}^T + \mathbf{E} \quad (1)$$

where $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_A] = \mathbf{X}\mathbf{P} \in R^{n \times A}$ is the score matrix, $\mathbf{P} \in R^{m \times A}$ is the loading matrix, $\mathbf{t}_i \in R^n$ is the latent variable known as the principal component, $\mathbf{E} \in R^{n \times m}$ is the residual matrix, A is the number of principle components and it could be determined according to the methods proposed by Valle et al. (1999).

It should be noted that all latent variables are uncorrelated in PCA, implying that there is no redundant information among them. It is very important in this work as we could assume that extracted uncorrelated features would show fewer causal relationships than the original variables with high correlation in industrial processes. Meanwhile, the introduction of residual matrix is to eliminate noise information from latent variables since the variance of noise is quite small so that features learned in PCA are of practical meaning.

However, dynamics are nonnegligible in industrial process while PCA is a static model. Ku et al. (1995) used ‘time-lag shift’ method to include dynamic information in PCA model, which is known as Dynamic PCA (DPCA). History observations with time lag d are contained so that the input matrix is changed as $\mathbf{X} \in R^{n \times m(d+1)}$.

2.2 Slow feature analysis

Slow feature analysis (SFA) focuses on the mining of features which change slow. Since industrial processes have significant inertial characteristics, dynamic characteristics are often slow and those who change fast are assumed as noise.

For a stochastic signal $X(t)$, the slowness could be defined as follows:

$$\Delta(X(t)) = \left\langle \dot{X}^2(t) \right\rangle_t \quad (2)$$

where $\langle X(t) \rangle_t = \frac{1}{N} \sum_{i=1}^N X(t_i)$ means the time averaging value of a certain time series with N observations and $\dot{X}(t) = X(t) - X(t-1)$ is the first-order derivative of X with respect to time.

Considering a given input time series signal $\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_m(t)]^T$ with m variables, the purpose of SFA is to find a set of slow features $\mathbf{s}(t) = [s_1(t), s_2(t), \dots, s_m(t)]^T$ so that the slowness of extracted feature is minimal. The objective of this optimal problem is:

$$\begin{aligned} \min_{g_i(\cdot)} \langle \dot{s}_i^2 \rangle_t, i = 1, 2, \dots, m \\ \text{s.t. } \langle s_i \rangle_t = 0; \langle \dot{s}_i^2 \rangle_t = 1; \langle s_i s_j \rangle_t = 0, \forall i \neq j \end{aligned} \quad (3)$$

where $s_i(t) = g_i(\mathbf{x}(t))$ and could be written as $s_i(t) = \mathbf{w}_i^T \mathbf{x}(t)$ for the linear situation.

Shang et al. (2016) gave a detailed introduction of the solutions of SFA problem and geometric interpretations. Two steps of singular value decomposition are performed accordingly to solve the SFA problem in this work.

Obviously, a good feature-learning method should avoid the influence of noise. SFA excludes those slow features that are faster than all input since they behave like noise as Shang et al. (2015) suggests. For robustness, a q -upper quantile value of the slowness of inputs was introduced. Denoting M_e as the number of features that are ought to be removed, M_e is defined as follows:

$$M_e = \text{card}\{s_i | \Delta s_i > \max_j^q \{\Delta x_j\}\} \quad (4)$$

As mentioned in section 2.1, since latent variables in SFA are uncorrelated, it is assumed SFA can also get better causal factors. To introduce dynamics in SFA, observations with time lag d are also included in inputs like DPCA. Therefore, input vector could be redefined in the following format: $\mathbf{x}(t) \triangleq [\mathbf{x}^T(t), \mathbf{x}^T(t-1), \dots, \mathbf{x}^T(t-d)]^T$.

3. GRANGER CAUSALITY ANALYSIS

The idea of Granger causality comes from Wiener’s notion of causality, ‘ X causes Y if the predictability of Y could be improved by introducing the information of X ’. Lacking the machinery of prediction, Granger (1969) used a bivariate autoregressive model to formalize the idea, that is, the introduction of the history information of X reduces the prediction error of Y in a bivariate autoregressive model comparing with the model only using the past information of Y . Supposing that two time series $X(t)$ and $Y(t)$ could be represented as the restricted model using autoregressive process:

$$\begin{aligned} X(t) &= \sum_{j=1}^p a_{1,j} X(t-j) + \varepsilon_{1x}(t) \\ Y(t) &= \sum_{j=1}^p b_{1,j} Y(t-j) + \varepsilon_{1y}(t) \\ \text{var}(\varepsilon_{1x}(t)) &= \Gamma_{1x}, \text{var}(\varepsilon_{1y}(t)) = \Gamma_{1y} \end{aligned} \quad (5)$$

Jointly, they could be described by the full model as:

$$\begin{aligned} X(t) &= \sum_{j=1}^p a_{21,j} X(t-j) + \sum_{j=1}^p a_{22,j} Y(t-j) + \varepsilon_{2x}(t) \\ X(t) &= \sum_{j=1}^p b_{21,j} X(t-j) + \sum_{j=1}^p b_{22,j} Y(t-j) + \varepsilon_{2y}(t) \\ \text{var}(\varepsilon_{2x}(t)) &= \Gamma_{2x}, \text{var}(\varepsilon_{2y}(t)) = \Gamma_{2y} \end{aligned} \quad (6)$$

Comparing (5) with (6), Γ_{1x} measures prediction accuracy of $X(t)$ using its own information while Γ_{2x} measures prediction accuracy of $X(t)$ using information of both X and Y . According to Granger causality, if Γ_{2x} is less than Γ_{1x} , then it is said Y causes X in Granger causality.

In other words, the causal influence can be measured as:

$$F_{X \rightarrow Y} = \ln \frac{\Gamma_{1y}}{\Gamma_{2y}} \quad (7)$$

However, statistical significance needs to be accessed before a causal link is established. F-test is applied in Granger causality:

$$\frac{(RSS_1 - RSS_2) / (p_2 - p_1)}{RSS_2 / (m - p_2)} \sim F_{p_2 - p_1, m - p_2} \quad (8)$$

where RSS_1 and RSS_2 are the residual sum of squares in the restricted model and the full model respectively. Variables p_1 and p_2 denote the number of parameters in different models accordingly and m is the number of observations. With a significance level α , a causal link could be established if p -value is larger than the F distribution value.

For multivariate cases, conditional Granger causality (Geweke, 1984) is applied, where the full model uses all measured variables and the restricted model excludes the information of the variable to be detected. A Matlab toolbox called Granger Causality Connectivity Analysis (GCCA) is available and is introduced by Seth (2010). In this work, conditional Granger causality analysis is performed according to GCCA toolbox.

4. PROPOSED METHODOLOGY

4.1 Feature based causality analysis

In spite of the high-dimensional measurements in industrial processes, they do not act independently and are highly correlated in contrast (Dong & Qin, 2018). As a matter of fact, there is a low-dimensional feature space explaining the most information of observations. In other words, features could be the real causes of the process instead of the original variables. Hence, a feature based causality analysis is proposed in this paper.

Figure 1 illustrates how feature based causality analysis is performed. A feature learning method is firstly performed on original variables (x) to extract essential latent variables (s) in the process. Then given a primary variable to be analyzed, causality analysis is conducted to identify the causal features(c). The numbers of original variables, latent features and causal features are m , A and B respectively and they are sorted in a descending order.

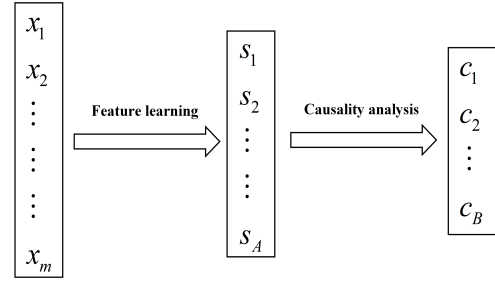


Fig. 1. Feature based causality analysis framework

4.2 Soft sensor design

For a high-dimensional industrial process, it's difficult to select appropriate secondary variables to predict primary variables since more variables mean the cost of complexity while fewer variables mean the loss of information. Obviously, a set of variables which have causal influence on primary variables are the most important variables. Enlightened by this idea, causal feature based soft sensor is proposed in this paper. Figure 2 shows the framework of causal feature based soft sensor.

As shown in Fig. 2, a feature learning method is applied to process data to extract features in the offline learning. Then given primary variables, feature based causality analysis is performed to obtain the causal features. Lastly, a soft sensor is trained to predict the primary variables. In online monitoring, extraction methods learned in offline learning stage are applied to process data and causal features could be extracted according to causal features' index. Assuming causal features as the inputs of the soft sensor that have been trained, predicted values of primary variables can be obtained.

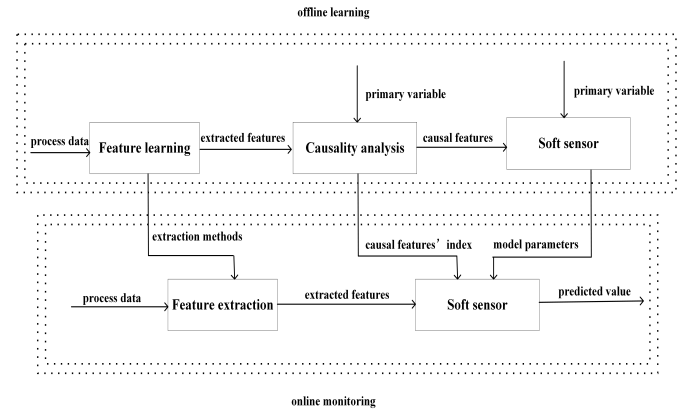


Fig. 2. Causal feature based soft sensor framework

The major characteristic is the introduction of causal features in soft sensor. On the one hand, the interpretability is enhanced since the inputs are causal variables. On the other hand, over-fitting could be reduced since the complexity is reduced when causal features are considered only.

5. CASE STUDY

In this section, the effectiveness of proposed methodology is presented through the Tennessee Eastman process (TEP), which was simulated to give a realistic industrial

process. Figure 3 (Chiang & Braatz, 2003) shows the diagram of the process. Relevant data are available in http://web.mit.edu/braatzgroup/TE_process.zip. It contains data under 22 different conditions, including one normal condition and other abnormal conditions. There are 52 different variables in this process, among which 33 variables can be measured in real time while the other 19 variables need to be analyzed respectively. Hence, 33 variables are chosen as the secondary variables and 19 variables are seen as the primary variables to be predicted. Detailed information is introduced in (Chiang & Braatz, 2003).

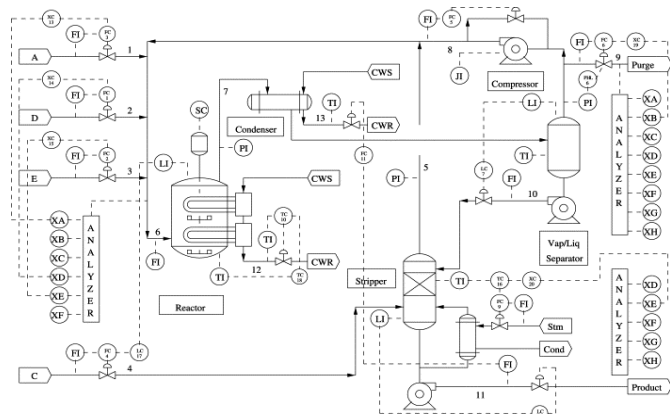


Fig. 3. Diagram of TE process

DPCA and DSFA methods are applied in feature extraction. The time lag d is set to 1 for both DPCA and DSFA. The number of principal components is set when cumulative percent variance (CPV) (Valle et al., 1999) is beyond 0.85 and the number of slow features is determined by q , which is set to 0.1. Causality analysis is performed using Granger Causality Connectivity Analysis toolbox where history length is set to 3 and the significance level is set to 0.05. Variable based causality analysis is also conducted for comparison with the feature based causality analysis.

5.1 Results of feature based causality analysis

In this part, we test whether the extracted features could be better causal factors than the original variables. Without loss of generality, all secondary variables are selected to extract the features, and all primary variables are chosen to test the performance of feature based causality analysis.

Figure 4 shows different numbers of causal features and causal variables when the primary variables are XMEAS(23-41). In most cases, the number of causal features is less than 10 while the number of causal variables is more than 10. It is obvious that inferred causal variables may be more than real ones for a complex system owing to the strong correlation and collinearity, which can be validated in the figure. However, since there is no correlation between each feature, numbers of causal features are much fewer than causal variables, which is more practical for causal analysis.

Then, causal influence is calculated according to (7). Figures 5 and 6 show average and maximum causal influence of causal features and causal variables respectively. It is expected high average and maximum causal influence

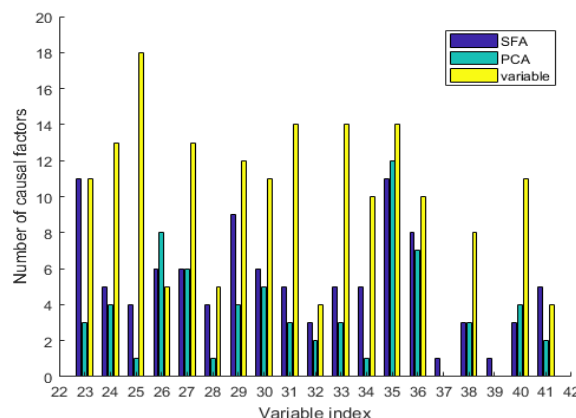


Fig. 4. Numbers of causal features and variables

should be reached when evaluating a good method. PCA based causality analysis outstands in most cases, while SFA based causality analysis can get a good result in some cases like XMEAS(39), XMEAS(40) and so on. Note that there are no causal factors of some primary variables in particular situations as shown in Fig. 4, corresponding causal influence are not shown in Figs. 5 and 6.

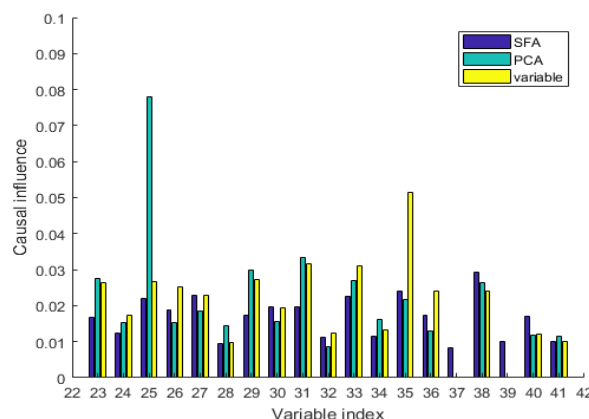


Fig. 5. Average causal influence

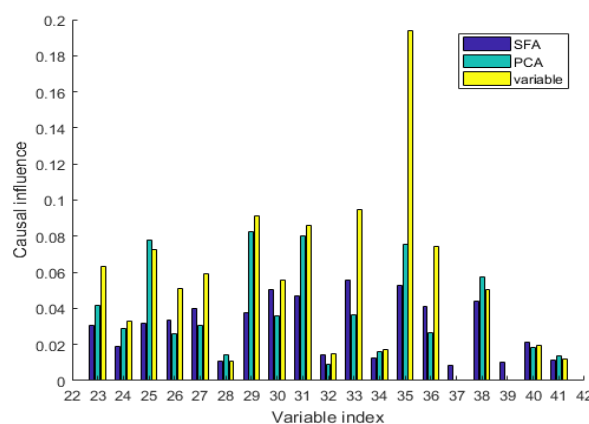


Fig. 6. Maximum causal influence

Additionally, some dominant features are expected to explain the whole process when applying the feature-extracted methods. Table 1 shows the ratio of top5 fea-

tures which carry the most variance information in PCA and change slowest in SFA respectively to causal factors detected by GCA. For example, the value ‘4/11’ in SFA about XMEAS(23) means that 11 slow features are detected as causal features while 4 of them are top5 slow features. Both PCA and SFA show a ratio over 0.4 and it is concluded that both features have a strong interpretability in the sense of causality.

Table 1. Ratio of Top5 features to causal factors

Primary variable	SFA	PCA
XMEAS(23)	4/11	2/3
XMEAS(24)	2/5	2/4
XMEAS(25)	3/5	1/1
XMEAS(26)	1/6	3/8
XMEAS(27)	3/6	2/6
XMEAS(28)	1/4	0/1
XMEAS(29)	4/9	3/4
XMEAS(30)	2/6	1/5
XMEAS(31)	3/5	1/3
XMEAS(32)	1/3	0/2
XMEAS(33)	4/5	2/3
XMEAS(34)	2/5	1/1
XMEAS(35)	4/11	5/12
XMEAS(36)	3/8	2/7
XMEAS(37)	0/1	0/0
XMEAS(38)	2/3	1/3
XMEAS(39)	0/1	0/0
XMEAS(40)	1/3	2/4
XMEAS(41)	1/5	0/2
Average	40.59%	40.58%

In conclusion, feature based causality analysis can find more practical causal factors in two aspects: an appropriate number of causal factors and appropriate causal influence, which may lead to a causal feature based analysis of complex systems in certain situations.

5.2 Results of soft sensors using causal features

In this part, causal feature based soft sensor is designed to show the feasibility. Component C in purge gas (XMEAS(31)) is chosen as the primary variable and PCA-based, SFA-based, variable-based soft sensors with causality analysis (CA) or not are tested.

The first soft sensor is based on least-squares regression (LSR). Two soft sensors based on recurrent neural networks (RNN) (Su et al., 1998) and artificial neural network (ANN) (Du, 2006) are tested in this paper. The number of cells and hidden neurons in RNN is set as 10, 20, 15 and 10 respectively, which means there are three RNN layers and one fully connected layer. There are 5, 10 and 5 hidden neurons in 3 hidden layers respectively in ANN. The activation function of two soft sensors in each layer is set as hyperbolic tangent function and Adam optimizer is used. It should be noted that all parameters of one certain soft sensor are set in the same way and the only difference between 6 submodels below is the input data. To avoid the occasionality in neural networks, one hundred repetitive testing experiments are conducted to get the average performance both in RNN and ANN. Correlation coefficient (r) between the predicted values and real values plus a root mean square error (RMSE) are calculated to show the performance.

Table 2 lists the prediction performance of different soft sensors. It can be vividly seen that although LSR without a causality analysis can always obtain a better performance in training data whether it is a feature based regression or a variable based regression, correlation coefficients of the regression using causality analysis are larger and root mean square errors are smaller in test data. What’s more, PCA based least-squares regression with causality analysis gets the best result and it corresponds to causal influence plot in Figs. 5 and 6 where PCA features get a good causal interpretability, which shows the feasibility of feature based causality analysis in soft sensor modeling. Similarly, results in RNN and ANN show soft sensors using causal factors get better results than soft sensors using untreated input factors while feature based methods get the best results in test data and it can also be concluded that causal feature based soft sensor get the best performance, which could be further applied into industrial processes.

Table 2. Prediction performance of different soft sensors

	model	r_{train}	$RMSE_{train}$	r_{test}	$RMSE_{test}$
LSR	PCA+CA	0.6679	0.2827	0.5206	0.2703
	PCA	0.7594	0.2472	0.5000	0.2787
	SFA+CA	0.7298	0.2597	0.5192	0.2714
	SFA	0.7987	0.2286	0.4875	0.2869
	Variable+CA	0.7527	0.2501	0.5115	0.2751
	Variable	0.7648	0.2448	0.5087	0.2758
RNN	PCA+CA	0.6787	0.2797	0.4920	0.2783
	PCA	0.7313	0.2594	0.3832	0.3099
	SFA+CA	0.7221	0.2646	0.4645	0.2882
	SFA	0.7147	0.2653	0.3264	0.3230
	Variable+CA	0.7203	0.2647	0.4634	0.2900
	Variable	0.7461	0.2535	0.4400	0.2977
ANN	PCA+CA	0.6486	0.2903	0.4941	0.2761
	PCA	0.7550	0.2493	0.4026	0.3135
	SFA+CA	0.7027	0.2717	0.4925	0.2803
	SFA	0.8052	0.2252	0.3735	0.3274
	Variable+CA	0.7015	0.2719	0.4792	0.2858
	Variable	0.7668	0.2440	0.4384	0.3039

Table 3 lists the number of input factors used in each soft sensor. It should be noted that fewer causal factors get the similar results in training data with better results in test data, which means a few input factors could represent the important behavior of the process and is necessary for the prediction of primary variables. Note that augmented data are used to extract dynamics, the number of input factors in SFA is larger than that in the original variables.

Table 3. Number of input factors in different soft sensors

Soft sensor	Number of input factors
PCA+CA	3
PCA	24
SFA+CA	5
SFA	38
Variable+CA	14
Variable	33

To evaluate the importance of causality analysis in soft sensor, random factors are selected as inputs. It is expected that prediction performance will decrease sharply since those inputs makes no sense for quality variables in most cases while those causal factors play an important role.

Table 4 lists the average performance of 100 times prediction when using random factors. It is vividly shown that prediction performances get worse in feature based soft sensors while prediction performances of variable based soft sensors do not decrease a lot. First, since most factors do not influence quality variables, random input factors will give a worse prediction performance than causal input factors. Second, variable based causality analysis is not so accurate since original data contain information of noise and are often correlated while feature based causality analysis may be more accurate since extracted features get rid of the noise and are orthogonal. Therefore, some variables not detected as causal factors may also contain information about quality variables, leading to the prediction performances of variable based soft sensors using random factors not decrease so sharply. And there are apparent differences between features so that causal information about quality variables is limited to a few factors, leading to a much worse prediction performance without them.

Table 4. Prediction performance of different soft sensors when using random factors

model		r_{train}	$RMSE_{train}$	r_{test}	$RMSE_{test}$
LSR	PCA	0.1779	0.3677	0.0794	0.3139
	SFA	0.2864	0.3555	0.1509	0.3154
	Variable	0.7213	0.2618	0.4853	0.2800
RNN	PCA	0.3128	0.3570	0.0683	0.3276
	SFA	0.3310	0.3584	0.0638	0.3358
	Variable	0.7146	0.2663	0.4296	0.2972
ANN	PCA	0.2099	0.3630	0.1021	0.3152
	SFA	0.2157	0.3706	0.0981	0.3236
	Variable	0.7016	0.2710	0.4448	0.2944

6. CONCLUSIONS

Feature based causality analysis is proposed in this paper to attenuate the influence of noise when using causality analysis. It outperforms the original variable based causality analysis in the case study and obtain more efficient causal factors. Moreover, the causal feature based soft sensor can avoid over-fitting. As the simulation shows, causal feature based soft sensor obtains the best performance and causal factors play an important role in predictions.

REFERENCES

Barnett, L., Barrett, A.B. and Seth, A.K. (2009). Granger causality and transfer entropy are equivalent for Gaussian variables. *Physical Review Letters*, 103(23), 238701.

Chiang, L.H. and Braatz, R.D. (2003). Process monitoring using causal map and multivariate statistics: fault detection and identification. *Chemometrics and Intelligent Laboratory Systems*, 65(2), 159–178.

Dong, Y. and Qin, S.J., (2018). Dynamic latent variable analytics for process operations and control. *Computers and Chemical Engineering*, 114, 69–80.

Du, D., Wu, C., Luo X., et al. (2006). Delay time identification and dynamic characteristics study on ANN soft sensor. *Sixth International Conference on Intelligent Systems Design and Applications*, Jinan, China. 1, 42–45.

Duan, P., Chen, T., Shah, S.L. and Yang, F. (2014). Methods for root cause diagnosis of plant-wide oscillations. *AIChE Journal*, 60, 2019–2034.

Geweke J.F.(1984). Measures of conditional linear dependence and feedback between time series. *Publications of the American Statistical Association*, 79(388), 907–915.

Granger, C.W.J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3), 425–438.

Ku, W., Storer, R.H. and Georgakakis, C. (1995). Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 30(1), 179–196.

Nalatore, H., Ding, M. and Rangarajan, G. (2007). Mitigating the effects of measurement noise on Granger causality. *Physical Review E*, 75(3 Pt 1), 031123.

Overbey, L.A. and Todd, M.D. (2009). Effects of noise on transfer entropy estimation for damage detection. *Mechanical Systems and Signal Processing*, 23(7), 2178–2191.

Schreiber T. (2000). Measuring Information Transfer. *Physical Review Letters*, 85, 461–464.

Seth, A.K., (2010). A MATLAB toolbox for Granger causal connectivity analysis. *Journal of Neuroscience Methods*, 186(2), 262–273.

Shang, C., Yang, F., Gao, X., et al. (2015). Concurrent monitoring of operating condition deviations and process dynamics anomalies with slow feature analysis. *AIChE Journal*, 61(11), 3666–3682.

Shang, C., Huang, B., Yang, F., Huang, D. (2016). Slow feature analysis for monitoring and diagnosis of control performance. *Journal of Process Control*, 39, 21–34.

Su, H.B., Fan, L.T. and Schlup, J.R. (1998). Monitoring the process of curing of epoxy/graphite fiber composites with a recurrent neural network as a soft sensor. *Engineering Applications of Artificial Intelligence*, 11(2), 293–306.

Valle, S., Li, W. and Qin, S.J. (1999). Selection of the number of principal components: The variance of the reconstruction error criterion with a comparison to other methods. *Industrial and Engineering Chemistry Research*, 38(11), 4389–4401.

Wang, K., Shang, C., Liu, L., et al. (2019). Dynamic soft sensor development based on convolutional neural networks. *Industrial and Engineering Chemistry Research*, 58(26), 11521–11531.

Yang, F., Duan, P., Shah, S.L. and Chen, T. (2014). Capturing connectivity and causality in complex industrial processes. *Springer*

Yuan Z. (2013). Combining independent component analysis and granger causality to investigate brain network dynamics with FNIRS measurements. *Biomedical Optics Express*, 4(11), 2629–2643.

Zhou, Z., Chen, Y., Ding, M., et al. (2009). Analyzing brain networks with PCA and conditional Granger causality. *Human Brain Mapping*, 30(7), 2197–2206.