Sensor Drift Detection in SNG Plant using Auto-Associative Kernel Regression

Jae-Min cha*, Taekyoung Lee, Joon-Young Kim, Junguk Shin, Jinil Kim, Choongsub Yeom
Plant SE Team
Institute for Advanced Engineering (IAE)
Yongin, Gyeonggi, South Korea
*dr.jaemin.cha@gmail.com

***Abstract*— With the rapid development of ICTs, condition monitoring has been used as a key technology in the plant industries. For reliable condition monitoring, sensors should output same values under same conditions regardless of time, but the sensitivity of sensors is gradually changed due to several factors such as temperature, humidity, contamination, aging, and etc. This type of situation is called as sensor drift problem. To solve this, several methods such as auto-associative neural network, auto-associative support vector regression, and etc. have been developed to detect sensor drifts earlier by estimating new input based on historical data. This study applied the auto-associative kernel regression model into a synthetic natural gas plant which produces synthetic natural gas from coals to detect sensor drifts during operation phase. To validate the auto-associative kernel regression model in the synthetic natural gas plant, a real data collected from an experimental operation are used. Based on the experimental results, the auto-associative kernel regression model can rapidly detect the sensor drift in the synthetic natural gas plant.

Keywords—sensor drift detection; auto-associative kernel regression; condition monitoring; synthetic natural gas

I. Introduction

Condition monitoring is a core technology to ensure safety and productivity by monitoring the conditions of equipment and detecting abnormal status earlier. For reliable condition monitoring, the sensors should output same values under the same conditions regardless of time. But the sensitivity of the sensors is gradually changed due to several reasons such as temperature, humidity, contamination, aging, and etc. This type of situation is called as sensor drift or sensor drift problem.

Sensor drifts can be critical for the reliable plant operation. Because if sensor drifts are occurred during the plant operation and not detected, the sensors send the drifted data to operators via the central control room. The users received the drifted data may adjust operating control parameters to maintain stable operation conditions. But sometimes the misadjusted controls can cause plant trip as well as accidents. That is, sensor drifts should be detected earlier.

To detect sensor drifts earlier, several auto-associative methods such as auto-associative neural network (AANN) [1], auto-associative support vector regression (AASVR) [2], auto-associative kernel regression (AAKR) [3], principal component support vector machine [4], and etc. have been developed by estimating new input based on historical data. 01 shows the basic scheme of condition monitoring for sensor drift detection using auto-associative methods.

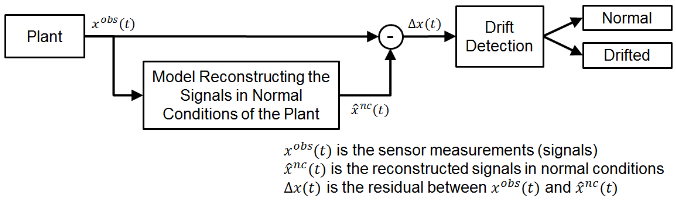

Fig. 1. Condition monitoring scheme for sensor drift detection using auto-associative methods

Meanwhile, synthetic natural gas (SNG) plants which produce natural gas from coals based on coal gasification process have also faced on sensor drifts during operations. Because variety of chemical substances are reacted and also reactants which can contaminate sensors are generated. Although sensor drifts are important issues in SNG plants, there are scarce studies focusing on sensor drift detection on SNG plants. This study applied the AAKR method into SNG plants. For this, real experimental operation data collected from a SNG plant of *A* company are used. The remainder of this paper are as follows. Section II simply presents a theoretical background. Experiments and results are presented in section III. Section IV concludes with summary.

.

II. Thereotical Background

A. Overview of SNG Process

This subsection briefly describes SNG process to understand SNG plants. The SNG process is generally composed of several processes from gasifier to SNG production via gas cooling, dust removal, acid gas removal (AGR), and methanation. For the methanation process, the SNG process generally consists of three to five reactors conducting following chemical reactions in eq. (1).

$$\begin{aligned} &CO + 3H_2 \rightarrow CH_4 + H_2O && \Delta H = -206\ \text{kJ/mol} \\ &CO_2 + 4H_2 \rightarrow CH_4 + 2H_2O && \Delta H = -165\ \text{kJ/mol} \\ &CO + H_2O \rightarrow H_2 + CO_2 && \Delta H = -41\ \text{kJ/mol} \end{aligned} \tag{1}$$

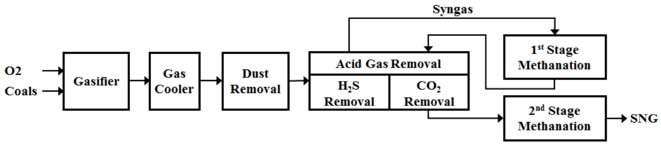

Fig. 2. Simplified block diagram of general SNG process

Firstly, coals and O_2 are entered into the gasifier process and syngas are generated via coal gasification. Next, syngas from gasifier enters into 1st stage methanation process after removing sulfur in the AGR process. Process gas from 1st stage methanation process enters into the AGR process to remove CO_2. In the AGR process, 98~99% of CO_2 in process gas is removed and process gas after removing CO_2 enters into 2nd stage methanation process. And then, SNG can be acquired. Fig. 2 shows a simplified block diagram of general SNG process.

B. AAKR

This subsection briefly describes of AAKR. AAKR is a type of similarity based model which is a non-parametric modeling technique that used the similarity of a query vector to memory vectors to infer the model's response. The basic idea of the AAKR is to perform the reconstruction of the current signal measurements, on the basis of historical signal observations. To explain the AAKR, detailed description from Hines and Garvey [4] are as follows.

To develop empirical model, memory vectors are stored in a matrix $\boldsymbol{X}$, where $X_{i,j}$ is the i^{th} observation of the j^{th} process variable. For n_m observations of p process variables, this matrix can be shown in eq. (2):

$$\boldsymbol{X} = \begin{bmatrix} X_{1,1} & \cdots & X_{1,p} \\ \vdots & \ddots & \vdots \\ X_{n_m,1} & \cdots & X_{n_m,p} \end{bmatrix} \tag{2}$$

A query vector can be represented by a $1 \times p$ vector of process variable measurements: $\boldsymbol{x}$.

$$\boldsymbol{x} = [x_1 \ x_2 \ \cdots \ x_p] \tag{3}$$

The corrected input is calculated as a weighted average of historical memory vectors: $\boldsymbol{X}_i$. The mathematical framework of this modeling technique is composed of three basic steps. First, the distance between a query vector and each of the memory vectors is computed. There are several distance functions such as Euclidean distance, Mahalanobis distance, Manhattan distance, and etc. that can be used, but the most commonly used function is the Euclidean distance.

For a single query vector, the distance is repeated for each of the n_m memory vectors, resulting in an $n_m \times 1$ matrix of distances: $\boldsymbol{d}$. Next, these distances are transformed to similarity measures used to determine weights by evaluating the Gaussian kernel, expressed by eq. (4):

$$\boldsymbol{w} = K_h(d) = \frac{1}{\sqrt{2\pi h^2}} e^{-d^2/h^2} \tag{4}$$

where is the kernel bandwidth, $\boldsymbol{w}$ are the weights for the n_m memory vectors.

Finally, these weights are combined with the memory vectors to make predictions according to eq. (5):

$$\hat{\boldsymbol{X}} = \frac{\boldsymbol{w^T X}}{a}, \text{ if } a = \sum_{i=1}^{n_m} w_i \tag{5}$$

III. Experiments

A. Data

For sensor drift detection experiment, real data are collected from experimental operation of the SNG plant of *A* company. Name of the company is hidden for confidentiality. The SNG plant has 83 sensor tags (process variables) in total. Sensor tags are composed of variety type of sensors such as temperature indicating alarms (TIAs), pressure indicating alarms (PIAs), pressure differential indicator (PDIs), and etc. Table I shows the sensor tags in the SNG plant. Among the tags, four PIA tags (PIA1114, PIA1201, PIA1302, and PIA1402) having high correlation (>0.9) are selected and used. Table II and fig. 3 shows correlation among the selected tags and Human Machine Interface (HMI) screen of the SNG plant showing where the selected tags are located.

To apply the AAKR into sensor drift detection problems, following data should be required as following: 1) train data to construct memory vectors, 2) test data to validate the performance of the method, and 3) drift data with drifted data for drift detection test. For this, the train data is collected during 9,000 seconds of normal operation with the selected tags. The test data is collected during 8,500 seconds of normal operation. The drifted data is also collected during 8,500 seconds having

TABLE I. SENSOR TAGS IN SNG PLANT

No.	Tag Name	No.	Tag Name	No.	Tag Name	No.	Tag Name
1	GAS1.H2	22	CO conversion	43	TIA1300F	64	PIA1302
2	GAS1.CO	23	20 CO conversion	44	TIA1300G	65	PIA1402
3	GAS1.CO2	24	CH4 selectivity	45	TIA1300H	66	PIA2501
4	GAS1.CH4	25	20 CH4 selectivity	46	TIA1300I	67	FIQ1201
5	MRU2.H2	26	TIA1200A	47	TIA1300J	68	FIQ1202
6	MRU2.CO	27	TIA1200B	48	TIA1300K	69	FIQ1501
7	MRU2.CO2	28	TIA1200C	49	TIA1300L	70	FIQ1502
8	MRU2.CH4	29	TIA1200D	50	TIA1400A	71	FIQ1502
9	GAS2.H2	30	TIA1200E	51	TIA1400B	72	MFC101.PV
10	GAS2.CO	31	TIA1200F	52	TIA1400C	73	MFC102.PV
11	GAS2.CO2	32	TIA1200G	53	TIA1400D	74	MFC103.PV
12	GAS2.CH4	33	TIA1200H	54	TIA1400E	75	MFC104.PV
13	GAS3.H2	34	TIA1200I	55	TIA1400F	76	MFC105.PV
14	GAS3.CO	35	TIA1200J	56	TIA2504A	77	Total input flow
15	GAS3.CO2	36	TIA1200K	57	TIA2504B	78	PCV1202
16	GAS3.CH4	37	TIA1200L	58	TIA2504C	79	PCV1302
17	GAS4.H2	38	TIA1300A	59	TIA2504D	80	PDI1201
18	GAS4.CO	39	TIA1300B	60	TIA2504E	81	PDI1301
19	GAS4.CO2	40	TIA1300C	61	TIA2504F	82	PDI1401
20	GAS4.CH4	41	TIA1300D	62	PIA1114	83	PDI1501
21	H2/CO	42	TIA1300E	63	PIA1201		

TABLE II. CORRELATION AMONG SELECTED SENSOR TAGS

Correlation	PIA1114	PIA1201	PIA1302	PIA1402
PIA1114	1.0000	0.992	0.998	0.990
PIA1201	0.9992	1.000	0.9988	0.9978
PIA1302	0.9998	0.9988	1.000	0.9997
PIA1402	0.9990	0.9978	0.9997	1.000

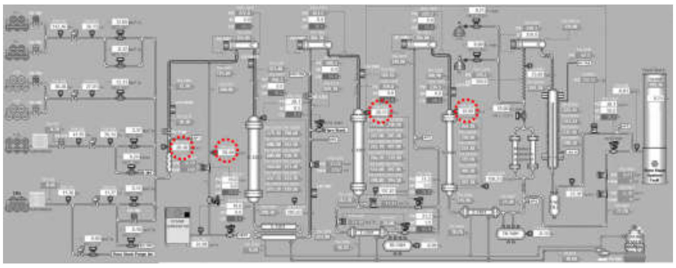

Fig. 3. HMI screen of the SNG plant and selected sensors (*red circle*)

same operation condition with the test data but PIA1201 tag is drifted out from around 2,000 second. For reference, the sampling time is one second. Fig. 4 shows the train, test, and drifted data each.

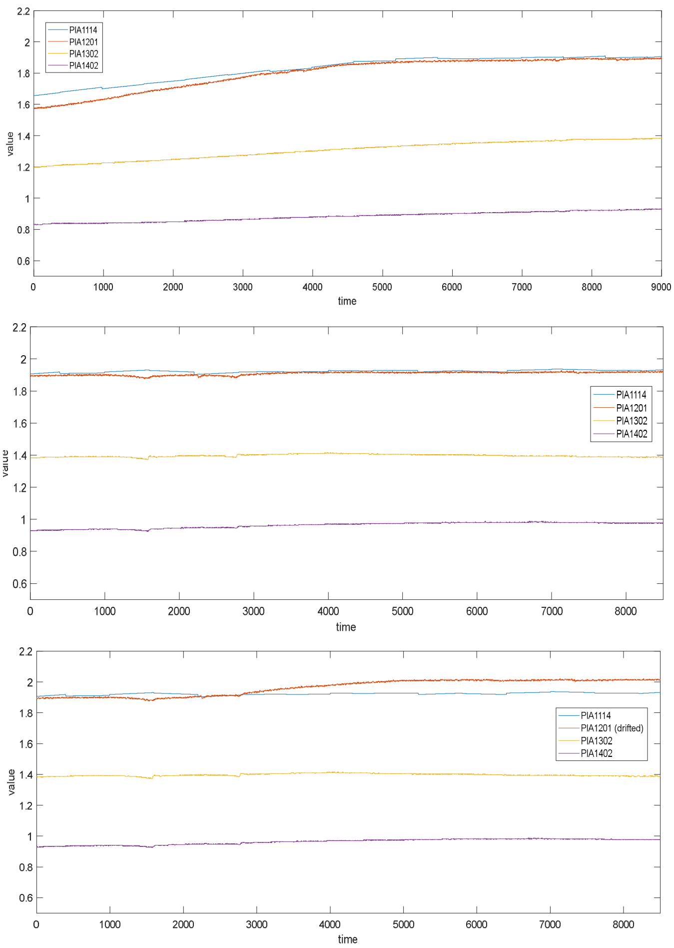

Fig. 4. Train, test, and drifted data

B. *Experiment and Results*

Sensor drift detection experiment was conducted as following steps. 1) The AAKR model was constructed using the train data. 2) The constructed model was validated with the test data. 3) Drift detection was tested with the drift data. More detailed description of each step is as follows. First of all, memory matrix were constructed with min-max vector selection method [5]. Size of the memory vectors is 300. After the memory vectors construction was complete, Euclidean distance was computed between the memory vectors and test data. Next, these distances were transformed to similarity measures and weights were determined using Gaussian kernel. Finally, these weights are combined with the memory vectors to make predictions. To validate how well the model predicts the correct values for the test data, accuracy between the predicted data and test data were measured. In this research, metric for accuracy is simply defined as the mean squared error (MSE) between the predicted data and test data. Equation for single variable is simply shown in eq. (6) [4].

$$A = \frac{1}{N}\sum_{i=1}^{N}(\hat{x}_i - x_i)^2 \tag{6}$$

where N is the total number of test observations, $\hat{x}_i$ is the model prediction of the i^{th} test observation, x_i is the i^{th}observation of the test data.

Table III and fig. 5 shows the accuracy of the prediction model using AAKR and comparison results. As shown in the table and figure, MSE is quiet small. It means that the prediction model is quiet accurate and the model is robust.

Since the prediction model was validated, drift detection was tested by comparing between the validated model and drift data. To detect the sensor drift, residuals between the drifted data and prediction model were calculated. A threshold for drift detection is 1% in this study. That is, if residuals between the prediction

TABLE III. ACCURACY OF THE PREDICTION MODEL USING AAKR

h (Gaussian kernel bandwidth)	PIA1114	PIA1201	PIA1302	PIA1402
0.3	4.1249e-04	3.9053e-04	2.7840e-04	0.0014
0.6	4.9937e-04	4.1588e-04	3.7772e-04	0.0016
0.9	5.3694e-04	4.5932e-04	4.8843e-04	0.0019

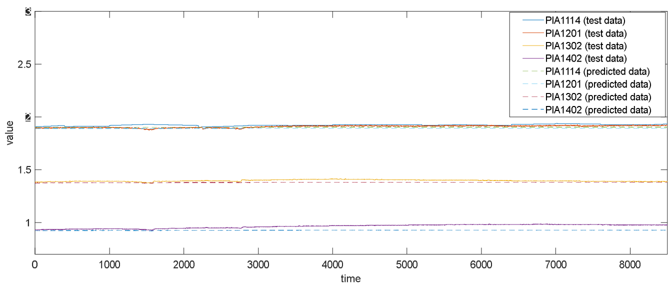

Fig. 5. Comparison between test data and validated model

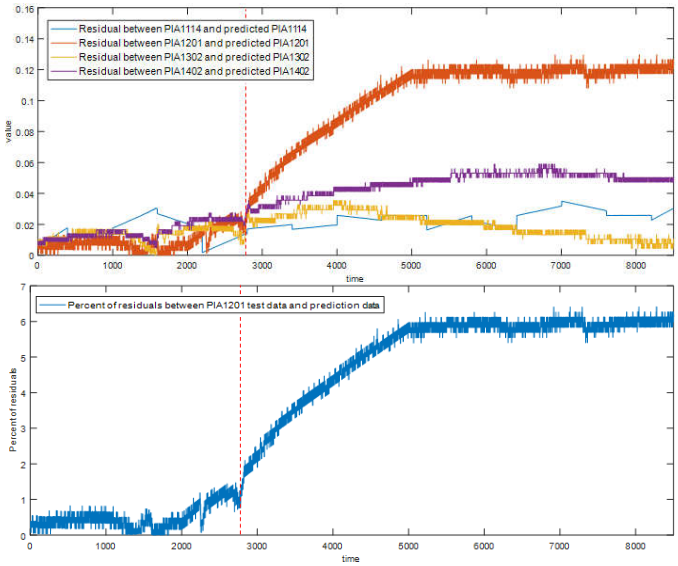

Fig. 6. Residuals between the predicted model and the drifted data (*top*) and percentage of residuals between the test data and the predicted model of PIA1201 (*bottom*)

model and drifted data are larger than 1%, sensor drifts are detected. Fig. 6 shows residuals between the prediction model and drifted data (*top*) and percentage of residuals between the test data and the prediction data of PIA1201 (*bottom*). As shown in the figures, residuals of the PIA1201 test data and prediction model are gradually getting larger from the point of 2,800 second. With the 1% of threshold criteria, the sensor drift can be detected after the sensor drift was occurred.

IV. Concluding Remarks

Sensor drifts have become an important issue in SNG plants. To solve sensor drift problems, this study applied the AAKR method into a SNG plant with collected data during experimental operations. Based on the AAKR method, the prediction model was robust and the sensor drift can be detected earlier after the sensor drift was occurred.

Acknowledgment

This work was supported by the Technology Innovation Program (10063187, Engineering Technique for Power Generation System Design using Industry Waste Heat, 10072058, Model Based Configuration Management System for Empowering Maintenance Engineering Capability of Small/Medium Size Plant) funded By the Ministry of Trade, Industry & Energy(MOTIE, Korea).

References

[1] J.W. Hines, R.E. Uhrig and D.J. Wrest, “Use of autoassociative neural networks for signal validation”, Journal of Intelligent and Robotic Systems, vol. 21, no. 2, pp. 143-154, 1998.

[2] S.-J. Kim, J.-G. Lim, and I.-Y. Seo, “A comparison of auto-associative models using support vector regression (SVR) and partial least square regression (PLSR) for online monitoring technique”, Proc. of the Korean Institute of Intelligent Systems Conference, vol. 20, no. 4, pp. 483-488, 2010.

[3] I.Y. Seo, B.N. Ha, S.W. Lee, C.H. Shin, and S.J. Kim, “Principal components based support vector regression model for on-line instrument calibration monitoring in NPPs”, Nuclear Engineering and Technology, vol. 42, no. 2, pp. 219-230, 2010.

[4] J.W. Hines and D.R. Garvey, “Development and application of fault detectability performance metrics for instrument calibration verification and anomaly detection”, Journal of Pattern Recognition Research, vol. 1, pp. 2-16, 2006.

[5] J.W. Hines and R. Seibert, “Technical review of on-line monitoringt techniques for performance assessment”, Vol. 3, 2007.