Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Adaptive Ensemble of Gaussian Processes Models for Prediction in Filamentous Sludge Bulking Recognition

Yiqi Liu*. Min Xie**

* *School of Automation Science & Engineering, South China University of Technology, Wushan Road, Guang Zhou 510640, China. (e-mail: aulyq@scut.edu.cn) .*
** *Department of Systems Engineering and Engineering Management, City University of Hong Kong, Kowloon, Hong Kong (e-mail: minxie@cityu.edu.hk)*

Abstract: Filamentous sludge bulking is considered as the most serious problem or fault usually happening in wastewater treatment plants (WWTPs) adopting activated sludge process (ASP). Proper process monitoring of sludge-bulking-related but hard-to-measure variables are nowadays one of bottlenecks limiting WWTP management with significant safety and efficiency implications. In this light, Global Gaussian Processes Regression (GPR) models and Local GPR models are learned by ensemble learning for quality-related but hard-to-measure variables prediction. Such coordination is able to capture global and local process behaviors properly, and then to obtain more robust and accurate prediction. To further approach the prediction deterioration as time evolutes, this paper proposed an adaptive ranking strategy to ensemble the sub GPR models. In this adaptive strategy, we used the moving-window technique to rank and to select few of the best sub-model predictions, and then average them together to make the final predictions. Also, due to the use of GPR as the sub-model, the proposed methodology is able to describe the uncertainties properly. The proposed prediction model has been validated in a real WWTP with filamentous sludge bulking. The results show that the proposed methodology is able to predict the quality-related variable with Root Mean Square Error (RMSE) being 25.6% and 21.6% better than the Bagging GPR model and the Average Ensemble GPR model, respectively.

Keywords: Ensemble learning, Soft-sensor, Gaussian Process Regression, Adaptive learning, Prediction

1. INTRODUCTION

Activated sludge process has been widely adopted to remove pollutants in wastewater treatment plants (WWTPs). However, stable operation of activated sludge process is often compromised by the occurrence of filamentous bulking. Filamentous bulking sludge, a term used to describe the excess proliferation of filamentous bacteria, often results in deteriorating sludge settleability, poorer operational performance and higher treatment cost (Martins, Pagilla, Heijnen and van Loosdrecht 2004, Olsson 2012). Filamentous bulking sludge is considered the most serious problem usually happening in WWTPs adopting ASP. It is documented that more than 50% of WWTPs encounter the filamentous bulking sludge problem worldwide (Martins, et al. 2004).

To prevent from serious deterioration of sludge settleability, soft-sensors have been proposed to achieve early warning for filamentous sludge bulking. Soft-sensor models can be dependent on mechanistic or data-driven models (Dürrenmatt and Gujer 2012). Data-driven soft-sensor gains popularity resulting from the fact that they do not require detailed understanding of the system (Kadlec, Gabrys and Strandt 2009). Popularity of Gaussian processes model provides an alternative to act as a soft-sensor to predict sludge bulking related variables. Gaussian processes models are, inherently, a global model, which fits a distribution over data and implies that the mathematical formulation of the assumed model governs the generation of data in the learning task. In the global learning, GPR model are usually approximated by matrix approximation, likelihood approximation methodologies (Snelson and Ghahramani 2006). Depending on the specific independence assumed, there are a number of variants to the approach, such as, partial independent conditional (PIC) (Snelson and Ghahramani 2007). One of the main disadvantages in global learning is the model selection problem. More precisely one still needs to select suitable and appropriate parameters to represent the observed data globally. Some researchers have argued that it is difficult if not impossible to obtain a general and accurate global learning. Hence, local learning has recently attracted much interest (Snelson, et al. 2007, Yuan, Ge, Huang, Song and Wang 2017). Local learning focuses on capturing only useful local information from the observed data. Chiwoo Park et al. proposed a DDM-GPR (Domain Decomposition Method-GPR) model, aiming to deal with non-stationary changes adaptively with cheap computation (Park, Huang and Ding 2011). Another way for localized regression of GPR models is to build multiple local predictors and to combine them by taking a weighted average of the local predictions, such as local probabilistic regression (LPR-GPR) (Urtasun and Darrell 2008), mixture of Gaussian process experts (Rasmussen and Ghahramani 2002). Because of the averaging mechanism, all these methods avoid the discontinuity problem of local kriging. Even though recent research progress and empirical studies demonstrate that the local learning paradigm is superior to global learning in some domains, employing only local information may lose the overall view of data. Local learning does not grasp the structure of the data, which may prove to be critical for guaranteeing better performance. In summary, there are complementary advantages for both local learning and global learning. Global learning summarizes the data and provides the practitioners with knowledge on the structure of data, since with the precise modelling of phenomena, the observations can be accurately regenerated and therefore thoroughly studied or analyzed.

To incorporate these two seemingly different yet complementary characteristics in an integrative framework that achieves good prediction accuracy, ensemble learning provides an alternative (Sagi and Rokach 2018). Ensemble learning is a simple but powerful strategy to improve soft-sensors prediction. The main bottleneck limiting the ensemble-learning-based soft-sensors performance is how to control individual model diversity and to operate the sub-model combination. To increase the sub-model diversity, three kinds of GPR models (Global GPR model: PIC-GPR models; Local GPR models: LPR-GPR and DDM-GPR) are implemented as the sub-models. Through coordinating the local and global GPR models, ensemble learning is able to capture global and local process behaviours. By doing so, we can obtain more robust and accurate prediction by making sub-model diverse. However, the standard ensemble-learning-based soft-sensors would deteriorate as the process evolutes, even though the diversity of GPR models is able to approach most of process conditions. It is unfortunate that the standard adaptive strategies usually achieve the highly accurate performance at the cost of intensive computation. It is mainly resulting from the fact that these strategies have to re-train the prediction model at each update. To deal with this issue, this paper proposed an adaptive ranking strategy to combine all the sub-models aiming to achieve more robust and accurate predictions. Different from the standard adaptive strategy, we used the moving-window technique to rank and to select few of the best sub-model predictions, and then average them together to make the final predictions. The motive behind this strategy is based on the assumption that few sub-models with the best performance in the past few steps will have similar prediction accuracy for the new coming data points. Also, due to the use of GPR as the sub-model, the proposed methodology is able to describe the uncertainties properly.

2. ADAPTIVE RANKING ENSEMBLE OF GPR MODELS (AR-EGPR)

In general, global GPR models, such as standard GPR or PIC-GPR models, are suitable for stationary processes, but are not able to approach abrupt local changes or non-stationary features. On the contrary, local GPR models, such as DDM-GPR and LPR-GPR, decompose the entire domain into smaller sub-domains and make a prediction at a test point using the related sub-domain training points, thus being able to have the adaptive to the non-stationary changes and efficient computation with operations for model training. Also, it is well known that local GPR models suffer from discontinuities in prediction on the sub-domain boundaries. Given the pros and cons corresponding to global and local GPR models, this paper proposed an ensemble learning to coordinate global and local models to ensure the predicted models able to adapt to both of stationary and non-stationary processes. Coordination of global and local models will, in turn, add more diversity of ensemble learning. Diversity is, essentially, a basic and necessary property to acquire acceptable accuracy. Different from the standard ensemble learning by using average strategy, a moving window is used to select few sub-models with the relatively best performance to serve as the baseline to justify which models are used for sequential prediction. The procedure of adaptive ranking ensemble learning is briefly described here.

Suppose the original testing data are $Z_T = \{(x_i, y_i)|_{i=1}^{l}\}$, the moving window length is s and the number of sub-model prediction values at each predicted step is m. During the current time, t, the moving window will envelop the data set $\{(x_i, y_i)|_{i=t-s}^{t-1}\}$. Since the predicted values for $x_{t-s,\cdots,t-1}$ have been known already, i.e., the predicted values for m sub-models, $\{(\hat{y}^{j}_{t-s,\cdots,t-1})|_{j=1}^{m}\}$ have been derived, *RMSE* value of each sub-model in the moving window can be obtained by using the following equation:

$$RMSE = \sqrt{\frac{1}{s}\sum_{j=1}^{s}(y_j - \hat{y}_j)^2} \qquad (1)$$

It is reasonable to assume that the prediction performance with respect to the time, t, is similar to the most recent prediction performance, such as, prediction at $t-s,\cdots,t-1$. Therefore, few sub-models with the relatively better performance at $t-s,\cdots,t-1$ can be used for prediction at time, t, if receiving the current input values, x_t. R sub-models with the best RMSE values in the moving window will be selected as the base models for the sequential prediction if given x_t. It is important to notice that the window length can be selected by checking Auto correlation function (ACF) and Partial correlation function (PCF) of the target variable, y. To assess the uncertainty level, the used indicator is the negative log predictive density (NLPD)

$$NLPD = \frac{1}{m}\sum_{i=1}^{m}\left[\frac{(\hat{y}^*(i)-\mu_*(i))}{2\sigma_*^2(i)} + \frac{1}{2}\log\left(2\pi\sigma_*^2(i)\right)\right] \qquad (2)$$

which considers the accuracy of the predictive variance σ_*^2 as well as the mean prediction μ_*. To further assess the prediction performance, the correlation coefficient, r, is defined and used as well.

The predictive mean and variance for the combined models can be calculated based on the property of the Gaussian mixture model:

$$\mu_* = E(y^*) = \frac{1}{R}\sum_{i=1}^{R}\hat{y}^*(i) \qquad (3)$$

$$\sigma_* = \mathrm{var}(y^*) = \frac{1}{R}\sum_{i=1}^{R}\sigma_{y^*}^2(i) + \frac{1}{R}\sum_{i=1}^{R}(\hat{y}^*(i) - E(y^*))^2 \qquad (4)$$

It has been observed that this simple averaging rule can significantly improve the model accuracy and robustness in various applications. It is important to notice that the standard variance can be achieved simultaneously with the mean values in the GPR model. σ_* can be used to indicate how likely of each prediction is of being correct. σ_*, $2\sigma_*$ and $3\sigma_*$ represent a confidence of 68.3%, 95.5% and 99.7%, respectively, meaning that the percentage of erroneous predictions in predicted data set will not exceed 31.7%, 4.5% and 0.3%. The resulting soft sensors using this technique can give confidence prediction values rather than bare predictions. Thus, this step is able to check how reliable the resulted predictive regions are.

3. CASE STUDY

3.1 Background of filamentous sludge bulking

The presented case is a full-scale WWTP (Beijing, China), which mainly treated municipal wastewater (480,000 population equivalents) with an Oxidation ditch (OD) process. OD process is a modified activated sludge biological treatment process that utilizes long solids retention time (SRT) to achieve good nitrogen removal performance. In this plant, the average influent flow was about 170,000 m^3/d, with an average OD hydraulic retention time (HRT) of 16.5 h. SRT was kept 15-22 d by withdrawing sludge from the secondary settler. Due to low COD loading rate (<0.25 kgCOD/kgMLSS/d), the occurrence of filamentous bulking sludge was observed in this plant. The phenomenon of bulking sludge lasted for about half a year. Lots of on-line measured parameters were recorded during the period of filamentous sludge bulking at one-day sample rate. These data were used to develop and validate the model in this study.

Filamentous bacteria are normal components of activated sludge biomass, where the existence of a fraction of filamentous bacteria is important and helpful to form flocs by serving as the floc-backbone for other bacteria to attach. Filamentous bulking sludge, a term used to describe the excess proliferation of filamentous bacteria, often results in slower settlement, poorer operational performance and higher treatment cost (Martins, et al. 2004, Olsson 2012). Sludge Volume Index (SVI) is an empirical measurement used to characterize the sludge bulking problem and hard-to-measured variable, being claimed that sludge bulking occurs when SVI is larger than 100 mL/g (Soyupak 1989). Some different values for SVI of 150, 220, even 280 mL/g are also documented (Rensink 1974). In this paper, 200 mL/g severed as the control limit for filamentous sludge bulking. Even proper control limit is obtained, it is still an open problem to seek accurate models to satisfy the complex characteristics of WWTP (Mogens, Willi;, Takashi; and Mark 2000). A range of factors, including feed quality (e.g. chemical oxygen demanding (COD) concentration), operational and environmental conditions (e.g. temperature, dissolved oxygen (DO) concentration, and COD loading rate), usually affect sludge settleability. Nonlinear dynamics, significant uncertainty (wastewater loads and weather) and multiple time scales further increase complexity to a sludge bulking modeling. Therefore, the development of proper models to prevent and control filamentous bulking is critical to ensure successful and stable operation of WWTPs using activated sludge process. One of plausible ways is to construct a model responsible to predict the filamentous sludge bulking related parameters SVI, thus appropriate actions can be promptly implemented to prevent the deterioration of sludge settleability. The selected input variables for model construction are shown as the paper (Liu, Guo, Wang and Huang 2016). 212 data points were sampled from the field. Data for the first 127 days were used for training, the remaining was for testing.

Scenario definition of predicted models: PIC (Partial Independent Conditional- GPR), DDM (Domain Decomposition Method-GPR), LPR (Local Probabilistic Regression-GPR), BGP (Bagging GPR), AGP (Average-GPR), R-EGP (Ranking ensemble-GPR) and AR-EGP (Adaptive Ranking ensemble-GPR)

3.2 Prediction performance

As profiled in Fig. 1 that AR-EGPR performs best for the SVI prediction with the *RMSE* and *r* being 0.29 and 0.95. This can be explained by the ability of AR-EGPR model to deal with the process data adaptively. It is also obvious that the deviation happens mainly in the significant variation stages. The main reason why averaging ensemble learning is able to better the prediction is that the ensemble generalization error is always smaller than the expected error of the individual models.

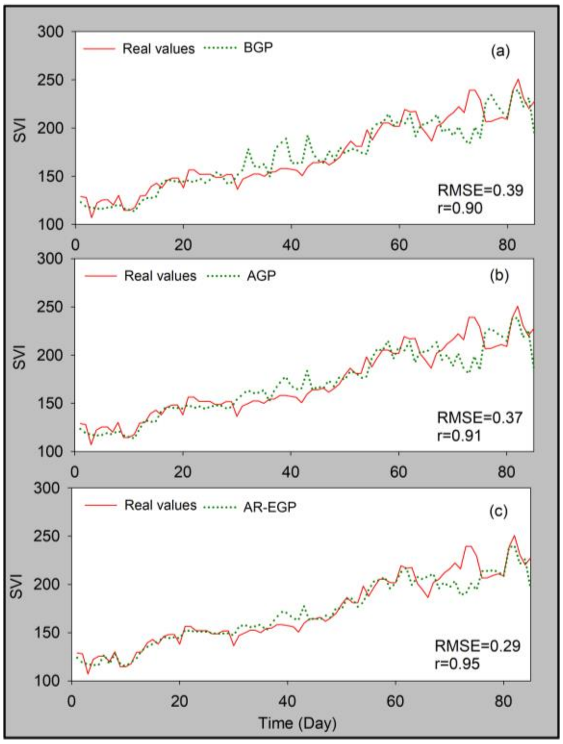

Fig.1 The SVI prediction performance of BGP, AGP and AR-EGP in the case study

As shown in Fig. 2 (a), the NLPD of the AR-EGP model is much smaller than that of other models, implying that the goodness of fit of AR-EGP is better. For this case study, PIC, DDM, LPR, BGP, AGP, R-EGP and AR-EGP have higher, and sometimes much higher, NLPD than AR-EGP. By the definition of NLPD, both a big RMSE and a small predictive variance will lead to a high NLPD. Thus, we can infer that, for the case study, the differences of NLPD between AR-EGP and AGP are mainly caused by too small predictive variances of AGP (i.e., AGP underestimates the predictive variances considerably), since the RMSEs produced by the two methods are very close for other data sets, the differences in NLPD come from the differences in predictive variance. For the slight non-stationary data set, the performance shape of NLPD with respect to all predicted models is very similar (Fig.2(a)). BGP has the highest RMSE and NLPD, suggesting that BGP might not be very competitive for both of stationary and non-stationary data sets. By comparing R-EGP and AR-EGP in Fig.2(a), it is also important to notice that they achieve very similar performance with NLPD being 0.38 and 0.36, respectively. On the contrary, the NLPD of AR-EGP is 19.4% better than that of R-EGP for the non-stationary scenario. This can be explained by the fact that moving-window-based adaptive ranking strategy is capable of enhancing prediction performance. Overall, AR-EGP outperforms all other methods even for the slightly non-stationary data sets.

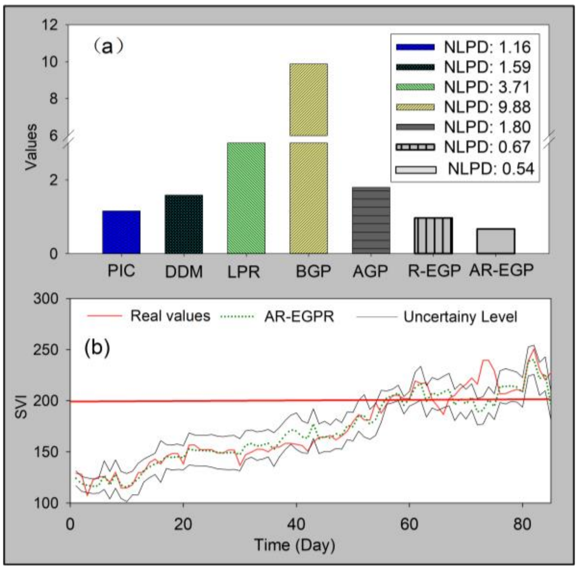

Fig.2 (a) The NPLD comparison of all GPR models; (b) The fitting of the predicted with the AR-EGP model and the true values for the testing samples [The red straight line is the SVI control limit]

To further check how reliable the predictive regions are, percentage of wrong predictive intervals is counted. The results in Table 1 confirm the validity of our algorithm: the rate of successful predictions is at least equal to the desired accuracy. Table 1 suggests that the predicted results with 90% confidence level are not acceptable with 18% predicted values out of control intervals. On the contrary, predicted results with 95% and 99% confidence level perform well. Fig. 2(b) complements the information given in Table 1 for prediction with 90% confidence. As profiled in Fig. 2(b), the fitting between prediction values and true values are acceptable, even though only with 82% confident level. The prediction deviation mainly happens in the places with the abrupt changes.

Table 1 Empirical Reliability of AR-EGP

Comments	Empirical Reliability (%)		
Empirical confidence	90%	95%	99%
Predicted confidence	82%	95.4%	99.7%

Fig. 2(b) displayed that the confidence would be widened due to deviation of steady state values. This is able further to check the tightness of our predictive regions for a specific significance level.

6. CONCLUSIONS

To monitor quality-related but hard-to-measure variables in industrial processes, this paper proposed an adaptive ensemble learning framework of Gaussian Process models. The framework is able to coordinate the local and global GPR models to capture process behaviours properly and to ensemble the sub GPR models adaptively, finally to obtain more robust and accurate prediction. Also, the proposed methodology is able to describe the uncertainties. The proposed prediction model, AR-EGPR, has been validated in a real WWTP with the filamentous sludge bulking fault. The results show that the proposed methodology is able to predict the quality-related variable with RMSE being 0.29 and 0.95 in the case study and being 25.6% and 21.6% better than the Bagging GPR model and the Average Ensemble GPR model, respectively. Ensemble learning achieved the better accuracy at the cost of increasing the computational consumption. Further theoretical study is needed to investigate this issue, which is an interesting on-going research topic.

REFERENCES

Dürrenmatt, D. J., and Gujer, W. (2012), "Data-Driven Modeling Approaches to Support Wastewater Treatment Plant Operation," *Environmental Modelling & Software*, 30, 47-56.

Kadlec, P., Gabrys, B., and Strandt, S. (2009), "Data-Driven Soft Sensors in the Process Industry," *Computers & Chemical Engineering*, 33, 795-814.

Liu, Y., Guo, J., Wang, Q., and Huang, D. (2016), "Prediction of Filamentous Sludge Bulking Using a State-Based Gaussian Processes Regression Model," *Scientific Reports*, 6, 31303.

Martins, A. M. P., Pagilla, K., Heijnen, J. J., and van Loosdrecht, M. C. M. (2004), "Filamentous Bulking Sludge—a Critical Review," *Water Research*, 38, 793-817.

Mogens, H., Willi;, G., Takashi;, M., and Mark, V. L. (2000), *Activated Sludge Models Asm1, Asm2, Asm2d and Asm3*, London: IWA publishing.

Olsson, G. (2012), "Ica and Me – a Subjective Review," *Water Research*, 46, 1585-1624.

Park, C., Huang, J., and Ding, Y. (2011), "Domain Decomposition Approach for Fast Gaussian Process Regression of Large Spatial Data Sets," *Journal of Machine Learning Research*, 12, 1697−1728.

Rasmussen, C., and Ghahramani, Z. (2002), "Infinite Mixtures of Gaussian Process Experts," *Advances in Neural Information Processing Systems*, 2.

Rensink, J. H. (1974), "New Approach to Preventing Bulking Sludge," *Journal (Water Pollution Control Federation)*, 46, 1888-1894.

Sagi, O., and Rokach, L. (2018), "Ensemble Learning: A Survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8, e1249.

Snelson, E., and Ghahramani, Z. (2006), "Sparse Gaussian Process Using Pseudo-Inputs," *Advances in Neural Information Processing Systems*, 18, 1257--1264.

Snelson, E., and Ghahramani, Z. (2007), "Local and Global Sparse Gaussian Process Approximations," *Journal of Machine Learning Research - Proceedings Track*, 2, 524-531.

Soyupak, S. (1989), "Effects of Operational Parameters on the Settling Properties of Activated Sludge," *Environmental Technology Letters*, 10, 471-478.

Urtasun, R., and Darrell, T. (2008), *Sparse Probabilistic Regression for Activity-Independent Human Pose Inference*,

Yuan, X., Ge, Z., Huang, B., Song, Z., and Wang, Y. (2017), "Semisupervised Jitl Framework for Nonlinear Industrial Soft Sensing Based on Locally Semisupervised Weighted Pcr," *IEEE Transactions on Industrial Informatics*, 13, 532-541.