Comparison of machine learning methods for multiphase flowrate prediction

Citation for published version:
Jiang, Z, Wang, H, Yang, Y & Li, Y 2020, Comparison of machine learning methods for multiphase flowrate prediction. in *2019 IEEE International Conference on Imaging Systems and Techniques (IST).* IEEE, pp. 1-6, 2019 IEEE International Conference on Imaging Systems and Techniques , ABU DHABI, United Arab Emirates, 8/12/19. https://doi.org/10.1109/IST48021.2019.9010450

Digital Object Identifier (DOI):
10.1109/IST48021.2019.9010450

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
2019 IEEE International Conference on Imaging Systems and Techniques (IST)

Comparison of machine learning methods for multiphase flowrate prediction

Zhenyu Jiang, Haokun Wang, Yunjie Yang
Agile Tomography Group, School of Engineering
University of Edinburgh
Edinburgh, UK
y.yang@ed.ac.uk

Yi Li
Division of Marine Science and Technology
Tsinghua Shenzhen International Graduate School
Shenzhen, Guandong,China
liyi@sz.tsinghua.edu.cn

***Abstract*—In this paper, three prevailing machine learning methods, i.e. Deep Neural Network (DNN), Support Vector Machine (SVM) and Gradient Boosting Decision Tree (GBDT) models were investigated and compared to estimate the flowrate of oil/gas/water three-phase flow. The time-series differential pressure signals collected from Venturi tube together with pressure and temperature measurements were utilized as input. Multiphase flow experiments were conducted on a laboratory-scale multiphase flow facility. Experimental results suggest that DNN and SVM based methods were able to achieve accurate and reliable estimation of multiphase flowrate, whilst GBDT failed to fit the estimation process well. Another finding emerged from this study is that volumetric gas phase flowrate can also be accurately predicted by implementing SVM model.**

***Index Terms*—deep neural network, flowrate, gradient boosting decision tree, machine learning, multiphase flow, support vector machine.**

I. Introduction

Accurate and instantaneous estimation of the multiphase flowrate is of significant importance to assist petroleum, gas and other multiphase flow industries to reduce cost, emission and enhance efficiency. Conventional studies suggest that separating each phase out of the mixture and then using single-phase flow meters to perform single-phase flow measurement is an achievable method [1]. However, this method always needs complex separating equipments and lack of real-time sensing capability.

Several attempts have been made to find potential measurement techniques for multiphase flow [2]. Differential pressure detection based devices such as Venturi tube and orifice plate, which perform well in single-phase flow metering, have been tested in multiphase flow measurement [3]. Such approaches, however, have failed to address accurately detecting multiphase flow, especially in high gas volume fraction (GVF) scenario. The failure of detecting high GVF multiphase flow by using Venturi tube is mainly due to the detected differential pressure crossing Venturi throat is sensitive to the instantaneous change of liquid amount. In detail, a small increasing of liquid amount usually leads to a large increasing of the measured differential pressure across Venturi tube and so causes gas flowrate overreading [4].

Recent evidence suggests that if the multiphase flowrate detection could be performed in scenario of assuming that there is no interaction between each flow phase, then, the experiment results on detecting GVF over 95% are acceptable [5]. However, in practical application, multiphase flow usually contains complex phase components and above assumptions cannot be satisfied.

To date, machine learning has been applied to explore new ways of predicting multiphase flowrate. For instance, Convolutional Neural Networks (CNN) and Flow Adversarial Networks (FAN) on predicting gas-liquid multiphase flows have been performed and the behaviour of the two models were evaluated by Hu [6]. The experimental results reveal that FAN has better performance on multiphase flowrate prediction than CNN. Meanwhile, Artificial Neural Network (ANN) was used to estimate flowrate of air-water two-phase flow with the measurement error less than 10% [7]. Independent component and principle component analysis were employed to reduce the dimensionality of the features and to improve the efficiency of ANN [8]. However, the ANN model needs to be specifically tuned for different flow regimes, which is inconvenient in practical applications. In addition, Support Vector Machine (SVM) was also applied in multiphase flow pattern recognition [9], indicating the possibility to train a more adaptive machine to fit the comprehensive flow regimes and predict instantaneous flowrate.

In this paper, the feasibility of estimating the flowrate of gas/oil/water three-phase flow based on Venturi tube and machine learning methods was investigated. Instantaneous data, such as the standard pressure, former and posterior differential pressure and temperature are collected from Venturi tube as the input of the machine learning models. Three prevailing machine learning methods, i.e. Deep Neural Network (DNN), SVM and Gradient Boosting Decision Tree (GBDT), are implemented and the performance was evaluated by analysing estimation error.

II. Methodology

A. Venturi tube for multiphase flowrate measurement

Venturi tube has been implemented for collecting multiphase flow data as input for DNN, SVM and GBDT models. The implemented Venturi tube with upstream diameter D and throat diameter d has been demonstrated in Fig. 1. It is able to provides the instantaneous standard pressure (P), former

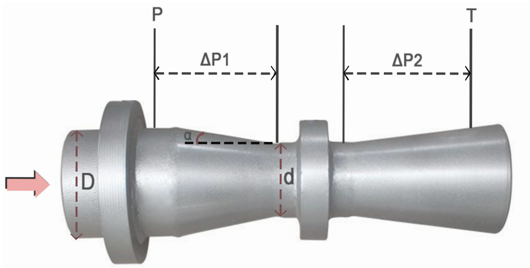


Fig. 1: Front view of conventional Venturi tube with upstream diameter D, throat diameter d and detecting parameters including standard pressure (P), former differential pressure (ΔP_1), posterior differential pressure (ΔP_2) and temperature (T).

and posterior differential pressure (ΔP_1 and ΔP_2) data as input to estimate the instantaneous gas and liquid flowrate of multiphase flow.

B. Deep neural network design and training

Instantaneous values of P, ΔP_1, ΔP_2 and T are selected as four input layers for two DNN models to estimate the instantaneous liquid and gas flowrate for water-oil-gas multiphase flow. The structure of the two DNN models are exactly the same, which contains 50 hidden layers, as shown in Fig. 2.

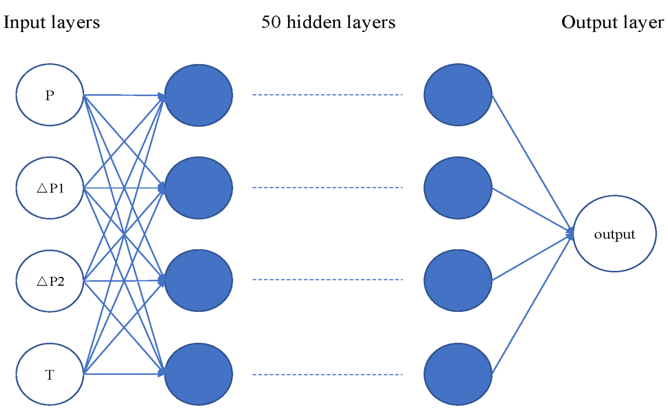


Fig. 2: Deep Neural Network (DNN) structure with four input parameters (P, ΔP_1, ΔP_2 and T), fifty hidden layers and one output layer for multiphase flowrate estimation.

In detail, liquid and gas instantaneous flowrate are estimated by two defined DNN models, f_l and f_g, respectively. It can be express as:

$$\begin{aligned} \tilde{r_l} &= f_l(P, \Delta P_1, \Delta P_2, T; \theta_l) \\ \tilde{r_g} &= f_g(P, \Delta P_1, \Delta P_2, T; \theta_g) \end{aligned} \tag{1}$$

where $\tilde{r_l}$ and $\tilde{r_g}$ are the estimated liquid and gas flowrate, respectively. θ_l and θ_g are the optimal parameters for liquid and gas estimation models.

Single phase flow meters are implemented to collect the liquid and gas flow data in upstream section of Venturi tube as reference for each model. The loss function for each model is defined as:

$$\begin{aligned} \mathcal{L}_l &= E[(r_l - \tilde{r_l})^2] \\ \mathcal{L}_g &= E[(r_g - \tilde{r_g})^2] \end{aligned} \tag{2}$$

where E is the expectation function.

For training process, the goal is to minimize the loss functions by obtaining optimal values of θ_l and θ_g. It can be achieved through back propagation which is detailed in (3):

$$\begin{aligned} \theta_l &= \underbrace{arg\ min}_{\theta_l} \mathcal{L}_l(\sim; \theta_l) \\ \theta_g &= \underbrace{arg\ min}_{\theta_g} \mathcal{L}_g(\sim; \theta_g) \end{aligned} \tag{3}$$

The exponential linear unit (ELU) function was selected as the activation function for each neuron except the output layer, which can help to accelerate training and prevent over-fitting to some degree [10]. The inputs will pass the DNN model with 50 hidden layers, and finally the predicted instantaneous liquid or gas phase flowrate can be obtained at output layer. During the training process, Levenberg-Marquardt method [11], which belongs to the "mountain climbing" method and uses gradient to find the maximum (minimum) value, was implied to obtain the best solution for flowrate estimation problem.

C. SVM

SVM as a generalized linear classifier working on binary classification with supervised learning has been proved that it is able to detect and recognise liquid-gas two phase flow pattern [9]. However, linear regression or binary classification model is not suitable for instantaneously detecting multiphase flow problem, which is a more sophisticated problem and cannot be regarded as a linear problem. Therefore, an improved SVM method with Gaussian kernel [12] is implemented in this paper. The decision function combined with Gaussian kernel can be formulated as shown in (4):

$$G_{SVM}(x) = sign(\sum_{SV} \alpha_n y_n K(x_n, x) + b) \tag{4}$$

where $G(x)$ is the decision function of SVM model, $sign(.)$ is signum function, SV is the set of support vector, α is a weight factor, y_n is the volumetric flowrate of each single phase flow, b is bias and $K(.)$ is the general form of Gaussian kernel with SVM parameter γ which can be expressed as:

$$K(x, x') = exp(-\gamma \left\| x - x' \right\|)^2 \tag{5}$$

where $K(.)$ is the kernel function and x is the estimation data set.

Since SVM performs a significant role in machine learning and data processing field, hence the evaluation of SVM to explore how machine learning works in flowrate estimation is necessary.

D. GBDT

GBDT, also known as Multiple Additive Regression Tree (MART) [13], is an iterative decision tree algorithm, which consists of multiple decision trees, is also chosen to be evaluated in this paper. To obtain the flowrate, conclusions of all trees are added together. GBDT is considered to perform the multiphase flowrate estimation mainly because that it performs well in dealing with both linear and non-linear regression problem [14]. The details of GBDT algorithm is shown in Algorithm. 1.

Algorithm 1 Gradient Boosting Decision Tree (GBDT) algorithm in multiphase flowrate detection.

Input:

The input data set, x;
The output data set, y;
The combined data set of x and y, $T = (x_1, y_1), (x_2, y_2)...(x_N, y_N)$;
The loss function, $L(y, f(x))$;

Output:

Ensemble of classifiers on the current batch of the regression tree, $F(x)$;

1: Initializing $f_0(x) = argmin \sum_{i=1}^{N} L(y_i, c)$;
2: Setting iteration number M for samples $i = 1, 2, 3...N$. The value of the negative gradient of the loss function in the current model is calculated and used as the residual estimate;
3: Calculate the approximate value $r_{mi} = -[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}]f(x) = fm-1(x), m = 1, 2...M$;
4: Basing on $(x_1, r_{m1}), (x_2, r_{m2}...(x_N, r_{mN})$, fitting a regression tree with leaf node region $R_{mj}, j = 1, 2...J$, where J is leaf node number of each tree;
5: For $j = 1, 2...J$, minimizing the loss function by estimating the value of leaf node region by linear search and calculating $c_{mj} = \underbrace{argmin}_{c} \sum_{x_i \in R_{mj}} L(y_i, f_{m-1}(x_i + c))$;
6: Updating $f_m(x) = f_{m-1}(x) + \sum_{J}^{j=1} c_{mj} I(x \in R_{mj})$;
7: Calculating $F(x) = \sum_{m=1}^{M} \sum_{j=1}^{J} c_{mj} I(x \in R_{mj})$;
8: **return** $F(x)$;

III. Results and Discussion

In this Section, experimental setup and estimate results generated by machine learning algorithms are presented in detail. The performance of the three methods on predicting multiphase flowrate were comprehensively evaluated by analysing the estimation error and deviation. The flowrate prediction accuracy of DNN, SVM and GBDT are compared as well.

A. Experimental setup

A multiphase flow experimental facility at Tsinghua University is utilized which enables wide-range combination of water, oil or gas single phase flow and separation of three phase flow. The volumetric flowrate of water, oil and gas single-phase flow are measured as reference to compare with the DNN,SVM and GBDT prediction results. Then, three single-phase flow is mixed as multiphase flow. The pressure and temperature parameters as training data are measured and recorded before the multiphase flow passing through into Venturi for DNN model. The experimental test matrix is demonstrated in Table. I.

TABLE I: Multiphase flow experimental test matrix

Liquid flowrate (m^3/h)	Water-Liquid ratio	Gas flowrate (m^3/h)	Gas volume fraction
1	0–100%	0–20	0–95%
2	0–100%	0–40	0–95%
3	0–100%	0–60	0–95%
4	0–100%	0–30	0–90%
6	0–100%	0–60	0–90%

To simulate the real industrial flow condition, various flow conditions are generated by the flow experimental facility to provide reliable samples assemble to industrial environment. Instantaneous P, ΔP_1, ΔP_2 values are provided from Venturi tube and T is measured by using temperature sensor of multiphase flow. Each parameter contains 40,000 data as input for DNN, SVM and GBDT models. Every 100 sets are packed as a batch, which means that there are 400 batches. Among the data sets, 75% random sample sets are used to train the models, 15% are used for validation and the remaining 15% are used to test the trained models.

B. DNN results without temperature T

Initially, only pressure parameters (without temperature) are considered as input for DNN model to estimate liquid and gas flowrate. The liquid and gas flowrate estimation results are given in Fig. 3.

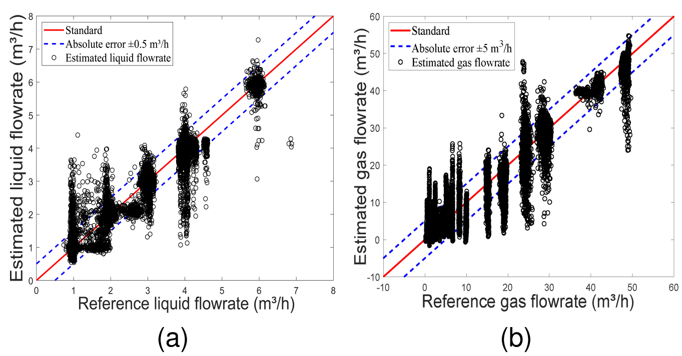


Fig. 3: Estimation results of (a) Liquid and (b) Gas flowrate without temperature of DNN.

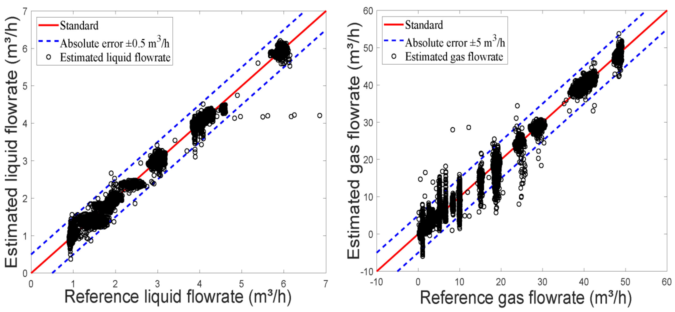
The estimation results in Fig. 3 appeared to be not satisfactory. In both regression results, the estimated outputs by DNN model and the references, which were collected by single phase flow meters in single phase flow pipes, cannot relate as a linear relationship, which means the estimated flowrate are far from the reference. The non-linearity in regression may be due to the rapid change of flow regimes in Venturi tube and Venturi tube is not sensitive enough to instantaneously measure the pressure.

Though by only using instantaneous measured pressure data, DNN cannot accurately predict multiphase flowrate, the results in Fig. 3 still indicate that there exists a relationship between the pressure parameters and flowrate. Therefore, in the following subsections, temperature as another parameter will be considered as an additional input to improve the estimation results. Meanwhile, the collected raw data are re-calculated by using moving average method [15] to generate a more smooth data set as input.

C. DNN results with temperature T

In order to mitigate the influence caused by the sudden change of flow regimes and to obtain more input data sets, moving average method are applied to P, ΔP_1, ΔP_2 and T data. The length of the 'window' is four, which means that the average is taken from first four samples, then, another average is taken from the second to fifth samples and so on.

The estimation results of trained DNN model by implementing the averaged data considering temperature are shown in Fig. 4.

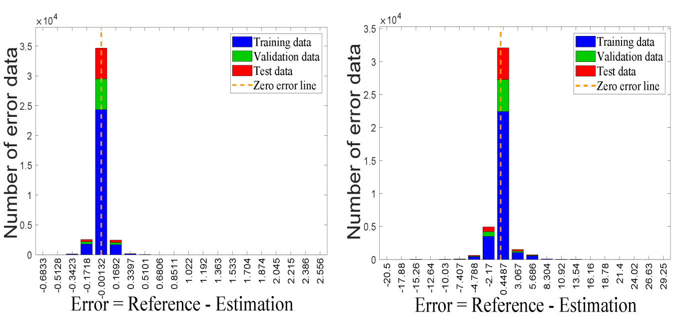


(a) (b)

Fig. 4: Estimation results of (a) Liquid and (b) Gas flowrate by moving time averaged data and including temperature of DNN.

Comparing the results in Fig. 4 and Fig. 3, it is obvious that by considering temperature factor and taking moving average of samples, the estimation results of liquid flowrate are more accurate. Likewise, the estimation results of gas flowrate are closer to the best fit line in Fig. 4 compared with the results in Fig. 3. Additionally, in Fig. 4, the liquid flowrate estimation results are more accurate than gas flowrate prediction.

Error histograms are calculated and plotted to evaluate the performance of DNN model, which are shown in Fig. 5.



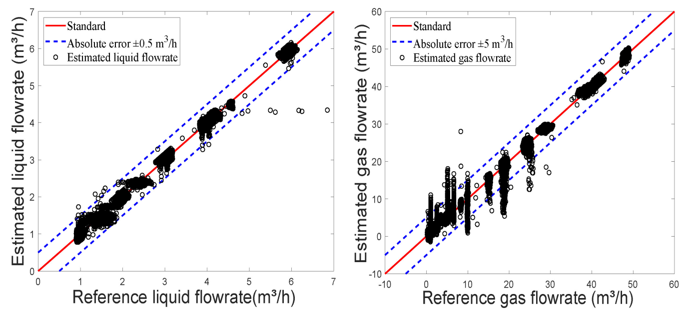
(a) (b)

Fig. 5: Error histograms of (a) Liquid and (b) Gas flowrate estimation using DNN.

The estimation errors for liquid phase are controlled within 6%, and 75% errors are within 0.07% range. Meanwhile, for gas phase estimation, about 80% estimated points range in 0∼1.5%, and almost all errors are below 8%. Small error range verifies that DNN model is able to accurately estimate liquid and gas flowrate with a proper input data set.

D. SVM results

The input data sets for SVM model are exactly the same as the data sets in Section III-C. It is supposed that SVM could also make successful prediction on multiphase flowrate estimation. Fig. 6 illustrates SVM model estimation results for liquid and gas flowrate.



(a) (b)

Fig. 6: Estimation results of (a) Liquid and (b) Gas flowrate by moving time averaged data and including temperature of SVM.

In liquid flowrate estimation process, 1273 support vectors are used to perform the estimation. The liquid flowrate prediction result reveals that the linearity between input sample sets and reference sets is acceptable. Although there are still some abnormal points deviating from the standard line, the correlation is strong.

In gas flowrate estimation process, 5367 support vectors are used in the training procedures. The regression result is in good linearity with some extreme points deviate away from the standard line, which is caused by the changeable gas rate in Venturi tube.

To evaluate the performance of SVM model, deviations are plotted in Fig. 7. Compared with DNN model, SVM model generates more estimated errors in amount in liquid and gas flowrate estimation. Especially for the gas phase, some estimated results have deviation more than 10 m^3/h and a part of results have deviation over 5 m^3/h. A possible explanation might be that the performance of SVM model is restricted by the algorithm principle.

Overall, SVM model is able to perform the estimation of multiphase flowrate whereas DNN model has better estimation results in terms of stronger correlation and smaller error range.

E. GBDT results

GBDT is selected as another method to solve flowrate estimation objective because of its capability in regression and

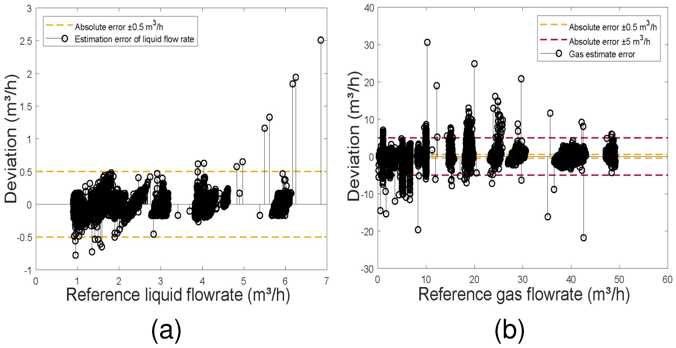


Fig. 7: Estimation error of (a) Liquid and (b) gas flowrate prediction results by using SVM model.

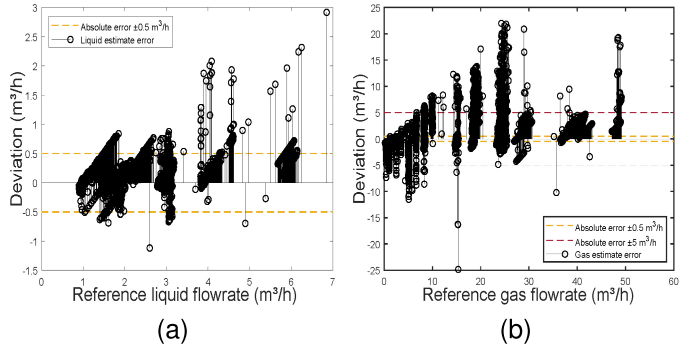


Fig. 9: Estimation error of (a) Liquid and (b) gas flowrate prediction results by using GBDT model.

classification. The performance of GBDT mainly depends on the selection of loss function and in this paper, regression tree is implemented to perform the estimation. The liquid and gas flowrate estimation results by using GBDT are shown in Fig. 8.

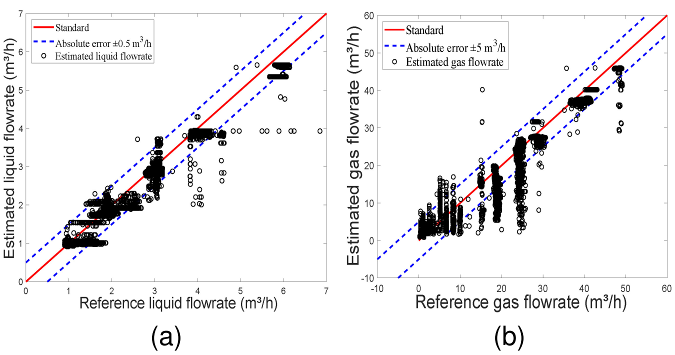


Fig. 8: Estimation results of (a) Liquid and (b) Gas flowrate by moving time averaged data and including temperature of GBDT.

Contrary to expectations, the fitness for both liquid and gas flowrate estimation is poor. This result is unexpected which suggests that GBDT is not a suitable method for estimating multiphase flowrate. It may because that the parameters measured by Venturi tube may not fit the GBDT model and the correlation between the input parameters and flowrate cannot be properly described by GBDT. It is further suggested that more sensor parameters may be required to achieve a more accurate estimation.

Deviation figures reflect the unreliable performance of GBDT in a more straightforward way, which is demonstrated in Fig. 9. Only a few estimates are within the acceptable range and larger deviation occurs, which matches the estimation performance in Fig. 8. It indicates that the multiphase flowrate predicted by GBDT model and existing data sets are not accurate and not reliable.

F. Machine learning models performance comparison and discussion

The estimation results by implementing DNN, SVM and GBDT models are evaluated by utilizing Mean Squared Error (MSE). The calculated MSE values are illustrated in Table. II.

TABLE II: MSE EVALUATION ON MULTIPHASE FLOWRATE PREDICTION FOR DNN, SVM AND GBDT MODELS

Data type	Input parameters	Model	MSE of flowrate estimation	
			Liquid	Gas
Raw data	$P, \Delta P_1, \Delta P_2$	DNN	0.0652	8.603
Moving averaged data	$P, \Delta P_1, \Delta P_2, T$	DNN	0.0049	2.0761
Moving averaged data	$P, \Delta P_1, \Delta P_2, T$	SVM	0.0169	2.3532
Moving averaged data	$P, \Delta P_1, \Delta P_2, T$	GBDT	0.0401	9.4521

Table. II is informative in several ways. First, it is apparent that averaged sample sets have more adaptivity for each flow condition than instant sample sets. For DNN model, it is found that a longer training time is required when taking the moving averaged sample sets as input in practice. However, MSE for DNN model suggests that better performance can be expected when we use moving averaged samples sets. Additionally, temperature as an important parameter helps to correct the model weights and bias. From the calculated MSE value for GBDT model, the highest MSE for gas flowrate estimation indicates that GBDT fails to tackle the multiphase flowrate prediction problem. Compared with the results in Section III-C and III-D, DNN has better performance with the same input data. Overall, both DNN and fine Gaussian SVM can achieve satisfactory results in gas flowrate estimation.

IV. CONCLUSION

This paper studied the estimation of instantaneous flowrate of gas/oil/water three-phase flow by combining Venturi tube with various machine learning methods, i.e. DNN, SVM and GBDT. The investigation of the input data sets has shown that the performance of DNN model on multiphase flowrate estimation can be improved by using modified data, such as moving averaged data rather than raw data, and by including more parameters, such as pressure and temperature as input. The experimental results demonstrate that DNN is effective for estimating the flowrate of both liquid and gas phase. It also obtained superior performance in estimating instantaneous flowrate whereas many previous studies only reported the estimation of time-averaged flowrate. This study also shows that GBDT is not a suitable approach for multiphase flowrate estimation in cooperation with Venturi tube. The results also suggest that SVM is effective for gas flowrate estimation. For liquid flowrate estimation, DNN has the best performance. In addition, the study has also shown that temperature is an important factor in multiphase flowrate estimation.

Future study will assess the effect of structure, activation function, and more pre-process operations on the performance of DNN model for multiphase flowrate estimation.

References

[1] D. H. Rothman and S. Zaleski, "Lattice-gas models of phase separation: interfaces, phase transitions, and multiphase flow," *Reviews of Modern Physics*, vol. 66, no. 4, p. 1417, 1994.

[2] V. Rajan, R. Ridley, and K. Rafa, "Multiphase flow measurement techniques–a review," *Journal of Energy Resources Technology;(United States)*, vol. 115, no. 3, 1993.

[3] P. Lin and T. Hanratty, "Detection of slug flow from pressure measurements," *International Journal of Multiphase Flow*, vol. 13, no. 1, pp. 13–21, 1987.

[4] Y. Xu, Q. Zhang, T. Zhang, and X. Ba, "An overreading model for nonstandard venturi meters based on h correction factor," *Measurement*, vol. 61, pp. 100–106, 2015.

[5] I. ISO, "Tr 11583-measurement of wet gas flow by means of pressure differential devices inserted in circular cross-section conduits," 2012.

[6] D. Hu, J. Li, Y. Liu, and Y. Li, "Flow adversarial networks: Flowrate prediction for gas-liquid multiphase flows across different domains," *IEEE transactions on neural networks and learning systems*, 2019.

[7] S. Cai and H. Toral, "Flow rate measurement in air-water horizontal pipeline by neural networks," in *Proceedings of 1993 International Conference on Neural Networks (IJCNN-93-Nagoya, Japan)*, vol. 2. IEEE, 1993, pp. 2013–2016.

[8] H. Shaban and S. Tavoularis, "Measurement of gas and liquid flow rates in two-phase pipe flows by the application of machine learning techniques to differential pressure signals," *International Journal of Multiphase Flow*, vol. 67, pp. 106–117, 2014.

[9] H. Wang and L. Zhang, "Identification of two-phase flow regimes based on support vector machine and electrical capacitance tomography," *Measurement Science and Technology*, vol. 20, no. 11, p. 114007, 2009.

[10] S. Qian, H. Liu, C. Liu, S. Wu, and H. San Wong, "Adaptive activation functions in convolutional neural networks," *Neurocomputing*, vol. 272, pp. 204–212, 2018.

[11] J. J. Moré, "The levenberg-marquardt algorithm: implementation and theory," in *Numerical analysis*. Springer, 1978, pp. 105–116.

[12] S. S. Keerthi and C.-J. Lin, "Asymptotic behaviors of support vector machines with gaussian kernel," *Neural computation*, vol. 15, no. 7, pp. 1667–1689, 2003.

[13] B. P. Roe, H.-J. Yang, J. Zhu, Y. Liu, I. Stancu, and G. McGregor, "Boosted decision trees as an alternative to artificial neural networks for particle identification," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 543, no. 2-3, pp. 577–584, 2005.

[14] S. A. Naghibi, H. R. Pourghasemi, and B. Dixon, "Gis-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in iran," *Environmental monitoring and assessment*, vol. 188, no. 1, p. 44, 2016.

[15] H. Akaike, "Maximum likelihood identification of gaussian autoregressive moving average models," *Biometrika*, vol. 60, no. 2, pp. 255–265, 1973.