LSTM and 1-D Convolutional Neural Networks for Predictive Monitoring of the Anaerobic Digestion Process

Mark McCormick^(✉) and Alessandro E. P. Villa

NeuroHeuristic Research Group, University of Lausanne, Quartier UNIL-Dorigny,
1015 Lausanne, Switzerland
mark.mccormick@unil.ch,
<http://www.neuroheuristic.org>

Abstract. Anaerobic digestion is a natural process that transforms organic substrates to methane and other products. Under controlled conditions the process has been widely applied to manage organic wastes. Improvements in process control are expected to lead to improvements in the technical and economic efficiency of the process. This paper presents and compares 3 different neural network model architectures for use as anaerobic digestion process predictive models. The models predict the future biogas production trend from measured physical and chemical parameters. The first model features an LSTM layer, the second model features a 1-D convolutional layer and the third model combines 2 separate inputs and parallel treatment using LSTM and 1-D convolutional layers followed by merging to produce a single prediction. The predictions can be used to adaptively adjust the substrate feeding rate in accordance with the transient state of the digestion process as defined by liquid feeding rate, the organic acid and ammonium ion concentrations and the pH of the digester liquid phase. The training and testing data were obtained during 1 year of continuous operation of a pilot-plant treating restaurant wastes. PLS regression and ICA were used to select the most relevant process parameters from the data. The 1-D Convolutional based model comprising 272 trainable parameters predicted the future biogas flow rate changes with accuracy as high as 89% and an average accuracy of 58%. The work-flow can be applied to optimize the control of the study digester and to control bioreactors in general.

Keywords: LSTM, 1-D Convolutional Neural Network · Bioprocess

1 Introduction

Restaurants, hotels, markets, fisheries and other small to medium size agro-food industries generate 88 million tonnes of organic waste per year in Europe of which 47 million tonnes per year are household food wastes and 17 million tonnes per year are food processing wastes [1]. This waste stream contains valuable components such as water (approximately 80% by mass) and valuable energy and

substances such as proteins and organic acids. In the context of the circular economy and current objectives to improve energy efficiency, recycle water and valuable substances, produce energy from renewable sources, reduce greenhouse gas emissions and close nutrient cycles, the waste management sector offers many opportunities to implement innovative technical solutions. Anaerobic digestion has been proposed as a technical solution to manage organic wastes. This study aims to contribute to the strategy used to control a high-performance anaerobic digester supplied by Digesto Sarl. The strength of the Digesto® digestion module is the ability to treat small to medium quantities of waste on their site of production, without any transport and in an automated mode (programmable logic controller for distance monitoring of the digester performances and for breakdown prevention). The elimination of transport costs and a high degree of automation make it the appropriate treatment solution for many end-users [2].

Currently, anaerobic digestion is successfully applied to reduce the organic matter content and to improve the dewatering properties of excess wastewater treatment sludge. Anaerobic digestion is also used to produce energy from municipal solid wastes, garden wastes and from energy crops. These processes are characterized by large bioreactors, long hydraulic retention times between 20 and 60 days and by regular feeding of well characterized and homogeneous substrates. In contrast, small-scale, on-site digestion of restaurant wastes would require an intensified bioprocess featuring a small digester, a short hydraulic retention time of less than 10 days and the capacity to tolerate wide variations in the feeding rate and the substrate composition.

Anaerobic digestion is a complex process characterized by non-linear functions and interactions between many different biochemical processes. A high-performance control strategy is thus required to intensify the anaerobic digestion process and to achieve the goals of small size, a short hydraulic retention time and long-duration autonomous operation. Artificial Neural Networks, and especially Deep Neural Networks, have demonstrated their capacity to map complex and non-linear relations between data. The purpose of this work is to construct models based on some popular Neural Network architectures and to assess their capacity to predict biogas production.

2 Materials and Methods

2.1 Machine Learning

Multivariate statistical analysis approaches have been used to detect faults and abnormal operation of anaerobic digestors treating waste activated sludge (WAS). Using Principal Component Analysis (PCA), Hotelling's T-squared and Shewhart control charts, the transitions to unstable periods were associated with accumulation of volatile fatty acids [3]. Computational self-adapting methods (Support Vector Machines, SVM) were compared with an analytical method to predict the total ammonia nitrogen (TAN) concentration in the effluent from a

two-stage anaerobic digestion (AD) process treating poultry wastes. The SVM-based model outperformed the analytical method for the TAN prediction, achieving a relative average error of 15.2% against 43% for the analytical method. Moreover, SVM showed higher prediction accuracy in comparison with Artificial Neural Networks [4]. Machine learning methods have been used to estimate the state of biogas plants, as defined by the ADM-1 model of the anaerobic digestion process, using on-line measurements of parameters such as biogas production, CH_4 and CO_2 content in the biogas, pH value and substrate feed volume [5].

In contrast to the conventional approach of using recurrent networks, especially LSTM layers, for sequence processing, convolutional architectures have demonstrated longer effective memory and have outperformed recurrent networks across a diverse range of tasks [6]. In the field of aircraft control, a “state-image” approach to capture the inflight state variables produced a feature map that contained the values of the in-flight parameters at a given moment. The control strategy also included historical data. A Convolutional Neural Network (CNN) and a Recursive Neural Network with Long Short-Term Memory (RNN (LSTM)) layers were implemented in parallel to output vectors that were subsequently merged and processed in fully connected layers. Inputting data to different branches to allow separate extraction of the time dependent and the current parameter values was expected to lead to improved accuracy when compared to sequential treatment by CNN and LSTM layers [7].

2.2 Experimental Set-Up

The study digester was a machine comprising tanks, pumps, heating elements, sensors and command and data acquisition capacities. The total liquid working volume was 630 liters. Shredded restaurant wastes were fed to the digester at regular intervals. The study data was acquired from on-line measurements and by analysis of samples collected at intervals between 1 and 7 days during 366 days of continuous operation. The study data included 22 parameters that characterize the input waste stream. Characterization of the digestion process included manual measurement of 44 parameters that describe the liquid and gas phases. The on-line measurements included redundant measurements of liquid and gas phase physical-chemical properties and the measurement of 7 mechanical parameters such as internal pumping and agitation. A total of 83 parameters were considered for the study. The digestion process was controlled manually using only the expert knowledge of the operators. The operating conditions were adjusted *ad hoc* during the study to optimize the rates of biogas production and the removal of the organic fraction of the waste. In this study, fixed duration feeding intervals were defined. At the start of each time interval a decision was made to feed or to not feed the substrate to the digestion machine.

2.3 Problem Solving Approach

Autonomous operation of the digester would require automated transfer of raw waste from a buffer tank to the digestion machine. The problem is to decide when

the raw waste should be transferred (fed) to the digestion machine. Overfeeding would upset the bioprocess and can lead to ceased waste degradation and biogas production. Underfeeding would lead to inefficient use of the digester. A useful model should make predictions that agree with the actual values that were measured after the prediction. The digester should be fed only when the biological state is conducive to waste degradation and biogas production. Therefore, the purpose of the neural network model is to automate decision making. To solve the problem, a supervised approach based on mapping measured process predictor variables to a single measured response variable was chosen. The models were trained and tested using simultaneously measured prediction and response values. The models were tested by comparing predicted to measured values of the response variable.

The proposed anaerobic digester control strategy is to compare the predicted future biogas flow rate trend to the actual biogas flow rate TREND. Increased biogas production rate is associated with improved bioprocess quality. Decreased or stable biogas production rate is associated with deteriorating bioprocess quality. If the predicted future biogas flow rate is higher than the actual measured biogas flow rate, then the decision is to feed the digestion machine. If the predicted biogas flow rate was the same as or lower than the actual measured biogas flow rate, then the decision was to not feed the digestion machine.

2.4 Data Analysis and Preprocessing

Analysis and data augmentation techniques were used to prepare a single dataset for use in developing all the Neural Network models. Raw data analysis aimed to identify the parameters to use to train and test the Neural Network model. Principal Component Analysis (PCA), Independent Component Analysis (FastICA) and Partial Least Squares (PLS) regression algorithms from Scikit-learn [8] were used for dimensionality reduction and to identify the parameters that make the greatest contribution to process stability. A plot of the weights of the first principal components shows that the waste loading rate, total solids, carbon and nitrogen content of the feed, and the total solids concentration in the of the digester make the greatest contributions to overall variance of the system. However, these parameters can be measured only manually and manipulation by the operators is difficult or impossible. Consequently, they are not practically useful in a neural network model for process control.

ICA is recommended when the data has a high degree of kurtosis. Since sharp peaks were observed for biogas H_2 content, and liquid phase ammonium and volatile fatty acid content, the raw data were analyzed using ICA. ICA analysis showed that the H_2 content of the biogas and the volatile fatty acid content of the digester were the most important independent components of the data. The raw data was used in the DNN models without pre-treatment to remove kurtosis because spiking of both H_2 and volatile fatty acids is not unusual in the anaerobic digestion process. The H_2 concentration in the biogas can be measured on-line. However, no sensors for volatile fatty acids are currently available at an affordable price. The method of Partial Least Squares Regression was used to

map the matrix of predictor variables to the biogas production rate. The results of PLS analysis showed that regression equation coefficients for the nitrogen content of the feed and the digester loading rate had the greatest magnitudes.

Considering the results of PCA, ICA and PLS regression analysis, knowledge of the anaerobic digestion process and the practical aspects of data acquisition, the following parameters were selected for inclusion in the dataset used to build the Neural Network models with 4 input features and 1 target variable.

- *Predictor variables (input features)*
 - Total mass of feed (water + waste + co-substrate) [kg/day]
 - pH
 - Ammonium concentration [mg N/liter]
 - Volatile fatty acid concentration [mg Acetic Acid/liter]
- *Response variable (target)*
 - Biogas flow rate [liter/hour]

Pre-processing included augmentation of the number of data points. In particular, the frequency of values obtained from off-line measurements of physical-chemical parameters was increased from approximately 1 per week to 1 per hour. The number of data points was increased by resampling to create one-hour time intervals followed by linear interpolation of the off-line data to fill the new sampling times. Interpolation was assumed to be valid because the changes in biological systems occur slowly over several days. The augmented dataset included 8733 rows of data representing the selected input features and the target. Using the Scikit-learn MinMaxScaler function, the previously standardized data were then normalized to values between 0 and 1. Figure 1 shows the pre-processed dataset that was used to build the Neural Network models. Five series of training and test data were obtained using the Scikit-learn TimeSeriesSplit function which respects the sequential order of the original data.

2.5 Neural Network Architectures

The study aimed to compare the recurrent, convolutional and combined neural network modeling approaches to solving a time series analysis problem having at 4 features. The models were built using Python 3.6.8 and the Keras 2.2.4 API running on top of TensorFlow 1.3.0 library for machine learning. The 3 architectures selected for development and the rationale for selection are summarized in Table 1 and the models are summarized in Fig. 2.

Long Short-Term Memory (LSTM). The sequential model was constructed using a single LSTM layer followed by a Dense layer. The input was a 3D tensor with shape (`batch_size`, `timesteps`, `input_dim`) where `batch_size` is the number of sample batches, `timestep` was set to four hours and `input_dim` was equal to the number of features. The LSTM layer had 160 units, used the `tanh` activation and `hard_sigmoid` recurrent activation functions with zero dropout. `Unit_forget_bias` was set to false since this setting was found to give a more

Fig. 1. Training and testing data after preprocessing. **(a)** Normalized training data of predictor variables. blue line: total volume of feed; red line: pH; green curve: ammonium concentration; orange line: volatile fatty acid concentration. **(b)** Normalized training data of the response variable, i.e. biogas flow rate. **(c)** Normalized test data of predictor variables. Same color code of panel (a). **(d)** Normalized test data of the response variable. (Color figure online)

accurate prediction. The model output was from a single dense layer having 1 unit, `relu` activation and `use_bias` set to False. The model had 105,760 trainable parameters.

1-Dimensional Convolutional Layer (Conv1-D). The sequential model was constructed using a single Conv1D layer followed by a MaxPooling1D layer followed by a Dense layer. The input was a 3D tensor with shape (`batch_size`, `timesteps`, `input_dim`), where `batch_size` is the number of sample batches, `timesteps` was set to four hours and `input_dim` was equal to the number of features. The Conv1D layer had 16 filters, a kernel size of 4, stride of 1, padding

Table 1. Evaluated Neural Network architectures

Neural network architecture	Rationale for selection
LSTM, biogas flow rate predictions	Expert knowledge that bioprocesses have long and variable lag-times between measured predictor values and responses. Historical information
1-D CNN, biogas flow rate predictions	Potential to extract useful information from detailed analysis of features at given moments. State image concept
LSTM/1-D CNN hybrid, biogas flow rate predictions	Potential to improve accuracy by separate processing of historical and state image information followed by merging the information

set to same, and used the `relu` activation function. The pool size of the Max-Pooling layer was set to 4. The model output was from a single dense layer having 1 unit, `relu` activation and and `use_bias` set to False. The model had 272 trainable parameters.

LSTM/Conv1-D Hybrid. The Keras Model Class API was used to construct a model having 2 separate input and processing branches and a single output. The first branch included an LSTM layer followed 2 dense layers. The LSTM layer had the same configuration as the layer described above. The first dense layer had 48 nodes and used the `relu` activation function. The final dense layer had 1 node and used the `relu` activation function. The second branch included a Conv1D layer followed by MaxPooling1D Dense and Flatten layers. The Conv1D layer had the same configuration as the layer described above. The MaxPooling1D layer strides were set to 1 and padding was set to same. The two branches were merged using the Keras concatenate layer. The merged output was further passed to a dense layer with 8 hidden nodes. Model output was from a dense layer having 1 node and using the `relu` activation. The model had 106,158 trainable parameters.

3 Results

The augmented dataset was split into 5 different pairs of training and test sets using the TimeSeriesSplit function. Each pair comprised separate sets of consecutive 1-h intervals and a total of between 5000 and 8500 timestamps. The LSTM, Conv1D and hybrid LSTM/Conv1D models were trained for 1000 epochs with a batch size of 4 h. Loss during training was assessed using mean square error. The loss during training of the 3 models is shown in Fig. 3.

The models were tested on data obtained previously using the TimeSeriesSplit function. The prediction accuracy was assessed visually by plotting the predicted values and the measured values at 4-h intervals corresponding to the

Fig. 2. Artificial Neural Network architectures: (a) LSTM, (b) Conv1D and (c) Hybrid LSTM/Conv1D architectures

Fig. 3. Comparison of training loss for the 3 ANN architectures.

batch-size of 4 times 1-h timestamps. The RMSE for the individual sample pairs was calculated at the same 4-h interval (batch).

The predicted and the observed biogas flow rates for the 3 different regression models are shown in Fig. 4abc. The RMSE of each batch is shown in Fig. 4def. To assess the accuracy of the 3 models over a long duration, the overall RMSE was also calculated for the entire population of batches.

Table 2 shows the overall RMSE of the 3 models.

Since the purpose of the model is to decide if the digester should be fed at the beginning of a time interval, a special evaluation procedure was developed to rate the models in terms of their capacity to make the correct decision regarding digester feeding. To rate the models, the consequences of the model's decision were compared to the actual change in measured biogas flow rate 4 h after the decision. Any decision to feed the digester, followed by an observed

Fig. 4. Predicted (dashed red curve) and observed (solid blue curve) biogas production for LSTM (panel a), 1-D CNN (panel b), and LSTM/1-D CNN hybrid (panel c). The corresponding errors (RMSE) are presented in panels d,e,f for LSTM, 1-D CNN and LSTM/1-D CNN hybrid ANN architectures, respectively. (Color figure online)

Table 2. RMSE of the entire population

Model (ANN architecture)	RMSE (%)
LSTM	1.7
1-D CNN	2.6
LSTM/1-D CNN hybrid	19.0

decrease in biogas production, was an incorrect decision. Any decision to feed the digester, followed by an observed stable or increased biogas production, was a correct decision. The model accuracy was simply the ratio of correct decisions to requested decisions expressed as a percentage. To validate the 1-D convolutional DNN model, 12 new data sets generated by splitting the original data were used to predict the biogas production trend 4-h in the future. The average accuracy of the prediction was 58% with a standard deviation of 24% and a range between 28% and 100%.

To investigate the response time between changes in the measured predictor and response variables, the error of the LSTM model predictions was evaluated at different lag-times after the feeding decision. In this case, error was defined as the RMSE of the measurements and predictions made at 1-h intervals during the time sequence. Population RMSE was calculated for lag-times from 1 to 100 h after a feeding decision. The results show that the model is most accurate when the lag-time between the feeding decision and the time off comparison of the predicted to the observed values is much greater than 4 h. This result suggests that the biological response time of biogas production to changes in the 4 measured process parameters used this study is longer than 4 h (Fig. 5) .

Fig. 5. RMSE of the predicted and observed biogas production at different times after measurement of predictor variables (LSTM model).

4 Discussion

The treatment of restaurant wastes is a particularly challenging application of anaerobic digestion because the loading rate and composition of the wastes vary over a relatively large range when compared to the treatment conditions for anaerobic digestion of waste activated sludge, manure and energy crops. Consequently, the feed controller must have the capacity to map a wide range of values of the relevant process parameters and account for long-duration trends. The results of this work show that a one-dimensional convolutional Deep Neural Network model trained using state image data obtained by measuring 4 parameters could be implemented as a control model to regulate substrate feeding to an anaerobic digester treating restaurant wastes. This results suggests that the convolutional network and state image approach leads to a more effective model because this approach is less influenced by the long and variable lag-times observed in a real bioprocess. In contrast, the LSTM model takes long range trends into account. Both the 1-D convolutional model and the LSTM models achieved low RMSEs between predicted and observed values of the first data set. However, model testing with new data showed that the LSTM did not have the ability to generalise. The hybrid LSTM/1-D convolutional network had the highest RMSE and the least ability to generalise.

The study demonstrated that the prediction accuracy of the neural network model is similar to the prediction accuracy of simply adjusting the feeding rate according to the observed trend in biogas production.

The average prediction accuracy of 58% should be compared to the accuracy of 63% obtained by simply following the trend in biogas production. The trend in biogas production was the result of *ad hoc* control of the digester by skilled operators working in a laboratory. It seems unlikely that a prediction accuracy

as high as 63% could have been achieved in an industrial setting without control by skilled operators. Importantly, the measured parameters of input liquid flow, output biogas flow, digester pH and digester ammonium content can be measured on-line using commercially available sensors. However, no sensor to measure volatile fatty acids is available commercially.

Considering that approximately 80% of the data points for pH, ammonium and volatile fatty acids concentrations were obtained by interpolation between values measured only 1 or 2 times per week, the accuracy of the model could be improved by increasing the sampling frequency thereby obtaining real data. Higher data resolution and consequently higher model accuracy could be achieved using on-line sensors.

5 Conclusion

This study evaluated 3 Neural Network architectures for use as supervised learning models to predict biogas production in an anaerobic digester. The models were constructed, trained and tested on the same time series dataset obtained from long-duration operation of an anaerobic digester treating restaurant wastes. Of the 3 models evaluated, the 1-D convolutional model was best able to accurately predict biogas production trends on new data as evaluated in terms of the ratio of correct predictions to the total number of requested predictions. The results suggest that the feeding rate of the study digester can be controlled using an 1-D Convolutional based DNN controller. Future work should aim to improve the resolution of the training data set by implementing on-line sensing of the relevant process parameters. During this study, the digester feeding rate was the only manipulated parameter. Future work should also aim to control additional manipulatable parameters and the mechanical parameters of the digestion process. Considering the wide range of variation observed in the input and target variables, the collection of time series data over a very long duration is required to make an accurate model.

Acknowledgement. This work was supported by Swiss Innovation cheque #32609 and by F.R. Mahrer of Digesto Sarl, Puplinge(GE), Switzerland.

References

1. Stenmarck, Å., et al.: FUSIONS: estimates of European food waste levels, Stockholm (2016). ISBN 978-91-88319-01-2
2. Capaccioli, S., et al.: Socio-economic benefits of small anaerobic digestion systems. In: Proceeding of the 24th European Biomass Conference and Exhibition, pp. 1530–1534, Amsterdam (2016)
3. Leite, W.R.M., Belli Filho, P., Gottardo, M., Pavan, P., Bolzonella, D.: Monitoring and control improvement of single and two stage thermophilic sludge digestion through multivariate analysis. *Waste Biomass Valorization* **9**(6), 985–994 (2016)

4. Alejo, L., Atkinson, J., Guzmán-Fierro, V., Roeckel, M.: Effluent composition prediction of a two-stage anaerobic digestion process: machine learning and stoichiometry techniques. *Environ. Sci. Pollut. Res.* **25**(21), 21149–21163 (2018)
5. Gaida, D., et al.: State estimation for anaerobic digesters using the ADM1. *Water Sci. Technol.* **66**(5), 1088–1095 (2012)
6. Bai, S., Kolter, J.Z., Koltun, V.: An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. [arXiv: e1803.01271](https://arxiv.org/abs/1803.01271) (2018)
7. Dong, Y.: An application of deep neural networks to the in-flight parameter identification for detection and characterization of aircraft icing. *Aerosp. Sci. Technol.* **77**, 34–49 (2018)
8. Pedregosa, F., et al.: Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**(Oct), 2825–2830 (2011)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.