Bioenergy recovery from sewage sludge in wastewater treatment plant towards energy neutrality: Machine learning and heterogeneous datasets

To-Hung Tsui [1, 2], Shuhan Lim [3], Jingxin Zhang [4], Yanjun Dai [5], Yen Wah Tong [1, 2, 3*]

1 Environmental Research Institute, National University of Singapore, 1 Create Way, 138602, Singapore
2 Energy and Environmental Sustainability for Megacities (E2S2) Phase II, Campus for Research Excellence and Technological Enterprise (CREATE), 1 Create Way, Singapore 138602, Singapore
3 Department of Chemical and Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, 117585, Singapore
4 China-UK Low Carbon College, Shanghai Jiaotong University, Shanghai, 201306, China
5 School of Mechanical Engineering, Shanghai Jiaotong University, Shanghai, 200240, China
Keywords: Anaerobic digestion, Sewage sludge, Pretreatment, Machine Learning
Presenting author email: eritht@nus.edu.sg
*Corresponding author email: chetyw@nus.edu.sg

Abstract

Sewage sludge management is nowadays one of the major challenges for wastewater treatment plants, and it could also account for up to 50% of the total operating costs. Over the years, anaerobic digestion is of great promise for sludge treatment as it stabilizes sludge, removes odors & pathogens, and produces renewable energy in the form of methane. However, the complex floc structure of waste activated sludge creates barriers to the penetration of hydrolytic enzymes for organics solubilization, and it limits the rate and extent of anaerobic microbial degradation. Due to the process constraint, it commonly results in anaerobic digestion with lower organic degradation efficiency (30–50%), longer sludge retention time (around 20–30 days) leading to larger digester volume requirement and higher capital cost for sludge treatment. Pretreatment technologies (esp. thermal pretreatment) have received huge attention for mitigating the constraints of the anaerobic digestion process and it commonly achieves: 1) enhanced destruction of sludge and biogas production; 2) reduced volume and better quality of biosolids, and 3) smaller digester volume or higher loading capacity (Kor-Bicakci and Eskicioglu, 2019). An energy feasibility study by Cano et al. (2015) suggested that most pretreatment approaches (other than thermal pretreatment) were not energy feasible when implemented at full-scale, due to their high energy requirements (generally in the form of electricity). Meanwhile, in terms of solubilization of intra-cellular organics for energy recovery, thermal treatment is the most energy-efficient and presents the highest potential to be applied at full-scale practice.

In the thermal pretreatment process, the initial sludge temperature is increased from the ambient temperature to a desired final temperature for a defined time period (minutes to hours). The system optimization (of thermal pretreatment and anaerobic digestion) involves complex tradeoffs and their many intertwined relationships are not linear in nature. For example, a higher temperature application is generally known to increase the sludge disintegration for higher methane production, but it also could generate a higher level of inhibitors (e.g. ammonia) and recalcitrant soluble compounds. Although many experimental studies have provided a valuable understanding of thermal pretreatment for anaerobic digestion, the qualitative implications were commonly drawn from the observation under defined experimental conditions (e.g. sludge properties, solid loads, thermal conditions, anaerobic digestion conditions) on final methane production. Meanwhile, for lower temperature (< 100°C) applications, it is common that several chemical types and different dosages were applied to assist the sludge disintegration. The incorporation of high-dimension variables makes the combined system (of thermal pretreatment and anaerobic digestion) more complicated to fully comprehend for the optimized design of energy efficiency. To ultimately achieve energy neutrality for wastewater treatment plants, it is necessary that the extra energy gained from anaerobic digestion should satisfy the total energy demand for thermal pretreatment and plant-wide wastewater treatment, which would necessitate numerical modeling for quantitative assessment. Therefore, to guide the complex system optimization, an insufficiency of current methodology is the over-reliance on laboratory testings and experiential learning by practitioners, which often creates confusion for understanding and decisions.

Recently, Machine Learning (ML) has started to be applied in the unit prediction of waste-to-energy processes (e.g. pyrolysis /gasification) (Li et al., 2021). Built on historical data from experiments or full-scale applications, the main advantage of ML application for sewage sludge treatment is that it can simulate the actual situation at reduced computational cost and less confined to the theoretical assumptions in traditional mechanistic models. Furthermore, in the majority of experimental research, those known controlling factors are often individually studied. When the scale-up application was subjected to more complicated scenarios, it was unknown which variables are more critical and to what extent can it affect the prediction of optimized energy recovery. Up to now, there is no system-level ML application (including thermal pretreatment and anaerobic digestion) that can encounter the complicated tradeoffs in engineering design. An outstanding challenge for such ML application is the consolidation of heterogeneous datasets that originated from many inconsistent experimental methodologies/designs. Furthermore, the unknown data quality from literature and the high-dimension of variables would also hamper the next-stage application of the generic algorithm. To address the methodological constraints

for efficient sewage sludge management (or similar combined system), we developed a unified framework that can advance the development of a generic ML algorithm for robust prediction of such system (i.e. anaerobic digestion inclusive of thermal hydrolysis and/or thermochemical pretreatment) involving complicated tradeoffs.

In this work, we will share the research efforts including dimensionality reduction for datasets, heterogeneous datasets integration, optimized generic algorithm. Furthermore, the next-stage advancement of ML applications will also be discussed. Figure 1 summarizes the overall framework of heterogeneous datasets integration and generic algorithm development in this study. The data collection was achieved from a total of >200 journal publications (of thermal pretreatment and anaerobic digestion for sewage sludge) over the last two decades. A literature search was conducted to collect the data from databases, including Web of Sciences, Google Scholar, and Scopus. In the bibliographic retrieval process, the keywords include thermal (pre)-treatment together with sludge and anaerobic digestion. The preliminary datasets strove to include a balanced representation (e.g. chemical type, temperature), while some studies with issues (e.g. missing core information, irrelevant experimental designs) were screened out. During the feature engineering stage, we carried pattern checking of data points at different dimensions, which aims to ensure quality datasets for the development of a generic algorithm. Briefly, the variables include sludge type (including primary sludge, waste activated sludge, and their sludge mixture), proximate composition (i.e. VS, TS, elemental composition, nutrient contents) at each treatment step, thermal conditions (i.e. temperature, duration, and pressure), chemicals (i.e. types and dosage), anaerobic digestion conditions (i.e. temperature, loading, duration, pH, C/N ratio) and final biogas production (i.e. control methane production and enhancement). Based on initial categorization, the data were first divided into two sub-datasets (i.e. with/without chemical dosage). Due to the heterogeneous features of data, the datasets were iteratively formatted and the continuous variables were standardized, where the inverse function of StandardScaler (from scikit-learn library) was used for inter-transformation between true and standardized values. In this work, Python programming was applied to major data pre-processing, data analysis and algorithms development. To handle the heterogeneous and high-dimensional datasets, the dimensionality reduction was assisted by Principal component analysis (PCA). Through PCA, the dataset is iteratively transformed into a reduced information space for features visualization and interpretation, while it still retains the inherent information as much as possible. These iterative steps also help evaluate the generalized dataset of its key features and informative regions, thereby enhancing faster algorithm convergence and more accurate prediction performance.

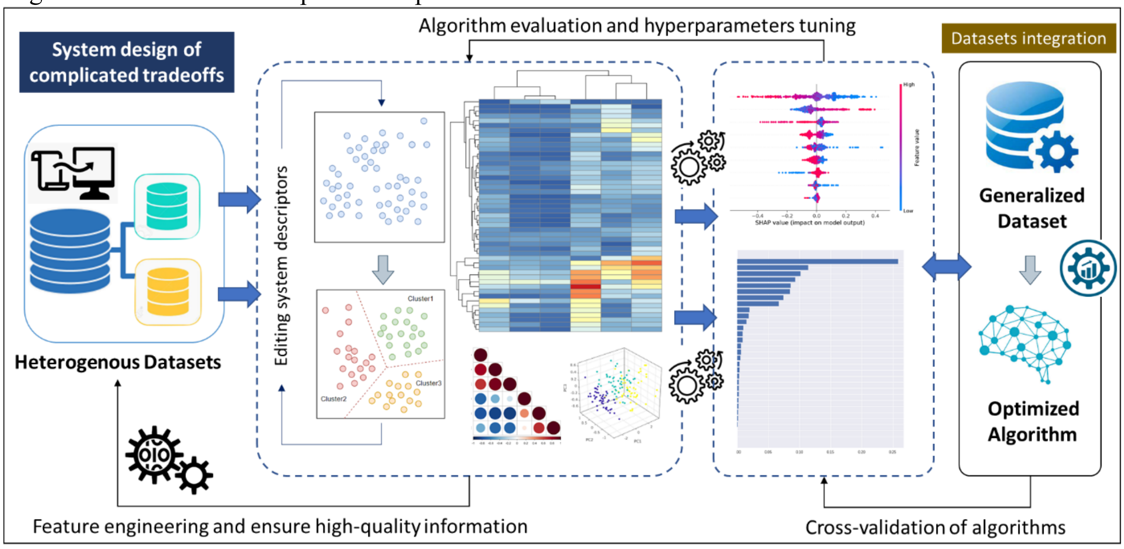

Figure 1. Overall framework of dimensionality reduction, datasets integration and generic algorithm development

Acknowledgements

This research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme

References

Cano, R., Pérez-Elvira, S. I., & Fdz-Polanco, F. (2015). Energy feasibility study of sludge pretreatments: a review. Applied Energy, 149, 176-185.

Kor-Bicakci, G., & Eskicioglu, C. (2019). Recent developments on thermal municipal sludge pretreatment technologies for enhanced anaerobic digestion. Renewable and Sustainable Energy Reviews, 110, 423-443.

Li, J., Zhu, X., Li, Y., Tong, Y. W., Ok, Y. S., & Wang, X. (2021). Multi-task prediction and optimization of hydrochar properties from high-moisture municipal solid waste: Application of machine learning on waste-to-resource. Journal of Cleaner Production, 278, 123928