CHEMICAL ENGINEERING TRANSACTIONS
VOL. 71, 2018
A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.aidic.it/cet
Guest Editors: Xiantang Zhang, Songrong Qian, Jianmin Xu

ISBN 978-88-95608-68-6; **ISSN** 2283-9216
DOI: 10.3303/CET1871118

Sewage Treatment Process Based on Big Data Management Mode

Ruifeng Ning

Ningxia Polytechnic, Yinchuan 750001, China
bdx513edu@126.com

With the rapid development of China's economy, people's living and industrial water consumption has increased significantly, and the resulting sewage discharge has also increased. At present, the source of sewage generation is widely distributed, the discharge time is not fixed, and there are many indicators for monitoring and monitoring of sewage. Traditional solutions have problems such as long collection period, large number of evaluation indicators, high dimension, and low system processing efficiency. In view of the above problems, this topic, based on an in-depth analysis of the characteristics of waste water data, combined with big data processing technology, researched and implemented a pollution index analysis system based on big data. On the basis of reducing the sewage data dimension, the k-means clustering algorithm is used to enhance the clustering effect of sewage big data, and the main over-standard value of waste water is clarified. The reference values of waste water treatment in this paper are COD, TN, TP, NH3-N. The IBR biological treatment process is selected to treat the sewage and the treatment results are obtained.

1. Introduction

Big data usually refers to a collection of information that generates too tremendously to be processed, analyzed and managed with the commonly used software in a timely manner. It is a new technology system for collecting, storing, analyzing, managing, mining and applying the huge mass of data (Jezowski et al., 2003). The sewage treatment process is a relatively complex project, in which, the applications of Big data will have more outstanding features than in other fields such as finance, electricity and biology. (Flores et al., 2015) Today, dramatically emerged computer technologies have ushered the human society into the new information era. In particular, with the digital technology, plenty of the knowledge and experience accumulated by human activities and social development have been heaped up into massive data resources, growing explosively in their species and scales. Today, the computer system is usually used to supervise the operation status of the sewage treatment process, and as necessary, it periodically collecs the system variables and the status of individual nodes to ensure a normal operation, display or control. Plenty of repetitive and redundant data are required for measuring, so that there is a huge mass of data in the long run. (Ali, 2014; Amir et al., 2016) Various variables for sewage treatment system are changed in real time in order to keep dynamic balance in the system. The measurements of these variables as changed can underlie an effective prediction for the system operation status and trends. There are data about normal operation, various abnormal conditions and fault status in the sewage treatment system, which are no doubt the key to the normal operation of sewage treatment system. Complex hydrodynamic properties also reward us a more complex biochemical effect. Different meteorological and soil conditions will interact with various boundary conditions in the water cycle process. Under the effect of human activities, there are quite different information about social water system. In the real world, the natural and the social water systems are intertwined and interact each other, greatly exacerbating the complexity of data. (Malghani et al., 2013; Mamais et al., 2015). Due to the dispersion of water resources in the spatial distribution, the dynamics in the temporal distribution, and the chronicity of the replacement cycle, the management of water system data is interdisciplinary an integrated. (Mona et al., 2011; Asgher, 2012) The complexity of the whole operation system diversifies the degrees and speeds of changes in various variables, resulting in different frequencies for collecting the signals, further in the asynchronism of data acquisition. In the process of data logging, multiple time-scale and incompleteness

Please cite this article as: Ning R., 2018, Sewage treatment process based on big data management mode, Chemical Engineering Transactions, 71, 703-708 DOI:10.3303/CET1871118

may cause partial data loss, resulting in the data incompleteness. Urban water system is dynamic in temporal and spatial distribution, and regenerative in the process of circulation. For the features of big data, see Figure 1.

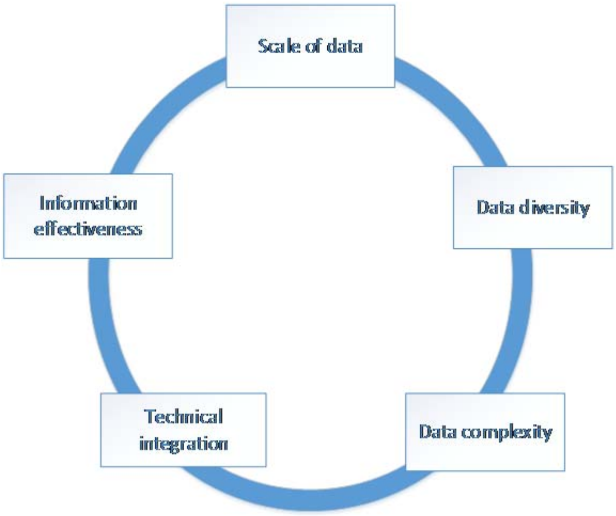

Figure 1: Big data feature

2. Experiments

2.1 Cloud platform for sewage treatment big data

The big data cloud platform for sewage treatment system mainly uses remote data acquisition device to timely collect and transmit data about the sewage treatment system, and the data transmission device to timely upload data to the cloud platform for management, wherein the remote data acquisition devices used here mainly include the industrial standard online and wireless sensors. It includes a variety of data analysis and treatment models, where appropriate indicator and standard limits can be set for sewage treatment system (Klangjorhor et al., 2014; Crini, et al., 2018). The platform can also automatically classify, aggregate, compare and analyze uploaded data, and then transmit the results to the client customers to realize real-time supervision and management on the sewage treatment status.

2.2 Building the cloud system

The system architecture includes five layers, i.e. data sensing layer, data aggregation layer, network interconnection layer, big data intelligent processing layer and data display layer, as shown in Figure 2.

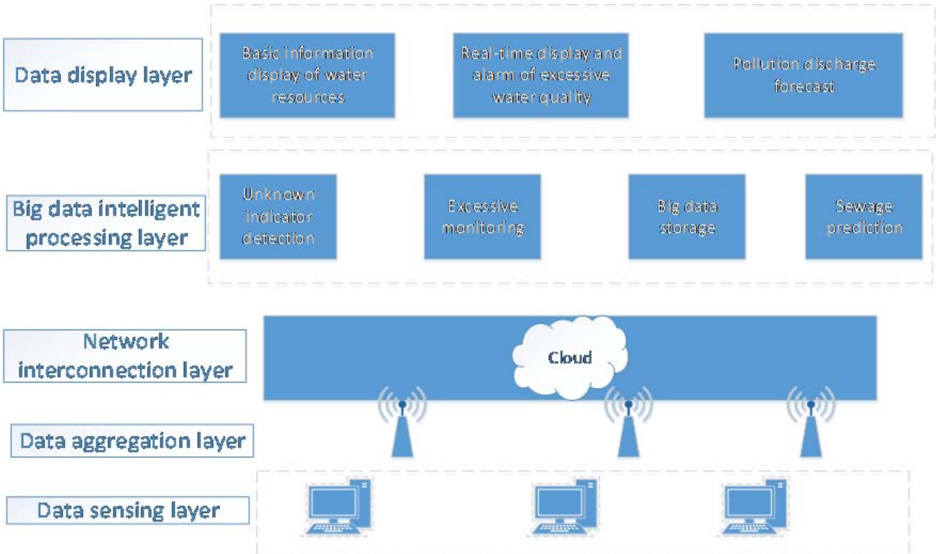

Figure 2: Sewage treatment cloud system mode

2.3 Sewage over standard algorithm

Up till now, there is still single indicator for testing sewage water quality, and in accordance with the international standards, it can only be judged whether a single indicator exceeds the standard limits. The national standard only specifies the value that a certain indicator exceeds the standard, but does not specifies which limits the individual indicator is allowed to exceed the standards. The limit of excessive discharge for a factory involves not only certain indicator, but also the number of indicators that exceed the standards. Given that the division of the factory's excessive discharge limits is fuzzy and complicated, this paper intends to use the data mining clustering algorithm to conduct a cluster analysis on the limits of plant’s pollution discharge standard. (Mamais et al., 2015)

2.3.1 Building the clustering model

Clustering refers to the classification of a class of things with greater similarity in the database. The distance between individuals classified into the specific category is closer, while the distance between individuals in different categories is too wide. The relationship between data in clustering analysis is described using the eigenvector distance between individuals. It is an important part of unsupervised learning method (Ali, 2014).

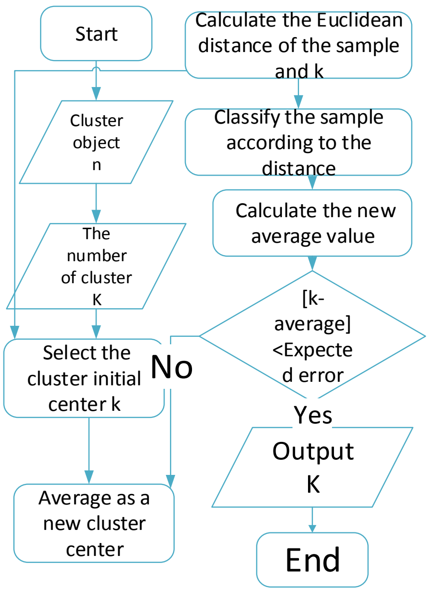

Figure 3: Flowchart of k-means cluster analysis algorithm

2.3.2 Principle of cluster analysis

Step 1: Select a certain distance as the measure between samples
The search process of the k-means clustering algorithm is limited to a part of all possible partition spaces. If the sample similarity between individual classes is very low, the k-means algorithm can often achieve a good result. Otherwise, the clusters should be further subdivided. It is possible to obtain a local rather than a global minimum solution for the scoring function because of the algorithm convergence (Crini et al., 2018).
Step 2: Selection criteria function
The k-means clustering algorithm is subjected to the selected similarity measures, among which, the commonly used similarity measure applies the sum-of-squared error criterion function to improve the clustering performance. Let the set

$$s = \{x_1, x_2, \cdots x_n\} \tag{1}$$

Which represent n points.

$$C_1, C_2, \cdots, C_k \tag{2}$$

Which represent k different partition sets, then satisfy

$$C_1 \neq \varnothing, i = 1, 2, \cdots, k \tag{3}$$

$$c_i \cap c_j = \varnothing, i = 1, 2, \cdots, k, \ j = 1, 2, \cdots, k \tag{4}$$

$$\bigcup_{i=1}^{k} c_i = s \tag{5}$$

 From

$$\{x_1, x_2, \cdots x_n\} \tag{6}$$

Randomly select k points Cl, C2, ..., Ck as the center point of the k cluster sets.
② The division is performed with Cl, C2, ..., Ck as the center point.

$$\|X_i - c_j\| < \|X_i - c_m\|, m = 1, 2, \cdots, k; \ j = 1, 2, \cdots, k \tag{7}$$

If so, it will be divided into set C.
③ Calculation criterion function - square error function:
This algorithm attempts to find the k partitions that minimize the value of the squared error function. When the resulting clusters are dense, and the clusters are distinct between clusters, it works particularly well
Then the sum-of-squared error criterion function is calculated by the formula:

$$E = \sum_{i=1}^{\mathrm{k}} \sum_{Xi \subset Cj} \|X_i - c_j\|^2 \tag{8}$$

Step3: Center of new cluster

$$c_1^*, c_2^*, \cdots, c_k^* \tag{9}$$

After the first iteration, the average error is better improved. Since the cluster center does not change during the two iterations, the iterative process is stopped and the algorithm is terminated (Lizarralde et al., 2015).

$$c_i^* = \frac{1}{n} \sum_{j=1} X_j, \ j = 1, 2, \cdots, k \tag{10}$$

With the cluster analysis, the out-of-limit test module in the sewage treatment process uses the appropriate test model where there are three types of inputs. The first type is the sewage indicator value measured directly, and the second type is the output of the sewage indicator test module, i.e. individual complex indicator measurements. The third type is the criterion for the determination of the out-of-limit extent, calculated by the improved k-means clustering model for sewage treatment system as described in Chapter 3. With the well-established out-of-limit test model, it is possible to determine the current sewage out-of-limit extent and output the final results (Ali 2014).

3. IBR sewage treatment process

Take data from practical sewage treatment process in a plant in recent two years as an example, the clustering analysis is carried out for the out-of-limit extent. In this process, assume 3 values k are specified, that is, it is required to divide the sewage out-of-limit extent into minor, medium, and critical levels. The COD indicator in the sewage treatment process represents a moderately polluted state, so that we choose the IBR biological treatment process for sewage. (Mamais et al., 2015)
The IBR biological treatment process (see Figure 4) is a cyclically activated sludge process that integrates anaerobic, anoxic, aerobic reactions and precipitation. It also has the advantages of a continuous flow activated sludge process divided by space and an intermittent activated sludge process divided by time interval. Compared with the former, the latter leaves out the external reflux of sludge, thus saving operating energy consumption and reducing processing facilities and investment. Multiple levels of A/A/O are formed in the reaction pool by adjusting the aeration-to-stop ratio to maximize the nitrogen and phosphorus removal (Asgher, 2012).

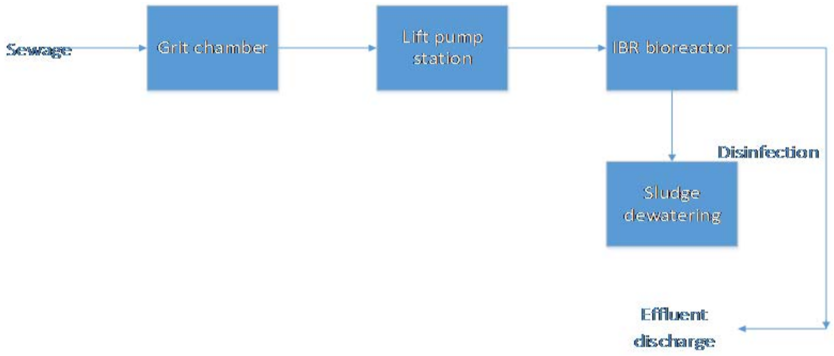

Figure 4: IBR biological processing technology flow

Table 1: IBR operation effect (mg/L)

Index		COD	TN	TP	NH3-N
Effluent	Maximum	185	43	5	39
	Minimum value	117	29	3	23
	Standard deviation	20	4	0.8	5
IBR effluent	Maximum	63	33	2	8
	Minimum value	29	117	1	3
	Standard deviation	7	2	0.3	1
	IBR removal rate (%)	72	44	63	78

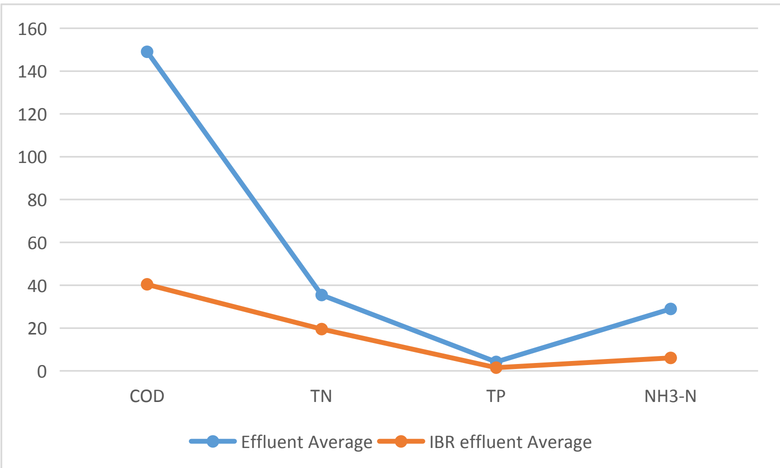

Figure 5: Average IBR Effect

The water quality of COD and TN, NH3-N can reach the first Classes A and B standard, respectively. since the influent COD and NH3-N are inherently low, and the aeration rate is sufficient and the sludge load is lower. Therefore, COD and TN, NH3-N can get a better effluent quality. Due to insufficient carbon source, TP cannot well fit the bill of the Class B standard as expected by the IBR process. This is because the nitrogen removal effect is not very good, which thus results in a bad removal effect of phosphorus.

4. Conclusion

This paper establishes a big data sewage treatment management cloud platform, uses K-Means algorithm cluster analysis to calculate the main over-standard value of wastewater, and finally removes COD efficiently through IBR sewage treatment process, and the removal effect of nitrogen and phosphorus is also good. At the same time, IBR/constructed wetland has the advantages of low energy consumption, good treatment effect, low investment and land area. After normal operation, combined with laboratory operation, COD, TN

NH3-N can reach Grade A emission standard when water temperature, water volume is low and influent nitrogen and phosphorus content is high, and TP can reach Level 1 emission. standard. For the problem that TP is difficult to meet the standard, a materialized strengthening method can be adopted.

Reference

Ali I., 2014, Water treatment by adsorption columns: evaluation at ground level, Separation & Purification Methods, 43(3), 175-205.

Amir M.N.I., Julkapli N.M., Hamid S.B.A., 2016, Incorporation of chitosan and glass substrate for improvement in adsorption, separation, and stability of tio 2, photodegradation, International Journal of Environmental Science & Technology, 13(3), 865-874.

Asgher M., 2012, Biosorption of reactive dyes: a review, Water Air Soil Pollut, 223, 2417–2435, DOI:10.1007/s11270-011-1034-z

Crini G., Lichtfouse É., Wilson L.D., Morin-Crini N., 2018, Adsorption-oriented using conventional and non-conventional adsorbents for wastewater treatment. In: Crini G, Lichtfouse É (eds) Environmental chemistry for a sustainable world, green adsorbents for pollutant removal—fundamentals and design, 1, 23–71, DOI:10.1007/978-3-319-92111-2_2.

Flores-Alsina X., Mbamba C.K., 2015, A plant-wide aqueous phase chemistry module describing pH variations and ion speciation/pairing in wastewater treatment process models, Water research, 85, 255-265, DOI:10.1016/j.watres.2015.07.014

Jezowski J., Poplewski G., Jezowska A., 2003, Simulated annealing optimization in chemical and process engineering i. optimisation method with the use of a simplex and simulated annealing, Inzynieria Chemiczna I Procesowa, 24(1), 47-62.

Klangjorhor J., Phitak T., Pruksakorn D., Pothacharoen P., Kongtawelert P., 2014, Comparison of growth factor adsorbed scaffold and conventional scaffold with growth factor supplemented media for primary human articular chondrocyte 3d culture, BMC Biotechnology, 14(1), 108.

Kyzas G.Z., Kostoglou M., 2014, Green adsorbents for wastewaters: a critical review, Materials, 7, 333–364, DOI:10.3390/ma7010333

Lizarralde I., Fernández-Arévalo T., Brouckaert C., Vanrolleghem P., Ikumi D. S., Ekama G. A., Grau P., 2015, A new general methodology for incorporating physico-chemical transformations into multi-phase wastewater treatment process models, Water research, 74, 239-256, DOI:10.1016/j.watres.2015.01.031

Malghani S., Gleixner G., Trumbore S. E., 2013, Chars produced by slow pyrolysis and hydrothermal carbonization vary in carbon sequestration potential and greenhouse gases emissions, Soil Biology & Biochemistry, 62(5), 137-146.

Mamais D., Noutsopoulos C., Dimopoulou A., Stasinakis A., Lekkas T. D., 2015, Wastewater treatment process impact on energy savings and greenhouse gas emissions, Water Science and Technology, 71(2), 303-308, DOI:10.2166/wst.2014.521

Mona S., Kaushik A., Kaushik C.P., 2011, Biosorption of reactive dye by waste biomass of nostoc linckia, Ecological Engineering, 37(10), 1589-1594.

Sabeen A.H., Ngadi N., Noor Z.Z., Raheem A.B., Agouillal F., Mohammed A.A., Abdulkarim B.I., 2018, Characteristics of the Effluent Wastewater in Sewage Treatment Plants of Malaysian Urban Areas, Chemical Engineering Transactions, 63, 691-696, DOI: 10.3303/CET1863116