Peertechz

ENGINEERING GROUP

ANNALS OF Robotics and Automation OPEN ACCESS JOURNAL

DOI: https://dx.doi.org/10.17352/ara

Research Article

The application of unsupervised machine learning to optimize water treatment membrane selection

Khaled Younes*, Omar Mouhtady and Hamdi Chaouk

College of Engineering and Technology, American University of the Middle East, Kuwait

Received: 21 May, 2021
Accepted: 02 July, 2021
Published: 03 July, 2021

***Corresponding author:** Khaled Younes, College of Engineering and Technology, American University of the Middle East, Kuwait,
E-mail: Khaled.younes@aum.edu.kw

https://www.peertechzpublications.com

Check for updates

Abstract

Artificial intelligence technologies have been extensively used to decipher water quality and characterization. Fewer studies have employed these techniques in the purpose of optimizing a water treatment process. Here, we apply unsupervised machine learning techniques for the optimization of the choice of membranes, following the different constraints and conditions encountered. The adopted data analysis techniques are the Principal Component Analysis (PCA) and the Hierarchical Cluster Analysis (HCA). Both methods showed their capacity to reveal resemblance and discrepancies between different membrane types and based on several properties. PCA is more appreciated than HCA as it removes any intercorrelation between factors and it helps in a better understanding of different trends of the dataset by establishing a Scores-Factors relation.

Introduction

With the increasing rate of worldwide population and the scarcity of natural water supply, due to the lack of global precipitation and climate change, the demand on drinking water is rising in an exponential way [1]. This increase has put pressure on water working treatment personal, as a need to boost drinking water production is more and more challenging. In order to overcome this problem, all applied fundamental scientific fields have found their place in the Water Treatment Plant's (WTP) optimization. Chemistry and Physics were involved in order to understand the behavior of water as a molecule and a mixture of different components, at the microscopic and macroscopic level, simultaneously [1]. The application of chemistry and physics in a single-handed way, has found its shortcomings in giving answers for the problems encountered in WTP [2-4]. It is not surprising that looking at a problem, in a monotonic way, will give less grasp to the phenomena occurred. In order to overcome this issue, the integration of applied mathematics to the applied sciences have gave a better understanding of the problem, hence, a better set of solutions for water treatment issues [1]. This combination of different scientific fields has put the artificial intelligence (AI) in the heart of water analysis and treatment [1].

In other hand, the application of AI technologies in water real-life problems have raised several concerns, because the mathematical models, employed, are usually based on assumptions that are difficult to implement in practice [1]. Moreover, modelling lacks of an overall understanding of the analyzed system requirements [2-4]. The non-linear relationships involved in water processing are challenging to fit [2-4]. This has raised the interest of using unsupervised machine learning approaches [1]. These techniques work on a given dataset (from chemistry or physics) and yields a certain trend, without a prior knowledge or any assumptions adopted [1].

The application of Machine Learning technologies has been more likely used to reveal unhidden pattern in analytical dataset, to better interpret the quality of analyzed water. In the process optimization side and material choice, extensive work can be executed. Therefore, our aim in this study is to perform two unsupervised machine learning techniques to better pick between filter types that are mostly used in WTPs. The two investigated methods are the Hierarchical Cluster Analysis (HCA) and the Principal Component Analysis (PCA).

Citation: Younes K, Mouhtady O, Chaouk H (2021) The application of unsupervised machine learning to optimize water treatment membrane selection. Ann Robot Automation 5(1): 030-033. DOI: https://dx.doi.org/10.17352/ara.000010

Machine learning methods

Principal Component Analysis (PCA): PCA is used in data exploration techniques used and for establishing descriptive models. It works on the dimensionality reduction. Data reduction is done, with two perspectives in mind: (1) the lower the dimensions yielded, and (2) the orthogonality of the new dimensions, which are the principal components (PCs). PCs are actually the direction of the maximal variance of the dataset. Performing this task for two factors is quite feasible, yet this issue gradually increases with the increase of the number of factors. Overcoming this problem is made easy with the development of sophisticated calculation algorithms [9–11]. Several studies have focused on the mathematical description of PCA. Hence, the theory is well developed. Other studies have used PCA as a tool to reveal some proxies in Geochemistry [12–15], Energy [16] and Biomass characterization [17,18].

Hierarchical Cluster Analysis (HCA): HCA is a classification technique of objects into different groups. It starts with one cluster, as individual item in its own cluster, and it iteratively merge clusters until all items belong to one cluster. It follows a bottom-up approach, where the clusters are merged together. Pictorially, dendrograms are used to represent the HCA. It can be represented using three techniques, the single-nearest distance or single linkage, the complete-farthest distance, and average-average distance or known as average linkage. The single linkage is described as the distance between the closest members of two clusters, the complete linkage as the distance between the members that are farther apart. The average linkage involves looking at the distances between all pairs and averages of these distances. This is also called the unweighted pair group mean averaging, that we have used in our study [9,10]. The application of HCA is more likely extended to molecular biology [19,20]. Other studies have traced the application of HCA in Biomass characterization [17,18].

The application of machine learning in membrane technology

Here, we will apply the above-described unsupervised machine learning tools for the elucidation of different trends that might occur in different type of water treatment membranes. Several features are employed to describe the membrane [21] (Table 1). Turbidity-Raw and turbidity-Effluent present the efficiency of the membrane in eliminating particles. Without any doubt, an efficient membrane requires low turbidity effluent. Some other features like Pre-Coat and Body Feed describe operating protocol. The pre-coat tank is filled with a given amount of water and with a given mass charge of diatomite that results in the specified pre-coat area density. Selection of DE grade and determination of concentration and mixing slurry are the tasks associated with body feed. Usually, the grade and concentration are determined as a first estimate during the design phase and refined during operation. The differential pressure (ΔP) states the energy required to perform filtration. The higher the differential is, more energy is required. Number of runs is an age-related factor; the higher it is, the longer the lifetime of the membrane will be.

The dendrograms in Figure 1 show similarities and dissimilarities between the investigated DE membranes. The Euclidean distance between strains was investigated based on the featured properties, shown in Table 1. Generally, two main clusters are observed. 900W(P), C-535 and FW12(DE) are the components of the cluster showing the highest similarity. Other membranes presented the components of the second cluster where lower similarity was observed. If one compares it with the components of the first cluster. In other words, the Euclidean distances between the components of the first cluster are relatively lower than those between the other components. The most similar membranes are FW-20(DE) and 4200(DE). These two membranes are exceptional in the second cluster, as higher discrepancy, between other components, is shown. Although the same features were presented for all membranes (Table 1), HCA allowed the distinction between two main patterns: 900W(P), C-535 and FW12(DE) compose one pattern and the rest compose the second patter. This data analysis technique has shown similarities and discrepancies between membrane types; a deeper investigation of this difference would be envisaged by PCA.

PCA was performed for the featured properties (Table 1), in order to test their distribution among different types of membranes. PCA also provides a general view of correlation and dissimilarity among the seven membrane type and the six investigated factors. The results were presented on two-dimensional perspective with a graph (plot for scores and loadings, simultaneously) obtained from Pearson correlation matrix for variables (Figure 2). The first and second PCs accounted for 75.57% of total variance in data set (PC_1, 52.07% and PC_2, 23.5%). This high value indicates that the comparison

Table 1: Description of seven Diatomaceous Earth (DE) plants and operating protocols.

Filter Aid - Grade	Pre-Coat	Body Feed	Turbidity - Raw	Turbidity - Effluent	Run	ΔP
FW-20 (DE)	0.11	10.8	0.75	0.1	14	54
4200 (DE)	0.08	17.7	0.4	0.12	10	57
700, 900, 1500 (P)	0.05	63	100	0.3	3.3	483
C-535	0.1	12	0.45	0.5	12	207
FW12 (DE)	0.26	40	50.5	0.09	3	186
900 W (P)	0.04	12.6	0.2	0.04	14	276
Hyflo Super-Cel (DE)	0.09	32.5	8.5	0.3	0.4	96

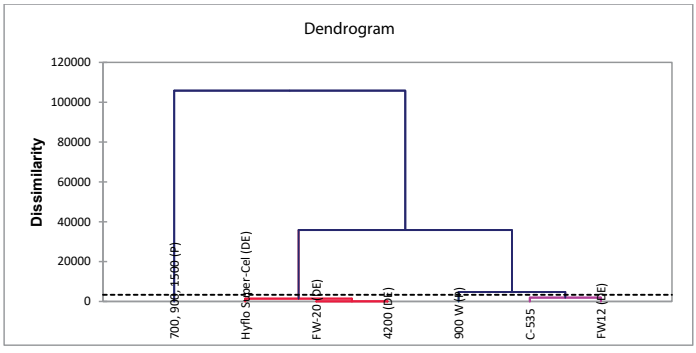

Figure 1 : Hierarchical cluster analysis, showing the complete linkage dendrograms of the data, based on the featured properties of different membranes (**Table 1**).

Citation: Younes K, Mouhtady O, Chaouk H (2021) The application of unsupervised machine learning to optimize water treatment membrane selection. Ann Robot Automation 5(1): 030-033. DOI: https://dx.doi.org/10.17352/ara.000010

of the employed parameters is statistically meaningful and reliable trends can be concluded from this dataset (Table 1). Four different clusters can be identified from the PCA approach (Figure 2). The first cluster contains FW-20(DE), 4200(DE), 900W(P) and C-535. The other three clusters contained FW12(DE), HyfloSuper-Cel(DE) and 700 900 1500(P), separately. Unlike HCA approach, 4200(DE) showed proximity to 900W(D) and its homologues.

Regarding factor loadings, PC_1 was most likely positively dominated by Body Feed, turbidity and ΔP. On the negative side of PC_1, the number of Runs has the major influence. Following these trends, it can be shown that PC_1 presents the "Efficiency Factors" of membranes. PC_2, was most likely positively dominated by Pre-Coat. This indicates that PC_2 is most likely an indicator of the conditioning state of the membranes. All membranes were found to be away from "Turbidity Raw" factor. This indicates that all investigated membranes present a high pollutant removal capacity. Cluster 1 of membranes (FW-20(DE), C-535, 4200(DE) and 900W(P), Figure 2) shows high correlation with the factor "Run". This indicates that these membranes could be more likely employed in conventional water treatment processes that require higher number of runs. FW12(DE) presented a high correlation to the "Pre-Coat" factor. This indicates that care should be taken when FW12(DE) is used in corrosive and highly reactant conditions (high acidity, alkaline conditions, temperature, pressure...). 700 900 1500 (P) membrane presents high correlation with ΔP. This indicates that this type of membrane requires and handles high-energy input. Interestingly, HyfloSuper-Cel(DE) have been projected near the node. This means that this membrane presented low influence to the investigated factors, or it presents intermediate behavior, if compared to the other membrane types.

Comparison between machine learning methods

The unsupervised machine learning approaches show interesting features of the compared properties, as trends of the relative properties are hardly seen, when analyzed independently. Both methods showed a clear dissimilarity between some of the membrane types. PCA showed higher efficiency rather than HCA; as along showing discrepancies, PCA allowed us to quantify the influence of the investigated factors. This feature makes it rather advantageous on the HCA, as it only distinguished, quantitatively, between membranes.

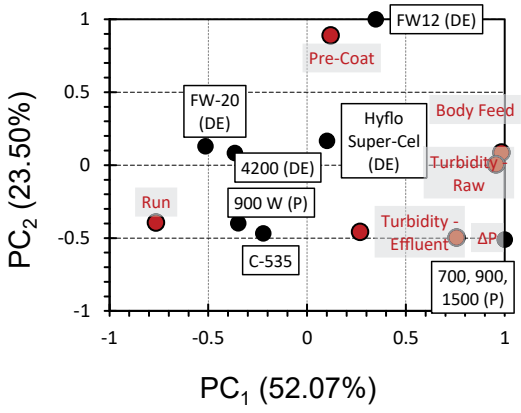

Figure 2: Principal component analysis: factor scores of the seven investigated DE membranes are presented with black bullets; factor loadings of the featured properties is presented with red bullets. The first two ordination axes (PC_1 and PC_2) are shown.

PCA simplifies the complexity of a dataset with high dimensionality while, at the same time, keeps the different patterns and highlights the significant trends. This yields a better interpretation and PCs act as the new factors representing the dataset. These factors are independent from each other, yet represent, in a single-handed way, a combination of all of the factors with a different proportion of influence.

Conclusion

This study only presents a small extent of the applicability of unsupervised machine learning to pick the required apparatus, for an investigated treatment. A small dataset has been purposely chosen, in order to reveal the correlations and discrepancies *via* simple data visualization. Hence, the proposed data mining approach have elucidated the efficiency of PCA and HCA to reveal trends between membrane materials. PCA found a better efficiency rather than HCA, as the first showed the influence and weight of each factor, in regard to the classification. The second was only restricted to classifying different membranes used, without deciphering the factors involved in this classification. Hence, we strongly recommend the application of PCA for depicting a better choice of equipment and to optimize water treatment process conditions.

References

1. Li L, Rong S, Wang R, Yu S (2021) Recent advances in artificial intelligence and machine learning for nonlinear relationship analysis and process control in drinking water treatment: A review. Chemical Engineering Journal 405: 126673. Link: https://bit.ly/3656tSM

2. Curcio S, Calabrò V, Iorio G (2006) Reduction and control of flux decline in cross-flow membrane processes modeled by artificial neural networks. Journal of Membrane Science 286: 125–132. Link: https://bit.ly/3ycAE6w

3. Ghandehari S, Montazer-Rahmati MM, Asghari M (2011) A comparison between semi-theoretical and empirical modeling of cross-flow microfiltration using ANN. Desalination 277: 348–355. Link: https://bit.ly/3dAO5W3

4. Liu QF, Kim SH (2008) Evaluation of membrane fouling models based on bench-scale experiments: A comparison between constant flowrate blocking laws and artificial neural network (ANNs) model. Journal of Membrane Science 310: 393-401. Link: https://bit.ly/3ycPbiE

5. Maier HR, Morgan N, Chow CWK (2004) Use of artificial neural networks for predicting optimal alum doses and treated water quality parameters. Environmental Modelling & Software 19: 485-494. Link: https://bit.ly/3hwbzwL

6. Ding S, Deng Y, Li H, Fang C, Gao N, et al. (2019) Coagulation of Iodide-Containing Resorcinol Solution or Natural Waters with Ferric Chloride Can Produce Iodinated Coagulation Byproducts. Environ Sci Technol 53: 12407-12415. Link: https://bit.ly/3xe8UOL

7. Adusei-Gyamfi J, Ouddane B, Rietveld L, Cornard JP, Criquet J (2019) Natural organic matter-cations complexation and its impact on water treatment: A critical review. Water Research 160: 130-147. Link: https://bit.ly/2UjAcog

8. Al'Aref SJ, Anchouche K, Singh G, Slomka PJ, Kolli KK, et al. (2019) Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging. Eur Heart J 40: 1975-1986. Link: https://bit.ly/3waaMqE

Citation: Younes K, Mouhtady O, Chaouk H (2021) The application of unsupervised machine learning to optimize water treatment membrane selection. Ann Robot Automation 5(1): 030-033. DOI: https://dx.doi.org/10.17352/ara.000010

9. Barnett TP, Preisendorfer R (1987) Origins and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by canonical correlation analysis. Monthly Weather Review 115: 1825–1850. Link: https://bit.ly/3jOainF

10. Hsu D, Kakade SM, Zhang T (2012) A spectral algorithm for learning hidden Markov models. Journal of Computer and System Sciences 78: 1460–1480. Link: https://bit.ly/2SFjoaU

11. Markopoulos PP, Kundu S, Chamadia S, Pados DA (2017) Efficient L1-norm principal-component analysis via bit flipping. IEEE Transactions on Signal Processing 65: 4252–4264. Link: https://bit.ly/3qGncoO

12. Younes K, Grasset L (2017) Analysis of molecular proxies of a peat core by thermally assisted hydrolysis and methylation-gas chromatography combined with multivariate analysis. Journal of Analytical and Applied Pyrolysis 124: 726-732. Link: https://bit.ly/3qlqgkh

13. Younes K, Grasset L (2018) Comparison of thermochemolysis and classical chemical degradation and extraction methods for the analysis of carbohydrates, lignin and lipids in a peat bog. Journal of Analytical and Applied Pyrolysis 134: 61-72. Link: https://bit.ly/3w6uiEk

14. Younes K, Grasset L (2020) The application of DFRC method for the analysis of carbohydrates in a peat bog: Validation and comparison with conventional chemical and thermochemical degradation techniques. Chemical Geology 545: 119644. Link: https://bit.ly/3hpyzxr

15. Younes K, Laduranty J, Descostes M, Grasset L (2017) Molecular biomarkers study of an ombrotrophic peatland impacted by an anthropogenic clay deposit. Organic Geochemistry 105: 20-32. Link: https://bit.ly/3w6dLQQ

16. Zhang F, Obeid E, Nader WB, Zoughaib A, Luo X (2021) Well-to-Wheel analysis of natural gas fuel for hybrid truck applications. Energy Conversion and Management 240: 114271. Link: https://bit.ly/2Ucx0uJ

17. Korichi W, Ibrahimi M, Loqman S, Ouhdouch Y, Younes K, et al. (2021) Assessment of actinobacteria use in the elimination of multidrug-resistant bacteria of Ibn Tofail hospital wastewater (Marrakesh, Morocco): a chemometric data analysis approach. Environ Sci Pollut Res 28: 26840–26848. Link: https://bit.ly/3dAwj5i

18. Ibrahimi M, Korichi W, Loqman S, Hafidi M, Ouhdouch Y, et al. (2020) Thermochemolysis–GC-MS as a tool for chemotaxonomy and predation monitoring of a predatory actinobacteria against a multidrug resistant bacteria. Journal of Analytical and Applied Pyrolysis 145: 104740. Link: https://bit.ly/3xcrTsV

19. Roufayel R, Murshid N (2019) CDK5: Key Regulator of Apoptosis and Cell Survival. Biomedicines 7: 88. Link: https://bit.ly/2Uk9MTD

20. Roufayel R, Mezher R, Storey KB (2021) The Role of Retinoblastoma Protein in Cell Cycle Regulation: An Updated Review. Curr Mol Med. Link: https://bit.ly/3dyKaZM

21. Hendricks D (2010) Fundamentals of water treatment unit processes: physical, chemical, and biological, Crc Press. Link: https://bit.ly/3hu3tEK

Citation: Younes K, Mouhtady O, Chaouk H (2021) The application of unsupervised machine learning to optimize water treatment membrane selection. Ann Robot Automation 5(1): 030-033. DOI: https://dx.doi.org/10.17352/ara.000010