DEVELOPING THE OPTIMIZED OCEAN CURRENT STRENGTHENING DESALINATION SEMI-PERMEABLE MEMBRANE DISTILLATION BASED ON DECISION TREE

Chaeyun Jing[1] , Seoyeon Shin, and Taeseon Yoon[2]

[1]Department of International, Hankuk Academy of Foreign Studies, Yong-In, Republic of Korea
[2] Hankuk Academy of Foreign Studies, Yong-In, Republic of Korea

ABSTRACT

Alongside improvements in desalination operation and development of new technologies, problems of weakened counter current and global warming have emerged. Therefore, our study suggests a new desalination model, based on the experimental Support Vector Machine (SVM) algorithm, for semi-permeable membrane separation. First, the reverse osmosis (RO) process used semi-permeable membrane and osmotic pressure to remove the solutes dissolved in seawater and obtain pure freshwater. The desalination process also applied MSF and MED, which are the best technologies developed through elimination of various problems that were previously experienced. This research is directed towards suggesting a model that can effectively create the semi-permeable membrane used in the desalination process. To efficiently prevent a counter current and safely obtain the water resources, an innovative technology is suggested by applying Genetic Algorithm (GA) to the SVM model for the semi-permeable membrane, and the effectiveness of the variables are proved with the help of the Decision Tree.

KEYWORDS

Decision Tree, Reverse osmosis, Support Vector Machine, MSF , MED

1. INTRODUCTION

Water shortage has long been raised as a serious global issue, andmany countries, including the Republic of South Africa, Egypt, Belgium, and Poland, have already been designated as countries with water shortage. The average annual precipitation of our country was 1283 mm, which is greater than the global average of 973 mm; however, because 70% of the land is composed of steep and mountainous landscape and most of the precipitation occurs in summer, much of the precipitation flows into the sea. In addition, because of the high population density, the per capita precipitation is only 12% of the global average. According to the Water Poverty Index (WPI) declared by the World Water Forum in 2006, South Korea is ranked 43 out of 147 countries [1], indicating a relatively adequate water suppl. However, a constructive solution is required for the upcoming crisis because the competition between the countries with high potential for water

industry growth is expected to be even greater in the future. Research on the water shortage problem has been carried out using different methods. One of these methods, desalination, which uses evaporation of salty seawater to produce drinking water, encountered various problems with on-site application, and reverse osmosis (RO) was suggested as a supplement, but was found to be cost-inefficient. Hence, new technologies such as MSF and RO have been applied. However, even with RO, which appeared to be the best option, desalination incorporating hybrid MSF, RO, and MED resulted in significant problems, such as worsening of the current and global warming. Therefore, it is considered a double-edged sword. During desalination, the density of water is lowered, and it becomes comparably light; therefore, the amount of seawater sinking into the lower layer reduces significantly in the polar regions, and the current along the equator weakens in the deep layer and is created on the surface, forming a counter current. Thus, the warm current forces towards the polar regions become weaker. This phenomenon is not a localized problem but a major cause for global warming. Hence, to solve this difficult problem, this study aims to develop an innovative desalination model that causes less global warming and efficiently utilizes water resources. The new model mediates SVM algorithm by applying the reverse osmotic pressure and temperature to the current's salinity and density of the existing RO using the algorithms of bioinformatics and incorporating them into MSF, RO, and Genetic Algorithm (GA) of MED to develop an innovative technology. In this study, a new desalination model is applied to the separation process through a semi-permeable membrane, based on the experimental SVM algorithm. Using a semi-permeable membrane and osmotic pressure, the RO is applied to MSF and MED to obtain pure freshwater by removing the solutes dissolved in the current. Hence, this research is directed towards suggesting a model that can help create an effective semi-permeable membrane for use in desalination. At this stage, where a technology previously considered the best creates counter currents, owing to the change in salinity and the resulting change in density, and causes global warming, we expect to efficiently prevent the density change caused by the change in salinity and the counter current by applying the principle of temperature and osmotic pressure. In addition, we expect to suggest an innovative technology by using the SVM model that incorporates GA for safely obtaining the water resources.

2. RESEARCH METHOD

2.1. Genetic Programming Utilizing Genetic Algorithm

Predicting results while carrying out desalination under complex plant conditions could be difficult in practical terms, and predictions cannot be made through simple equations because many variables with nonlinear relationships are involved. Therefore, different equations need to be formulated for various methodological steps to predict the results. However, this is highly time-consuming and only experts can perform the task. Hence, computers were programmed to generate equations to predict the results using GA. It is similar to combining basic functions (logarithmic function, trigonometric function, etc.) and arithmetic operators (+, –, *, /). For this, computers should first be trained if the Database (DB) is instructed in its final stage, although plant data can improve the program so that it can create the appropriate equations by itself. Subsequently, the formulated equation is validated, which is a step confirming that the formulated equation is applicable to different data. Finally, after the successful formulation of the equation, the actual application is executed to predict the data of a complex plant. The model that predicts the transmittance can be used later for process control and predicting the results in the future. The DB employed in our study used the data obtained from a desalination plant with a capacity of 1,000 m^3/day, currentlyoperating in Gi-jang, Busan.

2.2. Modeling Approach for Applying SVM to the Database

The SVM is utilized in mapping that involves finding the hyperplane located the furthest from the other hyperplanes and separated from the given data. Unlike other algorithms, the SVM finds the hyperplane that separates the dots with maximum-margin, in addition to finding the hyperplane that separates the dots. Consequently, out of countless hyperplanes that separate the dots from the two classes, the SVM finds the one that can maximally maintain the distance between the classes.

2.3. Modeling Approach for Applying Decision Tree to the Database

Decision Tree is a typical approach in data-mining analysis that is used for classification of the given data. The results to be analyzed are displayed in the form of a tree, so that the analysts can easily understand and explain them, which is a significant advantage. The results, thus analyzed, can be used directly in decision-making.

3. RESULTS

3.1 Genetic Programming Applying GA

As most mathematical models were, parameters for the adaptable model were experimentally derived; in the first case, the variables *Lp* and *a* were measured at 1.31 L/m^2-h-bar and 1.66-105, respectively. In the second case, *Lp* and *a* were measured at 5.85-101 L/m^2-h-bar and 9.24-106, respectively (Figure 1).

Insert Figure 1

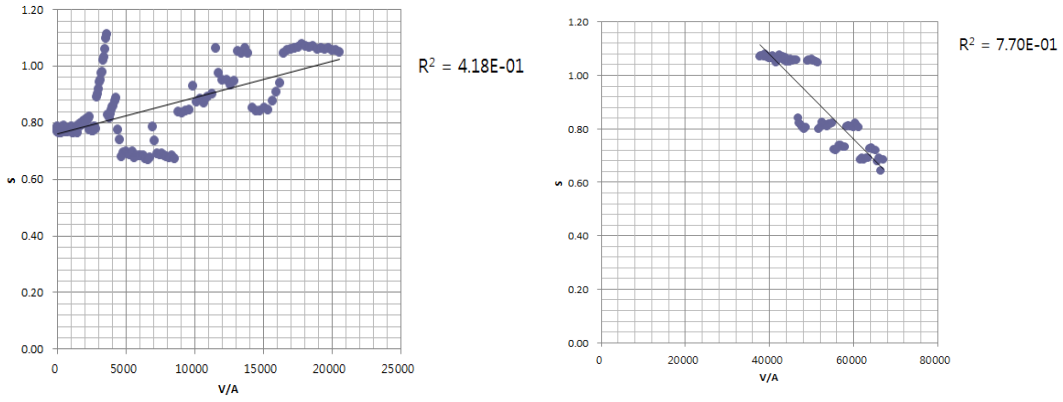


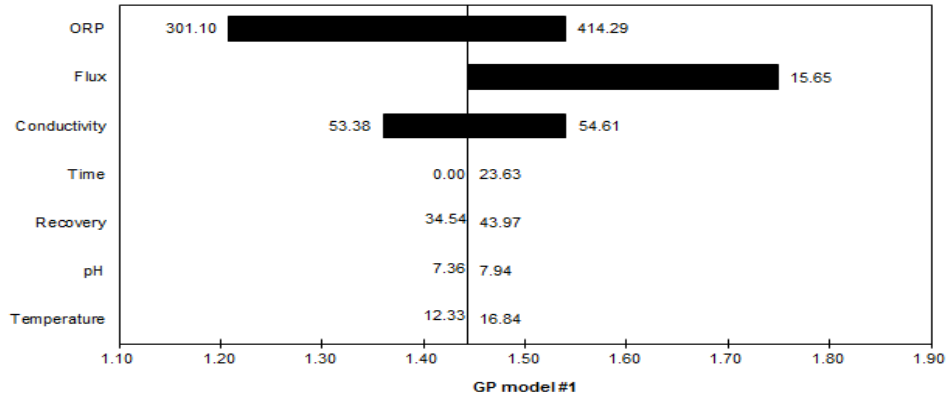
Figure 1. The relationship between VA and S (1) initial stage, (2) final stage

Looking at the reactivity of the Genetic Programming (GP) model to input variables, the GP model that adapted to the initial database had a high reactivity in productivity, influx, and ORP. On the other hand, GP model that adapted to the second database reacted to temperature,

regenerativity, time, and ORP (Figure 2). The mutual relationship between the difference in reactivity and input variables is not clear, but it is expected that future studies will utilize GP modeling to approach the information according to the plant operation.

Insert Figure 2

(1)



(2)

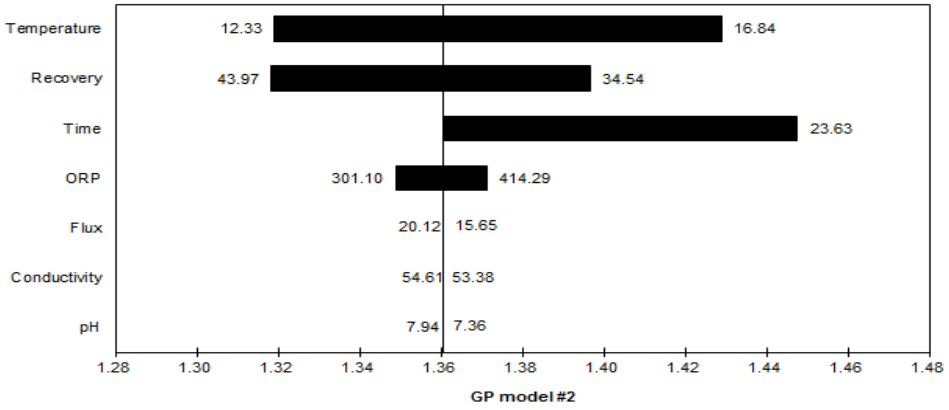


Figure 2. Analysis of the reactivity of GP models

(1) GP models adapted to the database in its initial stage
(2) GP models adapted to the final stage

3.2 Modeling Approach for Applying SVM to the Database

First, a training example from training, validation, and application, was used as the database, and the resulting value was derived through SVM parallel classification. Variables in each column used time, flux (water production rate of desalination plant), recovery rate, conductivity, pH, ORP, temperature, and transmittance of the membrane used in desalination. Using the first seven variables, the equation that predicts the eighth variable, the transmittance of the membrane, as the end product is the key point of the model. In the SVM algorithm, norm of longest example vector and |x|=420.66527 r_delta_sq =176959.27162 xisum=0.00000 asum=0.00000 were the key values for predicting the transmittance of the membrane (Figure 3).

3.3 Modeling Approach for Applying Decision Tree to the Database

From fold 0 to fold 8, there are 304 cases (Table 1); however, due to an extensive amount of data, a summary of the results is presented in Table 2.

Table 1. Experimental results of Decision Tree

Options: Rule-based classifiers Cross-validate using 10 folds Read 304 cases (7 attributes) from water_dat.data	
Rule 1: (3, lift 31.3)	pH >7.83 ; hwasi>15.16 -> class 1.7 [0.800]
Rule 2: (9/5, lift 24.9)	time <= 2.96; hwasi>14.69; hwasi<= 15.16 -> class 1.76 [0.455]
Rule 3: (2, lift 34.3)	time <= 4.63 flux >18.35 circle <= 41.85 pH <= 7.7; hwasi>14.21; hwasi<= 14.69 -> class 1.78 [0.750]
	. .
Default class: 1.32 Evaluation on hold-out data (30 cases): Rules ---------- No Errors 64 21(70.0%) <<	

Table 2. Decision Tree Summary

Fold	0	1	2	3	4	5	6	7	8	9	Mean	SE
Rules	64	61	62	61	62	63	70	60	65	63	**63.1**	**0.9**
Errors (%)	70.0	70.0	76.7	76.7	80.0	80.0	80.6	83.9	77.4	71.0	**76.6**	**1.5**

Decision Tree is a typical method used in data-mining analysis. We utilized the Decision Tree approach in order to obtain the maximum value for each category, and thereby finding the factor that influences the transmissibility of the semi-permeable membrane the most. For example, from Rule 8, (5, lift 47.0) time <= 4.63 flux >18.52 pH >7.7 pH <= 7.81 hwasi>14.21 hwasi<= 14.69-> class 1.79 [0.857]. When time, flux, pH, and temperature in degree Fahrenheit are above certain values, the probability of transmissibility of the semi-permeable membrane is 0.857, which

exceeds 80%. This can be a crucial factor for models that consider the transmission of the semi-permeable membrane.

Class	Cases	False Pos	False Neg
1.7	8	6	6
1.76	6	6	5
1.78	6	4	6
1.82	5	6	4
1.8	6	8	6
1.79	6	8	6
1.75	3	0	3
1.83	3	1	3
1.85	2	3	2
1.73	6	5	6
1.74	7	4	6
1.77	4	3	2
1.81	5	6	4
1.71	7	7	5
1.72	8	4	3
1.69	4	2	4
1.68	2	2	2
1.67	5	2	5
1.66	5	5	5
1.64	9	10	6
1.65	8	9	4
1.63	4	2	4
1.52	3	2	2
1.5	1	1	1
1.47	1	1	1
1.42	1	0	1
1.39	1	2	1
1.36	11	8	6
1.31	15	12	9
1.3	9	10	6
1.25	2	3	2
1.22	1	0	1
1.18	1	0	1
1.61	1	0	1
1.57	2	1	2
1.54	1	0	1
1.9	4	5	3
1.87	2	0	2
1.89	2	0	2
1.92	3	2	3

1.91	1	0	1
1.98	5	7	4
1.96	4	4	3
1.94	2	3	2
1.95	1	0	1
2	2	0	2
2.02	4	5	2
1.99	3	6	3
1.49	1	3	1
1.59	1	1	1
1.56	1	0	1
1.6	1	0	1
1.55	1	1	1
1.28	6	0	2
1.38	3	0	3
1.41	2	2	1
1.4	1	0	1
1.44	1	1	1
1.27	6	4	2
1.51	1	0	1
1.46	1	0	1
1.29	5	1	4
1.32	18	11	11
1.33	18	12	13
1.34	8	7	7
1.35	13	12	9
1.37	6	3	4
1.86	1	0	1
1.88	1	0	1
2.01	1	0	1
1.93	1	0	1
2.1	1	0	1
2.05	1	0	1
2.2	1	0	1
2.06	1	0	1

4. DISCUSSION

As the results of experiments concerned with Genetic Algorithm(GA), optimizing the unknown function Y=f(x), Support Vector Machine(SVM) with the furthest hyperplane separating the given data, and decision tree used in data-mining analysis turned out, it enabled us to propose semi-permeable membrane model used in desalination that are the most effective in preventing the reverse current flow and eventually to the global climate change. To explain the background, the recent technology of desalination is improving more and more, membrane itself during the desalination has a significant problem: It causes the weakened ocean current to further on reverse

direction and then causes global warming. According to three experiments, seven variables including flux, recovery rate, conductivity, PH, ORP, temperature, and transmittance of the membrane influence the transmittance of the semi-permeable membrane used. It was demonstrated by Decision Tree that this GP model, which incorporated SVM for experiment, could prove to be a better tool for making the RO more efficient, while being exposed to changing variables, than any other mathematical or mechanical model. In addition, being directly involved with the semi-permeable membrane, it is more suitable than the pressure of the membrane. This research not only contributes to the development of scientific desalination technology but also contributes to the international society in every scientific and humanitarian section because it states more efficient solution to Global warming, which is one of the significant problems globally.

5. CONCLUSION

RO is the most commonly employed method of desalination. It produces freshwater by transmitting seawater through a semi-permeable membrane while counter-utilizing osmosis for desalination. To suggest a more efficient and current-oriented model for RO using a semi-permeable membrane, three methods were used: GA, SVM classification, and Decision Tree.

It can be concluded that GA optimizes the unknown function Y = f(x) by finding x and building an algorithm of simulated evolution. SVM is the method used in mapping that finds the furthest hyperplane that separates the given data. Decision Tree is a typical method used in the data-mining analysis, which displays the results in the form of a tree, and therefore, analysts can easily understand and explain them, which is a significant advantage. These three experiments enabled us to propose models that are more effective.First, in the experiment where the GA was incorporated into the GP, variables *Lp* and *a* were measured in the adaptable model, to determine parameters that are based on the experiments in most of the mathematical models, at 1.31 L/m^2-h-bar and 1.66-105, respectively. For other parameters, the variables were measured at 5.85-101L/m^2-h-bar and 9.24-106, respectively. In addition, observations regarding the GP model indicated that there was a high reactivity in productivity, influx, and ORP in the initial database, while in the other database, temperature, regenerativity, time and ORP caused high reactivity. Second, in the experiment that applied the SVM algorithm, norm of longest example vector and |x|=420.66527 r_delta_sq =176959.27162 xisum=0.00000 asum=0.00000 result were the key values for predicting the transmittance of the membrane. Third, it was inferred through the Decision Tree that when time, flux, pH, and temperature in Fahrenheit are above certain values, the probability of transmittance is over 80%. It should be noted here that this could have a major impact on models that consider the transmission of the semi-permeable membrane.To emphasize the key point of this research, the three experiments and the effects of the variables demonstrate the change in counter current caused by the use and change of these input variables while carrying out desalination. To strengthen the current, flux, recovery rate, conductivity, pH, ORP, temperature, and transmittance of the membrane used for desalination should be carefully observed, as shown in the Decision Tree. These seven variables influence the transmittance of the semi-permeable membrane used in desalination; therefore, the values practiced in Decision Tree, i.e., the rule of the folds that are over 80% (fold4 (80%), fold5 (80%), fold6 (80.6%), and fold7 (83.9%)) should be followed for current-strengthening. The final value of time > 19.42, flux > 15.82, circle <= 34.98 and hwasi temperature <= 13.65 suggest the standard value of the reverse osmosis membrane model that we developed.It was demonstrated by Decision Tree that this GP model, which incorporated SVM for experiment, could prove to be a better tool for making the

RO more efficient, while being exposed to changing variables, than any other mathematical or mechanical model. In addition, being directly involved with the semi-permeable membrane, it is more suitable than the pressure of the membrane. Therefore, compared to all the new technologies currently under development, this model can most effectively create a current-oriented semi-permeable membrane.

REFERENCES

[1] World Water Forum. WPI (Water Poverty Index) (2006)

[2] Kim, Jang, Kim, Choi, Lee and Kim. Optimal Operation Condition Deduction of Pressurized MF process as desalination treatment, Daehan Environmental Engineering Journal 35(9) 624-629 1225-5025 KCI

[3] Addison Wesley. Genetic algorithms in search, optimization and machine learning, ISBN 978-0-201-15767-3

[4] Seung-Min Park, Jihee Han a, Sangho Lee a, JinsikSohn a, Young-Min Kim b, June-Seok, Choi b&Suhan Kim c, Analysis of reverse osmosis system performance using a genetic programming technique, Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK

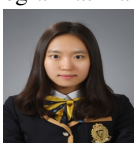
[5] Cristianini, Nello; and Shawe-Taylor, John; An Introduction to Support Vector Machines and other kernel-based learning methods, Cambridge University Press, 2000. ISBN 0-521-78019-5 ([1] SVM Book)

[6] Y. Yuan and M.J. Shaw, Induction of fuzzy decision trees. Fuzzy Sets and Systems 69 (1995), pp. 125–139

Authors

Chaeyun Jung

was born in Korea, in 1997. He is a junior, eleventh grade student in the international program at Hankuk Academy of Foreign studies since 2014. She graduated from Busan international middle school in 2013. She is the intern researcher of biological new medicine government institution lab,Dajeon and desalination-related gookmin university lab in Seoul. She has attended the international conference on Engineering and Applied Science(ICEAS), held in Osaka, with the paper of 'Analyzing ECG segmentation and rule of patterns by using Viterbi algorithm (HMMs)'



Seoyeon Shin

was born in Republic of Korea, in 1998. He is a tenth-grade student in the international program at Hankuk Academy of Foreign Studies.

Taeseon Yoon

was born in Seoul, Korea, in 1972. He was Ph.D. Candidate degree in Computer education from the Korea University, Seoul, Korea, in 2003. From 1998 to 2003, he was with EJB analyst and SCJP. From 2003 to 2004, he joined the Department of Computer Education, University of Korea, as a Lecturer and Ansan University, as a Adjunct professor. Since December 2004, he has been with the Hankuk Academy of Foreign Studies, where he was a Computer Science and Statistics Teacher. He was the recipient of the Best Teacher Award of the Science Conference, Gyeonggi-do, Korea, 2013