Application of Genetic Programming as a Powerful Tool for Modeling of Cellulose Acetate Membrane Preparation

Ehsan Saljoughi

Abstract— In the present study, genetic programming (GP) as a completely different approach in comparison with conventional methods, based on the imitation of natural evolution of living organism is proposed for the prediction of cellulose acetate (CA) polymeric membrane characteristics. The membrane preparation parameters of polymer concentration, additive concentration and coagulation bath temperature (CBT) were considered as model input parameters. GP model was developed to predict the pure water permeate flux of CA membranes based on the experimental data of a membrane pilot-scale system. Some statistical parameters were considered and calculated to investigate the reliability and accuracy of the proposed model. The model excellent general performance with $R^2=0.98$ and the low error values confirmed the accuracy of the model. According to the results, GP has displayed the potential to be used as a reliable tool for identifying the characteristics of CA membranes.

Keywords: cellulose acetate, genetic programming, membrane preparation, modeling

I. INTRODUCTION

Membrane separation processes present a number of important advantages such as low energy consumption, compact equipment, and less environmental impacts and capital investments. Thus, membranes have obtained an important place in chemical industries and have been used in a wide range of applications. Removal or recovery of toxic or valuable components, such as organic solvents and acids, from various industrial effluents; the production of high-quality water; and applications in the pharmaceutical and food industries are some examples of membrane technology applications in the chemical industries [1–4].

Among various membrane materials, polymers are the most widely used materials for membrane preparation owing to their proper mechanical properties, intrinsic transport property, flexibility in being formed into different shapes, and low cost [5]. Among different polymeric materials, cellulose acetate (CA) presents promising characteristics such as good toughness, noticeable biocompatibility, good desalting, low cost, and high potential flux. Thus it has been widely used in various membrane separation processes such as gas separation (GS), hemodialysis, reverse osmosis (RO), ultrafiltration (UF), and microfiltration (MF). Moreover, hydrophilic characteristic of CA membranes reduces fouling during a separation process [6-9].

There are different ways for the preparation of porous polymeric membranes including sintering, stretching, track etching, solution casting, and phase inversion [10]. Among these methods, phase inversion is the most extensively used technique for the preparation of asymmetric membranes. In this process, a homogeneous solution, which contains polymer, solvent and suitable additives, is cast onto a suitable support by using a casting knife, and then the support is immersed into a coagulation bath which commonly contains distilled water as a non-solvent. Subsequently, phase inversion occurs by exchange of solvent and non-solvent across the interface between the casting solution and the coagulation bath [11].

For a successful implementation of the membrane separation process it is necessary to obtain some information about the characteristics of a given membrane. On this basis, membrane preparation modeling based on fabrication parameters plays an important role in the characterization of a polymeric membrane and provides useful information about it. Mathematical models are usually complicated and do not provide sufficient information about the physical phenomena taking place. Moreover, such models are typically valid for specific conditions [12,13].

Artificial neural network (ANN) has shown the ability to model highly complex and non-linear systems without requiring detailed information about the physical parameters of the system. Madaeni et al. [14] and Tan et al. [15] successfully implemented ANNs to predict membrane characteristics and investigated the effects of different preparation parameters. Although ANNs have demonstrated to be an effective predictive tool for modeling of non-linear systems, its output for a specific system usually cannot be extended and used for other systems due to the specific ANN model structure and the dependence of the model to the initial assumptions of the parameters [16].

Therefore, alternative methods for predicting the characteristics of polymeric membranes by using available preparation data and extending it to a mathematical model, which can be applied simply on unavailable data, are so worthwhile. Genetic programming (GP), a branch of soft computing, is such a tool that can be widely used in various problems [17,18].

Because of the high degree of complexity and the high number of parameters that should be considered in prediction of membrane characteristics related to

E. Saljoughi is with the Chemical Engineering Department, Faculty of Engineering, Ferdowsi University of Mashhad, Mashhad, Iran. Correspondence should be addressed to E. Saljoughi (e-mail: saljoughi@um.ac.ir)
fabrication parameters, it seems that GP is a potential tool for reliable and accurate prediction of membrane characteristics, including CA membranes. This paper utilizes GP as a novel way to develop a mathematical model for CA membrane characteristics in terms of pure water permeate flux as a function of CA and the additive (polyethylene glycol) concentrations and coagulation bath temperature (CBT), based on the experimental preparation parameters. The goal is to exploit GP flexible tree structure to build a prediction model for characterizing CA polymeric membranes.

II. OVERVIEW OF GENETIC PROGRAMMING

Genetic algorithm (GA) is an optimization technique based on the principles of genetics and natural selection [19]. This technique allows a population composed of many individuals to evolve under specified selection rules to minimize the cost function. GP is an extension to conventional GA proposed by Koza [20] to solve a given problem by automatically generating algorithms and expressions on the basis of natural evolution. Unlike GA which uses a string of numbers to represent the solution, GP automatically creates several computer programs with a parse tree structure to solve the problem at hand. These programs are coded as a tree structure with its nodes (functions) and terminals (leaves) [20]. The chosen function set is \{+, -, *, /, sin, cos, exp, log, abs\}. The function set may also include other mathematical functions, conditional operators or any user-defined operators [21]. The chosen terminal set contains the independent variable X or an integer. The GP model is explicit and free from conceptual designs, thus it has few requirements about the domain knowledge of the problem and is therefore less problem-dependent [22,23]. The main advantage of GP is that it does not require any prior knowledge about the relations among parameters and their interactions. Fig. 1 shows a brief flowchart of GP. As experimental data, input parameters, introduced to the program, the GP starts to random creation of the initial GP model populations and evaluates their fitness value quantifying how well the created models solve the problem. New generations of programs are iteratively created by selecting parents based on their fitness andbreeding them via genetic operators including crossover, mutation and reproduction. Genetic operators introduce variability in the individuals and make evolution possible, which may produce better individuals in posterior generations. The crossover operator exchanges sub-trees from a pair of individuals, generating two offspring. The mutation operator replaces a randomly chosen sub-tree from an individual by a new sub-tree randomly generated. The reproduction operator simply copies individuals and inserts them in the next generation.

Because of selecting better individuals and passing their best characteristics to their offspring, the population tends to improve in quality along successive generations. This evolutionary process continues until a termination criterion, e.g. satisfactory fitness or maximum number of generation, is verified. At last, the best-so-far individual is designated as the output of the GP operation.

![Genetic Programming Flowchart](image)

Fig. 1. Schematic flowchart of genetic programming (GP) procedure.

III. EXPERIMENTAL

A. Materials

CA with an average molecular weight of 52,000 g mol\(^{-1}\) was purchased from Fluka for the preparation of CA polymeric membrane. 1-Methyl-2-pyrroldione (NMP) with analytical purity of 99.5% (Merck) and distilled water were used as solvent and non-solvent agents, respectively. Polyethylene glycol (PEG) with an average molecular weight of 400 g mol\(^{-1}\) was obtained from Loba Chemie Ltd and used as an additive.

B. Membrane Preparation

Various CA/PEG/NMP polymeric solutions were prepared with different CA concentrations ranging from 13.5 wt% to 17.5 wt% and PEG concentrations ranging from 0 wt% to 10 wt% at a coagulation bath temperature (CBT) of 0 and 25 °C. The details of the experimental procedure are provided in our previous work [4]. In brief, PEG was added to the homogeneous solution of CA in NMP, and was mixed by stirring for 12 h at room temperature. The prepared homogeneous solutions were cast using a film applicator with 180 µm clearance gap on a glass plate substrate. It was then moved to the distilled water coagulation bath for immersion precipitation. After primarily phase separation and membrane formation, the membrane was stored in water for 24 h to guarantee the complete phase separation. This allowed the water soluble components of the membrane to leach out.

C. Solution Viscosity Measurements

The viscosity of the prepared casting solution mixtures was measured using a Polyvisc digital rheometer (Model VISCO Star L) at a constant temperature of 25 °C.
D. Scanning Electron Microscopy (SEM)

The membranes were snapped under liquid nitrogen to give a generally consistent and clean break. The membranes were then sputter coated with thin film of gold. The membranes were mounted on a brass plate using double-sided adhesion tape in a lateral position. Cross sectional images of the membranes were obtained using Cam Scan SEM Model MV2300 microscope.

E. Experimental Setup

Permeation flux studies were carried out using a cross-flow filtration setup as shown schematically in Fig. 2. This apparatus consists of a UF disk module, a feed tank, a pump, pressure gauges installed on the feed and retentate pipes connected to the module, and some valves. Additionally, a bypass line was considered for the pump exit stream. A flat sheet membrane module made from stainless steel was used in all experiments. Effective filtration area of the membrane in the module was 24 cm². Permeate water flux (PWF) experiments were done at a transmembrane pressure of 0.2 bar and their values were calculated using the following equation:

\[ \text{Flux} = \frac{Q}{A \Delta t} \]  

where \( Q \) is the quantity of permeate (L), \( A \) is the effective membrane area (m²) and \( \Delta t \) is the sampling time (h) [4, 24].

![Fig. 2. Schematic diagram of the experimental set-up.](image)

The experiments were carried out in batch mode in which both permeate and retentate were returned to feed reservoir to keep a constant concentration. The feed reservoir temperature was maintained at approximately 30 °C.

IV. MODELING

The first population of mathematical models that simulated the investigated system in the first generation of GP was created randomly. Each model in the initial population consists of some nodes and terminals and represents a mathematical function. In this study, CA concentration, PEG concentration and CBT were considered as independent variables and permeate water flux of CA membrane was considered as a function of them. Furthermore, 20% of the experimental data were used as test data for the evaluation of the performance of each model.

The search space of GP is virtually unlimited and programs tend to grow in size during the evolutionary process. Code growth is a healthy result of genetic operators in the search of better solutions, but it also permits the appearance of pieces of redundant codes that increase the size of the programs without improving their fitness; a phenomena known as bloat. Bloating can cause stagnation in the evolutionary process [25]. Bloating can be defined as an increase in the individual growth without improvement in their fitness. In order to prevent bloating, maximum depth of individuals was considered 16 and code growth by crossover and mutation operators was acceptable just if the fitness of an individual was improved.

Fitness function evaluates the sum of absolute difference between the expected output values and the value returned by the individuals. In the GP model development, input and output variables should be normalized first to match the unit levels of individual factors [23].

The evolutionary processes were controlled by evolutionary parameters of: population size 250, maximum number of generations to be run 50, probability of crossover 0.5, probability of mutation 0.4, probability of reproduction 0.1, maximum depth for initial random individuals 8, maximum depth of mutation fragment 8, and maximum permissible depth of individuals after crossover 16. The initial populations were generated with the Ramped Half-and-Half procedure [20]. Moreover, the maximum number of generations was considered as termination criteria in each run.

Some statistical parameters were obtained for better evaluation of the model. These statistical parameters are Normalized Bias (NB), Standard Square Error (SSE), Mean Square Error (MSE), Root Mean Square Error (RMSE), and R-square (R²) which are described in Eqs. (2)–(6). They were used for the fitness investigation and error determination of the model.

\[
R^2 = \frac{\sum_{i=1}^{n} (Y_{\text{exp},i} - Y_{\text{model mean}})^2}{\sum_{i=1}^{n} (Y_{\text{exp},i} - \bar{Y}_{\text{exp}})^2} - \frac{\sum_{i=1}^{n} (Y_{\text{model},i} - \bar{Y}_{\text{exp}})^2}{\sum_{i=1}^{n} (Y_{\text{model},i} - Y_{\text{model mean}})^2} \tag{2}
\]

\[
\text{SSE} = \sum_{i=1}^{n} (Y_{\text{model},i} - Y_{\text{exp},i})^2 \tag{3}
\]

\[
\text{MSE} = \frac{\sum_{i=1}^{n} (Y_{\text{model},i} - Y_{\text{exp},i})^2}{n} \tag{4}
\]
where $Y_{\text{exp},i}$ is the experimental value of permeate water flux, $Y_{\text{model},i}$ is the value of permeate water flux predicted by the model, $Y_{\text{model,mean}}$ is the average value of the model permeate water flux prediction, and $n$ is the number of experiments.

V. RESULTS AND DISCUSSION

Some experiments were carried out in order to investigate the effect of CA concentration, PEG additive concentration, and coagulation bath temperature on the pure water permeation flux of the resultant CA membrane. The experimental results are shown in Fig. 3. As is observed, higher coagulation bath temperature causes more water permeation flux. This is more obvious in CA concentration of 13.5 wt%, where the water permeation flux is increased by more than two times via increasing the CBT from 0 to 25 °C. However, by increasing the CA concentration to 15.5 wt% and 17.5 wt%, this behavior becomes less significant. In all CA concentrations and CBTs, an increase in PEG concentration results in higher water permeation flux. In addition, at different levels of PEG concentration and CBT, the PWF (vertical axes in Fig. 3) tends to higher values regarding the membranes prepared from casting solutions containing less CA concentrations. Better presentation of above trend is observable in Table I. With respect to Fig. 3, in all CA levels, the flux is increased with increase in PEG concentration. This effect is more obvious in lower levels of CA.

<table>
<thead>
<tr>
<th>TABLE I</th>
<th>EFFECT OF CA CONCENTRATION ON THE PURE WATER FLUX (PWF) OF THE PREPARED MEMBRANES</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA concentration (wt%)</td>
<td>CBT (°C)</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>13.5</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>25</td>
</tr>
<tr>
<td>15.5</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>25</td>
</tr>
<tr>
<td>17.5</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>25</td>
</tr>
</tbody>
</table>

*The average PWF of the membranes prepared from different PEG concentrations

Cross-sectional SEM images of the selected membranes presented in Fig. 4 confirm the above explanation regarding the effect of the polymer and PEG concentration as well as coagulation bath temperature on the water permeation flux of the prepared membranes. As is observed, an increase in CBT and PEG concentration along with a decrease in CA concentration result in the formation of membranes with higher porosity. It is evident that the higher porosity of the prepared membranes reduces the resistance against water permeation flux [26].

![Fig. 3. PWF as a function of CBT and PEG concentration at: a) CA=13.5 wt%, b) CA=15.5 wt%, and c) CA=17.5 wt%.](image)

The CA and PEG concentrations along with the CBT were considered as independent variables to obtain a model for permeation flux of water by using GP. In the GP model development, input and output variables were normalized before running of the program by Eq. (7) in order to match the unit levels of individual factors. For simplicity, $X_{\text{min}}$ for all variables was set to zero. Indeed, its value hadn’t any effects on final results [23].

\[
X_{\text{Normalized}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}}
\]
SALIOUGHI: APPLICATION OF GENETIC PROGRAMMING AS A POWERFUL TOOL …

80% of the available experimental data was used for developing of GP model and remaining was used for validation of the developed model.

The execution time of running the evolutionary algorithm for this case was approximately 18 min on a Dell server (Core i3, RAM 2GB, Windows Seven). Fig. 3 illustrates the prediction of water fluxes using genetic programming, and a comparison is done between experimental data and results obtained from the GP model. As it can be clearly seen, there is very good agreement between the model and the experimental results. In other words, the model predicts the system behaviors well. However, in CA concentration of 17.5 wt%, some deviations are observed between the experimental and GP model results. This can be attributed to and interpreted by the noticeable increase in viscosity. The viscosity of the polymeric solutions is presented in Table II. As is observed, only a 2 wt% increase in CA concentration from 15.5 wt% to 17.5 wt% results in a twofold increase in viscosity which noticeably affects the membrane formation mechanism in the coagulation bath. The noticeable increase in the viscosity of the cast film slows the diffusional exchange rate of the solvent (NMP) and nonsolvent (water) during the solidification process and consequently results in formation of a denser structure with less PWF [7]. In other words, the reduction of agreement between the model and the experimental results in the membranes prepared from polymeric solutions containing 17.5 wt% CA can be attributed to the effect of the solution high viscosity which noticeably affects membrane morphology and permeability.

The best-so-far GP model (model-1) that was obtained after satisfying termination criterion is:

$$Y = \cos \left( \cos \sin X_2 \times X_1^2 \right) \times \left( X_1 + 2X_3 \right) \times \left\{ \log \left( X_1 X_2^2 + \cos X_2 \right) + X_2 + X_3 + \cos \sin X_2 \right\} \times \left( \cos X_1 \right)^6$$

where, $Y$, $X_1$, $X_2$, and $X_3$ denote the normalized water permeation flux, the normalized CA concentration, the normalized coagulation bath temperature, and the normalized PEG concentration, respectively.

### Table II

<table>
<thead>
<tr>
<th>solution composition</th>
<th>viscosity (cP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA (wt%)</td>
<td>NMP (wt%)</td>
</tr>
<tr>
<td>15.5</td>
<td>84.5</td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
</tr>
<tr>
<td>17.5</td>
<td>82.5</td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 5 shows the fitness value, the node numbers, and the depth of the best-so-far individual in each generation. With respect to this figure, GP provides more accurate predictions by increasing the node numbers and depth, which result from an increase in the degree of freedom of the individuals.
In order to investigate the reliability of the proposed model, some statistical parameters, Eqs. (2)–(6), which demonstrate the model performance in prediction of water permeation flux of the CA polymeric membranes, are presented in Table III.

<table>
<thead>
<tr>
<th>statistical parameter</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0.97807</td>
</tr>
<tr>
<td>SSE</td>
<td>$2.62 \times 10^{-2}$</td>
</tr>
<tr>
<td>MSE</td>
<td>$1.46 \times 10^{-3}$</td>
</tr>
<tr>
<td>RMSE</td>
<td>$3.82 \times 10^{-2}$</td>
</tr>
<tr>
<td>NB%</td>
<td>1.28695</td>
</tr>
</tbody>
</table>

The $R$-square (coefficient of fitting) value confirms that the GP model results are very well fitted to the experimental data. The values of standard square error, mean square error, and root mean square error in Table III indicate some insignificant errors of the model. With regard to NB value, the GP model slightly over predicts the water permeation flux.

All data presented in Table III confirm the accuracy of the proposed GP model for the prediction of pure water permeation flux of the CA membranes.

VI. CONCLUSIONS

In this study, the ability of GP approach in estimation of CA membrane characteristics in terms of pure water permeation flux as a function of CA concentration, PEG concentration and CBT was investigated. An important benefit of GP modeling is that the model is explicit and does not require underlying description of the physical process. The proposed GP model is an empirical model based on the parameters that influence preparation of the CA membranes and consequently affect performance of the prepared membranes in a pilot-scale system.

The experimental data were used for model development and validation. The performance of the GP model was evaluated by comparing the predictions with the experimental results. The proposed model showed a very good general performance with strong correlation. GP displayed a promising tool for the prediction of CA polymeric membrane characteristics. This study is evidence for great potential of GP in obtaining models for complex and high non-linear approximation problems.

REFERENCES