Neural Comput & Applic (2002)10:311–317

Neural
Computing
& Applications

Application of a Recurrent Neural Network to Prediction of Drug Dissolution Profiles

W. Y. Goh[1], C. P. Lim[1], K. K. Peh[2] and K. Subari[1]

[1]School of Electrical & Electronic Engineering; [2]School of Pharmaceutical Sciences, Universiti Sains Malaysia, Penang, Malaysia

The Elman Recurrent Neural Network was employed for the prediction of in-vitro *dissolution profiles of matrix controlled release theophylline pellet preparation, leading to the potential use of an intelligent learning system in the development of pharmaceutical products with desired drug release characteristics. A total of six different formulations containing various matrix ratios of substance to control the release rate of theophylline were used for experimentation. By using the leave-one-out cross-validation approach, the dissolution profiles of all the matrix ratios were consumed for training, except for one set that was taken as a reference profile, with which the network predicted profiles were compared. Performance of the network was assessed using the similarity factor, f_2, a criterion for dissolution profile comparison recommended by the United States Food and Drug Administration. Simulation results indicated that the Elman network was capable of predicting dissolution profiles that were similar to the reference profiles with an error of less than 8%. In addition, the Bootstrap method was used to estimate the confidence intervals of the f_2 values. The results revealed the potential of a neural-network-based intelligent system in solving nonlinear time-series prediction problems in pharmaceutical product development.*

Keywords: Bootstrap confidence interval; Elman network; Pharmaceutical product formulation; Prediction of drug dissolution profile; Recurrent Neural Networks; Similarity factor

Correspondence and offprint requests to: Dr C. P. Lim, School of Electrical & Electronic Engineering, Universiti Sains Malaysia, Engineering Campus, 14300 Nibong Tebal, Penang, Malaysia. Email: cplim@usm.my

1. Introduction

Artificial Neural Networks (ANNs) are simplified models of the central nervous system. They are networks of highly interconnected neural computing elements that have the ability to respond to input stimuli, and to learn to adapt to the environment. ANNs have been shown to be effective in handling various tasks, including pattern recognition and classification, modelling and forecasting, adaptive control, multisensor data fusion and noise filtering [1].

Lately, ANNs have been successfully applied as a problem-solving tool in the pharmaceutical industry [2–6]. The most widely used ANN architecture has been the Multi-Layer Perceptron (MLP) network [1]. Empirically, mathematical models are used to represent behaviours and dynamics between various interacting components in many pharmaceutical processes. ANNs can then be employed to predict the parameters of mathematical models that fit the behaviours of certain processes. For instance, the MLP network was applied [3] to analyse the quantitative relationship between several formulation factors and release parameters in a hydrophilic matrix capsule system containing cellulose polymers. In Takayama et al. [6], the applicability of the MLP network in the optimisation of pharmaceutical formulae for ketoprofen hydrogel ointment was demonstrated. In these cases, the MLP network was used, in general, to estimate the parameters of mathematical models that characterise the pharmaceutical formulae/processes.

In our work, a recurrent ANN model has been chosen as an alternative to a model-based approach for the prediction of drug dissolution profiles. However, instead of estimating parameters that fit models

of the dissolution profile as conducted by other researchers, we have treated the entire dissolution profile as a time-series curve. Each time point has been used as a dependent feature in which information contained in one time point affects further predictions, subject to subsequent inputs.

Among ANN architectures, recurrent networks are useful for storing information about time, and are particularly suitable for time series prediction [7]. The specialty in recurrent networks includes allowing internal network feedback. One salient property of the Elman recurrent network [8] is that the hidden unit activation functions (internal states) are fed back at every time step to provide an additional input in conjunction with other input features. This recurrence gives the network dynamic properties, which make it possible for the network to have internal memory. As a result, they can perform mappings that are functions of time.

In the present study, we have applied the Elman recurrent network to predict the dissolution profiles of a matrix controlled release theophylline pellet preparation. The similarity between the predicted and reference dissolution profiles was assessed using the similarity factor, f_2, a criterion recommended by the United States Food and Drug Administration (FDA) [9] for comparison between two dissolution profiles. In addition, a statistical evaluation of f_2 was conducted, in which the 95% confidence intervals for mean of the similarity factors were calculated using the bootstrap method [10], a popular data-resampling method in statistical analysis.

2. The Elman Network

The Elman network is a simple recurrent network that only involves partial feedback in its architecture, as depicted in Fig. 1. The outputs of the hidden layer are allowed to feed back to the context layer, and to augment additional units at the input level. The sequential input is processed in two time steps. During the first step at time $t-1$, the input units receive the first input vector from the input sequence. At the first time, the context units are set to 0. Both the input units and context units activate the hidden units. Since the context units are in the initial state, only the input units contribute to the activation of the hidden units at time $t-1$. The hidden units are then fed forward to activate the output units and, at the same time, fed back to activate the context units on the second step at time t. Now, the context units contain the exact values of those of the hidden units. At the next time step $t+1$, information in the context units is combined with the input units that receive the new input vector to activate the hidden units. The hidden units then activate the output units, as well as the context units at time $t+2$. The above sequence is repeated at the next time step. Thus, these context units provide the network with information that is recurrent in time, which is a desirable characteristic for handling time-series prediction problems.

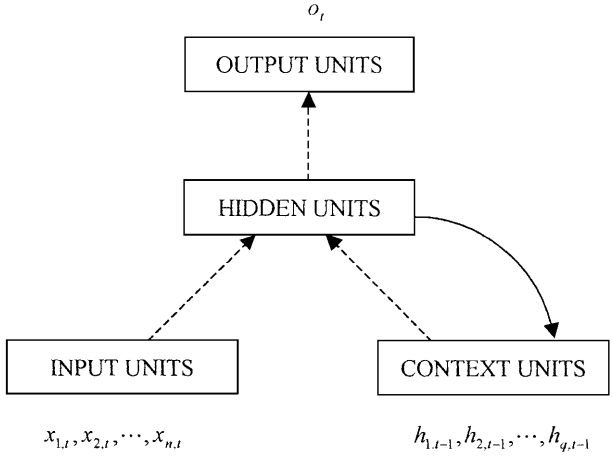

Fig. 1. Architecture of the Elman network.

2.1. Dynamics of the Elman Network

Given an input vector at time t, $\boldsymbol{x}_t = (x_{1t}, \cdots, x_{nt})$, the activation functions of the hidden and output layers of the Elman network with q hidden units can be represented, respectively, as follows:

$$h_{i,t} = \Psi\left(\gamma_{i0} + \sum_{j=1}^{n} \gamma_{ij} x_{j,t} + \sum_{l=1}^{q} \delta_{il} h_{l,t-1}\right) \quad (1)$$

$$= \Psi(net\ 1) =: \psi_i(x_t, h_{t-1}, \theta) \quad i = 1, \cdots, q$$

$$o_t = \Phi\left(\beta_0 + \sum_{i=1}^{q} \beta_i \psi_i(x_t, h_{t-1}, \theta)\right) \quad (2)$$

$$= \Phi(\text{net } 2)$$

where $h_{i,t}$ is the output of the i hidden unit, o_t is the output estimation of the target variable, θ is the vector of parameters containing all γ's and δ's, and γ's, δ's, β's are the weights from the input layer to the hidden layer, the context layer to the hidden layer, and the hidden layer to the output layer, respectively. The activation functions $\Psi(\cdot)$ and $\Phi(\cdot)$ in Eqs (1) and (2) are the hyperbolic tangent sigmoid function and log-sigmoid function, respectively, as follows:

$$\Psi = \frac{2}{1 + e^{-2(net1)}} - 1 \quad (3)$$

$$\Phi = \frac{1}{1 + e^{-(net2)}} \quad (4)$$

The recurrent connections from the hidden layer to the context layer are fixed at 1.0, but other connections in the network are adjustable by comparing with a target and updating the connections using back-propagation of an error [8]. Batch training is employed in which weights and biases are only updated after all of the inputs and targets have been presented. At each epoch, the entire input sequence is presented to the network, and its outputs are calculated and compared with the target sequence to generate an error sequence. For each time step, the error is back-propagated to find gradients of error for each weight and bias. In the present study, the mean squared error performance criterion is used, and weights are updated using gradient descent with an adaptive learning rate [11]. During the training process, the optimal learning rate changes as the algorithm moves across the performance surface. A proper setting of the learning rate can affect the performance of the algorithm: if the learning rate is too large, the algorithm may oscillate and become unstable; if the learning rate is too small, the algorithm will take a long time to converge. Thus, an adaptive learning rate will attempt to keep the learning step size as large as possible while keeping learning stable [11].

3. In-vitro Dissolution Profile Prediction

3.1. Materials and Preparations

Theophylline is a widely used drug in the management of asthma. However, it has a relatively narrow therapeutic index [12]. One of the reliable methods in sustaining the rate of drug release from the dosage form is by embedding the drug in nonsoluble matrix materials, which can produce more uniform serum concentrations with less fluctuation in peak-trough levels. In the previous study by Peh and Yuen [12], a novel multiparticulate matrix controlled release preparation of theophylline was formulated and evaluated in-vitro. In the preparation of the theophylline pellets, nonsoluble matrix materials microcrystalline cellulose (MCC) and glyceryl monostearate (GMS) were used in retarding the rate of drug release. In accordance with the physical experiments conducted [12], a series of formulations containing a constant proportion of theophylline, but different proportions of MCC and GMS, were prepared at ratios of 10:10:0, 10:8:2, 10:7:3, 10:6:4, 10:5:5 and 10:4:6.

The recurrent network analysis was performed with the MATLAB Neural Network Toolbox [11] package. Three inputs were used in the experiment. Since the proportion of theophylline was held constant, it was excluded as an input to the Elman network. Hence, the first and second inputs were the different proportions of microcrystalline cellulose (MCC) and glyceryl monostearate (GMS) at ratios of 10:0, 8:2, 7:3, 6:4, 5:5 and 4:6. The third input was the time points of the dissolution profile. There was a total of 11 time points (0, 0.25, 0.5, 1.0, 1.5, 2.0, 3.0, 4.0, 6.0, 8.0 and 10.0 hours) to form a complete profile. On the other hand, rates of drug release were supplied as target outputs of the network.

In determining the optimum network configuration, the procedure of cross-validation can be adopted for model selection by choosing one of several models that has the smallest estimated error function [13]. In k-fold cross-validation, the training data is divided at random into k subsets of (approximately) equal size. The network is trained k times, each time leaving out one of the subsets from training, and using the omitted subset to test its performance by evaluating the error function. In our experiments, since the data samples available are limited, we set k equal to the sample size, hence the 'leave-one-out' cross-validation training method. There was a total of 36 sets of data samples, as the dissolution studies were conducted six times for each of the six matrix ratios. All of the dissolution data were used for training (30 sets of samples), while the remaining six sets of samples containing the same proportion of MCC and GMS were used to compute the mean profile, which we refer to as the reference profile of the test set. Six dissolution profiles were predicted from the Elman network for each matrix ratio, with different random initialisation of the weights and biases of the network. Then, each profile was used to compare with the corresponding reference profile in order to assess the performance of the network.

3.2 Performance Measurement

Comparison between the predicted profiles from the Elman network and the reference dissolution profiles from the physical experiments were assessed using the similarity factor f_2. The equation of f_2 is as follows:

$$f_2 = 50 \times \log\left\{\left[1 + \left(\frac{1}{p}\right)\sum_{i=1}^{p}(\mu_{ri} - \mu_{ti})^2\right]^{-\frac{1}{2}} \times 100\right\} \quad (5)$$

where log is the logarithm based on 10, μ_{ri} and μ_{ti}

represent the percentage dissolved at time i for the predicted and reference profiles, respectively, and p is the number of time points tested. Indeed, f_2 is a function of the reciprocal of mean square-root transform of the sum of square distances at all points. The f_2 value lies between the interval of $(-\infty, 100)$. If the dissolution rates between μ_{ri} and μ_{ti} are identical at each time point (i.e. $\Sigma_{i=1}^{p} (\mu_{ri} - \mu_{ti})^2 = 0$), then $f_2 = 100$. In real situations, however, we do not expect to have the f_2 value anywhere near 100. Therefore, a range that defines a test profile which is accepted as 'similar' to a reference profile is needed. FDA defines that when f_2 falls between 50 and 100, then similarity of the dissolution profiles between two products can be claimed. The agreement of the lower limit ($f_2 = 50$) is because any sample time point of the profiles may be acceptable if an average difference of no more than 10% is achieved. When this 10% average difference is substituted into Eq. (5), f_2 becomes

$$f_2|_{10} = 50 \times \log\left\{\left[1 + (1/P)\sum_{i=1}^{P}|10|^2\right]^{-\frac{1}{2}} \times 100\right\} = 49.89 \approx 50$$

Therefore, a predicted profile is considered similar to a reference profile if the f_2 value is no less than 50.

3.4. Confidence Interval Estimation of the Similarity Factor

Owing to a variation in neural network predictions, similarity of the predicted and reference dissolution profiles may not be reflected statistically by only one f_2 value. In our experiment, six dissolution profiles were predicted by the network, and then compared with the corresponding reference profile. In addition to the mean f_2 value, the confidence interval for the mean f_2 value was computed by using the bootstrap method.

The bootstrap method has been an alternative data-resampling method to theoretical derivation in statistical analysis, by repeatedly resampling the original data and making inferences from the resamples [14]. This method is potentially superior to large sample techniques for small sample sizes [15]. Let $z = \{z_1, z_2, \cdots, z_n\}$ be a random sample from an unknown distribution. In the present study, since six dissolution profiles were predicted for every matrix ratio of MCC and GMS, n was equivalent to 6 and z is the random sample from distribution f_2. One needs to find an estimator and a $100(1 - \alpha)\%$ interval ($\alpha = 0.05$ for 95% confidence interval) for mean, μ. Let

$$\hat{\mu} = \frac{z_1 + z_2 + \cdots + z_n}{n}$$

The confidence interval for μ is found by determining the distribution of $\hat{\mu}$, and by finding such $\hat{\mu}_L$ and $\hat{\mu}_U$ that probability $\Pr(\hat{\mu}_L \le \mu \le \hat{\mu}_U) = 1 - \alpha$.

The distribution of $\hat{\mu}$ depends upon the distribution of the z_i's, which is unknown. The distribution of $\hat{\mu}$ could be approximated by the normal distribution as per the central limit theorem, only if n is large. For the case where n is small, the bootstrap resample z^* is done by drawing a sample of n values with replacement from z. Then, the mean of z^*, $\hat{\mu}^*$¶, is calculated. The above procedure is repeated until N bootstrap estimates are obtained, $\hat{\mu}_1^*, \cdots \hat{\mu}_N^*$. By sorting the bootstrap estimates in the order of $\hat{\mu}_{(1)}^* \le \hat{\mu}_{(2)}^* \le \cdots \hat{\mu}_{(N)}^*$, the $100(1 - \alpha)\%$ bootstrap confidence interval is $(\hat{\mu}_{(q_1)}^*, \hat{\mu}_{(q_2)}^*)$, where $q_1 = N\alpha/2$ and $q_2 = N - q_1 + 1$.

4. Results and Discussion

In ANN applications, a problem that often arises is the determination of the 'optimal' number of the hidden units. Normally, one has to resort to empirical methods to obtain the optimum network configuration that can achieve a good performance. In our case, leave-one-out cross-validation is preferred, as discussed in Section 3.1. After experimenting with various numbers of neurons in the hidden layer of the Elman network, we found that applying seven neurons in the hidden layer gave the best performance. Table 1 summarises the mean and standard deviation of f_2 values from six experiments with different matrix ratios of MCC and GMS using six, seven and eight neurons, in the hidden layer. The network with seven hidden units yielded the best results in means and standard deviations of f_2, except for a matrix ratio of 5:5, where the mean is slightly lower than that from eight hidden units, and for a matrix ratio of 6:4 where the standard deviation is larger than those from six and eight hidden nodes (although the mean value, in this case, is higher than others). However, the network with seven hidden units gave the best performance in an overall manner with the averages of means and standard deviations of f_2 for all matrix ratios being 79.02 and 3.40, respectively. As a result, the Elman network with seven hidden units was chosen for the rest of the experiments.

Table 2 shows the complete results comprising f_2 values from individual predictions, as well as the

Table 1. Means and standard deviations (stdev) of f_2 values from the Elman network with varying number of neurons in the hidden layer.

Matrix ratios (MCC:GMS)	6 hidden units		7 hidden units		8 hidden units	
	Mean	Stdev	Mean	Stdev	Mean	Stdev
10:0	50.38	8.30	59.12	2.73	50.29	5.94
8:2	80.60	5.00	81.28	3.56	75.11	10.04
7:3	81.39	5.07	84.50	2.41	78.07	3.71
6:4	80.90	4.85	86.06	5.30	85.98	3.45
5:5	85.59	3.44	86.89	1.27	87.21	4.46
4:6	68.21	5.21	76.28	5.13	76.11	5.30
Average	74.51	5.31	79.02	3.40	75.46	5.49

Table 2. The f_2 values of various proportion of MCC and GMS with seven hidden units in the Elman network.

MCC:GMS	P1	P2	P3	P4	P5	P6	Mean
10:0	60.92	62.20	61.28	56.63	55.53	58.18	59.12
8:2	83.84	76.38	77.38	81.81	84.91	83.35	81.28
7:3	83.86	84.12	82.22	82.82	88.99	84.96	84.50
6:4	96.43	83.36	81.54	84.13	86.11	84.81	86.06
5:5	87.93	87.14	84.77	88.11	87.37	86.05	86.89
4:6	86.40	75.39	74.38	73.85	72.00	75.68	76.28

Table 3. Average difference between two dissolution profiles.

Difference	2%	3.5%	6%	7%	8%	10%
f_2 limit	82.5	71.9	60.8	57.5	54.7	49.9

mean f_2 values from six trials using different matrix ratios of MCC and GMS. The results showed that the Elman network could achieve a good performance in *in-vitro* dissolution profile predictions. All of the f_2 values are above 50; thus, we can claim similarity between the predicted and the reference profiles. Table 3 provides error limits of f_2 from various average distances at multiple time points by appropriate substitution in Eq. (5). By using the f_2 error limits in Table 3, we can see that the difference between the predicted and the reference profiles in a matrix ratio of 10:0 was between 6% and 8%, while for the rest of the matrix ratios the difference was less than 3.5%.

The f_2 value of each trial is an estimate based on the dissolution profiles from the six predictions. Owing to sampling variations of the estimate, the true expected value is a function of both between trial differences as well as within trial prediction from one time point to another [16]. It is therefore a biased and conservative estimate of f_2. The bootstrap method can be used as a tool to estimate the confidence intervals. Table 4 shows the 95% confidence intervals for the mean f_2 values using the bootstrap

Table 4. Bootstrap for 95% Confidence Intervals (CIs) with seven neurons in the hidden layer.

Predicted ratio	Sample mean	200 bootstraps		400 bootstraps		500 bootstraps		1000 bootstraps	
		Mean	CI	Mean	CI	Mean	CI	Mean	CI
10:0	59.12	59.16	(57.24, 61.52)	59.10	(57.06, 61.01)	59.16	(57.09, 61.22)	59.06	(57.04, 60.96)
8:2	81.28	81.33	(78.78, 83.95)	81.24	(78.47, 83.95)	81.19	(78.63, 83.69)	81.37	(78.70, 83.86)
7:3	84.50	84.44	(82.98, 86.16)	84.58	(83.09, 86.48)	84.56	(83.18, 86.52)	84.47	(83.05, 86.47)
6:4	86.06	86.17	(83.04, 90.29)	86.12	(83.24, 90.59)	85.93	(83.12, 90.39)	86.12	(83.01, 90.59)
5:5	86.89	86.86	(85.91, 87.64)	86.88	(85.81, 87.70)	86.92	(85.81, 87.76)	86.92	(85.94, 87.71)
4:6	76.28	76.16	(73.23, 80.33)	76.42	(73.44, 80.64)	76.33	(73.58, 80.73)	76.32	(73.49, 80.61)

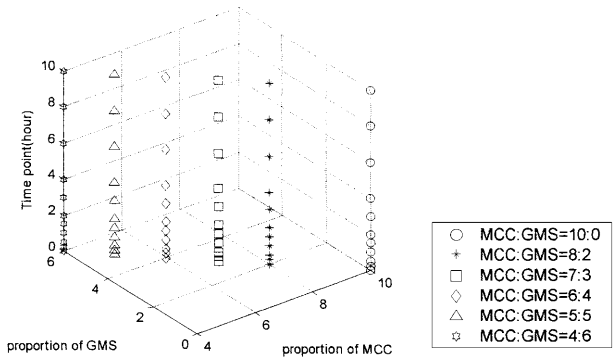

Fig. 2. A three-dimensional plot of the three inputs to the Elman network.

method with 200, 400, 500 and 1000 bootstrap samples. All of the calculated confidence intervals were within the acceptable range of f_2 specified by the FDA. It is indicated in Shah et al. [16] that in general, a total of 400 bootstrap samples should give precise estimates. However, we have generated up to a total of 1000 samples to ensure convergence of the confidence interval estimates. From the results in Table 4, we can see that 200 bootstrap samples have already given a stable estimate for the confidence intervals in our experiments.

Among all the results in Table 4, the predicted profiles of a matrix ratio of 10:0 yielded the lowest f_2 value (mean $f_2 = 59.12$), followed by the matrix ratios 4:6, 8:2, 7:3, 6:4 and 5:5. This observation could be due to the problem of interpolation and

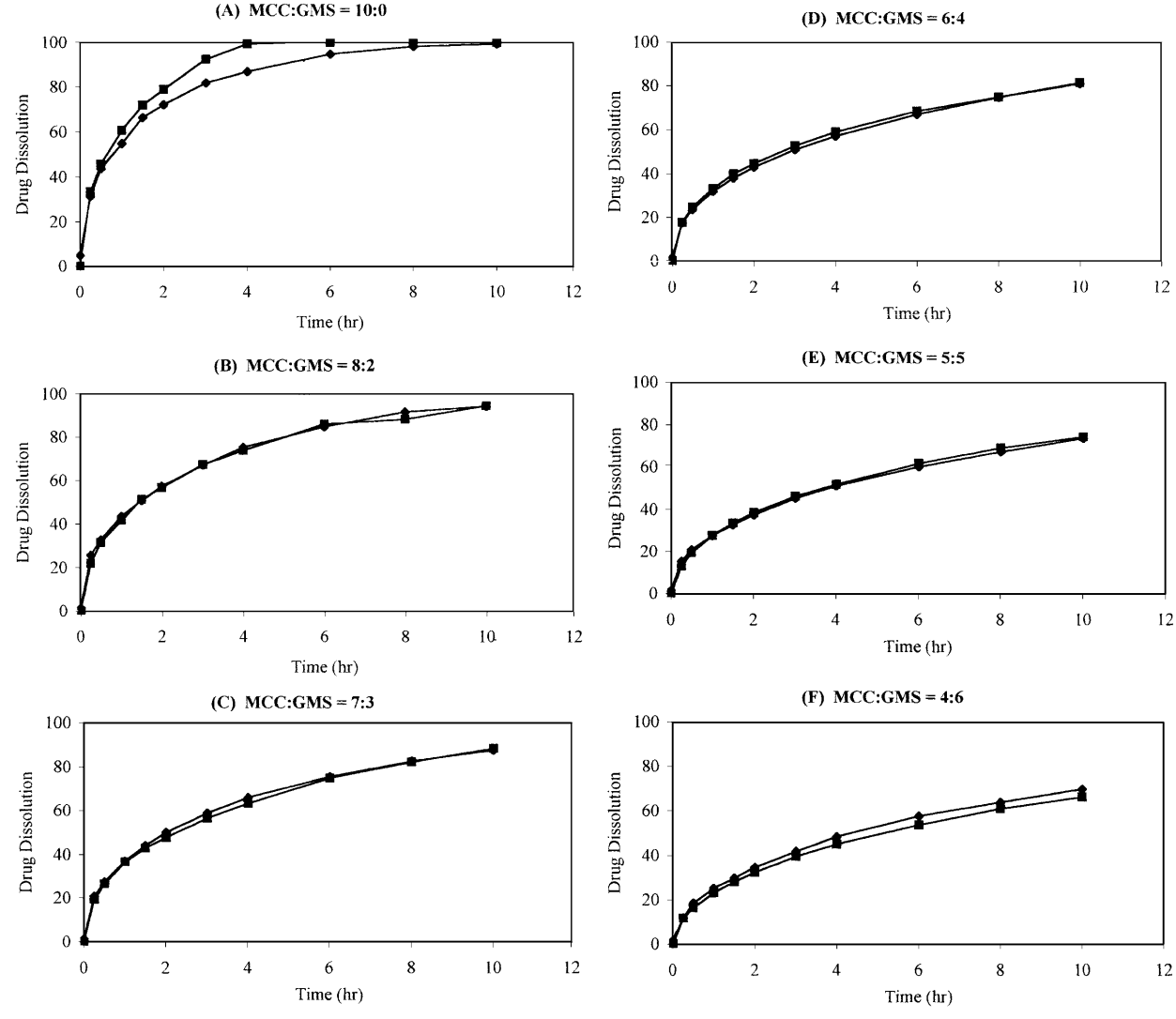

Fig. 3. Mean dissolution profiles of theophylline pellets with mixture of microcystalline cellulose (MCC) and glyceryl monostearate (GMS) in matrix ratios of (A) 10:0, (B) 8:2, (C) 7:3, (D) 6:4, (E) 5:5 and (F) 4:6 obtained from the Elman network (◆) and from physical experiments (■).

extrapolation. Interpolation refers to the process of predicting a function within the range of the original training data, while extrapolation refers to the process of predicting a function beyond the range of the original training data [17]. Basically, the network can give a better performance in interpolation compared with extrapolation of the input data. Figure 2 depicts distribution of the three inputs to the Elman network in a three-dimensional plot. We can see that the data sets of matrix ratios of 10:0 and 4:6 were located on the edges of the cube, encompassing all other input data. As a result, predictions of the matrix ratios of 10:0 and 4:6 actually amount to extrapolation by the network. This might be the reason for the interior results for matrix ratios of 10:0 and 4:6, compared with those from other matrix ratios. In general, the severity of extrapolation will increase if the distance between the subject data and the mean of the clustered data is increased. Therefore, profile predictions from a matrix ratio of 10:0 were interior to those from a ratio of 4:6, since the distance from the cluster of data for the former was further compared with the latter. For visual comparison, the mean predicted and reference dissolution profiles of various matrix ratios of MCC and GMS are illustrated in Fig. 3. As we can see, there was only a slight difference between the predicted and the reference dissolution profiles for all of the matrix ratios.

5. Conclusions

The present study has demonstrated a new application of the Elman recurrent neural network to the prediction of *in-vitro* dissolution profiles of matrix controlled release theophylline pellets formulation. The Elman network has performed satisfactorily in the dissolution profile prediction of controlled release theophylline pellets. Instead of conducting actual experiments for each and every matrix ratio to obtain the corresponding dissolution profiles, a suitable intelligent system can be used to predict the trend of the profile subject to different compositions of matrix substances. When a desired release profile is obtained, a confirmation test can then be conducted experimentally to verify the predicted profile. By using this approach, a lengthy and time-consuming experimentation to decide on the appropriate matrix ratio can be reduced. From our work, the potential of using an intelligent learning system for dissolution profile prediction is evident. Therefore, we believe that ANNs can be used as a powerful tool in pharmaceutical product formulation, as well as other areas in the pharmaceutical industry, so that the development tasks can be performed rapidly and efficiently with an increase of productivity, consistency and quality.

References

1. Patterson DA. Artificial Neural Networks Theory and Applications. Prentice Hall, Singapore, 1996
2. Hussain AS, Johnson RD, Vachharajani NN, Ritschel WA. Feasibility of developing a neural network for prediction of human pharmacokinetic parameters from animal data. Pharm Res 1993; 10: 466–469
3. Hussain AS, Yu X, Johnson RD. Application of neural computing in pharmaceutical product development. Pharm Res 1991; 8: 1248–1252
4. Opara J, Primožič S, Cvelbar P. Prediction of pharmacokinetic parameters and the assessment of their variability in bioequivalence studies by artificial neural networks. Pharm Res 1999; 16: 944–948
5. Brier ME, Zurada JM, Aronoff GR. Neural network predicted peak and trough gentamicin concentrations. Pharm Res 1995; 12: 406–412
6. Takayama K, Fujikawa M, Nagai T. Artificial neural network as a novel method to optimise pharmaceutical formulations. Pharm Res 1999; 16: 1–6
7. Rowe RC, Roberts RJ. Intelligent Software for Product Formulation. Taylor & Francis, London, 1998
8. Elman JL. Finding structure in time. Cognitive Science 1990; 14: 179–211
9. FDA guidance for industry: dissolution testing of immediate release solid oral dossage forms. August 1997
10. Efron B. Bootstrap methods: another look at the jackknife. Ann Statist 1979; 7: 1–26
11. Demuth H, Beale M. Neural Network Toolbox User's Guide. The MathWorks Inc., 1997
12. Peh KK, Yuen KH. Development and in vitro evaluation of a novel multiparticulate matrix controlled release formulation of theophylline. Drug Dev and Ind Pharmacy 1995; 21(13): 1545–1555
13. Bishop CM. Neural Networks for Pattern Recognition. Oxford University Press, New York, 1995
14. Shao J, Tu D. The Jackknife and Bootstrap. Springer-Verlag, New York, 1995
15. Zoubir AM, Boashash B. The bootstrap and its application in signal processing. IEEE Signal Process Mag 1998; 15(1): 56–76
16. Shah VP, Tsong Y, Sathe P, Liu JP. In vitro dissolution profile comparison – stastistics and analysis of the similarity factor, f_2. Pharm Res 1998; 15: 889–896
17. Leonard JA, Kramer MA, Ungar LH. A neural network architecture that computes its own reliability. Compt Chem Eng 1992; 16(9): 819–835