Unsupervised Dynamic Fuzzy Cognitive Map

Boyuan Liu, Wenhui Fan*, and Tianyuan Xiao

Abstract: Fuzzy Cognitive Map (FCM) is an inference network, which uses cyclic digraphs for knowledge representation and reasoning. Along with the extensive applications of FCMs, there are some limitations that emerge due to the deficiencies associated with FCM itself. In order to eliminate these deficiencies, we propose an unsupervised dynamic fuzzy cognitive map using behaviors and nonlinear relationships. In this model, we introduce dynamic weights and trend-effects to make the model more reasonable. Data credibility is also considered while establishing a machine learning model. Subsequently, we develop an optimized Estimation of Distribution Algorithm (EDA) for weight learning. Experimental results show the practicability of the dynamic FCM model. In comparison to the other existing algorithms, the proposed algorithm has better performance in terms of convergence and stability.

Key words: Fuzzy Cognitive Map (FCM); Estimation of Distribution Algorithm (EDA); nonlinear relation; machine learning

1 Introduction

Complexity systems such as dynamic social systems and integrated engineering systems have been puzzling decision makers for years, because the large number of highly correlated variables makes it difficult to grasp and manage the internal relationships of such systems. In the scenarios where such interdependencies are unknown, decision makers expect to predict the impact of their decisions. Fuzzy Cognitive Map (FCM) proposed by Kosko[1] is an effective tool that helps decision makers understand such systems. By describing the causal relationships between a set of concepts, FCM is capable of emulating human cognitive process[2]. Because of the great application values, FCMs have been successfully applied in various fields such as engineering[3], business[4], medicine[5],

• Boyuan Liu, Wenhui Fan, and Tianyuan Xiao are with the Department of Automation, Tsinghua University, Beijing 100084, China. E-mail: liuboyuan@126.com; fanwenhui@tsinghua.edu.cn.
* To whom correspondence should be addressed.

and social sciences[6].

An FCM model is composed of nodes and directed lines with weights. A typical graphical representation of FCM is illustrated in Fig. 1.

The nodes represent descriptive concepts, which can be attributes or behaviors of the system. The directed arc connecting two nodes represents a causal relationship between them, and the weight attached to the arc indicates the strength of the relationship between the two nodes. Once an FCM model is established, the inference process can be calculated by the following rule:

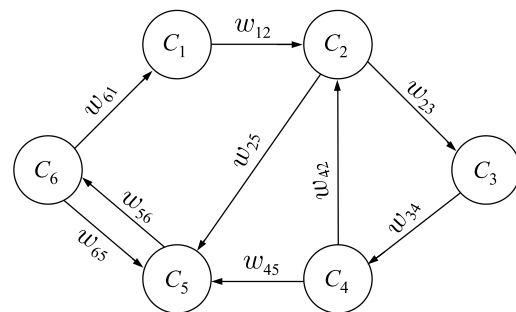$$A_j^t = f\left(\sum_{i=1, i \neq j}^{n} w_{ij} A_i^{t-1}\right) \quad (1)$$

Fig. 1   An FCM example.

where t is the inference step, A_j^t is the value of concept C_j at step t, A_i^{t-1} is the value of concept C_i at step $t-1$, and w_{ij} is the weight of the interrelation between concept C_i and C_j. f is a threshold function, which can squeeze the result of this formulation into the interval [0,1]. The range of w_{ij} belongs to the interval $[-1, 1]$, and a negative value indicates negative causality from C_i to C_j, which implies the increase in C_i leads to a decrease in C_j, and vice visa. The value zero indicates that there is no causal relationship between them.

As a generic model, FCM relies on several assumptions but some may be inappropriate. For example, it uses only simple monotonic and symmetric causal relationships between concepts. However, many real world causal relationships are neither symmetric nor monotonic, hence the FCM model is not robust enough to model a dynamic evolving system. Several methodologies are explored to overcome the shortcomings and improve the performance of FCMs. Hagiwara[7] extended FCMs (E-FCMs) by introducing nonlinear membership functions, time-delay weights, and conditional weights; however, the general expression of nonlinear relationships and conditional weights is not given. Neural Cognitive Map (NCM)[8] is another methodology that enhances the modeling and prediction capabilities. In Carvalho's Rule Based Fuzzy Cognitive Maps (RBFCM)[9] approach, fuzzy nodes are used instead of concepts, and each one contains several membership functions that represent the value or the change of concept. The fuzzy rule bases are used to relate and link concepts, and the fuzzy causality and fuzzy carry accumulation are introduced. Song et al.[10] extended the concept of a fuzzy event that can model the fuzzy probability of both cause and effect, and presented the Probabilistic Fuzzy Cognitive Map (PFCM) approach. The other extensions include Fuzzy Grey Cognitive Maps (FGCM)[11], Dynamic Cognitive Network (DCN)[12], and intuitionistic fuzzy cognitive maps[13]. These models have numerous features that enhance the capability of FCMs to describe the real world. The only disadvantage in these extended models is that they require participation from experts, which involves inevitable human errors, especially when facing a large-scale system.

In this paper, we focus our efforts on some drawbacks in classical FCMs, and propose an unsupervised dynamic fuzzy cognitive map model, which deals with dynamic behaviors and nonlinear relationships in systems. The model is discussed in detail and a formalized inference process is given. A corresponding learning algorithm that can be used to establish the proposed model is also given. Finally, some experiments are performed to validate the proposed algorithm.

## 2 Deficiencies in Classical FCM

2.1 Data credibility in the construction of FCM

In general, there are two main approaches to establish an FCM model: artificial construction approach and machine learning approach. With rapid advances in information technology and rising popularity of the simulation-based design and the evaluation methods, we can obtain and use more data than ever before. Hence, we can obtain a more accurate model by using enough data. The main drawback with large amount of data is that the process of data collection can be time consuming. In addition, the system properties change over time. For example, the cycle of economic data collection takes about one month. After obtaining enough data, we observe that the data credibility changes over time. For instance, the data collected at a later time in the cycle has significantly more value compared with the similar data collected earlier in the cycle. Therefore, if the establishment of an FCM model only relies on data processing and machine learning, a coordination mechanism should be developed to guarantee that dissimilar data is dealt in accordance with its time sequence during the learning process.

2.2 Nonlinear causal relationship between concepts

In most FCMs, the weight between the concepts is a fixed value, and is independent of the current state of the concept. In fact, the causal relationship is definitely nonlinear in some situations. In the example shown in Fig. 2, when the vehicle flow varies in a slight state, it exerts negligible influence on the traffic jam. Since the road has not attained saturation and cars on the road can run at their desired speed. With the increase in vehicle flow, as the road shows signs of saturation, the traffic jam becomes sensitive to the number of cars on the road. If the road is saturated, the change of vehicle flow in such a condition will not affect traffic condition anymore as the traffic is already paralyzed.

Situations like this are common in practical systems. So, the weight should be dynamic and take different
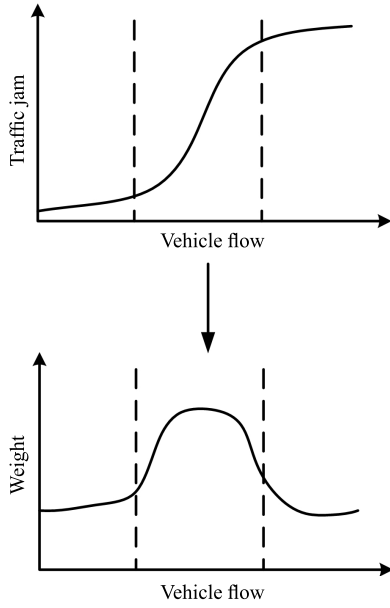
Fig. 2   Sample of nonlinear causal relationship.

values according to the current state. The best solution is to construct a function, which can be used to describe the relationship. Another solution is splitting the range of variable into intervals, and allowing each partition to possess its own weight.

2.3   The trend-effects of concepts

By applying dynamic weight to FCMs, we let the nonlinear relationship between concepts be in accordance with actual conditions. Next issue we concern with is that in classical FCMs, the influence of the concept itself is ignored. Groumpos and Stylios[14] took memory capabilities into consideration and presented a new calculation rule stated as follows:

$$A_j^t = f\left(\sum_{i=1,i\neq j}^{n} w_{ij} A_i^{t-1} + k_j A_j^{t-1}\right) \quad (2)$$

where k_j is the contribution of the previous value of concept A_j.

The memory capabilities can be integrated into the FCM model. As an example, we consider the speed of a moving car. The speed at the next step is affected by the other concepts (such as brake force and engine output), the concept itself (the current speed), and the acceleration state at the current step. Even though the values of all concepts remain unchanged, the speed will still be decided by its current acceleration state, which is the trend of speed. This implies that the causal relationship between the two concepts is as sensitive to the trend-effects as it is to the current state.

This phenomenon is called inertia in physics and it is referred to as trend-effect in this field. In order to obtain an accurate FCM model, this effect should not be neglected.

3   Dynamic FCM Model

To extend the capability of FCMs, we propose a Dynamic Fuzzy Cognitive Map (DFCM) model, which is capable of reflecting dynamic behaviors and modeling nonlinear relationships in systems.

As described in Section 2, the weight between concepts should be nonlinear. It is impractical to describe this relationship by a function; therefore, a triple is introduced. For the value of A_i, its relative position determines its specific weight to other concepts, as shown in Fig. 2. So w_{ij} is defined as follows:

$$w_{ij} = \begin{cases} w_{ij}^s, & A_i^t \in \text{domain(small)} \\ w_{ij}^m, & A_i^t \in \text{domain(medium)} \\ w_{ij}^l, & A_i^t \in \text{domain(large)} \end{cases} \quad (3)$$

In the light of the domain that a_i locates (small, medium, and large) at step t, w_{ij} can take three different values dynamically. The subscripts s, m, and l denote their domains respectively.

After analyzing the trend-effects of concepts, a new parameter is introduced in the model. It determines how trend-effect will affect the concept. The definition of this parameter is given as follows:

$$\beta_j(A_j) = \begin{cases} \beta_j^+, & \Delta A_j^t > 0; \\ 0, & \Delta A_j^t = 0; \\ \beta_j^-, & \Delta A_j^t < 0 \end{cases} \quad (4)$$

where ΔA_j^t indicates the change of concept A_j at step t.

The inference process of the proposed DFCM model can be determined by the next formula shown as follows:

$$A_j^t = f\left(\sum_{i=1,i\neq j}^{n} w_{ij} A_i^{t-1} + k_j A_j^{t-1} + \beta_j \Delta A_j'^{t-1}\right) \quad (5)$$

where $\Delta A_j'^{t-1}$ is the normalized value of ΔA_j^{t-1}. The most frequently used functions f include bivalent step function, trivalent function, sigmoid function, and hyperbolic function. A comparison of different threshold functions is given by Tsadiras[15]. Sigmoid function in common use is adopted with the following

definition:

$$f(x) = \frac{1}{1 + e^{-cx}} \qquad (6)$$

where the parameter c is used to determine the curve slope and $c = 5$ is frequently used.

4   The Optimized Estimation of Distribution Algorithm

4.1   Optimization model

In the DFCM model, we propose three types of parameters that are necessary to construct a complete model: weight parameters, memory capability parameters, and trend-effect parameters.

$$w = [w_{21}^{s}, w_{21}^{m}, w_{21}^{l}, \cdots, w_{n1}^{s}, w_{n1}^{m}, w_{n1}^{l},$$
$$\quad w_{12}^{s}, w_{12}^{m}, w_{12}^{l}, \cdots, w_{n-1\,n}^{s}, w_{n-1\,n}^{m}, w_{n-1\,n}^{l}],$$
$$K = [k_1, k_2, \cdots, k_n],$$
$$\beta = [\beta_1^+, \beta_1^-, \beta_2^+, \beta_2^-, \cdots, \beta_n^+, \beta_n^-].$$

For n concepts, the total count of the parameter to be evaluated is given by $3n(n-1) + 2n + n = 3n^2$. Therefore, an estimated value based on these parameters is given as follows:

$$A_j^{t\,(\text{est})} = f\left(\sum_{i=1, i \neq j}^{n} w_{ij} A_i^{t-1} + k_j A_j^{t-1} + \beta_j \Delta A_j^{\prime t-1}\right) \qquad (7)$$

The principle behind achieving these parameters is to make the gap between each estimated value and the real value as near as possible. For a given data item, the deviation is $\sum_{i=1}^{N}(A_i - A_j^{t\,(\text{est})})^2$. A general approach to solve this problem is to make the sum of all sample deviations as low as possible. Considering the data credibility analyzed in Section 2.1, the later the data collected, the higher the data credibility. So each data at different collection times should be given a weight, which represents its value. The data credibility function based on time is given as follows:

$$s\left(\frac{t}{T}\right) = \frac{0.5}{1 + e^{-10\left(2\frac{t}{T} - 0.5\right)}} + 0.5, \ t = 1, 2, \cdots, T \qquad (8)$$

where t is the current step in the data series, and T is the total data count. The curve image of this function is shown in Fig. 3.

As can be seen from the curve, the data becomes more valuable with time. We can tolerate a larger deviation at the early phase, but we desire a lesser deviation at the last phase. Therefore, the idea can be
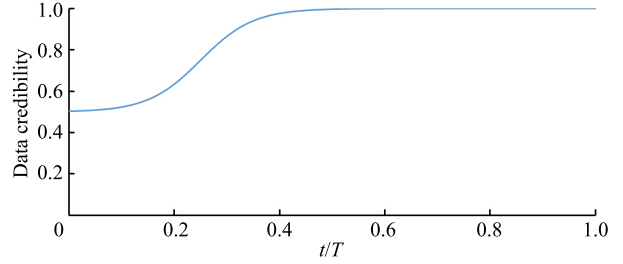

Fig. 3   Data credibility function.

expressed in the expression as follows:

$$\min \frac{1}{(T-1)n} \sum_{t=3}^{T} s\left(\frac{t}{T}\right) \sum_{i=1}^{n} \left(A_i^t - A_i^{t\,(\text{est})}\right)^2 \qquad (9)$$

After removing the constants, the optimization problems can be described by the following formulas:

$$\min \sum_{t=3}^{T} s\left(\frac{t}{T}\right) \cdot$$

$$\sum_{i=1}^{n} \left(A_i - \sum_{\substack{i=1 \\ i \neq j}}^{n} \frac{1}{1 + e^{-5(w_{ij} A_i^{t-1} + k_j A_j^{t-1} + \beta_j \Delta A_j^{\prime t-1})}}\right)^2 \qquad (10)$$

$$-1 \leqslant w_{ij} \leqslant 1 \ (1 \leqslant i, j \leqslant n, i \neq j) \qquad (11)$$
$$0 \leqslant k_j \leqslant 1 \ (1 \leqslant j \leqslant n) \qquad (12)$$
$$0 \leqslant \beta_j \leqslant 1 \ (1 \leqslant j \leqslant n) \qquad (13)$$

4.2   Optimized estimation of distribution algorithm

Weight learning of FCM is equivalent to the optimization problem of the connection matrix. Many studies have been conducted in the area of applying optimization algorithms to construct the weight matrix. Genetic Algorithm[16] is one of the earliest methods, which is subject to several revisions and improvements. Hebbian learning is another main stream algorithm, and its extensions include differential Hebbian Learning[17], Balanced Differential Algorithm, Nonlinear Hebbian Learning[18], and Active Hebbian Learning[19]. With the evolution of new optimization theories, some algorithms such as Particle Swarm Optimization (PSO)[20] and Differential Evolution (DE) algorithm[21] are also introduced gradually to the field of weight learning of FCMs.

The Estimation of Distribution Algorithm (EDA)[22] is a relatively novel evolutionary algorithm. In EDA, conventional genetic operators, such as crossover and mutation, are replaced by a statistical model to describe the probability distributions of solutions. By renovating probability model and sampling new individuals from the model, the solution gradually approaches to the

optima. The satisfactory performance of EDA has been proved in many fields of optimization, but its application in the field of FCM learning has not been explored. In this paper, we choose the EDA algorithm to solve the construction of connection matrix in FCM.

Probability model is the core in EDA. By employing the probability model and its renovation, EDA can describe the space distribution of solutions and control the overall evolution trend of the population. For the variables in continuous domains, a general approach is to assume that the variables obey Gaussian distribution. By considering the dependencies between variables in our problem space, an estimation of a multivariate normal density function is introduced. At each generation, the mean μ, and the covariance matrix Σ whose element is denoted by σ_{ij}^2 are estimated to describe the multivariate normal distribution. And new individuals are generated by this distribution.

Two main decisions are made in the proposed algorithm. First, we decide on how to generate samples from a multivariate normal density function. The sampling method used in the algorithm is a suitable method proposed by Scheuer and Stoller[23]. Given that Σ is positive-definite and its Cholesky decomposition is unique, the factorization can be written as $\Sigma = CC^{\mathrm{T}}$, where C is a lower triangular matrix. A multivariate normal distribution vector X can be calculated using $X = \mu + CZ$, where μ is the mean, C is the Cholesky decomposition of Σ, and Z is a vector consisting of standard normal random variables.

Next, we decide on how to avoid a local optimal solution. There is a drawback of premature convergence associated with continuous EDAs. The estimation of density function is based on a set of individuals, and the information of an individual itself is not reserved. Consequently, there is a possibility that a few excellent individuals strikingly different from others in this set may vanish during the process of averaging, which implies that a possible optimal solution is lost. In order to avoid this situation, at each generation, we pass on a proportion of the best individuals to the next generation. The rest of the individuals are generated by the probability model. And to maintain a high level of diversity, a dynamic variance is applied to control the generation of new samples. A higher variance can enhance the global searching ability at the early stage, and a lower variance can improve the convergence at the final stage. The specific method is adopted to adjust the variance of vector Z.

The proposed algorithm is named Optimized EDA (OEDA) and the complete algorithm is summarized in Algorithm 1 .

5 Experiments

5.1 Parameter selection for OEDA

The OEDA contains only one selection parameter, which is the population size. In most of the learning algorithms, the reference range of population size is provided. In this paper we establish a complete unsupervised DFCM learning algorithm, in which an exact value of population size is given. In the proposed algorithm, the time complexities of covariance matrix and Cholesky decomposition are sensitive to dimension, hence selection of a proper population size is crucial. An experiment is performed based on a dataset consisting of five attributes. As mentioned in Section

---

Algorithm 1   The optimized EDA algorithm.

Input: Dimension size n, population size m
Output: Best fitness f^*, best solution p^*
1: Initialize population P randomly
2: **while** stopping criterion is not met
3:    **for** $i \leftarrow 0$ to m do
4:       $f[i] \leftarrow$ Get fitness($P[i]$)
5:    **endfor**
6:    sort $P[i]$ by $f[i]$ in ascending order
7:    **for** i$\leftarrow 0$ to $m/2$ do
8:       $P_{\text{best}}[i] \leftarrow P[i]$
9:    **endfor**
10:   **for** $i \leftarrow 0$ to $m/10$ do
11:      $P_{\text{reservation}}[i] \leftarrow P[i]$ //keep the best individuals
12:   **endfor**
13:   $\mu \leftarrow$ Mean(P_{best})
14:   **for** $i \leftarrow 0$ to n do
15:      **for** $j \leftarrow 0$ to n do
16:         **for** $k \leftarrow 0$ to $m/2$ do
17:            $X \leftarrow P_{\text{best}}[k].\text{elem}[i]$ //the i-th element
18:            $Y \leftarrow P_{\text{best}}[k].\text{elem}[j]$ //the j-th element
19:         **endfor**
20:         $\sum_{ij} \leftarrow$ cov(X,Y)
21:      **endfor**
22:   **endfor**
23:   $C \leftarrow$ Cholesky decomposition(Σ)
24:   $r \leftarrow$ Current dynamic variance
25:   **for** $i \leftarrow 0$ to $m/2$ do
26:      $Z \sim N(0,r)$ //Generate a normal distribution
27:      $P_{\text{new}}(i) \leftarrow \mu + CZ$
28:   **endfor**
29:   $P \leftarrow P_{\text{new}} \cup P_{\text{reservation}}$
30: **endwhile**
31: $f^* \leftarrow f[0], p^* \leftarrow P[0]$

4.1, the dimension of the solution space with 5 attributes equals to 75; therefore, the population size increases from 35 to 150 to obtain the results. The OEDA runs 10 times in each ease, and the results are presented in Fig. 4.

As shown in Fig. 4, it is clear that as population size increases, the best fitness converges gradually. We choose a point at $4n^2$, which is 100, as the best population size.

5.2 Practicability of dynamic FCM model

In this section, the applicability of the DFCM model is verified. For this experiment, we use the Ozone Level Detection dataset taken from UCI Machine Learning Repository. The dataset comprises ground ozone level data organized in time-series. The original dataset consists of one-hour peak set with 2536 samples and 73 attributes. To increase the execution speed, we used 8 key attributes with 1000 samples. Besides DFCM, other three types of FCMs are construed: classical FCM, FCM with nonlinear relationships (expressed as FCM-NR), and FCM with trend-effects (expressed as FCM-TE). The connection matrixes of the four FCMs are constructed by the OEDA, and each runs three times. The computational results of average fitness values are presented in Table 1.

The smaller the fitness, the more accurate the FCM model. It can be observed from the table that by introducing either the nonlinear relationships or the trend-effects, we achieve the FCM model with better fitness. The nonlinear relationship causes a greater impact on the FCM, which reveals that the nonlinear

Fig. 4   Convergence of OEDA.

Table 1   Best fitness of different FCMs.

FCM	FCM-TE	FCM-NR	DFCM
230.34	209.32	194.80	179.11

relationships are prevalent in real world. The DFCM model with both nonlinear relationships and trend-effects can achieve the minimum value of fitness, which implies that it is the most accurate one.

5.3   Comparisons with other learning algorithms

In this section, the proposed OEDA algorithm is compared with some other known learning algorithms. The algorithms to be compared include DE algorithm, PSO algorithm, and the classical EDA algorithm EMNA$_{global}$. All these algorithms are implemented in C++ and deployed on the same platform. The dataset is created artificially. Dataset 1 contains 500 samples with 10 attributes, and dataset 2 contains 2000 samples with 5 attributes. Each algorithm runs 5 times on each dataset, and the evaluation targets contain the best fitness, iterations, and running times. The results are shown in Tables 2 and 3.

In general, the OEDA algorithm has good performance. The OEDA algorithm shows the following characteristics from the results: First, the calculation process is very stable. Both the fitness and iterations show that OEDA can provide a stable solution space, which guarantees the credibility of its solution. In comparison to EMNA$_{global}$, the probability of a local optimal solution is reduced. The second characteristic observed is its sensitivity to dimensions. For dataset 2 with 5 attributes, OEDA has a clear advantage in running time. For dataset 1 with 10 attributes, the running time increases when compared with the other algorithms. In spite of this, OEDA can achieve a stable, satisfactory solution. Moreover, the OEDA shows a distinct improvement on EMNA$_{global}$ in high dimension.

6   Conclusions

The value and potential of FCMs has become clear over the last several years. At the same time, some deficiencies have emerged as people begin to rely on them. In this paper, we focus on three clear defects of FCMs, and propose improvement approaches accordingly. An unsupervised dynamic fuzzy cognitive map is proposed as an extension of classical FCMs, which can not only describe the nonlinear relationships but also the trend-effects between concepts. It can be a reliable alternative to FCM when experts or decision makers try to get a better understanding of the causal relationships embedded in a complex system. Experimental results of several aspects show
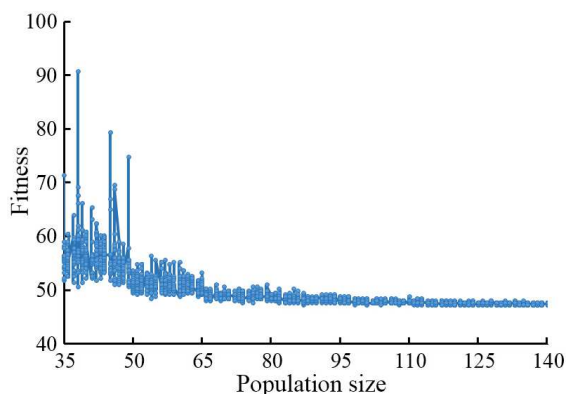
Table 2   Experiment results on dataset 1.

	Minimum fitness	Maximum fitness	Average fitness	Minimum iterations	Maximum iterations	Average iterations	Minimum time	Maximum time	Average time
DE	316.45	344.68	328.84	931	1886	1399.8	172 267	358 091	265 799
PSO	267.89	692.03	426	452	4704	1831.7	66 269	865 544	316 604
EMNA	268.85	272.08	270	408	448	425.2	594 802	662 588	630 934
ODEA	267.36	267.88	267.66	415	593	464.6	398 098	566 447	449 111

Table 3   Experiment results on dataset 2.

	Minimum fitness	Maximum fitness	Average fitness	Minimum iterations	Maximum iterations	Average iterations	Minimum time	Maximum time	Average time
DE	238.53	257.21	249.68	585	890	696.8	17 640	26 700	20 866.4
PSO	200.39	216.09	208.88	386	1276	832.4	11 181	36 272	23 861
EMNA	192.64	195.82	194.19	285	304	295.4	12 929	14 232	13 614.4
ODEA	192.6	194.74	193.63	264	282	274	9784	10 947	10 322.2

its advantage in comparison with others. But the experiments are not comprehensive and the FCM model still needs further verification in practical applications. Ongoing work includes finding a more convincing method to describe the nonlinear relationships between concepts, and improving the efficiency in a high dimension space.

References

[1] B. Kosko, Fuzzy cognitive maps, *International Journal of Man-Machine Studies,* vol. 24, pp. 65–75, 1986.

[2] A. K. Tsadiras and K. G. Margaritis, Cognitive mapping and certainty neuron fuzzy cognitive maps, *Information Sciences,* vol. 101, pp. 109–130, 1997.

[3] C. D. Stylios and G. Georgoulas, Modeling complex logistics systems using soft computing methodology of fuzzy cognitive maps, in *Automation Science and Engineering (CASE), 2011 IEEE Conference on*, Trieste, Italy, 2011, pp. 72–77.

[4] W. Shou, W. Fan, B. Liu, and Y. Lai, Knowledge map mining of financial data, *Tsinghua Science and Technology,* vol. 18, pp. 68–76, 2013.

[5] E. I. Papageorgiou, A new methodology for decisions in medical informatics using fuzzy cognitive maps based on fuzzy rule-extraction techniques, *Applied Soft Computing,* vol. 11, pp. 500–513, 2011.

[6] J. P. Carvalho, On the semantics and the use of fuzzy cognitive maps in social sciences, in *Fuzzy Systems (FUZZ), 2010 IEEE International Conference on*, Barcelona, Spain, 2010, pp. 1–6.

[7] M. Hagiwara, Extended fuzzy cognitive maps, in *Fuzzy Systems, 1992., IEEE International Conference on*, San Diego, CA, USA, 1992, pp. 795–801.

[8] T. Obata and M. Hagiwara, Neural cognitive maps (NCMs), in *Systems, Man, and Cybernetics, 1997. Computational Cybernetics and Simulation., 1997 IEEE International Conference on*, Orlando, FL, USA, 1997, pp. 3337–3342

[9] J. P. Carvalho and J. A. B. Tome, Rule based fuzzy cognitive maps and fuzzy cognitive maps-a comparative study, in *Fuzzy Information Processing Society, 1999. NAFIPS. 18th International Conference of the North American*, New York, NY, USA, 1999, pp. 115–119.

[10] H. J. Song, Z. Q. Shen, C. Y. Miao, Z. Q. Liu, and Y. Miao, Probabilistic fuzzy cognitive map, in *Fuzzy Systems, 2006 IEEE International Conference on*, Vancouver, BC, Canada, 2006, pp. 1221–1228.

[11] J. L. Salmeron, Modelling grey uncertainty with fuzzy grey cognitive maps, *Expert Systems with Applications,* vol. 37, pp. 7581–7588, 2010.

[12] Y. Miao, Z. Q. Liu, C. K. Siew, and C. Y. Miao, Dynamical cognitive network—An extension of fuzzy cognitive map, *Fuzzy Systems, IEEE Transactions on,* vol. 9, pp. 760–770, 2001.

[13] D. K. Iakovidis and E. Papageorgiou, Intuitionistic fuzzy cognitive maps for medical decision making, *Information Technology in Biomedicine, IEEE Transactions on,* vol. 15, pp. 100–107, 2011.

[14] P. P. Groumpos and C. D. Stylios, Modelling supervisory control systems using fuzzy cognitive maps, *Chaos, Solitons & Fractals,* vol. 11, pp. 329–336, 2000.

[15] A. K. Tsadiras, Comparing the inference capabilities of binary, trivalent and sigmoid fuzzy cognitive maps, *Information Sciences,* vol. 178, pp. 3880–3894, 2008.

[16] D. E. Koulouriotis, I. E. Diakoulakis, and D. M. Emiris, Learning fuzzy cognitive maps using evolution strategies: A novel schema for modeling and simulating high-level behavior, in *Evolutionary Computation, 2001. Proceedings of the 2001 Congress on*, Seoul, Korea, 2001, pp. 364–371.

[17] T. L. Kottas, Y. S. Boutalis, and A. D. Karlis, New maximum power point tracker for PV arrays using fuzzy controller in close cooperation with fuzzy cognitive networks, *Energy Conversion, IEEE Transactions on,* vol. 21, pp. 793–803, 2006.

[18] T. T. D. Gedeon, L. C. C. Fung, E. Papageorgiou, C. Stylios, and P. Groumpos, Fuzzy cognitive map learning based on nonlinear Hebbian rule, in *AI 2003: Advances in Artificial Intelligence*, T. T. D. Gedeon and L. C. C. Fung, Eds. Springer Berlin Heidelberg, 2003, pp. 256–268.

[19] E. I. Papageorgiou, C. D. Stylios, and P. P. Groumpos, Active Hebbian learning algorithm to train fuzzy cognitive maps, *International Journal of Approximate Reasoning,* vol. 37, pp. 219–249, 2004.

[20] K. E. Parsopoulos, E. I. Papageorgiou, P. P. Groumpos, and M. N. Vrahatis, A first study of fuzzy cognitive maps learning using particle swarm optimization, in *Evolutionary Computation, 2003. CEC'03. The 2003 Congress on*, 2003, pp. 1440-1447.

[21] W. Shou, W. Fan, and B. Liu, A data-based fuzzy cognitive map mining method using DE-SQP algorithm, in *Asia Simulation Conference, AsiaSim 2012, October 27, 2012 – October 30, 2012*, Shanghai, China, 2012, pp. 37–43.

[22] P. Larrañaga and J. A. Lozano, *Estimation of Distribution Algorithms: A New Tool for Evolutionary Computation,* vol. 2. Springer, 2002.

[23] E. M. Scheuer and D. S. Stoller, On the generation of normal random vectors, *Technometrics,* vol. 4, pp. 278–281, 1962.

Boyuan Liu is currently a PhD candidate at State CIMS Engineering Research Center, Tsinghua University, China. He received his bachelor degree and master degree from Harbin Institute of Technology in 2008 and 2010, respectively. His research interests include big data and data mining.

Wenhui Fan is a professor in Department of Automation, Tsinghua University, China. He received his BS degree from Northwestern Polytechnical University, China, in 1990, MS degree from Harbin Institute of Technology, China, in 1995, and PhD degree from Zhejiang University, China, in 1998. His research interests include collaborative simulation and multidisciplinary optimization.

Tianyuan Xiao is a professor at State CIMS Engineering Research Center, Tsinghua University, China. He received his bachelor degree from Tsinghua University in 1970. He has been researching into system simulation, virtual manufacturing, Computer Integrated Manufacturing (CIM), and networked manufacturing. Concurrently he has more than 200 publications in books, journals, and major conferences, and he won more than 10 awards in scientific research project.