

HIGH-ORDER FCM-BASED TIME SERIES PREDICTION WITH CONSTRAINED CONVEX OPTIMIZATION LEARNING

DAN SHAN¹, XIAOXU ZHANG¹, WEI LU² AND HUANMEIXI HOU¹

¹Department of Electronic Engineering
Dalian Neusoft University of Information
No. 8, Software Park Road, Dalian 116023, P. R. China
{ shandan; zhangxiaoxu; houthuanmeixi }@neusoft.edu.cn

²School of Control Science and Engineering
Dalian University of Technology
No. 2, Linggong Road, Ganjingzi District, Dalian 116024, P. R. China
luwei@dlut.edu.cn

Received October 2025; revised February 2026

ABSTRACT. *In the field of time series prediction, weight learning of Fuzzy Cognitive Maps (FCMs) has long faced two critical bottlenecks: first, limited predictive capability when confronted with long-term dependencies and complex patterns; second, although existing swarm intelligence and hybrid optimization methods have achieved certain breakthroughs, they suffer from difficulty in large-scale extension, and sensitivity to data noise. This study addresses these challenges from an optimization theory perspective, proposing a fast learning scheme for high-order FCMs. The core idea is to reformulate the weight learning problem as a constrained convex optimization task, which is then efficiently solved using classical interior-point algorithms. The method systematically optimizes key hyperparameters including the high-order decay factor, steepness coefficient, and number of nodes through parameter scanning. To comprehensively evaluate the proposed scheme, we constructed an experimental framework covering synthetic data, noisy data and ten publicly available benchmark datasets. Test results demonstrate that prediction accuracy improves significantly compared to state-of-the-art methods; meanwhile, the approach exhibits excellent performance in both small-scale and large-scale FCM learning as well as noisy data processing.*

Keywords: Fuzzy cognitive map, Convex optimization, Time series prediction, High-order FCM

1. Introduction. Time series prediction remains challenging in computational intelligence, particularly for systems with uncertainty, nonlinearity, and complex temporal dependencies. Fuzzy Cognitive Maps (FCMs), conceptualized by Kosko [1] and Hagiwara [2], combine fuzzy logic with graph-based causal reasoning, representing components as nodes and relationships as weighted directed edges. This intuitive yet rigorous framework enables successful deployment across diverse domains: healthcare diagnostics and pharmaceutical research [3, 4], time series forecasting [5, 6], educational assessment [7], epidemic control [8], and human-robot interaction modeling [9].

FCM effectiveness depends critically on accurate edge weight determination, quantifying causal influences between concepts. This weight learning problem has spawned three algorithmic philosophies. First, Hebbian learning approaches [10] adopt unsupervised strategies adjusting weights based on node activation correlations. Variants including Nonlinear Hebbian Learning (NHL) [11], Active Hebbian Learning (AHL) [12], and Data-Driven

Nonlinear Hebbian Learning (DDNHL) [10] demonstrate computational efficiency but are constrained by convergence to equilibrium states rather than replicating complete temporal trajectories. Second, bio-inspired computational intelligence recasts weight learning as global optimization. Population-based strategies – Genetic Algorithm (GA) [13], Particle Swarm Optimization (PSO) [14], Big Bang-Big Crunch optimization [15] – navigate high-dimensional weight spaces through evolutionary mechanisms. Recent innovations include memetic algorithms with mutual information [16], decomposition strategies for large-scale FCMs [16], and multi-agent genetic algorithms [13]. Third, hybrid algorithms fuse Hebbian and population-based approaches, initializing weights via Hebbian rules before evolutionary refinement [11, 12].

Subsequently, advanced frameworks have emerged: Quantum Fuzzy Cognitive Maps (QFCM) [14] harness quantum superposition for parallel exploration, yet their convergence behavior remains stochastic and hardware-dependent, offering no theoretical optimality guarantees; Interval Type-2 Fuzzy Cognitive Maps (IT2FCM) [13] address uncertainty through double-layered fuzzy representations, but the nested type-reduction operations escalate computational complexity to super-quadratic levels; Neuro Fuzzy Cognitive Maps (NFCM) [17] embed neural networks for gradient-based learning, introducing non-convex loss landscapes susceptible to local optima and requiring extensive hyperparameter tuning. Across all three variants, the absence of provable convergence guarantees and the exponential growth of computational cost with problem scale represent fundamental, unresolved barriers to reliable deployment.

More broadly, contemporary FCM weight learning faces three notable and interconnected limitations. **Limited temporal modeling capacity:** methods have demonstrated persistent difficulty in capturing long-term dependencies and nonlinear patterns, with population-based algorithms documented to exhibit premature convergence in high-dimensional spaces [13, 14]. **Escalating computational overhead:** parameters grow quadratically with node count [16], while quantum operations (QFCM), nested fuzzy evaluations (IT2FCM), and iterative neural training (NFCM) introduce substantially increased computational demands [13, 14, 17]. **Sensitivity to data noise:** learning algorithms have shown reduced robustness under noisy conditions, a notable limitation given FCMs' inherent uncertainty-handling design [18].

Beyond algorithmic challenges lies a fundamental limitation: conventional FCMs assume system state at time $t + 1$ depends solely on time t . This first-order Markov assumption contradicts dynamical systems theory where evolution exhibits memory effects and multi-scale correlations. Real-world processes show current behavior shaped by various historical horizons – recent events exert immediate effects while earlier conditions contribute attenuated yet significant lag influences.

Therefore, this paper introduces high-order FCM and a paradigm-shifting approach leveraging convex optimization theory [19]. In sharp contrast to the stochastic convergence of QFCM, the super-quadratic complexity of IT2FCM, and the non-convex optimization of NFCM, we reformulate high-order FCM learning as constrained convex optimization, enabling interior-point algorithms that provide rigorous global optimality guarantees and achieve polynomial-time $\mathcal{O}(n^3)$ complexity – independent of initialization and free from local optima traps. This principled formulation fundamentally distinguishes the proposed method from all existing FCM variants: whereas prior approaches sacrifice either convergence certainty or computational tractability, interior-point methods simultaneously guarantee both. Consequently, stochastic search inefficiencies are eliminated, deterministic solution quality is ensured, and computational requirements are dramatically reduced

relative to competing advanced FCM frameworks. Simultaneously, the framework systematically optimizes high-order decay factor, activation steepness, and model complexity through principled parameter scanning.

Section 2 establishes mathematical preliminaries and problem formulation. Section 3 presents convex optimization methodology, algorithm, and hyperparameter tuning. Section 4 reports experimental findings with comparative analysis. Section 5 concludes with insights and future directions.

2. Expansion of FCM from 1-Order to High-Order. The mathematical formulation for the dynamic behaviors of first-order FCMs is presented in Equation (1).

$$A_i(t+1) = f \left(\sum_{j=1}^m (w_{ij} A_j(t)) \right) \quad (1)$$

In this equation, $A_j(t)$ denotes the activation level of the j th concept at time t , $w_{ij} \in [-1, 1]$ represents the weight coefficient for causal interaction from concept j to concept i , m indicates the total number of concepts, and f is a nonlinear continuous non-increasing activation function – commonly the sigmoid function (Equation (2)) [20].

$$f(x) = \frac{1}{1 + e^{-\lambda x}} \quad (2)$$

Here, λ serves as the steepness factor: higher values create steeper activation curves, enhancing responsiveness to input changes.

High-order FCMs provide more accurate characterization of temporal behaviors in complex dynamic systems. A key principle is that recent states exert greater influence than earlier ones. The mathematical description is given in Equation (3).

$$A_i(t+1) = f \left(\alpha_t \sum_{j=1}^m (w_{1ij} A_j(t)) + \alpha_{(t-1)} \sum_{j=1}^m (w_{2ij} A_j(t-1)) + \cdots + \alpha_1 \sum_{j=1}^m (w_{tij} A_j(1)) \right) \quad (3)$$

Here, $w_{1ij}, \dots, w_{(t-1)ij}, w_{tij}$ correspond to causal weights from the j th node to the i th node at time steps $t, t-1, \dots, 1$. Once finalized, the FCM model performs successive iterations for predictive functionality.

The decay coefficients $\alpha_t, \alpha_{t-1}, \dots, \alpha_1$ regulate historical state influence across different time scales, following a decreasing sequence ($\alpha_t > \alpha_{t-1} > \cdots > \alpha_1$) where recent states have stronger impact on future state $A_i(t+1)$ than distant ones, capturing temporal attenuation in dynamic systems. This recency-biased weighting structure is well-motivated by the established principle in time series analysis that the predictive relevance of historical observations diminishes as they recede in time [21]; exponential decay in particular provides a parsimonious and mathematically tractable mechanism for implementing this attenuation, as it assigns geometrically decreasing weights to progressively older states while maintaining a non-zero contribution from the entire history.

The decay coefficient for the k th order is $\alpha_k = \alpha^{q-k-1}$, where $k \in \{0, 1, \dots, q-1\}$ indexes historical states, $\alpha \in (0, 1)$ controls decay rate, and q represents FCM order. The exponential form α^{q-k-1} is motivated by the analogy with exponential smoothing methods [22], where a single base parameter α governs the rate at which past influences are attenuated: a smaller α enforces rapid forgetting of distant states, while α close to unity allows long-range dependencies to persist. This assigns unity weight to the most recent state and exponentially diminishing weights to earlier states, implementing recency-biased memory for dynamic temporal dependencies.

3. Proposed Method. This section presents a comprehensive elaboration on the newly proposed approach designed for modeling and predicting large-scale, complex time series data. A key innovation of the proposed methodology lies in the efficient and accurate estimation of Fuzzy Cognitive Map (FCM) weight coefficients and hyperparameter optimization. The implementation steps of this approach are systematically outlined in Algorithm 1 below.

Algorithm 1 Learning weight matrices for high-order FCM via least squares

Require:

- 1: $TfType$: activation function type; $\lambda > 0$: activation parameter; $\alpha \in (0, 1)$: decay parameter
- 2: m : FCM nodes; $\beta = 0.0004$: regularization parameter; $\{\mathbf{x}_t\}_{t=1}^n$: time series data; q : FCM order

Ensure:

- 3: $\mathbf{W}_k \in \mathbb{R}^{m \times m}$: learned weight matrix for order k , $k = 1, \dots, q$
 - 4: **Initialization:** $\mathbf{W}_k \leftarrow \mathbf{0}^{m \times m}$ for $k = 1, \dots, q$
 - 5: **Data Preparation:** Construct feature matrices $\mathbf{H}_k$ from historical states at lag k for $k = 1, \dots, q$
 - 6: Construct target matrix $\mathbf{Y}$ from future states
 - 7: **Optimization:**
 - 8: **for** each node $j = 1$ to m **do**
 - 9: Transform target: $\tilde{\mathbf{y}}_j \leftarrow -\ln\left(\frac{1}{\mathbf{y}_{[:,j]}} - 1\right)$
 - 10: Solve convex optimization problem with variables $\{\mathbf{w}_{kj}\}_{k=1}^q$:
 - 11:
$$\min_{\{\mathbf{w}_{kj}\}} \left\| \lambda \sum_{k=1}^q \alpha_k \mathbf{H}_k \mathbf{w}_{kj} - \tilde{\mathbf{y}}_j \right\|_2 + \beta \sum_{k=1}^q \|\mathbf{w}_{kj}\|_1$$
 - 12: Subject to: $\|\mathbf{w}_{kj}\|_\infty \leq 1$ for $k = 1, \dots, q$
 - 13: Update $\mathbf{w}_{kj}$ for all $k = 1, \dots, q$ using interior-point solver
 - 14: **end for**
 - 15: **return** $\{\mathbf{W}_k\}_{k=1}^q$ where $\mathbf{W}_k = [\mathbf{w}_{k1}, \dots, \mathbf{w}_{km}]$
-

Given that a q -order FCM is employed in this framework, the weight structure consists of q distinct matrices $\{\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_q\}$, with each element constrained to the interval $[-1, 1]$.

Consider a fuzzy time series dataset comprising c training sequences, where each sequence contains n temporal observations. For the s th sequence, the state at time t is denoted as $\mathbf{A}_t^s$. The high-order FCM computes the next state through the sigmoid activation function as follows:

$$\mu_{(t+1)j}^s = \frac{1}{1 + e^{-\lambda(\alpha_t \mathbf{A}_t^s \mathbf{w}_{1j} + \alpha_{t-1} \mathbf{A}_{t-1}^s \mathbf{w}_{2j} + \dots + \alpha_{t-q+1} \mathbf{A}_{t-q+1}^s \mathbf{w}_{qj})}} \quad (4)$$

where $\lambda > 0$ controls the steepness of the sigmoid function, and $\mathbf{w}_{kj}$ represents the j th column of the weight matrix $\mathbf{W}_k$.

The learning objective is to minimize the discrepancy between predicted states $\mu_{(t+1)j}^s$ and observed states across all sequences, nodes, and time steps. This yields the mean squared error criterion:

$$\min_{\{\mathbf{W}_k\}} J = \frac{1}{cm(n-q)} \sum_{s=1}^c \sum_{t=q}^{n-1} \sum_{j=1}^m (\mu_{(t+1)j}^s - \hat{\mu}_{(t+1)j}^s)^2 \quad (5)$$

To facilitate numerical optimization, we apply the inverse sigmoid transformation to Equation (4), converting the nonlinear activation into a linear form:

$$\lambda (\alpha_t \mathbf{A}_t^s \mathbf{w}_{1j} + \alpha_{t-1} \mathbf{A}_{t-1}^s \mathbf{w}_{2j} + \cdots + \alpha_{t-q+1} \mathbf{A}_{t-q+1}^s \mathbf{w}_{qj}) = -\ln \left(\frac{1}{\mu_{(t+1)j}^s} - 1 \right) \quad (6)$$

Since the coefficient α_{t-k+1} depends only on the lag position k rather than on the absolute time t , we write $\alpha_k \equiv \alpha_{t-k+1}$ for brevity. Denoting the right-hand side as $\tilde{\mathbf{y}}_j = -\ln(1/\mu_{(t+1)j}^s - 1)$ and consolidating observations across all sequences and time steps, we obtain a compact matrix formulation:

$$\lambda \sum_{k=1}^q \alpha_k \mathbf{H}_k \mathbf{w}_{kj} = \tilde{\mathbf{y}}_j \quad (7)$$

where $\mathbf{H}_k$ aggregates the historical states at lag k from all training sequences.

Due to measurement noise and model approximation errors, Equation (7) generally lacks an exact solution. Therefore, we reformulate the problem as a constrained least squares optimization:

$$\begin{aligned} \min_{\{\mathbf{w}_{kj}\}} \quad & \left\| \lambda \sum_{k=1}^q \alpha_k \mathbf{H}_k \mathbf{w}_{kj} - \tilde{\mathbf{y}}_j \right\|_2 \\ \text{subject to} \quad & \|\mathbf{w}_{kj}\|_\infty \leq 1, \quad k = 1, 2, \dots, q \end{aligned} \quad (8)$$

The infinity-norm constraints ensure that all weights remain within $[-1, 1]$, preserving the interpretability of causal relationships in FCM.

For large-scale FCMs, sparse weight matrices confer substantial computational benefits. To promote sparsity, we augment the objective function with L_1 regularization penalties on each weight vector:

$$\begin{aligned} \min_{\{\mathbf{w}_{kj}\}} \quad & \left\| \lambda \sum_{k=1}^q \alpha_k \mathbf{H}_k \mathbf{w}_{kj} - \tilde{\mathbf{y}}_j \right\|_2 + \beta \sum_{k=1}^q \|\mathbf{w}_{kj}\|_1 \\ \text{subject to} \quad & \|\mathbf{w}_{kj}\|_\infty \leq 1, \quad k = 1, 2, \dots, q \end{aligned} \quad (9)$$

where β is a regularization hyperparameter controlling sparsity level. The value $\beta = 0.0004$ was determined empirically on the training set, yielding a favorable balance between sparsity promotion and prediction accuracy. As reported in [23], sensitivity experiments across varying $\beta \in \{0.0001, 0.0002, 0.0003, 0.0004\}$ show that RMSE fluctuations remain below 2.3% in all tested configurations, and below 0.5% in over 85% of cases, confirming that model performance is robust to small perturbations of β . Therefore, we fix $\beta = 0.0004$ throughout all experiments.

Equation (9) constitutes a convex optimization problem combining L_1 and L_2 norms. We employ interior-point methods to solve this problem efficiently. The convexity guarantees global optimum convergence for each node-specific weight vector $\{\mathbf{w}_{1j}, \dots, \mathbf{w}_{qj}\}$. By solving Equation (9) independently for all m nodes, we obtain the complete weight matrices $\{\mathbf{W}_1, \dots, \mathbf{W}_q\}$ that fully parameterize the high-order FCM for fuzzy time series forecasting.

4. Experiment. In this section, we conduct a comprehensive empirical study to validate the proposed approach (i.e., TLPHF) through a dual-faceted evaluation strategy. The first component examines the model's predictive capabilities across synthetic datasets

(both noise-free and noise-corrupted) as well as real-world temporal sequences. The second component provides a systematic benchmarking analysis against established baseline methods and contemporary algorithms to assess relative accuracy gains.

The convex optimization formulation presented in Equation (9) is solved using the ECOS solver, which is seamlessly integrated within the Python ecosystem and provides robust convergence guarantees for conic programs.

Our experimental evaluation uses ten real-world data sets spanning four canonical time series categories: trend-dominant, seasonally-varying, stochastic, and hybrid patterns to evaluate the effectiveness of the proposed method. The 10 data sets include the Soybean Price, Yahoo Stock, Sunspots, Image Recognition Task Execution Times (Mac1), Condition Monitoring of Hydraulic Systems (Cooling Efficiency), Appliances Energy Prediction (Press), Appliances Energy Prediction (Power), Power Consumption of Tetouan City (Humidity), Jena Climate (ρ), and Jena Climate (T_{pot}) time series. We benchmark the proposed TLPFH against nine competing approaches, including classical methods (AR, MR, ARIMA), neural networks (NARnet, LSTM), and FCM algorithms (TLSP-DE, QFCM, NFCM, IT2FCM).

Following standard practice, we employ temporal split validation with 85% training and 15% testing partitions. All experiments are conducted on uniform hardware (HP ProBook 440 G7, Intel Core i7-10510U, 8GB RAM) for fair comparison. Model performance is evaluated using Root Mean Square Error (RMSE) as defined in Equation (10), where lower values indicate superior forecasting accuracy.

$$RMSE = \sqrt{\frac{1}{r} \sum_{t=1}^r (x_t - \hat{x}_t)^2} \quad (10)$$

Among it, r is the prediction step, that is, the prediction level, x_t and $\hat{x}_t$ represent the actual observed value and predicted value at time t , respectively.

4.1. Synthetic data. To validate the proposed high-order FCM learning algorithm, we first generate synthetic data through pre-constructed simulated high-order FCMs. For a simulated FCM with m nodes and order q , we randomly select target nodes according to density level $\rho = 0.3$ and assign random weights in $[-0.8, 0.8]$ to connection edges. We generate q weight matrices $\{\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_q\}$ with exponential decay coefficients reflecting temporal decay characteristics. Dynamic stability is ensured by normalizing each weight matrix $\mathbf{W}_k$ whenever its spectral radius $\rho(\mathbf{W}_k)$ exceeds 0.9. This criterion is grounded in the stability theory of FCM-based dynamical systems: a sufficient condition for the iteration to converge to a unique fixed point is that the spectral radius of the weight matrix is strictly less than one [24]. The threshold of 0.9, rather than the theoretical boundary of 1.0, is deliberately adopted as a conservative safety margin. As demonstrated in [25], FCM dynamics are susceptible to slow convergence or non-convergent behavior (e.g., cyclic or chaotic patterns) when the spectral radius approaches unity; setting the threshold at 0.9 provides sufficient distance from this near-critical region to ensure reliable convergence throughout the synthetic data generation process.

We generate $c = 10$ training sequences, each containing n temporal observations. Initial states are created to base vectors from uniform distribution $[0.3, 0.7]$, keeping values within $[0.1, 0.9]$. Given sigmoid steepness $\lambda = 1.0$ and iterations $k = 100$, the simulated FCM performs iterative computation according to Equation (4), yielding c response matrices of size $n \times m$ where $n = k + q$.

Three systematic parameter scanning experiments evaluate algorithm performance: 1) fixed $m = 5$, $\alpha = 0.5$, scanning $\lambda \in [1.0, 4.5]$ with step 0.5; 2) fixed $\lambda = 1.0$, $\alpha = 0.5$,

testing $m \in \{5, 7, 8, 12, 15, 20\}$; 3) fixed $m = 5$, $\lambda = 1.0$, scanning 8 values in $\alpha \in [0.20, 0.90]$.

For each configuration, we extract training samples using previous q states as inputs and next-moment state as target. Each node's constrained convex optimization problem with L_1 regularization (Equation (9)) is solved using SLSQP. Learned weights are multiplied by decay coefficients and sparsified through adaptive thresholding (2% of maximum absolute weight).

Model performance is evaluated using prediction error RMSE. Results are visualized through tables and 3D surface plots showing model error variations with respect to steepness factor λ , node number m , and decay factor α versus FCM order.

Tables 1, 2, and 3 present the RMSE under different parameter configurations. Figure 1 illustrates the variation of RMSE performance under different parameter combinations.

It is worth noting that the key parameters (concept number m , steepness coefficient λ , and high-order decay factor α) are analyzed independently in the experiments; nevertheless, joint optimization of these parameters is expected to further refine their combinations and improve prediction accuracy.

4.1.1. Influence of sigmoid steepness parameter λ . According to Table 1, the steepness parameter λ significantly impacts learning accuracy for fixed network size ($m = 5$), decay factor ($\alpha = 0.5$). For low-order FCMs (Order=1), RMSE decreases from 1.36×10^{-1} at $\lambda = 1.0$ to 3.8×10^{-2} at $\lambda = 4.5$, a 72% reduction. High-order systems (Order=70) show similar trends with 78% improvement. At Order=100, RMSE converges to approximately 1.2×10^{-3} across λ values from 1.5 to 4.5, indicating reduced sensitivity to λ variations in higher-order systems. The reason for this is that larger λ enhances the system's nonlinear expression capability for capturing complex dynamics.

TABLE 1. Impact of steepness parameter λ and FCM order q on model performance with noise ($m = 5$, $\alpha = 0.5$)

λ	Order=1	Order=2	Order=5	Order=7	Order=10	Order=20	Order=30	Order=40	Order=50	Order=70	Order=100
1.0	1.36e-1	9.5e-2	3.7e-2	2.7e-2	1.7e-2	9.4e-3	6.9e-3	5.1e-3	6.6e-3	3.8e-3	1.3e-3
1.5	1.10e-1	6.4e-2	2.4e-2	1.8e-2	1.1e-2	6.2e-3	4.6e-3	3.5e-3	5.0e-3	2.6e-3	1.2e-3
2.0	8.7e-2	4.8e-2	1.8e-2	1.3e-2	8.4e-3	4.6e-3	3.5e-3	2.6e-3	3.6e-3	1.9e-3	1.2e-3
2.5	7.0e-2	3.8e-2	1.5e-2	1.1e-2	6.7e-3	3.7e-3	2.7e-3	2.0e-3	2.8e-3	1.6e-3	1.2e-3
3.0	5.8e-2	3.1e-2	1.2e-2	8.8e-3	5.5e-3	3.1e-3	2.3e-3	1.7e-3	2.3e-3	1.3e-3	1.2e-3
3.5	4.9e-2	2.7e-2	1.0e-2	7.5e-3	4.7e-3	2.6e-3	1.9e-3	1.4e-3	1.9e-3	1.1e-3	1.2e-3
4.0	4.3e-2	2.3e-2	8.9e-3	6.5e-3	4.1e-3	2.3e-3	1.7e-3	1.2e-3	1.6e-3	9.6e-4	1.2e-3
4.5	3.8e-2	2.1e-2	7.9e-3	5.8e-3	3.6e-3	2.0e-3	1.5e-3	1.1e-3	1.4e-3	8.4e-4	1.2e-3

4.1.2. Influence of network size m . Table 2 shows network size impact under fixed conditions ($\lambda = 1.0$, $\alpha = 0.5$). Increasing nodes from 5 to 20 reduces RMSE by 71% for Order=1 and 60% for Order=70, achieving minimum RMSE of 3.98×10^{-4} at Order=100 with $m = 20$. The reason for this is that increasing nodes improves state representation

TABLE 2. Impact of network size m and FCM order q on model performance with noise ($\lambda = 1.0$, $\alpha = 0.5$)

Nodes m	Order=1	Order=2	Order=5	Order=7	Order=10	Order=20	Order=30	Order=40	Order=50	Order=70	Order=100
5	4.43e-2	3.21e-2	2.73e-2	1.97e-2	1.50e-2	8.89e-3	5.33e-3	3.76e-3	3.28e-3	1.50e-3	1.39e-3
7	3.94e-2	2.86e-2	2.07e-2	1.61e-2	1.34e-2	7.42e-3	4.26e-3	3.00e-3	2.20e-3	1.76e-3	1.10e-3
8	3.29e-2	2.36e-2	1.85e-2	1.40e-2	1.23e-2	6.15e-3	3.80e-3	2.53e-3	2.10e-3	1.48e-3	1.03e-3
12	2.15e-2	2.01e-2	1.10e-2	1.03e-2	7.99e-3	3.95e-3	2.54e-3	1.88e-3	1.42e-3	9.70e-4	6.75e-4
15	1.81e-2	1.42e-2	9.50e-3	7.99e-3	6.16e-3	3.10e-3	1.92e-3	1.44e-3	1.14e-3	7.90e-4	5.51e-4
20	1.30e-2	1.01e-2	7.03e-3	6.48e-3	4.67e-3	2.43e-3	1.43e-3	1.05e-3	8.12e-4	5.95e-4	3.98e-4

capacity and parallel information processing. However, computational cost increases significantly.

4.1.3. Influence of temporal decay factor α . The decay factor α exhibits complex relationships with learning accuracy (Table 3). For fixed parameters ($m = 5$, $\lambda = 1.0$), low orders show minimal α sensitivity, while high orders demonstrate significant impact. At Order=50, RMSE varies from 2.06×10^{-3} ($\alpha = 0.20$) to 1.68×10^{-2} ($\alpha = 0.90$). Smaller decay factors consistently yield superior performance in high-order systems, with $\alpha = 0.20$ achieving minimum RMSE across most orders. The reason for this is that smaller α optimizes historical information weighting and reduces noise accumulation, especially beneficial for high-order systems.

TABLE 3. Impact of decay factor α and FCM order q on model performance with noise ($m = 5$, $\lambda = 1.0$)

α	Order=1	Order=2	Order=5	Order=7	Order=10	Order=20	Order=30	Order=40	Order=50	Order=70	Order=100
0.20	1.40e-1	6.97e-2	1.46e-2	1.07e-2	6.81e-3	2.89e-3	2.49e-3	2.02e-3	2.06e-3	1.37e-3	8.21e-4
0.30	1.40e-1	7.52e-2	1.72e-2	1.22e-2	7.84e-3	3.34e-3	2.89e-3	2.36e-3	2.72e-3	1.59e-3	8.28e-4
0.40	1.40e-1	8.07e-2	2.07e-2	1.43e-2	9.22e-3	3.94e-3	3.32e-3	2.82e-3	2.90e-3	1.88e-3	8.42e-4
0.50	1.40e-1	8.62e-2	2.50e-2	1.72e-2	1.12e-2	4.80e-3	4.02e-3	3.48e-3	3.76e-3	2.26e-3	8.63e-4
0.60	1.40e-1	9.17e-2	3.06e-2	2.13e-2	1.42e-2	6.12e-3	5.05e-3	4.49e-3	4.70e-3	2.82e-3	8.93e-4
0.70	1.40e-1	9.72e-2	3.78e-2	2.72e-2	1.90e-2	8.43e-3	6.81e-3	6.21e-3	5.61e-3	3.74e-3	9.51e-4
0.80	1.40e-1	1.03e-1	4.68e-2	3.58e-2	2.69e-2	1.32e-2	1.03e-2	9.58e-3	9.42e-3	5.55e-3	1.10e-3
0.90	1.40e-1	1.08e-1	5.81e-2	4.82e-2	4.03e-2	2.52e-2	2.08e-2	1.97e-2	1.68e-2	1.03e-1	1.55e-3

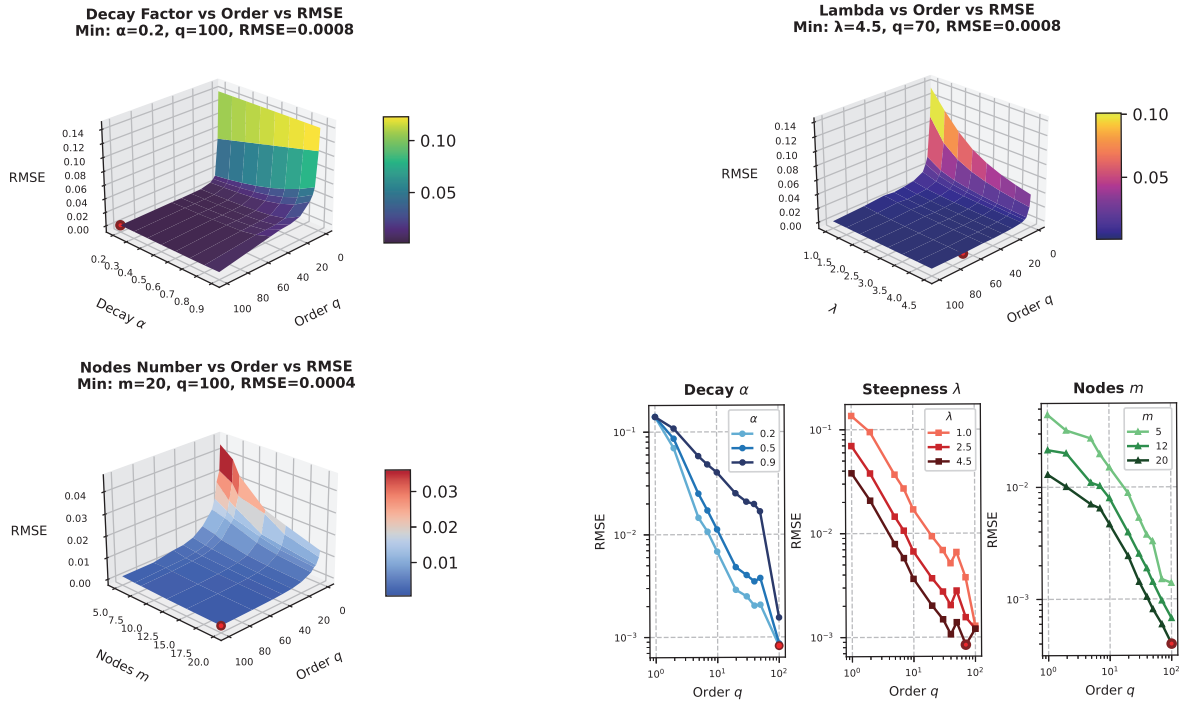


FIGURE 1. Plots of RMSE trend for parameters as FCM order increasing for synthetic data

4.1.4. Optimal parameter configuration. The optimal configuration for minimum RMSE is $m = 20$, $\lambda = 4.5$, $\alpha = 0.20$, and $q = 70$ -100, achieving RMSE below 6.0×10^{-4} . For practical applications balancing accuracy and efficiency, we recommend $m = 8$, $\lambda = 3.0$ -4.0, $\alpha = 0.30$ -0.40, and $q = 30$ -50, providing RMSE of 2.0 - 3.0×10^{-3} with learning time under 1000 seconds.

4.2. Synthetic noisy data. To emulate measurement errors, additive Gaussian noise ($\sigma = 0.02$) is applied to synthetic data, creating noise-corrupted response matrices. Experimental results are also visualized through tables and 3D surface plots showing model error variations with respect to λ , m , and α versus FCM order. The parameter impacts are similar to noise-free synthetic data.

Figure 2 illustrates RMSE performance under different parameter combinations. The optimal configuration for minimum RMSE is $m = 20$, $\lambda = 1.0$, $\alpha = 0.5$, and $q = 70$, achieving RMSE of 6.0×10^{-4} .

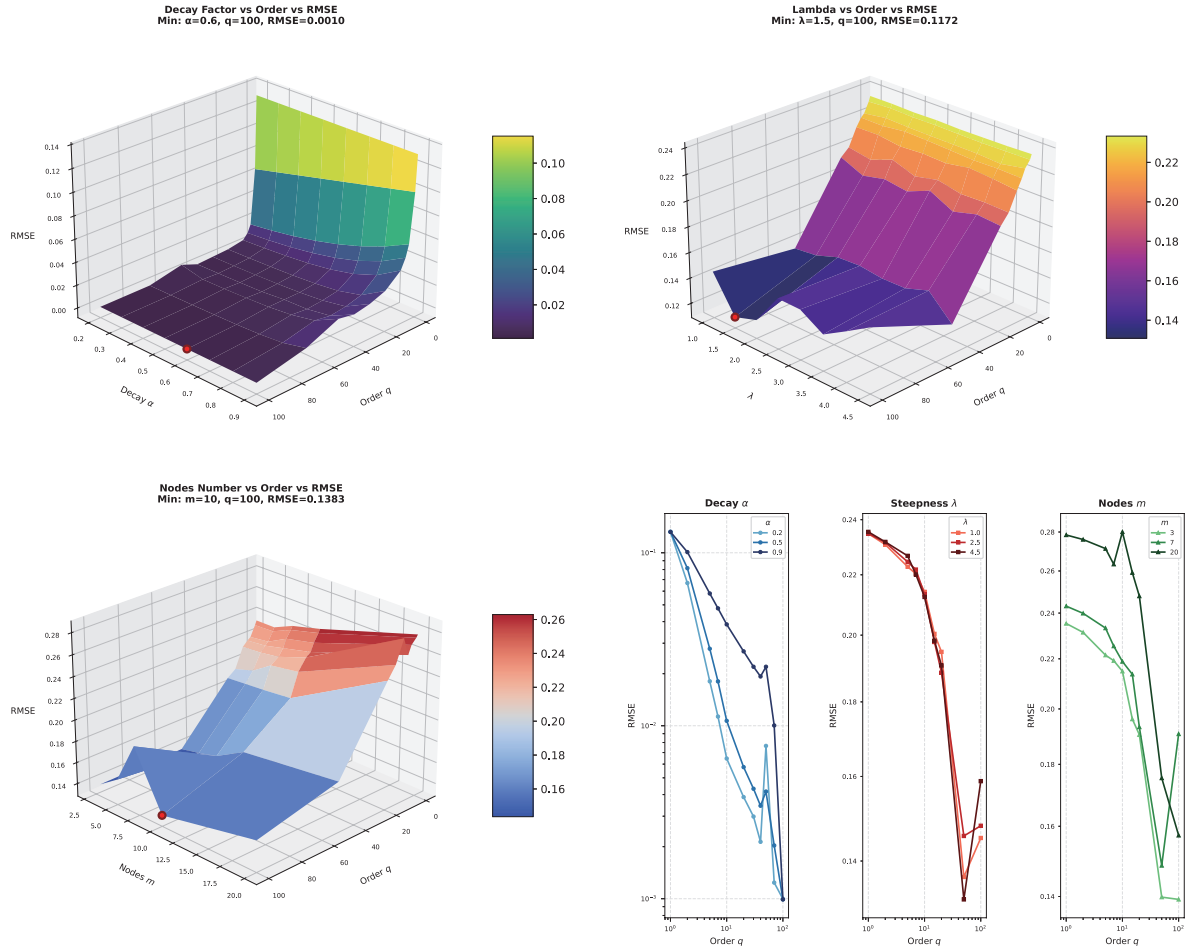


FIGURE 2. Plots of RMSE trend for parameters as FCM order increasing for synthetic data with noise

For lambda effects, increasing λ from 1.0 to 4.5 consistently reduces RMSE across all orders: 76% improvement at Order=1 and 67% at Order=70. At Order=100, RMSE converges to 1.0×10^{-3} regardless of λ , indicating diminished sensitivity in extremely high-order systems.

Network size significantly impacts accuracy: expanding from $m = 5$ to 20 nodes yields 14% RMSE reduction at Order=1 and 67% at Order=70, achieving minimum error of 8.4×10^{-4} at $m = 20$, Order=50. However, computational cost increases.

The decay factor exhibits order-dependent behavior, with smaller values ($\alpha = 0.20$ - 0.40) consistently outperforming larger values in high-order systems. At Order=50, RMSE varies 2.9-fold from 7.6×10^{-3} ($\alpha = 0.20$) to 2.2×10^{-2} ($\alpha = 0.90$), while low-order systems show minimal α sensitivity. This amplifies at Order=70 with 8.3-fold variation, though Order=100 converges to 9.9×10^{-4} across all α values.

4.3. Real data. Since parameter scanning results across all 10 datasets exhibit consistent trends, the Yahoo Finance stock dataset – being highly stochastic and volatile – is selected as a representative example for detailed parameter analysis to further demonstrate the prediction accuracy of high-order FCM, as it poses greater predictive challenges and more effectively illustrates the sensitivity and robustness of the proposed method under varying parameter configurations. A comprehensive parameter scanning approach covers order values from 1 to 100, with $\lambda \in [1.0, 4.5]$, $\text{nodes_num} \in [3, 20]$, and $\text{decay_factor} \in [0.20, 1.00]$. Figure 3 illustrates RMSE performance under different parameter combinations. Results show maximum RMSE of 0.2354 at Order=1 and minimum RMSE of 0.1172 at $\lambda = 1.5$, Order=100.

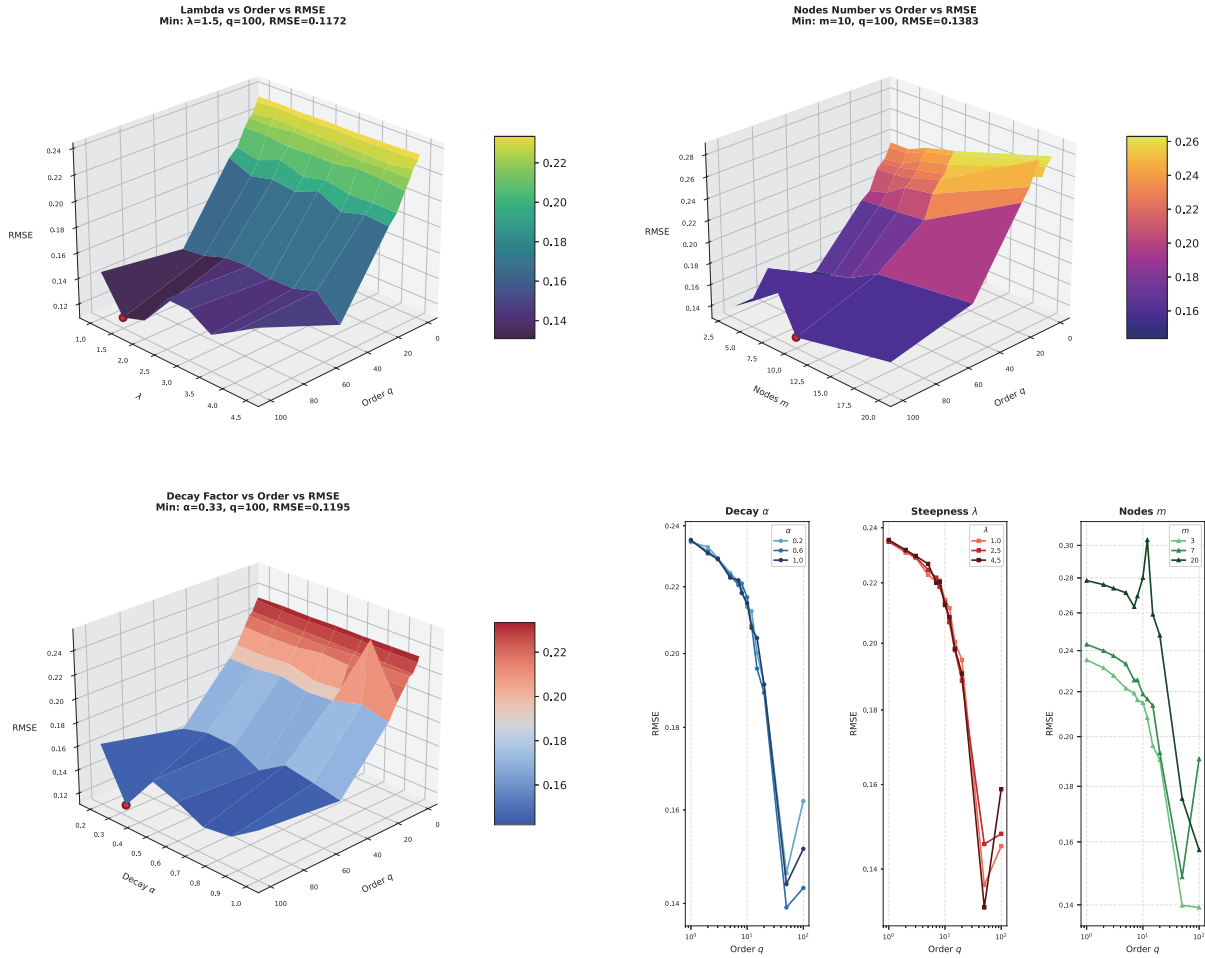


FIGURE 3. Plots of RMSE trend for parameters as FCM order increasing for real data

Five key trends emerge: 1) RMSE generally decreases with increasing FCM order across all parameters; 2) λ parameter exhibits moderate impact on RMSE: low-order FCMs show minimal λ sensitivity, while $\lambda = 1.5$ achieves optimal performance (RMSE = 0.1172) for high-order systems; 3) $m = 3$ provides optimal efficiency, achieving RMSE = 0.1398 at Order=50, while higher node numbers increase costs. Increasing model order is more computationally efficient than increasing nodes, offering lower cost and faster computation; 4) Moderate α values (0.33-0.60) generally yield better performance; 5) Among all examined parameters, variations in FCM order exhibit the most pronounced impact on RMSE performance.

The recommended balanced configuration is $\lambda = 1.5$, $m = 3$, $\alpha = 0.33$, and Order=50, achieving RMSE = 0.1386, representing excellent accuracy-efficiency trade-off for real-time applications.

4.4. Comparison with other methods. Table 4 evaluates TLPHF against nine competing approaches across ten real-world datasets. TLPHF consistently delivers competitive performance across diverse domains, achieving average accuracy enhancement of approximately 63% over traditional methods (AR, MA, ARIMA, NARnet, LSTM) and 27% over advanced FCM-based approaches (TLSP-DE, QFCM, NFCM, IT2FCM).

TABLE 4. Experimental results obtained by TLPHF, AR, MA, ARIMA, NARnet, LSTM, TLSP-DE, QFCM, NFCM and IT2FCM

Data set (Train/Test)	TLPHF	AR	MA	ARIMA	NARnet	LSTM	TLSP-DE	QFCM	NFCM	IT2FCM
Soybean Price (1148/203)	14.32	67.14	66.68	67.997	58.81	66.96	42.75	38.49	28.34	33.25
Yahoo Stock (1069/189)	0.12	1.599	1.568	1.896	1.455	1.449	1.331	0.605	0.768	1.361
Sunspots (2797/493)	7.11	15.39	14.51	16.08	9.998	9.84	7.57	8.53	6.47	7.26
Image Recognition Task Execution Times (Mac1) (850/150)	0.0027	0.0052	0.0049	0.0068	0.0036	0.0035	0.0029	0.0029	0.0027	0.0029
Hydraulic Systems Cooling Efficiency (1874/331)	0.13	0.256	0.239	0.285	0.195	0.176	0.136	0.117	0.196	0.120
Appliances Energy Prediction (Press) (16775/2960)	0.009	0.042	0.039	0.046	0.026	0.018	0.013	0.015	0.014	0.012
Appliances Energy Prediction (Power) (16775/2960)	85.6	262.08	224.29	269.35	184.56	101.08	86.34	79.93	107.66	86.18
Power Consumption of Tetouan City (Humidity) (44553/7862)	0.49	1.540	1.489	1.605	1.013	0.808	0.598	0.586	0.576	0.551
Jena Climate (ρ) (60473/10672)	0.28	0.847	0.790	0.934	0.678	0.331	0.321	0.337	0.301	0.299
Jena Climate (T_{pot}) (2797/493)	0.09	0.313	0.326	0.306	0.288	0.176	0.176	0.104	0.150	0.115

Performance differences reflect fundamental approach characteristics. Traditional linear models offer computational efficiency and interpretability but struggle with nonlinear dynamics in complex datasets. Advanced approaches better model nonlinear patterns, yielding improved outcomes. However, these approaches typically use metaheuristic optimization for parameter learning, potentially converging to local optima. In contrast, TLPHF's convex optimization guarantees global optimum convergence, explaining its superior performance versus other FCM variants and consistently higher prediction accuracy.

5. Conclusion. This paper proposes a high-order FCM-based prediction method for large-scale complex time series using constrained convex optimization for weight learning. This approach overcomes traditional FCM limitations in capturing long-term dependencies while addressing computational costs, scalability issues, and noise sensitivity of existing swarm intelligence methods. The experimental framework covers synthetic data, noisy

data, and ten public benchmark datasets across four time series types: trend, seasonal, random, and composite. The procedure includes 1) data normalization; 2) FCM weight computation via constrained convex optimization; 3) model prediction; and 4) comparative validation. Results demonstrate 63% improvement over traditional methods and 27% over state-of-the-art methods, with superior computational efficiency across different FCM scales and noisy scenarios. The study analyzes key parameter impacts (concept number m , steepness coefficient λ , and high-order decay factor α) but treats parameters independently without joint optimization. Future work will develop automated hyperparameter tuning mechanisms (Bayesian optimization), adaptive parameter selection strategies, and extensions to multivariate and streaming data scenarios to enhance prediction accuracy and practicality.

REFERENCES

- [1] B. Kosko, Fuzzy cognitive maps, *International Journal of Man-Machine Studies*, vol.24, no.8, pp.65-75, 1986.
- [2] M. Hagiwara, Extended fuzzy cognitive maps, *Proc. of IEEE International Conference on Fuzzy Systems*, pp.795-801, 1992.
- [3] G. Nápoles, I. Grau, R. Bello and R. Grau, Two-steps learning of fuzzy cognitive maps for prediction and knowledge discovery on the HIV-1 drug resistance, *Expert Systems with Applications*, vol.41, no.3, pp.821-830, 2014.
- [4] A. Amirkhani, M. Kolahdoozi, C. Wang and L. Kurgan, Prediction of DNA-binding residues in local segments of protein sequences with fuzzy cognitive maps, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol.17, no.4, pp.1372-1382, 2020.
- [5] M. Bose and K. Mali, Designing fuzzy time series forecasting models, *International Journal of Approximate Reasoning*, vol.111, no.C, pp.78-99, 2019.
- [6] Y. Liu and L. Wang, Long-term prediction of time series based on fuzzy time series and information granulation, *Granular Computing*, vol.9, no.2, pp.1-13, 2024.
- [7] N. H. Shafii, R. Alias, S. R. Shamsuddin and D. S. M. Nasir, Fuzzy time series for projecting school enrolment in Malaysia, *Journal of Computing Research and Innovation*, vol.6, no.1, pp.11-21, 2021.
- [8] S. Mei, Y. Zhu, X. Qiu, X. Zhou, Z. Zu, A. V. Boukhanovsky and P. M. A. Sloom, Individual decision making can drive epidemics: A fuzzy cognitive map study, *IEEE Transactions on Fuzzy Systems*, vol.22, no.2, pp.264-273, 2014.
- [9] D. T. Huyen, N. T. Mai, S. Tanaka, K. Hashikura, M. A. S. Kamal, I. Murakami and K. Yamada, Vietnamese-focused color-based models for dyads expressing for robots, *International Journal of Innovative Computing, Information and Control*, vol.21, no.2, pp.287-305, 2025.
- [10] D. Shan, W. Lu and J. Yang, The data-driven fuzzy cognitive map model and its application to prediction of time series, *International Journal of Innovative Computing, Information and Control*, vol.14, no.5, pp.1583-1602, 2018.
- [11] G. Feng, W. Lu and J. Yang, The modeling of time series based on least square fuzzy cognitive map, *Algorithms*, vol.14, no.3, pp.69, 2021.
- [12] J. Zhang, Z. Tao, J. Liu, X. Liu and H. Chen, A hybrid interval-valued time series prediction model incorporating intuitionistic fuzzy cognitive map and fuzzy neural network, *Journal of Forecasting*, vol.44, no.1, pp.93-111, 2025.
- [13] A. Amirkhani, M. Shirzadeh, T. Kumbasar and B. Mashadi, A framework for designing cognitive trajectory controllers using genetically evolved interval type-2 fuzzy cognitive maps, *International Journal of Intelligent Systems*, vol.37, no.1, pp.305-335, 2022.
- [14] M. Kolahdoozi, A. Amirkhani, M. H. Shojaeefard and A. Abraham, A novel quantum inspired algorithm for sparse fuzzy cognitive maps learning, *Applied Intelligence*, vol.49, no.10, pp.3652-3667, 2019.
- [15] E. Bozdog and C. Kadaifci, A distance-based approach to fuzzy cognitive maps using Pythagorean fuzzy sets, *International Journal of Fuzzy Systems*, vol.27, no.1, pp.93-109, 2025.
- [16] X. Zou and J. Liu, A mutual information-based two-phase memetic algorithm for large-scale fuzzy cognitive map learning, *IEEE Transactions on Fuzzy Systems*, vol.26, no.4, pp.2120-2134, 2018.
- [17] A. Amirkhani, M. Shirzadeh, M. H. Shojaeefard and A. Abraham, Controlling wheeled mobile robot considering the effects of uncertainty with neuro-fuzzy cognitive map, *ISA Transactions*, vol.100, pp.454-468, 2020.

- [18] G. Feng, W. Lu, W. Pedrycz, J. Yang and X. Liu, The learning of fuzzy cognitive maps with noisy data: A rapid and robust learning method with maximum entropy, *IEEE Transactions on Cybernetics*, vol.51, no.4, pp.2080-2092, 2021.
- [19] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [20] A. Rehman, Neural computing for online arabic handwriting recognition using hard stroke features mining, *International Journal of Innovative Computing, Information and Control*, vol.17, no.1, pp.177-191, 2021.
- [21] O. Orang, P. C. de Lima e Silva and F. G. Guimarães, Time series forecasting using fuzzy cognitive maps: A survey, *Artificial Intelligence Review*, vol.56, no.8, pp.7733-7794, 2023.
- [22] E. S. Gardner, Exponential smoothing: The state of the art – Part II, *International Journal of Forecasting*, vol.22, pp.637-666, 2006.
- [23] D. Shan, L. Wang, W. Lu and J. Chen, Convex optimization based high-order fuzzy cognitive map modeling and its application in time series predicting, *IEEE Access*, vol.12, pp.12683-12698, 2024.
- [24] I. Á. Harmati, M. F. Hatwagner and L. T. Kóczy, Global stability of fuzzy cognitive maps, *Neural Computing and Applications*, vol.35, no.10, pp.7283-7295, 2022.
- [25] T. Koutsellis, G. Xexakis, K. Koasidis, A. Nikas and H. Doukas, Parameter analysis for sigmoid and hyperbolic transfer functions of fuzzy cognitive maps, *Operational Research*, vol.22, no.5, pp.5733-5763, 2022.

Author Biography



Dan Shan received M.S. degree in Control Theory and Control Engineering from Dalian University of Technology, China, in 2004.

She joined Dalian Neusoft University of Information, China, in 2009. She is currently a professor in Department of Electronic Engineering. Her current research interests include computational intelligence, fuzzy modeling and granular computing, knowledge discovery and data mining, fuzzy intelligent systems and low-cost integrated automation systems.



Xiaoxu Zhang received the M.S. degree in System LSI from Waseda University, Japan, in 2010.

She joined Dalian Neusoft University of Information, China, in 2016. She is currently an assistant professor in Department of Electronic Engineering. Her current research interests include computational intelligence and low-cost integrated automation systems.



Wei Lu received the M.S. degree and the Ph.D. degree in Control Theory and Control Engineering from Dalian University of Technology, China, in 2004 and 2015, respectively.

He joined Dalian University of Technology in 2004. He is currently a professor in the School of Control Science and Engineering, Dalian University of Technology. His current research interests include computational intelligence, fuzzy modeling and granular computing, knowledge discovery and data mining, fuzzy intelligent systems and low-cost integrated automation systems. He also serves as a frequent reviewer for many international journals including IEEE Transactions on Fuzzy Systems, IEEE Transactions on Cybernetics, Knowledge-Based Systems, Applied Soft Computing, Experts Systems and Applications, Int. J. of Approximate Reasoning, Int. J. of Granular Computing (Springer) as well as international conferences.



Huanmeixi Hou is an undergraduate student majoring in Integrated Circuit Design and Integrated System at Dalian Neusoft University of Information, China.

Her current research interests include computational intelligence and low-cost integrated automation systems.