

3-22-2022

Modeling Time Series Using Multi-Modality Fuzzy Cognitive Maps

Guoliang Feng

1.School of Control Science and Engineering, Dalian University of Technology, Dalian 116023, China,;
fengguoliang911@foxmail.com

Wei Lu

1.School of Control Science and Engineering, Dalian University of Technology, Dalian 116023, China;2.Liaoning Industrial Equipment Distributed Control Technology Innovation Center, Dalian 116023, China,; luwei@dlut.edu.cn

Jianhua Yang

1.School of Control Science and Engineering, Dalian University of Technology, Dalian 116023, China;2.Liaoning Industrial Equipment Distributed Control Technology Innovation Center, Dalian 116023, China;

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the Artificial Intelligence and Robotics Commons, Computer Engineering Commons, Numerical Analysis and Scientific Computing Commons, Operations Research, Systems Engineering and Industrial Engineering Commons, and the Systems Science Commons

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation.

Modeling Time Series Using Multi-Modality Fuzzy Cognitive Maps

Abstract

Abstract: A multi-modality modeling method for time series data based on fuzzy cognitive maps is proposed to address the problem that a single model is difficult to accurately reflect the multi-modal characteristics of time series. *The bootstrap method is used to select multiple sub-sequences from the original time series which contain the diverse modality in the original time series. The fuzzy cognitive map sub-models are constructed on each sub-sequences respectively. The formed sub-models are further merged by means of granular computing method and the merging performance with different weighting strategies is analyzed.* The developed multi-modal model not only has prediction abilities at the numeric and interval level, but also has better interpretability. Experimental results on public datasets exhibit the usefulness and satisfactory efficiency of the proposed approach.

Keywords

fuzzy cognitive maps, granular computing, time series, model weight

Recommended Citation

Guoliang Feng, Wei Lu, Jianhua Yang. Modeling Time Series Using Multi-Modality Fuzzy Cognitive Maps[J]. Journal of System Simulation, 2022, 34(3): 543-554.

用于时间序列数据建模的多模态模糊认知图

冯国亮¹, 卢伟^{1,2*}, 杨建华^{1,2}

(1. 大连理工大学控制科学与工程学院, 辽宁大连116023; 2. 辽宁省工业装备分布式控制专业技术创新中心, 辽宁大连116023)

摘要: 针对单一模型难以准确反映时间序列多种变化模态的问题, 提出了一种基于模糊认知图的时间序列数据多模态建模方法。该方法使用随机自助法选取多个子序列, 以包含各种变化模态。在各个子序列上分别建立子模糊认知图模型。使用粒计算方法对子模型进行有效融合; 并分析了不同权重策略融合的性能。所建立的模型不仅可以对时间序列数据进行数值及区间预测, 还具有更好的语义解释性。在公开数据集上的实验结果证明了该方法的有效性和实用性。

关键词: 模糊认知图; 粒计算; 时间序列; 模型权重

中图分类号: TP274; TP391.9 文献标志码: A 文章编号: 1004-731X(2022)03-0543-12

DOI: 10.16182/j.issn1004731x.joss.20-0834

Modeling Time Series Using Multi-Modality Fuzzy Cognitive Maps

Feng Guoliang¹, Lu Wei^{1,2*}, Yang Jianhua^{1,2}

(1. School of Control Science and Engineering, Dalian University of Technology, Dalian 116023, China;

2. Liaoning Industrial Equipment Distributed Control Technology Innovation Center, Dalian 116023, China)

Abstract: A multi-modality modeling method for time series data based on fuzzy cognitive maps is proposed to address the problem that a single model is difficult to accurately reflect the multi-modal characteristics of time series. The bootstrap method is used to select multiple sub-sequences from the original time series which contain the diverse modality in the original time series. The fuzzy cognitive map sub-models are constructed on each sub-sequences respectively. The formed sub-models are further merged by means of granular computing method and the merging performance with different weighting strategies is analyzed. The developed multi-modal model not only has prediction abilities at the numeric and interval level, but also has better interpretability. Experimental results on public datasets exhibit the usefulness and satisfactory efficiency of the proposed approach.

Keywords: fuzzy cognitive maps; granular computing; time series; model weight

引言

时间序列数据建模与预测一直是一个经典研究课题, 它通过对历史数据分析、理解、建模从而对未来作出推测。时间序列预测已经广泛用于多种领域, 诸如经济、工业、气候预测等领域, 其意义在于能够有效地对这些领域的决策提供依

据^[1]。研究人员利用线性系统理论、随机过程理论、黑箱理论等方法开发了许多经典的时间序列数值模型, 如差分移动平均自回归模型^[2]、神经网络模型^[3-4]、支持向量机^[5-6]、贝叶斯网络^[7]等模型。这些模型已广泛应用于各个领域, 在数值预测上显示出了良好的预测性能。然而这些数值模型由于可解释性低, 难以被人们理解。为了使时

收稿日期: 2020-10-30 修回日期: 2021-01-09

基金项目: 国家自然科学基金(62073056, 61876029); 国家重点研发计划(2019YFB1705103); 中央高校基本科研业务费专项资金(DUT20LAB129)

第一作者: 冯国亮(1984-), 男, 博士生, 研究方向为模糊认知图学习及应用。E-mail: fengguoliang911@foxmail.com

通讯作者: 卢伟(1976-), 男, 博士, 教授, 研究方向为计算智能和模糊智能系统。E-mail: luwei@dlut.edu.cn

间序列模型具有可解释性,并保持对不确定性的鲁棒性,模糊集理论提供了一种可行的选择。模糊化被引入到了时间序列建模中,继而构建模糊时间序列模型^[8]。模糊时间序列模型的主要步骤有:定义域区间的划分、区间的语言描述、时间序列的模糊化、建立模糊时间序列变量间的模糊逻辑关系、计算预测值并解模糊化^[9-10]。此外,在一些领域只需知道近似值即可,获得时间序列的精确值并不是必须的。如对于每日气温的预报,气温的最低值与最高值更具有指导意义;在股票市场中,投资者对未来一段时间的下降或上升的区间范围更感兴趣。在这些领域中,人们更希望在符号或信息粒度级别上对时间序列进行建模预测,而不仅仅是预测时间序列的精确数值。为了构造更高层次的时间序列模型的表示方法,研究者又进一步提出了基于粒计算的时间序列建模方法^[11-12]。粒计算^[13]是一种以人为中心的信息处理框架平台,它采用“合理信息粒化原则”^[14-15]将繁杂抽象的信息按照一定的规则划分成简单易于理解的信息粒。该方法不过度追求模型的精确数值,而是在模型的“精确性”与“可解释性”之间进行调和处理,从而使时间序列的动态行为更容易被理解。

作为一种软计算工具,模糊认知图(fuzzy cognitive maps, FCMs)以知识表示为基础来实现推理过程,能够捕捉给定动态系统的行为,从而被用来对复杂的系统进行描述和建模^[16-17]。FCMs由概念节点和有向权重组成,节点代表了系统的主要行为特征,权重反映了概念节点之间的因果关系。基于FCMs的知识推理能力,Stach等^[18]利用实数编码遗传算法来学习模糊认知图,对时间序列进行数字和语言层面的预测。为了进一步提高FCMs模型的预测精度,Lu等^[19]设计了一种高阶模糊认知图对时间序列进行建模和预测;采用模糊c-均值聚类算法构建FCMs框架,利用遗传算法学习FCMs的权值。高阶模糊认知图的概念节点不仅依赖于所有概念节点的最后状态,还依赖于节点的多个超前状态,从

而提高了FCMs的逼近能力。Pedrycz等^[20]设计了一种使用信息粒来描述数值时间序列的框架。在时间序列的幅值与幅值变化的空间中,利用模糊c-均值聚类算法对数据进行聚类,每个聚类中心作为FCMs的一个节点,对时间序列进行建模预测。Salmeron等^[21]使用动态优化的策略,对FCMs的权重进行动态更新,以提高时间序列的预测精度。该方法根据时间序列的局部特性动态调整模糊认知图的权值。Froelich等^[22]使用合理粒度原则和模糊c-均值聚类原理对时间序列进行粒度近似。通过这种方式,在生成时间序列的近似值时设计了特殊性和普遍性之间的折衷方式。另外,在对得到的粒子进行粒度化和语言项赋值后,时间序列可以更容易地被人类理解,这是直接分析原始数据所不具有的特性。

对时间序列进行建模分析时,现有的方法通常对整个时间序列进行整体建模,即假设时间序列只有一个变化模态。然而,时间序列在不同时间段内的变化模式、不同区间内的变化模式都是不尽相同的;此外,主体时间序列也可能是来源于多个客体。因此直接对整体时间序列进行建模,挖掘出时间序列的所有变化特征比较困难。如果能对时间序列的每种变化模态进行建模,则对序列的知识表达将更全面。本文在给定的时间序列中使用有放回随机策略选取大量短序列,以包含时间序列中的各种变化模态。然后,使用模糊认知图工具在每个子序列上进行建模。对每个模型的输出,借助粒化的方法进行融合,最终获得时间序列的输出。所提出的方法不仅能获得时间序列的数字预测值,还能获得预测区间;并具有语义的描述,使人们更容易理解时间序列的变化特征。

1 背景知识

1.1 模糊认知图

模糊认知图 FCMs (fuzzy cognitive maps) 的拓扑结构是一种带反馈环的权重有向图,它通过节点和权重边来描述物理系统的特性。如图1所示,

一个典型的 FCMs 由概念节点和节点之间的权重边组成, 这些节点代表待研究系统中的主要变量, 权重边表示各个节点之间的相互影响程度。

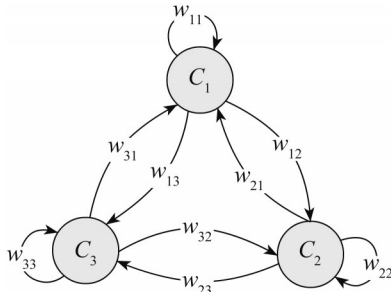


图 1 含有 3 个节点的模糊认知图
Fig. 1 FCMs with three nodes

标准 FCMs 所蕴含的语义可以由一个 4 元组 (C, W, A, f) 来表示^[23]。假设 FCMs 有 n 个节点, $C = \{C_1, C_2, \dots, C_n\}$ 是 n 个概念节点的集合, W 为 $n \times n$ 维的权重矩阵:

$$W = \begin{bmatrix} w_{11} & \dots & w_{1n} \\ \vdots & & \vdots \\ w_{n1} & \dots & w_{nn} \end{bmatrix} \quad (1)$$

式中: $w_{ij} \in [-1, 1]$ 其值的正负号与大小分别反映了节点间因果关系的影响方向和程度。 $w_{ij} > 0$ 表示随着 C_i 的增加会导致 C_j 的增加, 且影响程度为 w_{ij} , 反之亦然; $w_{ij} < 0$ 表示随着 C_i 的增加会导致 C_j 的减少, 且影响程度为 $|w_{ij}|$, 反之亦然; 而 $w_{ij} = 0$ 表示 C_i 与 C_j 之间没有因果关系。 $A: C_i \rightarrow A_i(t)$, $A_i(t)$ 是节点 C_i 在 t 时刻的状态值, 该值是动态变化的。 f 为激活函数, 它是一个非线性单调递增的函数, 将与目标节点有关联的所有节点的输入状态值整合到激活函数的定义域内。节点 C_j 在 $t+1$ 时刻的状态值可以通过式(2)计算:

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

一个常用的激活函数为 Sigmoid 函数:

$$f(x) = \frac{1}{1 + e^{-\lambda x}} \quad (3)$$

该激活函数将节点的状态值限定在区间 $[0, 1]$, 其中 $\lambda > 0$ 决定了激活函数的陡峭程度。

1.2 信息粒化

信息粒由 Zadeh 最先提出^[24-25], 并给出模糊信息粒的一般性定义:

$$g = (x \text{ is } G) \text{ is } \mu$$

式中: x 为论域 U 上的变量; G 为 U 的一个模糊子集; μ 为 x 属于模糊集 G 的概率; g 为信息粒, 其包含的语义是“ x 属于模糊集 G 的概率水平为 μ ”。信息粒化体现在数据的抽象与信息知识表示过程中, 不同层次的抽象水平可以由不同大小的信息粒来表示。

Pedrycz 等^[26-27]在此基础上进一步发展了信息粒的可解释性研究, 将信息粒化看作“以人为中心”的信息处理过程, 用于对实际模型以及决策过程的知识表示。信息粒化遵循合理性原则, 即合理的信息粒化一方面要尽可能地表征所研究的数据; 另一方面, 应尽可能地对数据进行抽象, 并进行语义表示。简单来讲, 粒化的原则表现在 2 个方面: 合理性和特殊性^[28]。合理性表示有多少数据被构造的信息粒所覆盖, 要求信息粒覆盖尽可能多的数据; 而特殊性则表示产生的信息粒应具有清晰的语义含义, 要求构造的信息粒尽量小, 这样语义也就越清晰。可以看出, 这二者的要求是相互矛盾的。

假设有一组数据集 $Y = \{y_1, y_2, \dots, y_p\}$ 。通过确定信息粒所覆盖数据集中元素的个数, 来量化信息粒的合理性。令 $i \in [1, p]$, 定义信息粒化的合理性 cov 为

$$cov = card\{y_i | y_i \in g\} \quad (4)$$

式中: $card(\cdot)$ 表示基数, 即数据集 Y 有多少元素属于信息粒 g 。通过信息粒 g 的宽度来量化特殊性, 特殊性 sp 定义为

$$sp = \exp(-\alpha|U-L|) \quad (5)$$

这里, U 和 L 分别表示信息粒所覆盖数据幅度的上界与下界, 上下界之间的距离 $|U-L|$ 越小, 则说明信息粒越具有特殊性。 $\alpha > 0$ 信息粒度水平, 为构造的信息粒提供了灵活性, 其值大小影响信

息粒的特殊性。

信息粒化的关键是构建一个函数，使得信息粒的合理性与特殊性都优。然而，增大信息粒的合理性必然导致特殊性减小，反之亦然。为了平衡这2个特性，将构造信息粒的问题转换为如下优化问题：

$$\arg \max_{u,L} J = cov \cdot sp \quad (6)$$

2 多模态模型的建立

使用模糊认知图进行时间序列建模时，一般遵循以下4个步骤：时间序列的模糊化、模糊认知图学习、建模及预测、预测值的解模糊化^[29]。本节将详细阐述整个建模的过程。

2.1 时间序列的模糊化

对于一时间序列 $X = \{x_1, x_2, \dots, x_m\}$ ， m 为该时间序列所包含的样本数据的个数。采用三角形模糊集对时间序列 X 进行模糊化处理。三角模糊集的隶属度函数公式为

$$\mu(x; a, b, c) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x < b \\ \frac{c-x}{c-b} & b \leq x < c \\ 0 & x \geq c \end{cases} \quad (7)$$

式中：参数 a 和 c 分别对应三角形的左右2个顶点， b 为三角形中间顶点，且 $a \leq b \leq c$ 。

假设将时间序列 X 的定义域用 n 个点 $\{P_1, P_2, \dots, P_n\}$ 划分成连续且不相交的区间，每个分割点是一个模糊语言变量。则样本数据 x_i 对这些模糊集的隶属度表示为 $\{\mu_1(x_i), \mu_2(x_i), \dots, \mu_n(x_i)\}$ ，令 $k=1, 2, \dots, n$ ，那么模糊集的隶属度满足以下约束条件：

- (1) $\mu_k(x_i) \geq 0$;
- (2) $\sum_{k=1}^n \mu_k(x_i) = 1$;
- (3) $\mu_k(x_i) = 1$;
- (4) 如果 $x \notin [P_{k-1}, P_{k+1}]$ ，那么 $\mu_k(x_i) = 0$;

(5) $\mu_k(x_i)$ 在 $[P_{k-1}, P_k]$ 内单调递增，在 $[P_k, P_{k+1}]$ 内单调递减。

图2给出了一个三角形模糊集划分的例子。样本值被归一化到了区间 $[0, 1]$ 之内，样本空间使用 $\{0, 0.25, 0.5, 0.75, 1\}$ 这5个点划分。则每个样本值对应这些分割点的隶属度为 $\{\mu_1, \mu_2, \mu_3, \mu_4, \mu_5\}$ ，每个样本值所对应的5个隶属度的和为1。

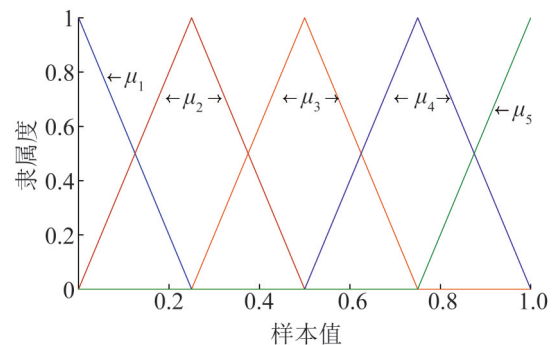


图2 三角形模糊集划分
Fig. 2 Triangular fuzzy partition

时间序列 $X = \{x_1, x_2, \dots, x_m\}$ 经过三角形模糊化后，得到模糊时间序列形成的矩阵 U ：

$$U = \begin{bmatrix} u_{11} & \dots & u_{1n} \\ \vdots & & \vdots \\ u_{m1} & \dots & u_{mn} \end{bmatrix} \quad (8)$$

式中： m 为时间序列 X 的样本数量； n 为划分的模糊集的个数； U 每行的元素为对应时间序列样本的模糊隶属度值，并且每行元素的和为1。三角模糊语言变量的向量为

$$P = (P_1, P_2, \dots, P_n) \quad (9)$$

2.2 模糊认知图建模

模糊时间序列确定后，即可用于FCMs的学习。FCMs作为一种符号表示的推理机制，其推理式(2)可以写成如下向量的形式：

$$A_j(t+1) = f(A(t)W_j) \quad (10)$$

式中： $A(t) = (A_1(t), A_2(t), \dots, A_n(t))$ 为FCMs的所有节点在 t 时刻的状态值， $A_j(t+1)$ 为FCMs的第 j 个节点在 $t+1$ 时刻的状态值； $W_j = (w_{1j}, w_{2j}, \dots, w_{nj})^T$ 为权重矩阵 W 的第 j 列。作为激活函数的 Sigmoid

函数具有唯一的逆函数, 因此将式(7)进行逆变换后, 可以表示为

$$f^{-1}(A_j(t+1)) = A(t)W_j \quad (11)$$

显然, 式(11)为一个线性方程。其中, $f^{-1}(y) = -\frac{1}{\lambda} \ln \frac{1-y}{y}$ 为 Sigmoid 函数的逆函数。

时间序列模糊化处理后, 得到了模糊时间序列矩阵 U , 用于 FCMs 权重 W 的学习。令 $Z = (U_1, U_2, \dots, U_{m-1})^T$, 其中, $U_i = (u_{i1}, u_{i2}, \dots, u_{im})$; $Y_j = f^{-1}((u_{2j}, u_{3j}, \dots, u_{mj})^T)$, 将 Z 与 Y_j 带入式(11), 可得:

$$Y_j = ZW_j \quad (12)$$

Z 与 Y_j 可以分别看作历史数据的自变量与因变量, 因此, FCMs 权重的学习问题可以转化为最小二乘问题的求解。为了求解 W_j 构造如下目标函数:

$$\begin{aligned} \arg \min_{W_j} & \|ZW_j - Y_j\|_2 \\ \text{s.t.} & \|W_j\|_\infty \leq 1 \end{aligned} \quad (13)$$

式中: 目标函数中的 $\|ZW_j - Y_j\|_2$ 为最小化实际值与预测值之间的最小二乘误差, 以期获得 W_j 的合理近似解。约束项 $\|W_j\|_\infty \leq 1$ 保证了权重矩阵每一个元素的值均在区间 $[-1, 1]$ 内。其中, 无穷范数 $\|W_j\|_\infty = \max\{|w_{1j}|, |w_{2j}|, \dots, |w_{mj}|\}$ 。式(13)是一个带有线性约束的标准凸优化问题, 当前已有多种凸优化技术来求解该问题, 如常用的内点法、原始对偶法^[30]。

时间序列的预测是基于历史数据来推测未来的数据。本文中 FCMs 的推理机制, 被用于时间序列的预测。已知 t 时刻 FCMs 的状态向量为 $U_t = (u_{t1}, u_{t2}, \dots, u_{tm})$, 一旦权重矩阵 W 被确定, 那么 $t+1$ 时刻状态向量的预测值 \hat{U}_{t+1} 可以通过式(14)计算所得:

$$\hat{U}_{t+1} = f(U_t W) \quad (14)$$

通过式(14)可以顺序地获得模糊认知图下一时刻的状态向量, 因此得到了状态向量的预测矩阵 \hat{U} 。为了获得最终的时间序列预测的数值, 还需

要将状态向量进行解模糊化:

$$\hat{x}_{t+1} = \frac{\sum_{j=1}^n \hat{u}_{(t+1)j} P_j}{\sum_{j=1}^n \hat{u}_{(t+1)j}} \quad (15)$$

式中: n 为 FCMs 的节点数; P_j 为 FCMs 的第 j 节点的三角模糊语言值, 见式(9)。

2.3 多模态建模

时间序列中数据随着时间的变化以及来源可能是多样的, 即时间序列中通常包含多个模态的特征, 为了更深入地理解时间序列的特征, 需要建立的模型能挖掘并利用这些模态特征进行建模。而且对大型的时间序列直接建模, 会导致模型过于复杂, 不利于人们的理解。因此, 采用多模态的建模方法对时间序列进行分析, 从而挖掘出时间序列子系统的变化特征, 最终来表征整个时间序列的变化特征。

图 3 给出了 FCMs 多模态建模系统的基本结构, 使用自助法(Bootstrap)从时间序列 X 中随机地抽取 p 个子集, 每个子集上训练一个 FCMs 子模型, 然后对各个子模型的输出利用信息粒化的方法进行融合, 获得最终的输出数值以及输出数值的上下界限值。图中, T 为时间子序列。

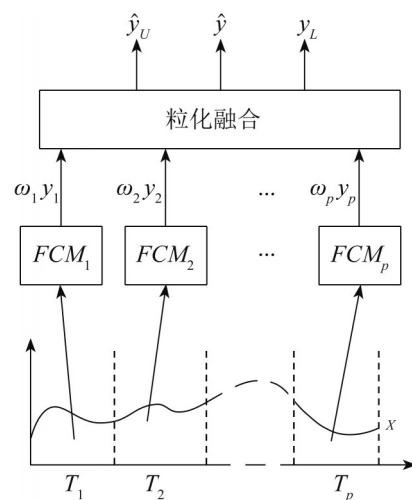


图 3 模型框架
Fig. 3 Modeling framework

多模态FCMs模型的具体步骤如下:

(1) 对长度为 m 的时间序列 X , 使用模糊集数为 n 的三角模糊集进行模糊化处理, 得到模糊时间序列 U 。

(2) 将长度为 m 的模糊时间序列 U , 使用自助法(Bootstrap)随机地选取 p 个子集, 需要注意的是所分割的子序列不一定是顺序相邻的。本研究中每个子序列的起点均是随机选取的, 并且子序列的长度是相同的。设定每个子序列的长度为 k , 则第 $i(i=1, 2, \dots, p)$ 个子序列表示为

$$\bar{U}^i = \begin{bmatrix} \bar{u}_{11}^i & \cdots & \bar{u}_{1n}^i \\ \vdots & & \vdots \\ \bar{u}_{ki}^i & \cdots & \bar{u}_{kn}^i \end{bmatrix} \quad (16)$$

(3) 对于每个子序列 \bar{U}^i , 使用2.2节所述方法进行FCMs建模, 得到 p 个FCMs模型。

(4) 如何对子模型的输出进行融合是实现多模态模型的关键。假设 p 个模型的权重分别为 $\omega_1, \omega_2, \dots, \omega_p$, 模型输出与其权重形成了一组带权重的数据集 $Y = \{(y_1, \omega_1), (y_2, \omega_2), \dots, (y_p, \omega_p)\}$, 这里, $\omega_i \in [0, 1]$ 是数据 y_i 的权重 $i=1, 2, \dots, p$ 。对数据集 Y 进行信息粒化处理, 构造如下优化目标函数:

$$\arg \max_{U, L} J = cov \cdot sp \cdot flag \quad (17)$$

这里, 构造的合理性函数为 $cov = \Sigma \omega_i, y(i) \in [\hat{y}_L, \hat{y}_U], i \in [1, p]$, 特殊性函数为 $sp = \exp(-\alpha|\hat{y}_L - \hat{y}_U|)$, \hat{y}_L 与 \hat{y}_U 分别表示信息粒所覆盖数据幅值的上界与下界。用于时间序列预测时, $flag$ 代表上一时刻实际值是否在上下界内:

$$flag = \begin{cases} 1, & y \in [\hat{y}_L, \hat{y}_U] \\ 0, & \text{other} \end{cases}$$

构造目标函数的目的是使构造的信息粒所覆盖数据的权重尽可能地大, 同时信息粒的区间尽可能地小。在对时间序列进行区间预测时, 总是希望所得到区间值越小越好, 但是同时也要尽量令实际值在预测区间内, 式(17)所求得的区间的物理意义即为预测区间的上下界限。

(5) 假设上一步获得的信息粒的区间内有 q 个

数据, 重新计算这些数据的权重为 $\lambda_1, \lambda_2, \dots, \lambda_q$, 最终预测值计算公式为

$$\hat{y} = \sum_{i=1}^q \lambda_i y_i \quad (18)$$

式中: $y_i \in [\hat{y}_L, \hat{y}_U]$, \hat{y}_L 与 \hat{y}_U 则分别作为预测区间的下界与上界。

对于模型的预测性能评价, 使用均方根误差(RMSE)进行评价:

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N \|x(t) - y(t)\|^2} \quad (19)$$

式中: $x(t)$ 表示 t 时刻的实际值; $y(t)$ 为 t 时刻的模拟值。显然, RMSE值越小, 则表示模型的预测性能越好, N 为数据的个数。

为了获得多模态模型最终的输出结果, 需要确定各子模型的权重, 权重策略采用以下3种方法:

(1) 平均权重

使用每个子模型输出的算术平均值作为多模态系统的输出, 即 p 个子模型的权重均为 $\frac{1}{p}$ 。第 i 个子模型的权重 ω_i 为

$$\omega_i = \frac{1}{p} \quad (20)$$

(2) 模型权重

计算出每个模型的RMSE, RMSE越小则表明其对应的子模型的权重应越大。取每个子模型RMSE的倒数, 并归一化。第 i 个子模型的权重 ω_i 为

$$\omega_i = \frac{1/RMSE_i}{\sum_{i=1}^p 1/RMSE_i} \quad (21)$$

(3) 动态权重

由于模型上一个时刻的预测值与实际值的绝对差值, 来确定下一时刻该模型的权重。这种方式下, 每个子模型的权重是实时变化的。第 i 个子模型的权重 $\omega_i(t)$ 为

$$\omega_i(t) = \frac{1/|x_i(t-1) - y_i(t-1)|}{\sum_{i=1}^p 1/|x_i(t-1) - y_i(t-1)|} \quad (22)$$

子模型的权重确定后, 多模态模型的输出结

果计算为

$$\hat{y} = \sum_{i=1}^p \omega_i y_i \quad (23)$$

3 实验结果与分析

3.1 评价指标

使用RMSE对所构建模型的数值预测性能进行评价, 见式(19)。对于区间预测性能的评估定义预测区间覆盖概率PICP (prediction interval coverage probability)和预测区间归一化平均宽度PINAW (prediction interval normalized average width)进行评价^[31-32], 其公式为

$$PICP = \frac{1}{N} \sum_{i=1}^N \delta_i \quad (24)$$

$$PINAW = \frac{1}{NR} \sum_{i=1}^N (\hat{y}_U - \hat{y}_L) \quad (25)$$

式中: N 为预测样本数; δ_i 为一个布尔值: $\delta_i =$

$$\begin{cases} 1, & x \in [\hat{y}_L, \hat{y}_U] \\ 0, & x \notin [\hat{y}_L, \hat{y}_U] \end{cases}$$

x 是实际值, \hat{y}_L 与 \hat{y}_U 分别为预测

区间的下界与上界; R 为时间序列数据幅值变化的范围。综合这2个评价指标, 构造基于宽度的覆盖度指标CWC (coverage width-based criterion):

$$CWC = PINAW(1 + PICP \cdot e^{-PICP}) \quad (26)$$

这些评价指标中, RMSE用于评价预测值与真实值之间的误差, RMSE越小, 表明预测模型的性能越好。PICP与PINAW用于评价预测区间的指标, PICP越大表明预测区间包含实际样本值的概率越高; PINAW用于评估预测区间的大小, 该指标越小说明预测区间越小, 一个优秀的预测模型应具有PICP高, PINAW, CWC与RMSE小的特点。

3.2 数据实验

为了展示本文所提方法的工作过程并验证其有效性, 使用人工与实际数据集对提出的方法进行实验。用到的数据集既有小型也有大型时间序列, 用到的序列如下:

(1) Mackey-Glass (MG)序列。MG是一个混

沌、非周期、不收敛的序列, 该序列通过以下时滞微分方程产生:

$$\frac{dx_t}{dt} = \frac{0.2x_{t-\tau}}{1+x_{t-\tau}^{10}} - 0.1x_t \quad (27)$$

令 $x_0 = 1.2$, $\tau = 17$, 且当 $t < 0$ 时, $x_t = 0$; 将这些参数代入式(27), 产生的序列含有1 201个样本, 该序列如图4(a)所示。

(2) AUD/USD汇率。澳元与美元每月汇率的时间序列, 该数据集包含1969年7月—1995年8月期间的每月汇率, 共有314个样本, 如图4(b)所示。

(3) Vatnsdalsa河流量。1972年1月1日—1974年12月31日期间, Vatnsdalsa河每日的平均流量(m^3/s), 共有1 095个样本, 如图4(c)所示。

(4) 风速时间序列。该数据集的时间间隔为10 min, 共有39 195个样本, 如图4(d)所示。

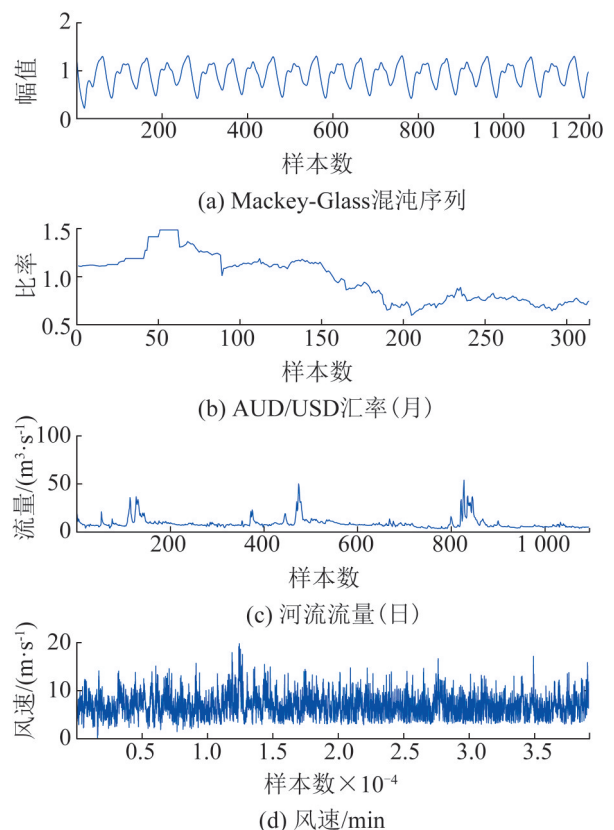


图4 实验数据集
Fig. 4 Experimental datasets

实验中,上述时间序列数据集样本的前80%作为训练集,后20%作为测试集。模型的参数如表1所示,为了降低模型的复杂性,每个子FCMs模型的节点数 $n=3$,每个子序列的样本数 $k=5$,使用自助法随机选取的子序列个数 $p=100$,粒度水平 $\alpha=1$ 。

表1 模型参数设置
Table 1 Model parameter setting

序号	主要参数	量值
1	λ	5
2	子模型数量 p	100
3	节点数 n	3
4	子序列长度 k	5
5	粒度水平 α	1

实验平台为CPU 2.3 GHz,内存4G的笔记本电脑,软件使用Matlab R2018a。实验中采样的数据每次选择的起点是不同的,每段子序列的变化也是不尽相同的。因此,使用这些子序列所训练的FCMs模型也是不尽相同的。使用大量的随机选择方法的目的是为了尽可能地包含到时间序列数据间的变化情况。

FCMs的3个节点依照数据论域的模糊集所赋予的语义分别定义为幅值低、幅值中、幅值高。对时间序列使用三角模糊集进行模糊化处理时,分别选取时间序列的最小值、中值以及最大值作为模糊集的语言变量。表2给出了各个数据集对应的FCMs模型中节点语义所对应的具体数值。以MG序列为例,语义“幅值低”对应的数值为0.219 2,“幅值中”对应的数值为0.766 5,“幅值高”对应的数值为1.311 7。当给出语言变量的语义描述后,模型输出的预测值在获得数值的同时,也可以根据隶属度的大小被分类为对应的语义上。

图5~8分别画出了这4个数据集的测试集,采用动态权重方式所获得的预测曲线及预测区间情况。图5~7为对应数据集的全部测试集预测情况;由于风速数据量较大,图8截取了部分数据进行

展示。黑色曲线为实际值,红色曲线为预测值,青色区域为预测区间;图中灰色区、白色以及黄色带状区域分别对应预测值的语义为“幅值低”、“幅值中”、“幅值高”。观察这些图可知所提出模型不仅能够得到预测值的精确数值,还可以获得预测值的区间值,同时可以对预测值进行语义的描述。从而从多个维度对数据进行描述,更便于人们的理解。

表2 FCMs节点的语义描述
Table 2 Semantics of FCMs nodes

数据集	幅值低	幅值中	幅值高
MG	0.219 2	0.766 5	1.313 7
AUD/USD 汇率	0.598 0	1.043 0	1.488 0
Vatnsdalsa 河流量	3.670 0	28.835 0	54.000 0
风速时间序列	0.090 0	9.940 0	19.790 0

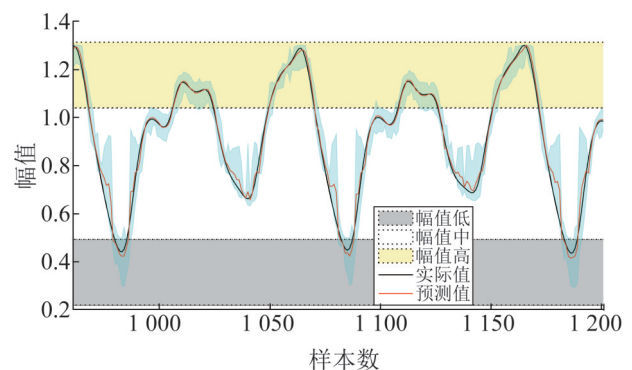


图5 MG序列预测结果

Fig. 5 Prediction results of MG time series

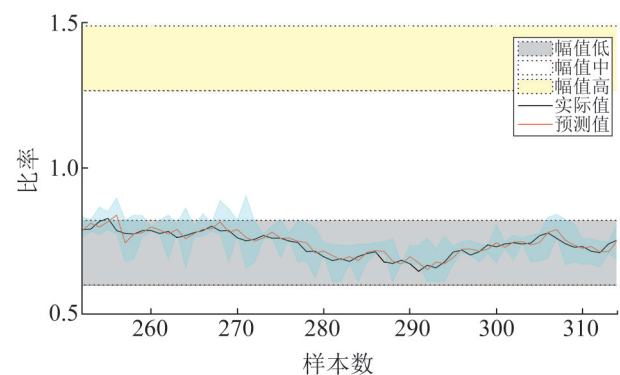


图6 AUD/USD时间序列预测结果

Fig. 6 Prediction results of AUD/USD time series

为了定量说明多模态预测模型的性能,表3给出了各数据集在相应评价指标下的实验结果。

观察表中数据可以发现, 使用平均权重与模型权重的方法具有较高的PICP值, 这说明预测区间包含实际值的概率更高。然而平均权重与模型权重方法的PINAW指标较高, 即区间距离比较宽, 也就是牺牲了区间的精度从而提高覆盖性。动态权重方式虽然PICP指标稍微低些, 但是PINAW指标更小, 说明这种方法生成的预测区间更窄, 精度更高。综合指标CWC的值更能反映每种方法的优劣, 从表中可以看出所有的数据集实验中动态权重的方法具有更小的CWC值。这说明给予所有子模式相同的权重或者权重不变是不合适的。进一步观察表中各个权重方法产生的数值预测精度指标RMSE, 也能发现类似的结论。动态权重情况下产生的RMSE值最小, 说明预测的数值精度更高。

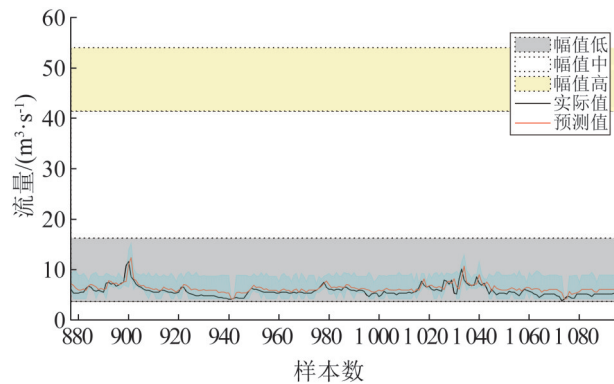


图7 Vatsndalsa 河流量序列预测结果
Fig. 7 Prediction results of Vatsndalsa time series

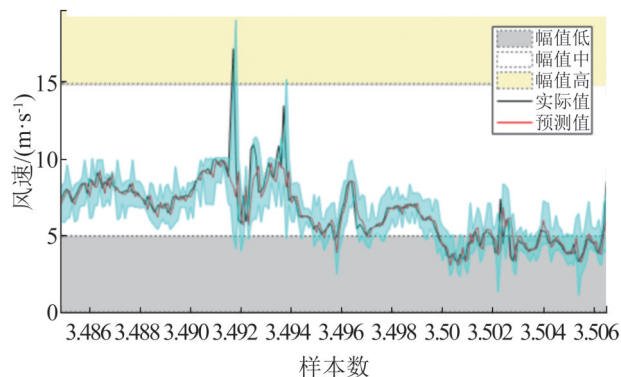


图8 风速序列预测结果
Fig. 8 Prediction results of wind speed

表3 预测结果

Table 3 Forecast results

数据集	权重方法	PICP	PINAW	CWC	RMSE
MG	动态权重	0.912 5	0.147 2	0.201 1	0.026 5
	模型权重	0.920 8	0.292 5	0.399 8	0.116 9
	平均权重	0.950 0	0.357 1	0.488 3	0.107 4
汇率	动态权重	0.873 0	0.095 9	0.130 9	0.016 8
	模型权重	1.000 0	0.662 2	0.905 8	0.176 2
	平均权重	1.000 0	0.673 6	0.921 4	0.203 1
河流量	动态权重	0.867 6	0.068 5	0.093 4	0.760 0
	模型权重	0.904 1	0.135 5	0.185 2	1.549 6
	平均权重	0.908 7	0.148 0	0.202 2	1.738 5
风速	动态权重	0.870 1	0.128 1	0.174 8	0.762 6
	模型权重	0.976 1	0.314 1	0.429 7	1.261 1
	平均权重	0.976 3	0.339 2	0.463 9	1.528 7

为了进一步分析所提模型的性能, 与使用全体数据训练FCMs的方法进行了对比。图9为使用整体数据训练FCMs的RMSE曲线图, 其中遍历节点数的范围是3~20。从图9中可以看出, 随着节点的增加预测RMSE先减小, 但是当节点数超过某一值时, 预测值的RMSE反而会增大, 说明一直增加FCMs的节点数并不能提高预测精度, 过多的节点数反而增加了模型的复杂性。由图9可知, 使用整体数据进行FCMs建模, 4个数据的最佳节点数分别是8, 7, 5和10。

表4给出了所提出的方法与整体数据建模方法以及长短期记忆网络LSTM (long short-term memory)方法进行时间序列数据预测精度的RSME对比结果。其中, 整体建模方法所选节点数为最佳节点数, 选用的LSTM具有200个隐藏单元, 最大迭代次数250次。由表4可知多模态方法取得了较高的预测精度。

表5给出了各个算法在不同数据上运算的时间。整体数据训练模型用时最少。多模态模型执行粒化过程中耗时较多, 需要对所有子模型预测的序列数据遍历计算, 以获得令公式(17)最大化的区间, 其时间复杂度为 $n(n-1)/2$, 即算法的时间复杂度为 $O(n^2)$ 。LSTM需要训练的模型较大, 因此时间消耗最多。

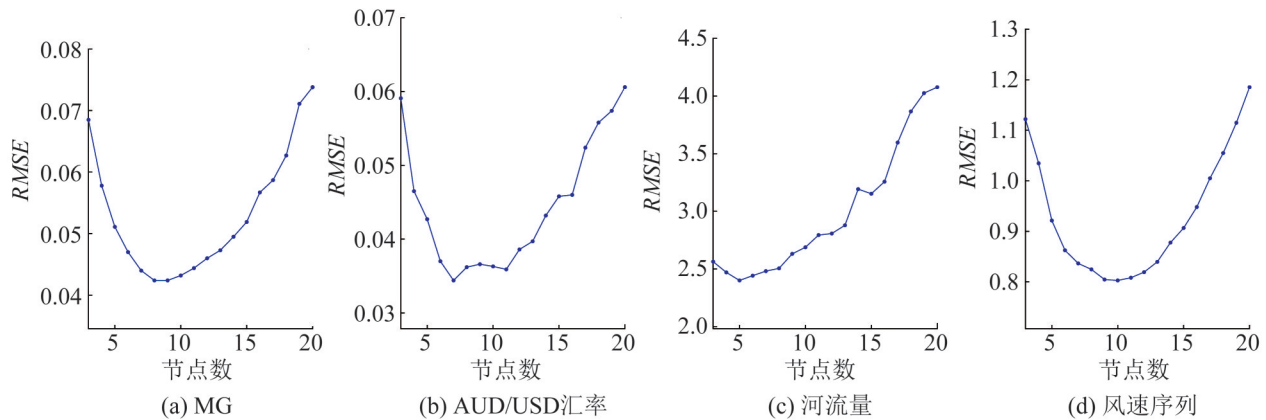


图9 不同节点数的整体数据训练FCMs的RMSE
Fig. 9 RMSE of overall data with varying number of nodes

表4 预测RMSE对比
Table 4 Comparisons of RMSE

数据集	多模态模型	整体模型		LSTM
		RMSE	节点数	
MG	0.026 5	0.042 4	8	0.034 8
汇率	0.016 8	0.034 4	7	0.022 5
河流量	0.760 0	2.400 6	5	1.204 1
风速	0.762 6	0.802 8	10	0.841 6

表5 算法运行时间
Table 5 Algorithms runtime s

数据集	多模态模型	整体模型	LSTM
MG	51.95	1.47	179.41
汇率	46.04	1.16	57.44
河流量	51.26	0.92	157.23
风速	335.05	4.06	4 137.44

4 结论

时间序列往往具有多种变化特征，即序列的变化是多模态的。如果对时间序列的每个变化模态分别进行建模，并对子模型进行有效的融合，将能更有效地反应出时间序列的变化规律。本文提出了基于模糊认知图的一种时间序列处理框架结构。该方法通过大量随机选取时间片段的方法，以提取时间序列可能的模态变化特征。对时间序列进行模糊化处理，赋予时间序列语义描述，从而更利于人们的认知与决策。对各个子模型的输出进行粒化处理进行整合；在粒化处理中，通过满足粒化区间内子模型的权重最大化，粒化区间

的最小化，从而获得时间序列的预测区间。研究了多种权重选取方式用于子模型的有效融合，实验结果表明动态权重的方法获得的预测区间指标更优，预测精度更高。

所提出的方法中采用自助法大量选取时间片段，目的是通过大量随机选取时间段，来保证选取到时间序列的所有变化模态，然而这种方法必然会产生冗余时间段，会降低模型的效率。因此后续需要进一步研究更有效的时间序列变化模态的提出方法，如通过时间序列的自动分割及聚类，主动提取序列的变化模型形式。对于子模型的权重，本文验证了3种基本权重方式，可以进一步研究相关权重选取方法，以提高模型的性能。

参考文献:

- [1] Andreas S Weigend, Neil A Gerschenfeld. Time Series Prediction: Forecasting the Future and Understanding the Past[J]. International Journal of Forecasting (S0169-2070), 1994, 10(1): 463-466.
- [2] Lee Y S, Tong L I. Forecasting Time Series Using a Methodology Based on Autoregressive Integrated Moving Average and Genetic Programming[J]. Knowledge-Based Systems (S0950-7051), 2011, 24(1): 66-72.
- [3] Zhang G P. Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model[J]. Neurocomputing (S0925-2312), 2003, 50: 159-175.
- [4] Khashei M, Bijari M, Ardali G A R. Hybridization of Autoregressive Integrated Moving Average (ARIMA) with

- Probabilistic Neural Networks (PNNs) [J]. *Computers & Industrial Engineering* (S0360-8352), 2012, 63(1): 37-45.
- [5] Cao L J, Tay F E H. Support vector Machine with Adaptive Parameters in Financial Time Series Forecasting [J]. *IEEE Transactions on Neural Networks* (S1045-9227), 2003, 14(6): 1506-1518.
- [6] Chaâbane N. A Novel Auto-Regressive Fractionally Integrated Moving Average-Least-Squares Support Vector Machine Model for Electricity Spot Prices Prediction[J]. *Journal of Applied Statistics* (S0266-4763), 2014, 41(3): 635-651.
- [7] Guo H, Liu X, Sun Z. Multivariate Time Series Prediction Using a Hybridization of VARMA Models and Bayesian Networks[J]. *Journal of Applied Statistics* (S0266-4763), 2016, 43(16): 2897-2909.
- [8] Song Q, Chissom B S. Fuzzy Time Series and its Models [J]. *Fuzzy Sets and Systems* (S0165-0114), 1993, 54(3): 269-277.
- [9] Singh P. A Brief Review of Modeling Approaches Based on Fuzzy Time Series[J]. *International Journal of Machine Learning and Cybernetics* (S1868-8071), 2017, 8(2): 397-420.
- [10] Bose M, Mali K. Designing Fuzzy Time Series Forecasting Models: a Survey[J]. *International Journal of Approximate Reasoning* (S0888-613X), 2019, 111: 78-99.
- [11] Lu W, Chen X, Pedrycz W, et al. Using Interval Information Granules to Improve Forecasting in Fuzzy Time Series[J]. *International Journal of Approximate Reasoning* (S0888-613X), 2015, 57: 1-18.
- [12] Froelich W, Pedrycz W. Fuzzy Cognitive Maps in the Modeling of Granular Time Series[J]. *Knowledge-Based Systems* (S0950-7051), 2017, 115: 110-122.
- [13] Bargiela A, Pedrycz W. Toward a Theory of Granular Computing for Human-Centered Information Processing [J]. *IEEE Transactions on Fuzzy Systems* (S1063-6706), 2008, 16(2): 320-330.
- [14] Pedrycz W. The Principle of Justifiable Granularity and an Optimization of Information Granularity Allocation as Fundamentals of Granular Computing[J]. *Journal of Information Processing Systems* (S1976-913X), 2011, 7 (3): 397-412.
- [15] Pedrycz W, Homenda W. Building the Fundamentals of Granular Computing: a Principle of Justifiable Granularity[J]. *Applied Soft Computing* (S1568-4946), 2013, 13(10): 4209-4218.
- [16] Papageorgiou E I, Salmeron J L. A Review of Fuzzy Cognitive Maps Research During the Last Decade[J]. *IEEE Transactions on Fuzzy Systems* (S1063-6706), 2012, 21(1): 66-79.
- [17] Felix G, Nápoles G, Falcon R, et al. A Review on Methods and Software for Fuzzy Cognitive Maps[J]. *Artificial Intelligence Review* (S0269-2821), 2019, 52 (3): 1707-1737.
- [18] Stach W, Kurgan L A, Pedrycz W. Numerical and Linguistic Prediction of Time Series with the Use of Fuzzy Cognitive Maps[J]. *IEEE Transactions on Fuzzy Systems* (S1063-6706), 2008, 16(1): 61-72.
- [19] Lu W, Yang J, Liu X, et al. The Modeling and Prediction of Time Series Based on Synergy of High-Order Fuzzy Cognitive Map and Fuzzy c-Means Clustering[J]. *Knowledge-Based Systems* (S0950-7051), 2014, 70: 242-255.
- [20] Pedrycz W, Jastrzebska A, Homenda W. Design of Fuzzy Cognitive Maps for modeling time series[J]. *IEEE Transactions on Fuzzy Systems* (S1063-6706), 2016, 24 (1): 120-130.
- [21] Salmeron J L, Froelich W. Dynamic Optimization of Fuzzy Cognitive Maps for Time Series Forecasting[J]. *Knowledge-Based Systems* (S0950-7051), 2016, 105: 29-37.
- [22] Froelich W, Pedrycz W. Fuzzy Cognitive Maps in the Modeling of Granular Time Series[J]. *Knowledge-Based Systems* (S0950-7051), 2017, 115: 110-122.
- [23] G Felix, G Nápoles, R Falcon, et al. A Review on Methods and Software for Fuzzy Cognitive Maps[J]. *Artificial Intelligence Review* (S0269-2821), 2019(52): 1707-1737.
- [24] Zadeh L A. Toward a Theory of fuzzy Information Granulation and its Centrality in Human Reasoning and Fuzzy Logic[J]. *Fuzzy Sets and Systems* (S0165-0114), 1997, 90(2): 111-127.
- [25] Zadeh L A. Is There a Need for Fuzzy Logic? [J]. *Information Sciences* (S0020-0255), 2008, 178(13): 2751-2779.
- [26] Bargiela A, Pedrycz W. Toward a Theory of Granular Computing for Human-Centred Information Processing [J]. *IEEE Transactions on Fuzzy Systems* (S1063-6706), 2008, 16(2):320-330.
- [27] Pedrycz W. Granular Computing for Data Analytics: A Manifesto of Human-Centric Computing[J]. *IEEE/CAA Journal of Automatica Sinica* (S2329-9266), 2018, 5(6): 1025-1034.
- [28] Pedrycz W, Vukovich G. Abstraction and Specialization of Information Granules[J]. *IEEE Transactions on Systems, Man and Cybernetics - PartB: Cybernetics* (S1083-4419), 2001, 31(1): 106-111.
- [29] W Stach, L A Kurgan, W Pedrycz. Numerical and Linguistic Prediction of Time Series with the Use of

- Fuzzy Cognitive Maps[J]. IEEE Transactions on Fuzzy Systems (S1063-6706), 2008, 16(1): 61-72.
- [30] S Boyd, L Vandenberghe. Convex Optimization[M]. London: Cambridge University Press, 2004.
- [31] Khosravi A, Nahavandi S, Creighton D. Prediction Intervals for Short-Term Wind Farm Power Generation Forecasts[J]. IEEE Transactions on Sustainable Energy (S1949-3029), 2013, 4(3): 602-610.
- [32] Quan H, Srinivasan D, Khosravi A. Short-Term Load and Wind Power Forecasting Using Neural Network-Based Prediction Intervals[J]. IEEE Transactions on Neural Networks and Learning Systems (S2162-237X), 2013, 25 (2): 303-315.