



基于演化聚类分析的组合预测改进熵权模型及其应用

陶志富, 朱家明, 刘金培, 陈华友

引用本文:

陶志富, 朱家明, 刘金培, 等. 基于演化聚类分析的组合预测改进熵权模型及其应用[J]. 控制与决策, 2020, 35(2): 410–416.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2018.0824>

您可能感兴趣的其他文章

Articles you may be interested in

基于灰熵模型的区间型指标和权重的不确定多属性决策方法及其应用

The uncertain multi-attribute decision making methods and application based on grey entropy model with interval-type attribute values and weights

控制与决策. 2020, 35(3): 657–666 <https://doi.org/10.13195/j.kzyjc.2018.0543>

新信息优先原理下非等间距GM(1,1)模型优化研究

Research on optimization of non-equidistant GM(1, 1) model based on the principle of new information priority

控制与决策. 2019, 34(10): 2221–2228 <https://doi.org/10.13195/j.kzyjc.2018.0163>

基于时滞效应的多变量离散灰色预测模型

Multi-variable time-delayed discrete grey model

控制与决策. 2017, 32(11): 1997–2004 <https://doi.org/10.13195/j.kzyjc.2016.1022>

基于匹配属性相似度的应急决策方案推荐方法

Emergency decision response plan recommendation method based on similarity of matched attributes

控制与决策. 2016, 31(7): 1247–1252 <https://doi.org/10.13195/j.kzyjc.2015.0725>

基于信息熵的改进人工蜂群算法

Improved artificial bee colony algorithm based on information entropy

控制与决策. 2015(6): 1121–1125 <https://doi.org/10.13195/j.kzyjc.2014.0377>

基于折衷型变权向量的直觉语言决策方法

Compromise type variable weight vector based method Intuitionistic linguistic making decision

控制与决策. 2015, 30(12): 2233–2240 <https://doi.org/10.13195/j.kzyjc.2014.1517>

基于具有自适应分段损失函数支持向量机的产品销售预测模型

Product sale forecasting method based on support vector machine with adaptive segmented loss function

控制与决策. 2015(10): 1803–1809 <https://doi.org/10.13195/j.kzyjc.2013.1541>

基于T-S模糊模型的网络控制系统非脆弱 ∞ 跟踪控制

Non-fragile ∞ tracking control for networked control systems based on T-S fuzzy model

控制与决策. 2015(1): 110–116 <https://doi.org/10.13195/j.kzyjc.2013.1264>

基于演化聚类分析的组合预测改进熵权模型及其应用

陶志富¹, 朱家明², 刘金培³, 陈华友^{2†}

(1. 安徽大学 经济学院, 合肥 230601; 2. 安徽大学 数学科学学院,
合肥 230601; 3. 安徽大学 商学院, 合肥 230601)

摘要: 针对组合预测传统熵值赋权方法可能存在大样本下权重差异不大的不足, 引入演化聚类分析技术, 提出一类组合预测改进的熵值赋权模型. 在确定好单项预测方法和结果后, 通过对所有单项预测结果给出的绝对误差信息进行聚类分析, 得到单项预测方法下误差信息在不同类中的离散分布, 给出单项预测结果在整体和局部两个层面的差异比较工具, 进而利用得到的离散分布应用熵权模型给出改进的熵值赋权方法. 结合美元兑日元的汇率数据分析表明了所提出方法的可行性和有效性.

关键词: 改进熵权; 聚类分析; 组合预测; 汇率

中图分类号: O211.67; TP277

文献标志码: A

A modified entropy weighting method in combination forecasting based on evolutionary clustering analysis and its application

TAO Zhi-fu¹, ZHU Jia-ming², LIU Jin-pe³, CHEN Hua-you^{2†}

(1. School of Economics, Anhui University, Hefei 230601, China; 2. School of Mathematical Sciences, Anhui University, Hefei 230601, China; 3. School of Business, Anhui University, Hefei 230601, China)

Abstract: In order to overcome lower difference among weights in combination forecasting with large sample determined by traditional entropy weighting methods, a modified entropy weighting method is developed by introducing the evolutionary automatic clustering analysis. When single forecasting approaches and predictions are fixed, all absolute errors are clustered and diverse discrete distributions of absolute errors corresponding to single forecasting approaches among clusters are obtained, which are tools for further comparisons among the single forecasting methods. Then, a modified entropy weighting method is given using the obtained discrete distributions. By using a numerical study with exchange rate between dollar and yen, the feasibility and effectiveness of the developed method are illustrated.

Keywords: modified entropy weighting method; cluster analysis; combination forecasting; exchange rate

0 引言

组合预测方法于 1969 年由 Bates 等^[1]提出, 通过对单项预测结果进行适当的赋权和信息融合以达到提高预测精度的效果. 此类对单项预测值序列进行赋权和组合的方式称为结果的组合. 随着研究的广泛和深入开展, 组合预测还表现在数据层面的组合和方法层面的组合两个方面, 例如, 某些行业综合指数的编制^[2]和若干能够相互支撑的技术构成的混合预测技术^[3]等. 事实上, 无论是结果的组合还是方法的组合, 单项预测方法的遴选涉及组合预测的机理与内在的机制, 是组合预测的核心问题. 文献[4]和文献[5]分别利用冗余信息、预测有效性和群组一致性的概念对组合预测的上述核心问题进行了广泛和深入

的讨论. 遵循 Bates 等^[1]给出的组合预测范式, 本文仍着重考虑基于结果的组合预测问题.

赋权问题是结果层面组合预测的关键问题, 权重的分配直接关系到组合预测的实现和效果. 现有的赋权方法从权重的确定方式上可以分为非最优赋权、最优赋权和组合赋权 3 种^[4,6]. 非最优赋权依据预测科学的基本原理, 力求以简便的规则确定权重, 例如熵权. 最优赋权依据一定的目标构建最优化模型以确定权重, 如组合预测误差平方和达到最小等. 面对数据量较大的观测时间序列预测问题, 最优化赋权往往在及时性上存在一定的不足, 因此在实际的经济和金融数据预测中常常使用等权分配的办法^[5,7], 从而考虑简单易行和有效的非最优化赋权方法仍是值得

收稿日期: 2018-06-15; 修回日期: 2018-09-12.

基金项目: 国家自然科学基金项目 (71701001, 71771001, 71871001, 61502003); 安徽省哲学社会科学规划项目 (AHSKQ2016D13); 安徽大学博士科研启动基金项目.

[†]通讯作者. E-mail: huayouc@126.com.

探讨的理论和应用课题。

熵权应用于组预测理论与实践已经有很多相关报道,通过熵测度量数据波动的大小给出单项预测方法权重^[8-9]。但当观测样本量和单项预测方法较多时,应用熵权时可能会存在两个弊端:1)由于数据量较大,残差分布相对一致,容易导致权重差异偏小;2)数据计算量偏大。因此,本文拟将聚类分析技术引入到组预测问题中,与熵权模型相结合,给出一类改进的组预测熵值赋权方法。

聚类分析基于相似性对数据进行分类,其在预测问题中的应用已有相关报道。王永县等^[10]利用聚类分析将观测时间序列依据划分为若干局部特征相似的子类,实现对数据降维和预测。王晓原等^[11]通过聚类分析对数据进行分类,进而利用集对分析实现预测。Goia等^[12]基于功能聚类实现预测数据的分类,在分类数据的基础上应用函数线性回归模型进行数据预测。而在Song等^[13]提出的模糊时间序列模型以及随后的模糊时间序列成果报道^[14-17]中,聚类以及模糊聚类方法在数据类型划分以及基于类型划分的预测技术中都有直接和广泛的应用。可见,现有的预测文献报道中聚类分析主要应用于观测数据的分类识别与划分,将其应用于组预测权重的划分鲜有报道。

本文针对各单项预测的残差数据,通过引入聚类分析模型对残差数据进行分类,进而考虑某单项预测结果下残差数据属于各聚类中心的频率分布。在数据量偏大时,可以将上述频率视为一组离散概率分布,一方面通过聚类分析,可以从整体上对残差数据的相似性进行度量;另一方面,通过导出的离散概率分布可以从微观上对比不同单项预测结果的差异。因此,本文将通过组预测问题引入聚类方法给出一类客观的赋权模型和算法,并结合数据分析表明所提出方法的可行性和有效性。

1 基于聚类分析的组预测改进熵权模型

设 $\{y_t, t = 1, 2, \dots, n\}$ 为某社会经济现象的指标序列的观察值,现有 m 种单项预测方法对其进行预测,预测结果分别记为

$$\{\hat{y}_{it}, i = 1, 2, \dots, m, t = 1, 2, \dots, n\}.$$

考虑各单项预测结果之间的差异性,对其进行赋权和结果的组合。设各单项预测方法对应的权重向量为

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_m\},$$

满足 $\omega_i \geq 0, i = 1, 2, \dots, m$ 且 $\sum_{i=1}^m \omega_i = 1$ 。记

$$e_{it} = y_t - \hat{y}_{it}, i = 1, 2, \dots, m, t = 1, 2, \dots, n,$$

则 e_{it} 表示第 i 种预测方法在第 t 时刻的预测误差。下面引入聚类分析模型,给出一类组预测改进的熵权方法。

现有预测文献中应用的聚类技术包括集对分析聚类预测^[18]、模糊 C 均值聚类^[19]和动态聚类技术^[20]等,通过对传统的聚类算法(如 K 均值聚类、谱聚类)等进行改进,演化聚类算法可以克服传统聚类算法在类中心数需要人为确定以及容易陷入局部最优等局限。因此,本文主要针对预测误差数据应用演化自动聚类算法进行分类,并基于分类结构构造一类改进的熵权模型,基本流程如下。

step 1: 为反映预测结果的波动水平,对 e_{it} 取绝对值操作,即 $|e_{it}|, i = 1, 2, \dots, m, t = 1, 2, \dots, n$ 。

$|e_{it}|$ 表示不同预测方法在不同时点的绝对预测误差,传统的熵权归一化每组单项预测的绝对误差数据,进而通过计算熵值信息反映各单项预测结果的波动情况并进行赋权。可见,经典的熵权为分别考虑单项预测误差数据的波动大小进行赋权,缺乏单项预测结果之间整体性的差异分析。因此,给出step 2进行所有单项预测绝对误差数据的分类操作。

step 2: 对 $|e_{it}|(i = 1, 2, \dots, m, t = 1, 2, \dots, n)$ 进行演化聚类分析,得出各时点不同预测方法给出的绝对预测误差的划分情况。

Chakrabarti等^[21]首次提出演化聚类问题并提出时间平滑性框架,基于这一框架实现了演化版的 K 均值聚类算法和演化版的层次聚类算法,随后被Xu等^[22]、Lee等^[23]和Kim等^[24]加以改进^[25]。本文以Yarpiz网站提供的演化聚类算法为基础,对绝对误差数据进行聚类。可以看出,针对 $|e_{it}|$ 的聚类结果考虑了每个具体的预测绝对误差所处的误差水平。另一方面,还可以看出当某单项预测给出的绝对误差数据集中于某一类中心时,表明该单项预测结果的波动较小;而当其绝对误差数据相对均匀地分布于各类中心时,表明该单项预测结果的波动较大。

step 3: 统计各单项预测方法下绝对误差数据分布于不同聚类中心的情况。

设绝对误差数据经演化聚类后的类中心数为 K ,有

$$\text{Ind}_{it}^k = \begin{cases} 1, & |e_{it}| \in C_k; \\ 0, & |e_{it}| \notin C_k. \end{cases}$$

$$i = 1, 2, \dots, m, t = 1, 2, \dots, n, k = 1, 2, \dots, K.$$

第 i 个预测结果的绝对误差信息中隶属于第 k 个类中心的数量为

$$\text{num}_i^k = \sum_{t=1}^n \text{Ind}_{it}^k, k = 1, 2, \dots, K.$$

对固定的 i 对应的 num_i^k 进行归一化处理,第 i 项单项预测方法的绝对误差数据分布于各类中心的情况可归纳为 $p_i = (p_i^1, p_i^2, \dots, p_i^K)^T, i = 1, 2, \dots, m$.

依据前述分析可知, $p_i(i = 1, 2, \dots, m)$ 的数值分布越均匀,第 i 项单项预测结果的波动会越大,反之亦然.从预测角度看,预测结果波动较大,其预测效果相对于波动较小的预测方法往往不佳,需要赋予较小的权重,因此给出如下改进熵值赋权方法.

step 4: 依据绝对误差数据的分布情况给出改进的依据熵值的权重.

由上述符号和结论,给出第 i 项单项预测方法的权重 $\omega_i(i = 1, 2, \dots, m)$ 为

$$\begin{cases} E(p_i) = -\frac{1}{\ln K} \sum_{k=1}^K p_i^k, \\ \omega_i = \frac{1 - E(p_i)}{\sum_{i=1}^m (1 - E(p_i))}. \end{cases}$$

可以看出,区别于传统的熵权模型,权重向量 Ω 虽然也是依据熵值加以确定,但这里确定熵值的离散概率分布是由聚类分析给出的统计数据,因此称为改进的熵权模型.

2 基于改进熵权的组合预测模型构建

基于上述讨论,应用聚类分析技术可以给出组合预测的单项预测方法赋权,进而进行组合预测.预测流程如下.

step 1: 针对观测时间序列 $\{y_t\}$,构建 m 种单项预测方法得到相应的预测值序列 $\{\hat{y}_{it}\}$,并计算每种预测结果的绝对误差序列 $|e_{it}|, i = 1, 2, \dots, m, t = 1, 2, \dots, n$.

step 2: 依据基于聚类分析的改进熵权模型确定 m 种单项预测方法对应的权重向量 $\Omega = \{\omega_1, \omega_2, \dots, \omega_m\}$.

Step 3: 依据权重向量 Ω 对单项预测结果进行集成,即 $\hat{y}_t^c = \sum_{i=1}^m \omega_i \hat{y}_{it}, t = 1, 2, \dots, n$.

step 4: 对基于聚类分析改进熵权的组合预测结果进行评价.

依据上述步骤可知,观测样本量为 n 的改进的熵权模型计算复杂度为 $O(kmn)$.其中: k 为迭代的次

数, m 为单项预测方法个数, n 为观测样本数.以最简单的组合预测误差平方和极小为目标的最优化赋权方法^[4]的复杂度为 $O(m^2n^2)$.可以看出,上述改进的熵权方法较最优化赋权方法在时间复杂度上更低.

3 实例分析

为验证本文所提出的赋权方法的有效性,以2016年1月1日~2018年6月12日的日元兑美元日汇率收盘价观测值为例,考虑模型的具体实施与建立.选取截止2018年5月31日的630个数据作为样本,建模和分析过程如下.

step 1: 针对选取的630个观测样本,构建如下单项预测方法:

1) Holt-Winters两参数指数平滑模型.

Holt-Winters两参数指数平滑适用于对具有线性趋势的序列进行修匀,其基本思想是假定序列有一个比较固定的线性趋势,每期有一个随机波动序列 r_t ,平滑公式为

$$\begin{cases} \hat{y}_{t+1} = \alpha y_t + (1 - \alpha)(y_{t-1} + r_{t-1}), \\ \hat{r}_t = \beta(\hat{y}_{t+1} - \hat{y}_t) + (1 - \beta)r_{t-1}. \end{cases}$$

通过应用上述平滑公式,向前 l 期的预测值为

$$\hat{y}_{t+l} = \hat{y}_{t+1} + (l - 1)\hat{r}_t.$$

应用Eviews软件,确定参数值为 $\alpha = 0.99, \beta = 0.00$,相应的误差平方和为300.944 4,均方根误差为0.6911.其预测方程为

$$\hat{y}_{t+T} = 108.8206 - 0.0242T.$$

2) BP神经网络预测.

神经网络技术应用不同的拓扑结构实现不同的计算功能,其中基于误差反向传播算法的多层前馈网络(BP神经网络)模型因为具有良好的函数逼近特性,且解决了多层网络模型中隐含层的连接权问题,进而具有提高学习和记忆的功能,使得BP神经网络成为应用最广泛的经典模型之一. Kolmogorov定理和BP定理指出3层神经网络可以在任意希望的精度上实现任意连续函数构成的映射,因此可以模拟任意非线性问题^[26],由输入层、隐含层(可以是多层)和输出层构成.

基本的BP算法为一类简单的最速下降法,有

$$\begin{cases} W_{ij}(t) = W_{ij}(t-1) - \Delta W_{ij}(t), \\ \Delta W_{ij}(t) = \frac{\eta \partial E}{\partial W_{ij}(t-1)} + \alpha \Delta W_{ij}(t-1). \end{cases}$$

其中: W_{ij} 为节点 i 与 j 之间的联接权值; $\eta \in (0.01,$

0.1) 为学习速率; α 为冲量系数, 取值约为 0.9; t 为网络迭代步数; E 为定义在权空间 $W(t)$ 上的误差超曲面。

一般地, 在具体设计时, 首先依据经验公式^[27-28]初步确定隐含层神经元个数, 如 $s = \log_2^N$ (N 为模式数), $s = 2n + 1$ (n 为输入层结点数), $s = (0.43mn + 0.12nn + 2.54m + 0.77n + 0.35)^{1/2} + 0.51$ (m 为输入层个数, n 为输出层个数), 再通过对不同神经元数的网络进行训练和对比, 最终确定神经元数。

本文基于 Matlab 编程实现应用 BP 神经网络的预测, 选取前 70% 数据作为训练样本, 中间 15% 数据作为测试数据, 最后 15% 数据作为验证数据。经过测算, 确定 BP 神经网络模型由 3 个输入层神经元、12 个隐层神经元和 1 个输出层神经元组成, 从输入层到隐含层以及从隐含层到输出层的传递函数分别采用 S 型函数和线性函数。运行结果如图 1 所示。

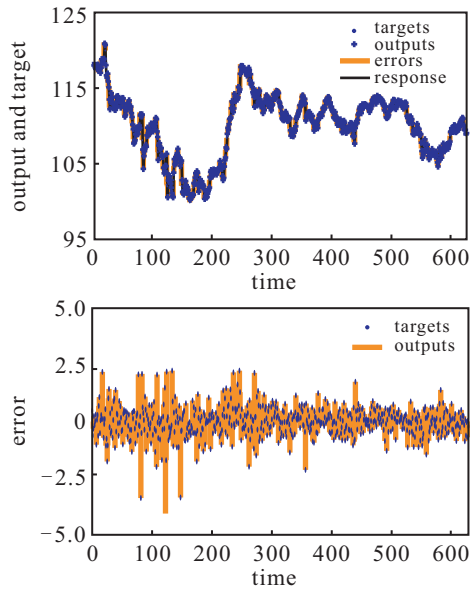


图 1 BP神经网络预测效果

3) 非线性自回归神经网络。

非线性自回归神经网络 (nonlinear autoregressive neural network, NARnet) 是一类有记忆功能的神经网络, 其典型的结构包括输入层、隐含层、输出层以及输入和输出延时, 网络的输出取决于当前输入和过去输出, 其方程为

$$y(t) = f[y(t-1), \dots, y(t-d)].$$

其中: $y(t)$ 为输出, d 为延时阶数, f 为应用神经网络实现的非线性函数。

为避免过拟合, 待拟合的样本通常分为训练样本 (80%)、检验样本 (10%) 和测试样本 (10%), 通过反复测算确定延时阶数和隐含层神经元个数。NARnet 采

用 Levenberg-Marquardt (LM) 算法进行训练, 依据误差自相关图、拟合效果误差图和拟合值与实际值的相关系数选择合适的模型。通过 Matlab R2014a 神经网络工具箱中时间序列分析建模工具实现。

分析发现, 当隐含层神经元个数为 16、延时阶数为 12 时, 模型效果较为理想, 应用 NARnet 拟合效果如图 2 所示。

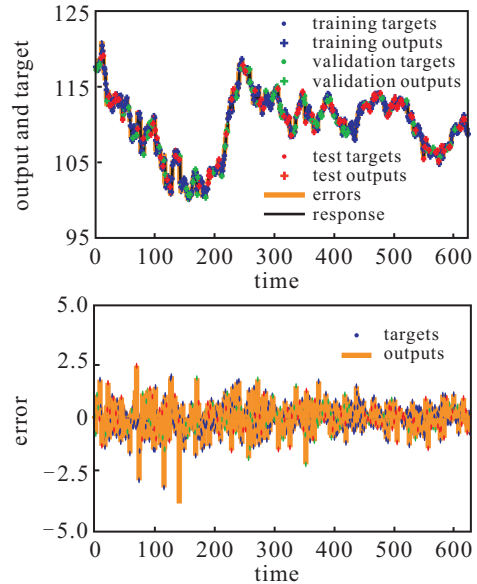


图 2 NARnet神经网络预测效果

step 2: 依据 step 1 确定的 3 个单项预测结果对应的绝对误差数据, 经实验考虑对 Holt-Winters 指数平滑、BP 神经网络和 NARnet 预测应用基于聚类分析的赋权模型, 得到具体的权重向量, 计算过程如下。

考虑到单项预测方法产生的信息损失, 实际有效绝对误差 1 881 项, 应用演化聚类分析后分为 3 类, 3 种单项预测结果的绝对误差属于 3 个类的频率分别为:

Holt-Winters 指数平滑

$$p_1 = (0.149\ 9, 0.012\ 8, 0.837\ 3)^T;$$

BP 神经网络预测

$$p_2 = (0.154\ 7, 0.008\ 0, 0.837\ 3)^T;$$

NARnet 预测

$$p_3 = (0.419\ 5, 0.250\ 4, 0.330\ 1)^T.$$

上述频率可以视为 3 种单项预测方法给出的绝对误差信息在 3 个类中的离散分布, 应用改进的熵权模型, 计算得到 3 种单项预测方法的权重为

$$\Omega = (0.431\ 9, 0.440\ 5, 0.127\ 6)^T.$$

step 3: 依据 step 2 给出的权重对单项预测结果进行集结, 得到基于改进熵权模型的组合预测值序列

$\hat{g}_t^c, t = 1, 2, \dots, 630$. 预测值与实际值的对比如图3所示.

step 4: 为反映不同单项预测结果和组合预测结果的效果,从平均绝对误差(MAE)、均方根误差(MSE)、平均绝对百分比误差(MAPE)和均方百分比误差(MSPE)4个指标对不同预测结果进行评价,结果如表1所示.

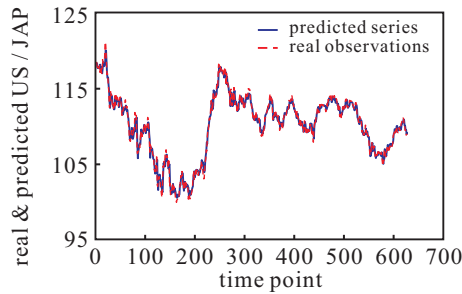


图3 美元兑日元汇率预测效果对比

表1 参数设置

	MAE	MSE	MAPE	MSPE
Holt-Winters	0.502 9	0.027 6	0.004 6	2.528 9e-04
BP神经网络	0.493 5	0.027 3	0.004 5	2.513 0e-04
NARNet预测	1.628 9	0.084 2	0.014 9	7.701 8e-04
普通熵权组合	0.619 4	0.031 9	0.005 7	2.914 6e-04
改进熵权组合	0.492 8	0.026 3	0.004 5	2.404 3e-04

除与各单项预测的结果进行比较外,为了体现改进熵权的优势,给出普通熵权的组合预测结果及其在4个指标下的水平(如表1所示). 依据普通熵权计算的权重向量为

$$\Omega = (0.350\ 2, 0.355\ 1, 0.294\ 7)^T.$$

应用上述模型,对未来10个时点的汇率数据进行预测,效果如图4所示.

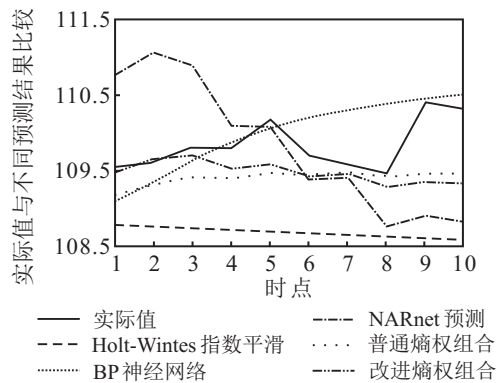


图4 汇率实际值与不同预测结果对比

结合表1和图4可以看出:

1) 就预测精度而言,两种组合给出的结果均优于单项预测结果,但是两种组合的结果相对于单项预测

结果改进的幅度偏小,其原因在于本文所选取的单项预测结果其拟合精度均保持了较高的水平,组合方法给出的改进空间是有限的.

2) 本文给出的改进的熵权组合结果优于普通熵权给出的结果.

3) 判断预测效果的好坏还应关注到外推的预测准确性上. 由图4可见,在保持略高水平预测精度的同时,对比本文所给出的组合结果和普通熵权给出的组合结果与实际值之间的相似性,可以得出普通熵权所给出的组合结果与实际值之间的相似性为 $-0.165\ 0$,而本文给出的相似性为 $0.475\ 1$,表明就整体趋势而言本文给出的结果更优. 事实上,能够在保持预测精度的同时还能更加准确地把握时序变化的趋势对于很多实际的金融和经济预测而言往往更加重要.

4) 整体而言,基于结果的组合预测可以在一定程度上消除由于方法的选择导致和存在的预测风险.

4 对比分析

为给出更多论据,在现有结果的基础上,引入基于自适应神经模糊系统^[29](adaptive network based fuzzy inference system, ANFIS)、基于数据处理组合算法^[30](group method of data handling, GMDH)和灰色预测方法^[4],并考虑如下几类组合模式,结果如表2所示.

表2 预测效果对比

组合	方法	MAE	MSE	MAPE	MSPE
G_1	普通熵权	0.635 2	0.032 7	0.005 8	2.995 7e-04
	改进熵权	0.455 4	0.024 7	0.004 2	2.274 2e-04
G_2	普通熵权	0.834 0	0.042 6	0.007 6	3.950 2e-04
	改进熵权	0.804 3	0.041 2	0.007 4	3.812 8e-04
G_3	普通熵权	0.489 4	0.027 0	0.004 5	2.474 3e-04
	改进熵权	0.489 4	0.027 0	0.004 5	2.474 3e-04
G_4	普通熵权	0.619 4	0.031 9	0.005 7	2.914 6e-04
	改进熵权	0.492 8	0.026 3	0.004 5	2.404 3e-04

改进的熵权中残差信息均被聚为5类进行模型计算, $M_1 \sim M_6$ 分别表示Holt-Winters指数平滑、BP神经网络预测、NARNet预测、ANFIS预测、GMDH预测和灰色预测. $G_1 = \{M_1, M_2, M_4\}$, $G_2 = \{M_1, M_2, M_4, M_6\}$, $G_3 = \{M_1, M_2, M_5\}$, $G_4 = \{M_1, M_2, M_3\}$ 分别表示由不同单项预测方法结果进行组合的群组. 其中,GMDH选取截止2017年12月31日,630个

观测数据作为输出神经元(y_t), 滞后期为1~5的汇率数据($y_{t-1}, y_{t-2}, \dots, y_{t-5}$)为5个输入神经元。

由表2给出的实验结果可见: 首先, 改进熵权给出的组合结果相对于普通熵权给出的组合结果是非劣的; 其次, 与文中所给结果相类似, 当单项预测的整体精度较高时, 两种熵权给出的结果对于精度的提高是有限的; 再次, 具体到不同的组合模式下, 两类熵权给出的组合结果均受到单项预测效果的影响, 而改进熵权相对于普通熵权在一定程度上能够放大精度高的单项预测效果, 但是当单项预测方法的精度比较接近时, 两类熵权给出的组合结果差异性也极小。

5 结 论

目前, 针对组合预测的非最优赋权模型研究较少, 且在观测样本量较大时, 由于各单项预测给出的误差信息具有一定的相似性, 导致传统以熵值确定的权重差异性不大。为克服这一不足, 本文将聚类分析引入到组合预测的权重确定问题中。通过对所有单项预测方法给出的绝对误差信息进行聚类, 进而考虑不同单项预测方法下绝对误差在不同类中的分布差异, 并应用熵权模型给出一类新的权重确定模型, 其优势在于能够从整体和局部两个维度同时考虑不同预测方法给出的结果差异。实验结果表明所提出方法可行、合理、有效。

参考文献(References)

- [1] Bates J M, Granger C W J. The combination of forecasts[J]. Journal of the Operational Research Society, 1969, 20(4): 451-468.
- [2] Babu C N, Reddy B E. A moving-average filter based hybrid ARIMA-ANN model for forecasting time series data[J]. Applied Soft Computing, 2014, 23(1): 27-38.
- [3] 陈荣, 梁昌勇, 陆文星, 等. 面向旅游突发事件的客流量混合预测方法研究[J]. 中国管理科学, 2017, 25(5): 167-174.
(Chen R, Liang C Y, Lu W X, et al. The research of tourist flow hybrid forecasting model for tourism emergency events[J]. Chinese Journal of Management Science, 2017, 25(5): 167-174.)
- [4] 陈华友. 组合预测方法有效性理论及其应用[M]. 北京: 科学出版社, 2008: 55-75.
(Chen H Y. Validity principle theory of combination forecasting and its application[M]. Beijing: Science Press, 2008: 55-75.)
- [5] Clemen R T, Winkler R L. Combining economic forecasts[J]. Journal of Business & Economic Statistics, 1986, 4(1): 39-46.
- [6] 陈华友, 朱家明, 丁珍妮. 组合预测模型与方法研究综述[J]. 大学数学, 2017, 33(4): 1-10.
(Chen H Y, Zhu J M, Ding Z N. A survey of researches on combination forecasting models and methodologies[J]. College Mathematics, 2017, 33(4): 1-10.)
- [7] Lamberson P J, Page Scott E. Optimal forecasting groups[J]. Management Science, 2012, 58(4): 805-810.
- [8] 熊国强, 潘泉, 张洪才. 基于熵权Markov链的商品零售价格指数预测方法[J]. 软科学, 2008, 22(4): 13-16.
(Xiong G Q, Pan Q, Zhang H C. A research of Markov chain method based on entropy right to forecast commodities' retail price index[J]. Soft Science, 2008, 22(4): 13-16.)
- [9] 苑泽明, 金字, 王天培. 上市公司无形资产评价指数研究——基于创业板上市公司的实证检验[J]. 会计研究, 2015, 36(5): 72-79.
(Yuan Z M, Jin Y, Wang T P. Listing corporation intangible assets index researching-based on the gem listing corporation empirical test[J]. Accounting Research, 2015, 36(5): 72-79.)
- [10] 王永县, 詹一辉, 张少. 降水时间序列的聚类分析和预测[J]. 系统工程理论与实践, 1994, 14(11): 67-71.
(Wang Y X, Zhan Y H, Zhang S. Clustering analysis and prediction for precipitation time series[J]. System Engineering — Theory & Practice, 1994, 14(11): 67-71.)
- [11] 王晓原, 张敬磊. 区域物流需求分析集对聚类预测模型研究[J]. 软科学, 2004, 18(5): 11-13.
(Wang X Y, Zhang J L. Study on the set pare analysis classified prediction model for analyzing regional logistics demand[J]. Soft Science, 2004, 18(5): 11-13.)
- [12] Goia A, May C, Fusai G. Functional clustering and linear regression for peak load forecasting[J]. International Journal of Forecasting, 2010, 26(4): 700-711.
- [13] Song Q, Chissom B S. Fuzzy time series and its models[J]. Fuzzy Sets and Systems, 1993, 54(3): 269-277.
- [14] Chen S M. Forecasting enrollments based on fuzzy time series[J]. Fuzzy Sets and Systems, 1996, 81(3): 311-319.
- [15] Chen S M, Hwang J R. Temperature prediction using fuzzy time series[J]. IEEE Transactions on Systems, Man, and Cybernetics-Part B, 2000, 30(2): 263-275.
- [16] Sadaei H J, Guimarães F G, da Silva C J, et al. Short-term load forecasting method based on fuzzy time series, seasonality and long memory process[J]. International Journal of Approximate Reasoning, 2017, 83(1): 196-217.
- [17] Chen S M, Phuong B D H. Fuzzy time series forecasting based on optimal partitions of intervals and optimal weighting vectors[J]. Knowledge-Based Systems, 2017, 118(1): 204-216.
- [18] 高洁, 盛昭瀚. 集对分析聚类预测法及其应用[J]. 系统工程学报, 2002, 17(5): 458-462.
(Gao J, Sheng Z H. Method and application of set pair analysis classified prediction[J]. Journal of System Engineering, 2002, 17(5): 458-462.)
- [19] 杨世坚, 贺国光. 基于模糊C均值聚类和神经网络的短时交通流预测方法[J]. 系统工程, 2004, 22(8): 83-86.
(Yang S J, He G G. A short-term traffic flow forecasting

- method based on combination of fuzzy C -mean clustering and neural network[J]. System Engineering, 2004, 22(8): 83-86.)
- [20] 王珏, 齐琛, 李明芳. 基于 SSA-ELM 的大宗商品价格预测研究[J]. 系统工程理论与实践, 2017, 37(8): 2004-2014.
(Wang J, Qi C, Li M F. Prediction of commodity prices based on SSA-ELM[J]. Systems Engineering — Theory & Practice, 2017, 37(8): 2004-2014.)
- [21] Chakrabarti D, Kumar R, Tomkins A. Evolutionary clustering[C]. Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2006: 554-560.
- [22] Xu K S, Kliger M, Hero Iii A O. Adaptive evolutionary clustering[J]. Data Mining and Knowledge Discovery, 2014, 28(2): 304-336.
- [23] Lee H, Lee S, Yoon B. Technology clustering based on evolutionary patterns: The case of information and communications technologies[J]. Technological Forecasting and Social Change, 2011, 78(6): 953-967.
- [24] Kim M S, Han J. A particle-and-density based evolutionary clustering method for dynamic networks[J]. Proceedings of the VLDB Endowment, 2009, 2(1): 622-633.
- [25] Mukhopadhyay A, Maulik U, Bandyopadhyay S. A survey of multiobjective evolutionary clustering[J]. ACM Computing Surveys (CSUR), 2015, 47(4): 1-46.
- [26] 史忠植. 知识发现[M]. 北京: 清华大学出版社, 2002: 230-264.
(Shi Z Z. Knowledge discovery[M]. Beijing: Tsinghua University Press, 2002: 230-264.)
- [27] 张德丰. Matlab 神经网络应用设计[M]. 北京: 机械工业出版社, 2009: 207-243.
(Zhang D F. Application and design of Matlab neural network[M]. Beijing: Mechanical Industry Press, 2009: 207-243.)
- [28] 郑绪枝, 雷靖, 夏薇. 基于快速确定隐层神经元数的 BP 神经网络算法[J]. 计算机科学, 2012, 39(b06): 432-436.
(Zheng X Z, Lei J, Xia W. Algorithm for BP neural network by identifying numbers of hidden layer neurons quickly[J]. Computers Science, 2012, 39(b06): 432-436.)
- [29] Jiang J R. ANFIS: Adaptive-network-based fuzzy inference system[J]. IEEE Trans on Systems, Man and Cybernetics, 1993, 23(3): 665-685.
- [30] Ivakhnenko A G, Savchenko E A. Problems of further GMDH algorithm development[J]. System Analysis Modelling Simulation, 2003, 43(10): 1301-1309.

作者简介

陶志富(1985—), 男, 副教授, 博士, 从事经济预测与决策分析、模糊集理论及其应用等研究, E-mail: Taulzf@126.com;

朱家明(1990—), 男, 博士生, 从事时间序列分析和组合预测的研究, E-mail: 18355150818@163.com;

刘金培(1984—), 男, 副教授, 博士, 从事经济预测与决策分析、信息融合和模糊集合理论等研究, E-mail: liujinpei2012@163.com;

陈华友(1969—), 男, 教授, 博士生导师, 从事经济预测与决策分析、信息融合、模糊集理论及其应用等研究, E-mail: huayouc@126.com.

(责任编辑: 郑晓蕾)

下 期 要 目

- 基于联合子空间模型的模拟信息转换器研究进展.....刘铁锋, 等
- 智能制造系统基于数据驱动的车间实时调度.....吴秀丽, 等
- 基于 RBF 神经网络补偿的一种绳牵引并联机器人支撑系统的力/位混合控制.....王宇奇, 等
- 基于非线性能量收集的多用户 MIMO 认知无线供电通讯网络.....王 哲, 等
- 基于余弦控制因子和多项式变异的鲸鱼优化算法.....黄清宝, 等
- 基于视觉传感器的小型多 AUV 系统队形排列.....解则晓, 等
- 带有非匹配扰动的连铸结晶器振动位移系统自适应反步滑模控制.....李建雄, 等
- 基于高效回环检测和重定位的 SLAM 算法.....佟国峰, 等
- 基于群灰狼优化的光伏逆变器最优无源分数阶 PID 控制.....杨 博, 等
- 基于自适应差分进化的常压塔轻质油产量多目标优化.....丁进良, 等
- 敏捷凝视卫星密集点目标聚类与最优观测规划.....耿远卓, 等