

关于股票价格的二阶模糊时间序列

刘智^{1,2}, 张铁¹, 董莹³, 徐爽爽²

(1. 东北大学理学院, 辽宁沈阳 110819; 2. 沈阳工业大学基础部, 辽宁辽阳 111003;
3. 大连民族大学理学院, 辽宁大连 116600)

摘 要: 由于股票价格的时间序列具有不确定性, 股市的真实模型不容易建立, 而模糊时间序列在解决模糊性数据和不确定性数据方面具有较大优势; 因此, 本文首先将数据进行预处理并改进论域划分的方法, 然后利用三角隶属度函数进行数据的模糊化处理, 再利用模糊化后的数据建立三层 BP 神经网络, 最后, 应用广义的逆模糊数公式将预测模糊集进行逆模糊化, 从而得到预测结果. 应用本文方法对印度国家银行(SBI)股票价格和 Alabama 大学的入学人数进行预测, 预测结果精度较高.

关 键 词: 二阶模糊时间序列; BP 神经网络; 逆模糊数; 模糊时间序列; 股票价格

中图分类号: O 159 **文献标志码:** A **文章编号:** 1005-3026(2019)02-0300-05

A Second-Order Fuzzy Time Series Model for Stock Price Analysis

LIU Zhi^{1,2}, ZHANG Tie¹, DONG Ying³, XU Shuang-shuang²

(1. School of Sciences, Northeastern University, Shenyang 110819, China; 2. Department of Basics, Shenyang University of Technology, Liaoyang 111003, China; 3. College of Science, Dalian Nationalities University, Dalian 116600, China. Corresponding author: LIU Zhi, E-mail: liuzhi2099@sut.edu.cn)

Abstract: It is difficult to model stock market because of its uncertainty, while fuzzy time series has its advantages in dealing with fuzzy and uncertainty data. Accordingly, the data was first preprocessed and a new way to divide the universe of discourse was given, after which the data was fuzzified using the triangular membership function, and a three-layer BP neural network was then established according to the fuzzified data. Finally, the generalized inverse fuzzy number formula was used to defuzzify the fuzzy relation, with the prediction results obtained. The method was used for predicting the stock price of State Bank of India (SBI) and the enrollment of the University of Alabama, and the results showed that the prediction accuracy is higher than that of the related previous methods.

Key words: second-order fuzzy time series; BP neural network; inverse fuzzy number; fuzzy time series; stock price

1965 年, Zadeh^[1] 提出了模糊集的概念. 由于模糊性事件在现实生活中是普遍存在的, 从而模糊数学一经产生, 便显示出旺盛的生命力, 得到了广泛的应用. 1972 年, Chang 和 Zadeh^[2] 又提出了模糊数的概念. 时至今日, 模糊数学在实际中的应用几乎涉及到了国民经济的各个领域及部门^[3-10].

1927 年, 英国统计学家尤尔 (Yule) 首创的 AR(2) 模型, 标志着时间序列分析的开端. 随后, 时间序列的理论和实践得到了飞速发展, 在现代社会中的应用也日益广泛^[11-12].

模糊时间序列的概念首次由 Song Q 和 Chissom B S 在 1993 年提出, 用模糊关系方程作为模型结合时间序列理论, 将其应用到大学注册

收稿日期: 2017-12-06
基金项目: 国家自然科学基金专项基金天元访问学者项目(11726616); 辽宁省博士科研启动基金资助项目(201501164); 中央高校基本科研业务费专项资金资助项目(DC201501040).
作者简介: 刘智(1979-), 女, 辽宁辽阳人, 沈阳工业大学讲师, 东北大学博士研究生; 张铁(1956-), 男, 辽宁沈阳人, 东北大学教授, 博士生导师.

学生数的预测和温度预测等领域,他们所发展的模糊时间序列预测方法在处理数据资料不完整或者不合理的问题时,通过模糊理论的特性就可以得到有效的解决,所以预测结果较传统模型准确率有了一定的提高.

模糊时间序列的关键步骤就是论域的划分、模糊逻辑关系的建立和逆模糊化. 本文改进了划分论域的方法,利用 BP 神经网络将模糊化的数据建立模糊关系,并应用该关系进行预测得到预测模糊集,最后应用逆模糊数公式得到预测数据. 通过比较,本文提出的方法具有更高的精度,且应用起来更方便有效.

1 二阶模糊时间序列

1.1 模糊时间序列的基本概念

定义 1 ^[13] 对实数域 \mathbf{R} 的任意一个固定的子集 $Y(t) (t=0,1,2,\cdots)$, $f_i(t) (i=1,2,\cdots)$ 是定义在 $Y(t)$ 的一组模糊集,且 $F(t)$ 是 $f_i(t) (i=1,2,\cdots)$ 的集合,则称 $F(t)$ 为定义在 $Y(t)$ 上的模糊时间序列.

定义 2 ^[13] 设 $F(t)$ 仅由 $F(t-1)$ 确定,这种关系可以通过模糊关系 $F(t) = F(t-1) \cdot R(t, t-1)$ 得到,其中 $R(t, t-1)$ 表示 $F(t)$ 到 $F(t-1)$ 的模糊逻辑关系,则被称为定义在 $F(t)$ 上的一阶模糊关系,记为: $F(t-1) \rightarrow F(t)$.

定义 3 ^[14] 设 $F(t)$ 由 $F(t-1), F(t-2), \cdots, F(t-n)$ 确定,则 n 阶模糊时间序列被定义为 $F(t-n), \cdots, F(t-1) \rightarrow F(t)$.

1.2 构建二阶模糊时间序列

1.2.1 数据预处理

根据观测数据,利用式(1)计算出数据的变化率 y_{t+1} :

$$y_{t+1} = \frac{x_{t+1} - x_t}{x_t} \times 100\% \quad (1)$$

式中 x_t 为原始数据.

1.2.2 改进的聚类分析算法

将经过预处理后的数据按照升序排列后设为 a_i ,按照以下步骤进行处理:

①计算每两个数据之间距离 d_i 和 n 个数据之间的平均距离 d_{mean} :

$$d_i = a_{i+1} - a_i. \quad (2)$$

$$d_{\text{mean}} = \frac{\sum_{i=1}^{n-1} (a_{i+1} - a_i)}{n-1}. \quad (3)$$

②按如下规则聚类.

规则一:将每个数据归为一类,形成初始聚类.

规则二:将每两个数据之间距离 d_i 从小到大排列,下标保持不变.从最小的 d_i 开始,进行如下处理:

情况一:如果 a_i 及 a_{i+1} 所处的聚类只有一个元素,形如 $\cdots, \{a_i\}, \{a_{i+1}\}, \cdots$, 且 $d_i < d_{\text{mean}}$, 则将 a_i 及 a_{i+1} 归为一类 $\{a_i, a_{i+1}\}$, 否则保持原聚类不变.

情况二:如果 a_i 或 a_{i+1} 所处的聚类有 m 个元素 ($m > 1$), 则保持原聚类不变.

重复以上过程直到 $d_i > d_{\text{mean}}$, 聚类结束.

③划分论域:利用聚类的结果,按照以下规则划分论域.

规则一:假设当前聚类有两个或两个以上的元素,则保留最小和最大的元素去掉其他的元素,得到论域 $[a_m, a_n]$.

规则二:假设当前聚类只有一个元素 a_i , 它前面为论域 $[a_k, a_l]$, 它后面为论域 $[a_m, a_n]$, 计算 $a_i - d_{\text{mean}}$ 和 $a_i + d_{\text{mean}}$, 如果 $a_i - d_{\text{mean}} > a_l$, 且 $a_i + d_{\text{mean}} < a_m$, 则得到论域 $[a_i - d_{\text{mean}}, a_i + d_{\text{mean}}]$; 否则, 计算 $d_1 = d_{\text{mean}}, d_2 = a_i - a_l, d_3 = a_m - a_i$, 设 $d_k = \min\{d_1, d_2, d_3\}$, 得到论域 $[a_i - d_k, a_i + d_k]$.

规则三:假设当前聚类只有一个元素 $a_i, \{a_i\}$ 前为论域 $[a_k, a_l], a_i$ 后为聚类 $\{a_m\}$, 则计算 $d_1 = d_{\text{mean}}, d_2 = a_i - a_l, d_3 = (a_m - a_i)/2$, 设 $d_k = \min\{d_1, d_2, d_3\}$, 得到论域 $[a_i - d_k, a_i + d_k]$.

规则四:假设当前聚类只有一个元素 $a_i, \{a_i\}$ 前为聚类 $\{a_l\}, a_i$ 后论域为 $[a_m, a_n]$, 则计算 $d_1 = d_{\text{mean}}, d_2 = a_m - a_i, d_3 = (a_i - a_l)/2$, 设 $d_k = \min\{d_1, d_2, d_3\}$, 得到论域 $[a_i - d_k, a_i + d_k]$.

规则五:将论域补充完整,构成连续的区间.

1.2.3 建立二阶模糊关系链

利用三角隶属度函数公式得到模糊集. 将预处理后的数据分别对应到相应的模糊集,60% 作为训练数据,其余的作为测试数据. 构建三层 BP 神经网络. 利用模糊集的中点坐标作为神经网络的输入、输出,建立模糊时间序列.

1.2.4 利用逆模糊数公式逆模糊化

利用 1.2.3 节输出的结果,假设 m_{i-1}, m_i, m_{i+1} 分别为 A_{i-1}, A_i, A_{i+1} 的中点; $\mu_{i-1}, \mu_i, \mu_{i+1}$ 分别为 A_{i-1}, A_i, A_{i+1} 的隶属度; b_i 为变化率的预测值,则逆模糊数公式为

$$b_i = \begin{cases} \frac{1 + \mu_{i+1}}{\frac{1}{m_i} + \frac{\mu_{i+1}}{m_{i+1}}}, & i = 0; \\ \frac{\mu_{i-1} + 1 + \mu_{i+1}}{\frac{\mu_{i-1}}{m_{i-1}} + \frac{1}{m_i} + \frac{\mu_{i+1}}{m_{i+1}}}, & 1 \leq i \leq n-1; \\ \frac{\mu_{i-1} + 1}{\frac{\mu_{i-1}}{m_{i-1}} + \frac{1}{m_i}}, & i = n. \end{cases} \quad (4)$$

再利用式(4)计算出原始数据的预测值:

$$p_{t+1} = x_t \cdot (1 + b_t\%) . \quad (5)$$

2 仿 真

应用本文提出的模糊时间预测方法,对 SBI 日股票价格和 Alabama 大学的入学人数进行预测,并将预测结果与其他方法预测结果进行比较.

2.1 预测股票价格

本文选用从 2012-06-05 到 2012-07-31 SBI 日股票价格(数据来自 <http://in.finance.yahoo.com/>),应用本文提出的方法对股票价格进行预测.

①用 x_t 表示日股票价格(见表 1 第 2 列),利用式(1)对数据进行预处理,预处理后的数据用 y_i 表示,见表 1 第 3 列.

②划分论域. 利用式(3)计算出每个数据之间的平均距离 $d_{\text{mean}} = 0.214\,9$,根据 1.2.2 的算法将数据划分为 43 个论域,再根据股票的熔断机制(股票的日变化率不得超过 7%),在最前面加上论域 u_0 ,在最后面加上 u_{44} ,将其补充完整. 结果如下:

$$\begin{aligned} u_0 &= [-7, -4.656\,1), \\ u_1 &= [-4.656\,1, -4.062\,1), \\ u_2 &= [-4.062\,1, -3.468\,2), \\ u_3 &= [-3.468\,2, -3.298\,1), \\ &\vdots \\ u_{40} &= [2.893\,7, 3.039\,9), \\ u_{41} &= [3.039\,9, 3.041\,8), \\ u_{42} &= [3.041\,8, 3.552\,0), \\ u_{43} &= [3.552\,0, 4.062\,4), \\ u_{44} &= [4.062\,4, 7). \end{aligned}$$

计算出相应的区间中点,结果如下:

$$\begin{aligned} m_0 &= -5.828\,1, m_1 = -4.359\,1, \\ m_2 &= -3.765\,2, m_3 = -3.383\,1, \\ m_4 &= -3.060\,0, m_5 = -2.583\,7, \\ &\vdots \end{aligned}$$

$$\begin{aligned} m_{38} &= 1.900\,9, m_{39} = 2.425\,1, \\ m_{40} &= 2.966\,8, m_{41} = 3.040\,8, \\ m_{42} &= 3.296\,9, m_{43} = 3.807\,2, \\ m_{44} &= 5.531\,2. \end{aligned}$$

表 1 SBI 股票价格数据模糊化
Table 1 Data fuzzification of the stock price of SBI

月/日	股价/卢比	变化率/%	模糊集	模糊集中点
6/5	2 080.25	—	—	—
6/6	2 159.45	-3.807	A_{43}	3.807
6/7	2 167.85	0.389	A_{28}	0.476
6/8	2 180.05	0.563	A_{29}	0.605
\vdots	\vdots	\vdots	\vdots	\vdots
7/24	2 094.80	0.109	A_{23}	0.108
7/25	2 070.65	-1.153	A_{12}	-1.114
7/26	2 017.15	-2.584	A_5	-2.584
7/27	1 941.20	-3.765	A_2	-3.765
7/31	2 005.20	3.297	A_{42}	3.297

③利用划分论域将股票变化率数值进行模糊化,结果见表 1 第 4 列,并找到相应模糊集的中点,结果见表 1 第 5 列.

④利用三层 BP 神经网络建立二阶模糊关系链. 将模糊化的数据分为两组,前 24 个数据作为训练数据,后 15 个数据作为测试数据. 假设要预测的时刻为 t ,其对应的模糊集作为输出节点数据;设其前 2 个时刻分别为 $t-1, t-2$,其相应模糊集分别作为 2 个输入节点数据. 利用训练数据训练并得到相应的神经网络,再利用测试数据进行测试,得到其对应的模糊集(见表 2 第 3 列).

⑤利用式(4)进行逆模糊化得到相应变化率的预测值(见表 2 第 4 列). 再利用式(5)计算出股票价格的预测值(见表 2 第 5 列).

表 2 SBI 股票价格预测结果
Table 2 Prediction results of the stock price of SBI

月/日	股价/卢比	预测 模糊集	预测 变化率/%	预测股 票价格	AER/%
6/8	2 180.05	A_{30}	0.659	2 182.1	0.094
6/11	2 164.55	A_{15}	-1.073	2 156.7	0.363
6/12	2 206.90	A_{36}	1.266	2 191.9	0.680
6/13	2 222.25	A_{28}	0.469	2 217.2	0.227
6/14	2 154.25	A_6	-2.190	2 173.6	0.898
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
7/5	2 232.20	A_{27}	0.362	2 232.3	0.005
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
7/24	2 094.80	A_{24}	0.157	2 095.8	0.048
7/25	2 070.65	A_{12}	-1.128	2 071.2	0.027
7/26	2 017.15	A_6	-2.190	2 025.3	0.404
7/27	1 941.20	A_3	-3.380	1 949.0	0.402
7/31	2 005.20	A_{40}	2.826	1 996.1	0.454

为了证明本文方法的优越性,将本文方法的预测值与文献[15]、文献[16]、文献[17]和文献[18]方法的预测值进行比较,结果见图 1. 可以看出本文的方法最接近真实值. 为了更直观地比较,利用式(6)计算出各种方法的均方误差 RMSE,利用式(7)计算出绝对误差率 AER,本文提出的方法绝对误差率最大值为 0.898%,最小值为 0.005%,可见本文方法预测值更加精确,更值得推广.

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (p_t - x_t)^2}{n}} . \tag{6}$$

$$AER = \frac{|x_t - p_t|}{x_t} \times 100\% . \tag{7}$$

式中: p_t 表示 t 时刻的预测值; x_t 表示 t 时刻的真实数据; n 表示数据总数.

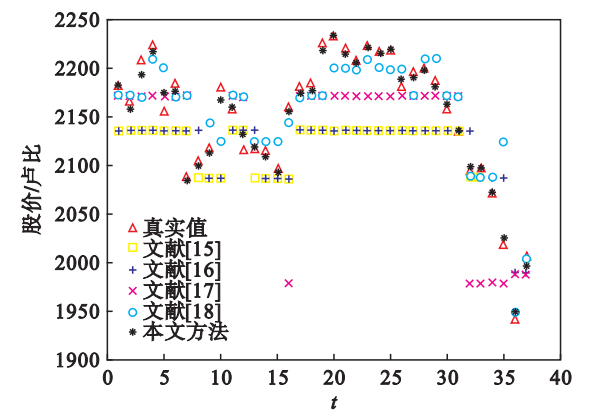


图 1 采用不同方法的预测值

2.2 预测 Alabama 大学的入学人数

应用本文提出的方法对 Alabama 大学从 1971 到 1992 年的入学人数进行预测.

首先对入学人数数据进行预处理,计算出年变化率;然后,计算变化率间的平均距离 $d_{\text{mean}} = 0.6745$,根据 1.2.2 的算法将数据划分为 27 个论域,并计算出每个论域的中点,将所有数据模糊化;利用三层 BP 神经网络建立二阶模糊关系链.

将模糊化的数据分为两组,利用训练数据训练并得到相应的神经网络,再利用测试数据进行测试得到其对应的模糊集;最后,利用式(4)进行逆模糊化得到相应变化率的预测值(见表 3 第 4 列),再利用式(5)计算出预测的入学人数(见表 3 第 5 列).

表 3 的最后一列为每个预测值与真实值的绝对误差率,通过表 3 可以看出,本文提出的模糊时间序列方法得到的绝对误差率最大为 3.65%,表

明本文方法预测精度高. 为了进一步证明这一点,在表 4 中将本文预测结果与文献[13]、文献[18]和文献[19]方法的预测结果进行比较. 通过表 4 可以看出,由本文方法得到的预测值的均方误差和绝对误差率都远远小于其他方法.

表 3 Alabama 大学入学人数预测
Table 3 Predicted enrollment of the University of Alabama

入学时间	人数 $x(t)$	变化率 $y(t)/\%$	预测变 化率/ $\%$	预测值 $f(t)$	AER/ $\%$
1971	13 055	—	—	—	—
1972	13 563	3.89	—	—	—
1973	13 867	2.24	2.374 0	13 885	0.13
1974	14 696	5.98	6.593 4	14 781	0.58
1975	15 460	5.20	4.223 0	15 317	0.92
⋮	⋮	⋮	⋮	⋮	⋮
1979	16 807	5.96	9.829 1	17 420	3.65
⋮	⋮	⋮	⋮	⋮	⋮
1988	18 150	7.66	7.220 6	18 076	0.41
1989	18 970	4.52	4.779 6	19 017	0.25
1990	19 328	1.89	1.852 6	19 321	0.04
1991	19 337	0.05	3.113 4	19 367	0.16
1992	18 876	-2.38	-2.243 0	18 903	0.14

表 4 不同方法预测结果比较
Table 4 Comparison of predicted results from different methods

入学时间	人数 $x(t)$	文献[13] 预测值	文献[18] 预测值	文献[19] 预测值	本文 预测值
1971	13 055	—	—	14 464	—
1972	13 563	14 000	—	14 464	—
1973	13 867	14 000	—	14 464	13 885
1974	14 696	14 000	14 500	1471 0	14 781
1975	15 460	15 500	15 500	15 606	15 317
⋮	⋮	⋮	⋮	⋮	⋮
1988	18 150	16 813	18 500	18 473	18 076
1989	18 970	19 000	18 500	18 473	19 017
1990	19 328	19 000	19 500	19 155	19 321
1991	19 337	19 000	19 500	19 155	19 367
1992	18 876	—	18 500	18 473	18 903
RMSE		775 687	86 694	227 194	666.40
AER/ $\%$		4.380 0	1.529 4	2.386 5	0.101 0

3 结 语

本文提出了一种新的二阶模糊时间序列. 它首先计算出数据的变化率,再利用聚类分析方法将数据变化率划分成相应的论域;然后利用三角

隶属度函数将变化率进行模糊化,建立三层的 BP 神经网络,并据此构造二阶模糊关系链;最后利用逆模糊数公式得到预测的数据的变化率并计算出预测值.为了体现本文方法的鲁棒性及精确性,本文将提出的方法应用于股票价格和入学人数预测,实验结果精度较高.

参考文献:

[1] Zadeh L A. Fuzzy sets[J]. *Information Control*,1965,8(3): 338 – 353.

[2] Chang S S L, Zadeh L A. On fuzzy mapping and control[J]. *IEEE Transactions on Systems, Man, and Cybernetics*,1972, 2:30 – 34.

[3] Dubois D J. Fuzzy sets and systems:theory and applications [M]. [S. l.]:Academic Press,1980:10 – 79.

[4] Saeidifar A, Pasha E. The possibilistic moments of fuzzy numbers and their applications[J]. *Journal of Computational and Applied Mathematics*,2009,223(2):1028 – 1042.

[5] Xu J P, Liu Q, Wang R. A class of multi-objective supply chain networks optimal model under random fuzzy environment and its application to the industry of Chinese liquor [J]. *Information Sciences*, 2008, 178 (8): 2022 – 2043.

[6] Panigrahi M, Panda G, Nanda S. Convex fuzzy mapping with differentiability and its application in fuzzy optimization[J]. *European Journal of Operational Research*,2008,185(1): 47 – 62.

[7] Lee C S, Wang M H. Ontology-based intelligent healthcare agent and its application to respiratory waveform recognition [J]. *Expert Systems with Applications*, 2007, 33 (3): 606 – 619.

[8] Yager R R. On ordered weighted averaging aggregation operators in multicriteria decision making [J]. *IEEE Transactions on Systems, Man, and Cybernetics*, 1988, 18 (1):183 – 190.

[9] Chrysafis K A, Papadopoulos B K. On the fuzzy difference equations of finance[J]. *Fuzzy Sets and Systems*,2008,159 (24):3259 – 3270.

[10] Aldrich J. Correlations genuine and spurious in Pearson and Yule[J]. *Statistical Science*,1995,10(4):364 – 376.

[11] Box G E P, Jenkins G M, Reinsel G C, et al. Time series analysis: forecasting and control [J]. *Journal of the Operational Research Society*, 1975,22(2):199 – 201.

[12] Anderson O. The elimination of spurious correlation due to position in time or space[J]. *Biometrika*,1914,10(2/3): 269 – 279.

[13] Song Q, Chissom B S. Forecasting enrollments with fuzzy time series—part I [J]. *Fuzzy Sets and Systems*,1993,54 (1):1 – 9.

[14] Hebb D. The organization of behavior[M]. [S. l.]:Wiley, 1949:1 – 48.

[15] Chen S M. Forecasting enrollments based on fuzzy time series [J]. *Fuzzy Sets and Systems*,1996,81(3):311 – 319.

[16] Chen S M. Forecasting enrollments based on high-order fuzzy time series[J]. *Cybernetic Systems*,2002,33(1):1 – 16.

[17] Yu H K. Weighted fuzzy time series models for TAIE X forecasting[J]. *Physica A*,2005,349(3/4):609 – 624.

[18] Singh P, Borah B. High-order fuzzy neuroexpert system for time series forecasting[J]. *Knowledge Based Systems* 2016, 46:12 – 21.

[19] Jilani T A, Burney S M A, Ardil C. Fuzzy metric approach for fuzzy time series forecasting based on frequency density based partitioning [J]. *Proceedings of World Academy of Science: Engineering and Technology*, 2007, 23 (2): 112 – 117.

(上接第 299 页)

[4] Vapnik V N. The nature of statistical learning theory [M]. Berlin:Springer-Verlag,1995:1 – 50.

[5] Hsu C W, Chang C C, Lin C J. A practical guide to support vector classification [R]. Taipei: University of National Taiwan,2003:1 – 12.

[6] Huang C L, Wang C J. A GA-based feature selection and parameters optimization for support vector machine [J]. *Expert Systems with Applications*,2006,31(2):231 – 240.

[7] Zhang K, Li C, Wang Y, et al. Collaborative support vector machine for malware detection [J]. *Procedia Computer Science*,2017,108:1682 – 1691.

[8] Tomar D, Agarwal S. Twin support vector machine;a review from 2007 to 2014[J]. *Egyptian Informatics Journal*,2015, 16(1):55 – 69.

[9] Bamakan S M H, Wang H D, Ravasan A Z. Parameters optimization for nonparallel support vector machine by particle swarm optimization[J]. *Procedia Computer Science*, 2016,91:482 – 491.

[10] 陈建飞,蒋刚,杨剑锋.改进 ABC – SVM 的参数优化及应用[J]. *机械设计与制造*,2016,52(1):24 – 28.

(Chen Jian-fei, Jiang Gang, Yang Jian-feng. Improved ABC-SVM parameter optimization and application[J]. *Machinery Design and Manufacture*,2016,52(1):24 – 28.)