DOI: 10.11992/tis.201707023
网络出版地址: http://kns.cnki.net/kcms/detail/23.1538.tp.20180419.0901.002.html

标签传播时间序列聚类的股指期货套期保值策略研究

李海林[1,2]，梁叶[1]
（1. 华侨大学 信息管理系，福建 泉州 362021；2. 华侨大学 现代应用统计与大数据研究中心，福建 厦门 361021）

摘　要：利用时间序列聚类方法进行股指期货的套期保值，关键要选择合适的聚类方法。本文从新的视角来研究并提高时间序列聚类方法在金融数据分析领域的应用性能，提出一种基于标签传播时间序列聚类的股指期货套期保值模型。该模型以动态时间弯曲为相似性度量方法来构建现货股票网络空间结构，将每只股票看作一个节点，利用标签传播方法将节点划分到不同的簇中，最终实现股票数据聚类。另外，构建最小追踪误差优化模型来确定每支股票在现货组合中的最优权重，从而得到最优组合。实验分别比较新方法和传统聚类方法确定现货组合的追踪误差，结果表明新方法能够提高现货组合的追踪精度，为丰富金融市场投资和管理方式提供新的研究思路。

关键词：标签传播；时间序列；聚类；动态时间弯曲；套期保值
中图分类号：TP391　　**文献标志码：**A　　**文章编号：**1673−4785(2019)02−0288−08

中文引用格式：李海林，梁叶．标签传播时间序列聚类的股指期货套期保值策略研究 [J]．智能系统学报，2019，14(2)：288–295.
英文引用格式：LI Hailin, LIANG Ye. Research on the stock index futures hedging strategy using label propagation time series clustering[J]. CAAI transactions on intelligent systems, 2019, 14(2): 288–295.

Research on the stock index futures hedging strategy using label propagation time series clustering

LI Hailin[1,2], LIANG Ye[1]
(1. Department of Information Systems, Huaqiao University, Quanzhou 362021, China; 2. Research Center of Applied Statistics and Big Data, Huaqiao University, Xiamen 361021, China)

Abstract: Choosing a suitable clustering method is crucial in using time series clustering in stock index futures hedging. This study aims to investigate and improve the application performance of time series clustering in the financial data analysis field from a new perspective. We propose a model of stock index futures hedging based on label propagation time series clustering. In the model, a network space of spot stock was built using dynamic time warping as similarity measure. Each stock in the network was treated as a node, which would be divided into different clusters using label propagation, and finally, the stock data was clustered successfully. An optimization model for minimizing tracking error was constructed to obtain the optimal weight of each stock in the spot portfolio. Finally, we obtained the optimal spot portfolio. The tracking errors of the portfolio of the proposed method and that of the traditional clustering method were compared by tracking the index in the experiment. The proposed method showed the ability to improve tracking accuracy, providing a new way to enrich the investment and management of financial market.

Keywords: label propagation; time series; clustering; dynamic time warping; hedging

收稿日期：2017−07−11.　**网络出版日期：**2018−04−20.
基金项目：国家自然科学基金项目 (71771094, 61300139)；福建省社科规划项目 (FJ2017B065)；华侨大学中青年教师科研提升计划项目 (ZQN-PY220).
通信作者：李海林. E-mail: hailin@mail.dlut.edu.cn.

股指期货是基于股票指数的金融衍生品，不仅能够对现货资产进行对冲来降低系统风险，还能作为一种投机套利工具来丰富资产组合，获得良好收益。股指期货与现货市场存在一定的关

联, 正常市场条件下双方表现出同涨同跌的情况[1]。现货组合与股指期货表现出的趋同程度越大, 套期保值的效果越明显。标的指数是一篮子股票, 在实际操作中, 是无法使用所有现货股票来构建组合的。此外, 随机或者人工指定少数股票来构建投资组合, 不仅无法充分把握整体信息, 又会产生一定的模拟误差, 增加投资风险。因此, 用较少的股票来构建现货投资组合, 降低同步买卖的成本, 并反映该投资组合与标的指数之间的相关性, 是本文的研究关键。

目前, 针对股指期货的套期保值研究有不少的研究和应用成果[2-6], 其中通过指数复制和组合追踪来研究股指期货的套期保值也日益得到国内外研究者的关注, 采用新的量化研究方法成为研究指数复制问题的一个方向。苏治等[7] 建立混合整数线性规划模型并引入内核搜索分析框架, 通过实证分析发现, 增强型内核搜索法在成分股很大时才能够得到高质量的解, 考虑现实交易成本特征的模型具有更好的稳健性, 在指数追踪时投资组合的动态调整具有一定的必要性。倪禾[8] 提出一种基于启发式遗传算法的寻优方案, 通过最大效用函数来寻找一个最为经济的组合, 该组合拥有最少的资产数、较少的权重调整次数和尽可能的接近或超越标的指数收益的优势, 具有较强的实用性。胡春萍等[9] 构建时间加权 SVN 的指数优化复制模型, 不仅能够考虑时间因素对历史追踪误差的影响, 还能够提高追踪组合的追踪效果。刘睿智等[10] 使用自适应套索方法构造稀疏投资组合, 使得该组合对指数的追踪效果良好。Chen 等[11] 设计了一种稳健的跟踪市场指数投资组合模型, 该模型旨在利用 0-1 规划找到组合和标的指数之间的最大相似性, 且在实验中, 该模型得到的投资组合具有更低的跟踪误差和投资风险。Guastaroba 等[12] 为了平衡最小跟踪误差和超额收益的关系, 设计出混合整数线性规划方法, 利用内核搜索方法寻求最优解。该方法可以引入真实交易的特征, 获得的最优组合也具有较好的表现, 相比于传统的启发式方法具有更优的性能。Filippi 等[13] 针对指数追踪问题中获得超额收益和最小跟踪误差双的目标问题, 提出了双目标整数线性规划模型。该模型旨在获得的组合在指数追踪上具有最小的追踪误差, 并且能够获得超额收益。

时间序列聚类在金融领域的应用越来越受到研究者的关注, 对探究金融市场的发展规律、把握市场信息起到重要的作用。Dose 等[14] 利用基于随机优化技术的时间序列聚类来增强指数跟踪, 通过层次聚类得到不同的簇, 每个簇中选择股票的子集并确定相应的权重后作为投资组合。该方法表明, 通过聚类方法要比随机选择股票组合具有更少的噪声和更好的稳健性。Nanda 等[15] 利用 K-means 聚类将股票划分成不同的簇, 并从不同簇中选择一些股票作为组合。由于组合来自不同的簇, 分散投资使得风险最小化。柴尚蕾等[1] 利用基于独立主成分分析和模糊 C 均值聚类的两阶段优化方式构建现货组合, 相比随机方法得到更低的跟踪误差, 提高了组合对大盘动荡的抵抗力。Lemieux 等[16] 分别探讨了通过 3 种传统聚类方法获得的投资组合上在实际交易中的应用, 分析不同聚类技术是如何影响分析师对不同风险组合的看法。

时间序列聚类在金融领域得到了充分的利用, 但传统时间序列聚类方法一般先指定初始簇中心, 也不能充分反映其空间组织联系。社区发现是复杂网络研究中的重要研究工具, 根据某种规则将网络中的节点划分为若干个社区, 每个社区内部的节点连接较为紧密, 而社区之间连接稀疏。因此, 社区发现与聚类是不谋而合的。目前结合社区发现探索时间序列聚类的有关成果仍旧不足[17], 标签传播[18] 作为简便的划分算法, 可以在相连的节点中传播有用的信息。将标签传播应用到时间序列聚类分析, 是一种新的研究视角。此外, 动态时间弯曲是一种度量准确性高的度量方法, 是各个领域理论和应用研究的热点。综上, 本文提出一种基于标签传播的时间序列聚类方法来对股指期货套期保值策略进行研究。利用动态时间弯曲构建股票池网络, 将每只股票看作网络中独立的节点; 通过标签传播方法将网络中的节点划分成不同的社区, 实现时间序列聚类; 构建最小追踪误差模型, 优化每只股票在组合中的权重。

1　相关理论基础

标签传播是社区发现领域的重要方法之一, 其简单高效的思想使其得到了广泛的关注。动态时间弯曲是泛化能力强的度量方法, 是各个领域的研究热点。为了充分理解标签传播方法计算原理, 对标签传播和动态时间弯曲的基本原理进行阐述, 明确两者在本研究中的重要作用。并给出了优化权重过程所使用的最小追踪误差模型。

1.1　标签传播

标签传播 (label propagation algorithm, LPA)

是根据网络的局部信息结构，利用节点的连接关系自动传播信息，最终得到社区的划分结果。LPA 不需要事先指定社区规模、个数、优化目标函数，算法简单又容易实现，同时具有接近线性的时间复杂度。因此，获得了社区发现领域的广泛关注。设无向无权网络$G(V,E)$，V为节点集，E为边集。利用 LPA 获取节点的标签集L，以c_x表征节点x的标签。先随机为每个节点分配唯一的标签作为其所在的社区的标识，其标签的决定取决于邻接节点标签的分布状况。该算法如下。

算法　标签传播算法 (label propagation algorithm, LPA)

输入　网络$G(V,E)$;

输出　L。

1) $t=0$，首次定义全部节点的标签。节点x有$L_x(0)=x$。$L_x(t)$为x于第t次计算时的标签。

2) $t=1$。

3) 打乱节点顺序，获取打乱顺序的节点集V'。

4) 对于各个节点$x\in V'$，设$L_x(t)=f(L_{Nb(x)}(t))$，其中$Nb(x)$表示节点x的邻接点集，$f(\cdot)$得到个数最大的标签。

5) 若$L(t)$与$L(t-1)$相等，则算法停止；反之，$t=t+1$，返回 3)，算法重复进行 ($L(t)$为第t次计算获取的标签集)。

1.2　动态时间弯曲

动态时间弯曲 (dynamic time warping, DTW) 为一种鲁棒性强的度量方法，最早用于语音识别[19]。与欧氏距离相比，DTW 能够弯曲时间轴达到不等长时间序列的度量，充分反映时间序列的形态[20]。特别地，在时间数据挖掘[21]领域中，动态时间弯曲具有较为广泛的应用。利用欧氏距离构建两条时间序列$S=\{s_1,s_2,\cdots,s_n\}$和$Q=\{q_1,q_2,\cdots,q_m\}$的距离矩阵$\boldsymbol{D}_{n\times m}$，$\boldsymbol{D}$中的元素为数据点的欧氏距离。DTW 的目的就是在$\boldsymbol{D}$中找到一条弯曲路径$P=\{p_1,p_2,\cdots,p_H\}$使得$S$和$Q$之间的累积代价最小，则$S$和$Q$的 DTW 距离为

$$\mathrm{DTW}(S,Q)=\min_P\left(\sum_{h=1}^{H}p_h\right) \tag{1}$$

式中：P是由H个距离元素的集合，代表着S和Q之间数据点的最佳匹配关系。最优弯曲路径可以利用动态规划方法来构造一个累积代价矩阵$\boldsymbol{\gamma}$获得，即

$$\gamma(i,j)=D(i,j)+\min\begin{cases}\gamma(i,j-1)\\ \gamma(i-1,j-1)\\ \gamma(i-1,j)\end{cases} \tag{2}$$

式 (2) 表示当前的累积代价为当前距离加上相邻 3 个最小的累积代价。最后$\gamma(n,m)$为起点反向搜索路径，以相邻最小的累积代价元素作为下一个路径节点，直到搜索至$\gamma(1,1)$。那么，$\gamma(n,m)=\mathrm{DTW}(S,Q)$即为两条时间序列的 DTW 距离。

1.3　最小追踪误差模型

为得到最优现货组合，从标的指数的N只成分股中选出$k(k<N)$只来构建现货投资组合，使得该现货投资组合的股票收益率与标的指数的收益率误差最小。根据最小误差追踪模型 (minimum error tracking model, METM) 来得到各个股票所占比例，得到最优现货组合。设有T个交易日的追踪组合，在第$t(t\in[1,T])$个交易日的收益率为

$$r_t=\ln\left(\sum_{i=1}^{n}p_{i,t}\times w_i\right)-\ln\left(\sum_{i=1}^{n}p_{i,t-1}\times w_i\right) \tag{3}$$

标的指数的收益率为

$$R_t=\ln(P_t)-\ln(P_{t-1})=\ln\left(\frac{P_t}{P_{t-1}}\right) \tag{4}$$

式中：$p_{i,t}$为组合的第$i(i=1,2,\cdots,n)$只股票于第t天的价格；w_i为第i只股票在组合中所占比例；P_t为标的指数在第t天的指数值。追踪误差TE为追踪组合的收益率和标的指数收益率的误差平方和的均方根：

$$\mathrm{TE}=\sqrt{\frac{1}{n}\sum_{t=1}^{T}(r_t-R_t)^2} \tag{5}$$

优化现货投资组合需要结合以下约束：一是资本的预算；二是投资比例的最低要求和最大限制，表示对风险的控制，也可根据投资者的风险偏好来制定。综合目标函数和约束条件，METM 如下：

$$\begin{aligned}&\min\mathrm{TE}=\\&\sqrt{\frac{1}{n}\sum_{t=2}^{T}\left[\ln\left(\sum_{i=1}^{n}p_{i,t}\times w_i\right)-\ln\left(\sum_{i=1}^{n}p_{i,t-1}\times w_i\right)-\ln\left(\frac{P_t}{P_{t-1}}\right)\right]}\\&\text{s.t.}\sum_{i=1}^{n}w_i=1,\alpha\leqslant w_i\leqslant\beta,i=1,2,,n\end{aligned} \tag{6}$$

2　标签传播时间序列聚类

聚类是数据挖掘重要任务之一，根据一定规则将数据划分为若干个簇，簇内的对象保持着高度相似性，簇之间的对象尽可能不同。在金融领域中，聚类对于板块分析、投资组合分析有着重要的意义。LPA 的便捷高效，应用到时间序列聚类则是一种新兴尝试。为了充分反映时间序列的网络空间结构，并且能够根据时间序列之间的关系相互影响自动实现聚类，提出一种基于标签传播的时间序列聚类方法，并将其应用到股指期货套期保值优化策略中。

算法　标签传播时间序列聚类 (time series clustering based on label propagation, TCLP)

输入　时间序列数据集;

输出　时间序列聚类结果。

1) 对时间序列进行标准化处理;

2) 利用 DTW 度量每条时间序列之间的距离;

3) 将时间序列视为节点, 指定距离阈值ε, 距离小于阈值ε的时间序列创建连接, 得到时间序列网络$G(V,E)$;

4) 利用 LPA 对时间序列网络进行划分, 实现聚类。

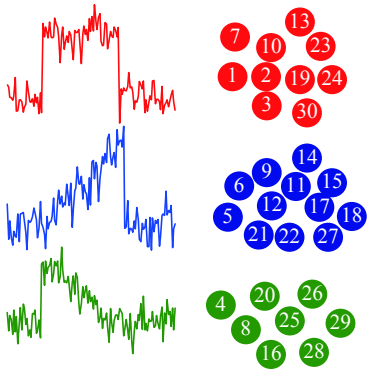
为消除量纲对度量的影响, 对时间序列进行标准化处理。例如, 由于公司资本大小不同, 经营业绩不同等原因, 使得不同股票价格千差万别, 若没有对股票价格进行标准化处理, 那么时间序列相似性度量结果则不准确, 使得两条形态很相似的时间序列之间的距离很大。步骤 2) 是时间序列相似性度量, DTW 可以实现数据点“一对多”匹配, 从而实现两条不等长时间序列的度量。由于股票数据存在可能停牌、数据缺失、数据错误等原因, 利用 DTW 可以有效地解决数据清洗之后带来的序列不等长问题。使用 LPA 之前, 需要创建好时间序列网络, 而构建时间序列网络的方法通常有两种, 即$k-NN$和$\varepsilon-NN$。$k-NN$是节点连接与之距离最短的前K个节点, $\varepsilon-NN$为给定阈值, 满足阈值ε要求的节点进行连接。相关研究表明, 以$\varepsilon-NN$创建的网络后获得的聚类结果好于$k-NN$[17], 因此本文使用$\varepsilon-NN$方法构建时间序列网络结构。在步骤 4) 中使用 LPA 对以每条时间序列为节点的网络进行划分, 最终达到时间序列聚类目的。

3　数值实验

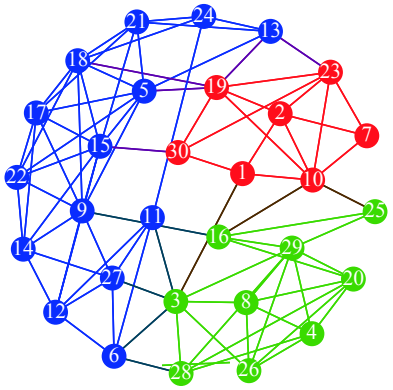
3.1　仿真实验

为验证方法的有效性, 通过对 Keogh 教授提供 CBF 数据集[22] 进行检验。CBF, 即 Cylinder-Bell-Funnel, 它是一种人工数据集, 每个类的数据都是服从标准正态分布的噪声加上一个不同于每个类的偏移量。CBF 共有 3 个类别, 每个类别时间序列形态各不相同, 可以直观地体现聚类效果。度量每条时间序列之间的 DTW 距离, 以时间序列S与它的前K条相似序列距离之和的均值作为S的距离阈值ε, 若Q与S之间的距离小于ε, 则Q与S相连。在本实验中K取 10, 表示每条时间序列与之最为相似的 10 条时间序列距离之和的均值作为连接阈值。可以发现, 每条时间序列都具有不同的连接阈值, 那么与之相连的时间序列个数则不同。CBF 数据集中各个时间序列的形态以及聚类结果如图 1 所示。

图 1(a) 左侧给出 CBF 三种形态相异的时间序列, 右侧是对应形态的时间序列的 ID。图 1(b) 展示新方法的聚类效果, 用不同颜色代表不同簇, 发现新方法能够成功将数据集划分为 3 个簇。具体分析图 1(b), 尽管节点 3 与节点 1 相连且真实情况也同属一类, 但是节点 3 与更多的绿色节点相连, 导致在标签传播过程中节点 3 接受了绿色节点传过来的标签, 被划分到了绿色的簇中。同理可分析节点 13 和节点 24。度量方法和网络构建方式影响着时间序列空间网络布局和最后的聚类结果。然而, 通过构建时间序列网络并利用标签传播方法实现聚类, 为聚类分析提供了一种新的研究模式。



(a) CBF 数据集



(b) TCLP-DTW 在 CBF 中的聚类结果

图 1　CBF 数据集和 TCLP-DTW 聚类效果

Fig. 1　CBF dataset and the clustering result of TCLP-DTW

3.2　实证分析

为了检验方法的真实有效性, 使用真实股票数据来进行实证分析。采用金融行业的股票数据, 采用 2014 年 1 月 2 日至 2014 年 12 月 31 日沪深 300 股成分股的日收盘价作为实验数据, 数据

从锐思数据库下载获得。为保证后续挖掘的质量，对数据进行清洗，即剔除 7 天以上未开盘股票，默认值为该股票的平均收盘价。数据清洗后剩余 265 只股票数据。

对每只股票数据做标准化处理：

$$Y_i = \frac{X_i - \mu}{\delta} \tag{7}$$

式中：X_i为股票第i个时刻的价格，μ为这支股票的价格均值，δ为价格方差。

利用基于 TCLP 时间序列聚类方法，对实验股票数据集进行聚类。TCLP 不必先指定簇数，能依据度量公式的特点自行划分数据集。本实验在构建股票网络结构之前，先利用 DTW 度量每条时间序列的距离，接着由前 50 相似序列距离之和的均值确定距离阈值，最后利用标签传播方法对股票网络结构进行划分，划分结果如图 2。

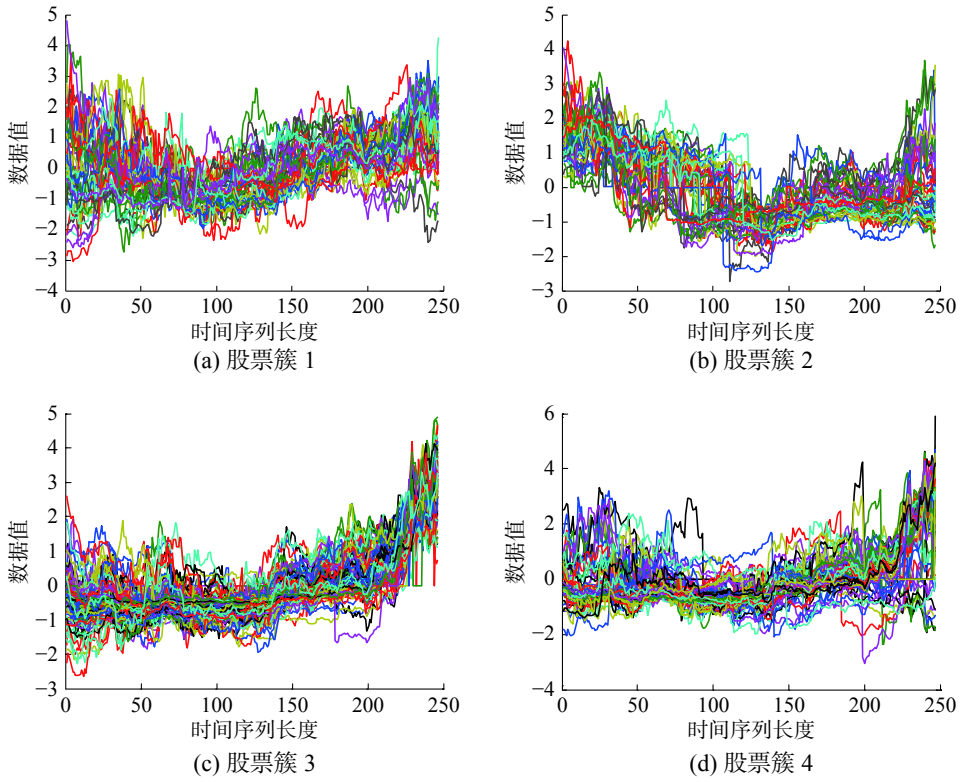


图 2　标签传播股票聚类效果

Fig. 2　Stock clustering produced by label propagation

TCLP-DTW 将股票数据集划分为了 4 个形态不同的股票簇，并且簇内的股票时间序列表现出类似形态，而不同股票簇之间的形态有着明显的区别，说明聚类效果较好。

从每个股票簇中选定追踪组合成分股，并通过 METM 计算每只追踪成分股的权重，得出追踪误差。首先，从每个股票簇中确定成分股。选择能呈现该股票簇走势的几只股票，在建立组合时表征了整体情况，分散风险。

各个股票簇的股票依据在沪深 300 指数中的比重从大到小排列，采用前两只比重大的股票，获得 8 只股票，选取结果如图 3 所示。

可以发现，各个簇选择股票的走势是相似的，而不同股票簇的成分股走势也不同，充分体现TCLP－DTW方法的聚类效果。

其次，利用 METM 计算每只成分股的权重。观察 4 种投资约束条件下 TCLP-DTW 得到的现货组合追踪误差情况。

如表 1 所示，得到了 TCLP-DTW 构建优化现货组合的 TE。尽管不同约束条件下能够获得最小的 TE，但是部分股票的权重占比很大，不利于风险分散。此外，尽管前面两种约束条件得到的追踪误差并非最低的，然而贵州茅台等股票价格较高，因此在组合中所占权重相对少一些。TCLP-DTW 构建优化现货组合是比较倾向于低投资比例的，并且股价较高的成分股占比较少，有利于投资者控制成本风险。

由于 K-means 聚类方法简单、高效，成为金融领域中应用得较多的聚类方法。为对比 K-means 和 TCLP-DTW 创建投资组合的效果，利用 K-means 将股票池划分成 4 个簇，同时对比同一条件下双方的TE，如图 4 所示。

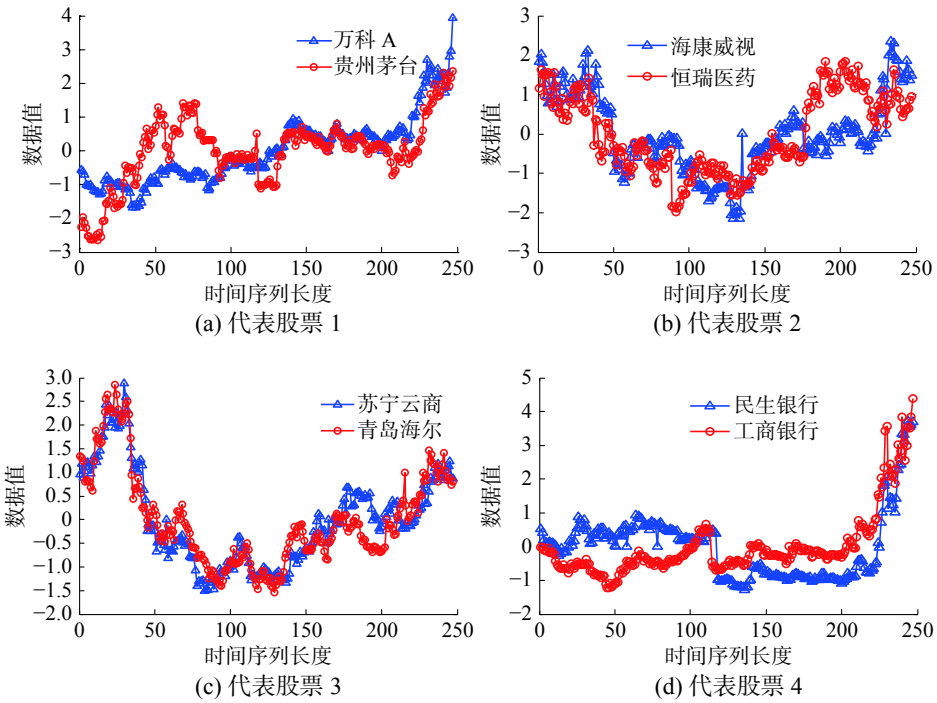


图 3　8 支成分股选取情况

Fig. 3　Eight constituent stocks were pick out

表 1　TCLP-DTW 构建优化现货组合成分权重

Table 1　Constituent stock weight optimally created by TCLP-DTW　　%

约束条件	成分股权重								追踪误差TE
	000002	600519	002415	600276	002024	600690	600016	601398	
0.5~25	9.00	9.00	9.00	19.72	9.00	17.61	17.67	9.00	1.40
1~20	20.00	1.3	1.68	5.8	19.31	11.91	20.00	20.00	1.17
5~50	5.00	5.00	5.00	9.77	5.00	10.91	14.64	44.67	1.28
0~100	54.62	0.68	0.62	1.71	3.95	3.92	3.31	31.36	0.99

从图 4 中发现, K-means 方法在投资比例不设限的条件下同样得到最低的 TE。由于投资比例不设限, 就能在更大的范围中搜索最优解。对比 TLPC-DTW、K-means 和随机这 3 种选股方式所构建优化组合的 TE。随机选股为对于各个约束比例模拟追踪 10 次, 每回随机抽取 8 只股票创建组合。利用 METM 得到股票的权重, 以 10 次 TE 的均值当做该种投资约束下的 TE。尽管 K-means 在初始化时随机选取簇中心, 但整个聚类过程是一个迭代优化的过程, 使得每个簇中对象尽可能相似, 簇间的对象尽可能相异, 这与随机选取方式有着本质的区别。为了比较聚类方式构建最优现货组合和随机构建最优现货组合, 利用

$$\mathrm{TE}_{\mathrm{others}}=\frac{E_{\mathrm{TLPC-NSM}}-E_{\mathrm{others}}}{E_{\mathrm{others}}} \tag{8}$$

作为相对追踪误差率。E_{others}为其他方法得到的 TE。若$\mathrm{TER}_{\mathrm{others}}$是负数, 表明 TLPC-DTW 构建组合的 TE 获得了优化; 反之, 说明误差更大。结果如表 2 所示。

从表 2 给出的 TE 以及 TER 发现, TCLP-DTW 构建的现货组合在 3 种投资比例约束条件下现货组合的 TE 较于 K-means、随机方法的 TER 均为负数, 表明误差均得到了改进。K-means 只对等长的时间序列聚类, 若股票长度不同, 则先预处理达到等长效果。K-means 的簇中心是簇内的对象的平均值, 不是实际中的股票。TCLP-DTW 是根据股票的相互影响进行划分, 并能从中得到股票连接关系。随机方式的随机性容易存在收益波动大的情况。通过对比发现, TCLP-DTW 构建的组合得到更小和稳定的 TE, 为研究套期保值提供新的视角。

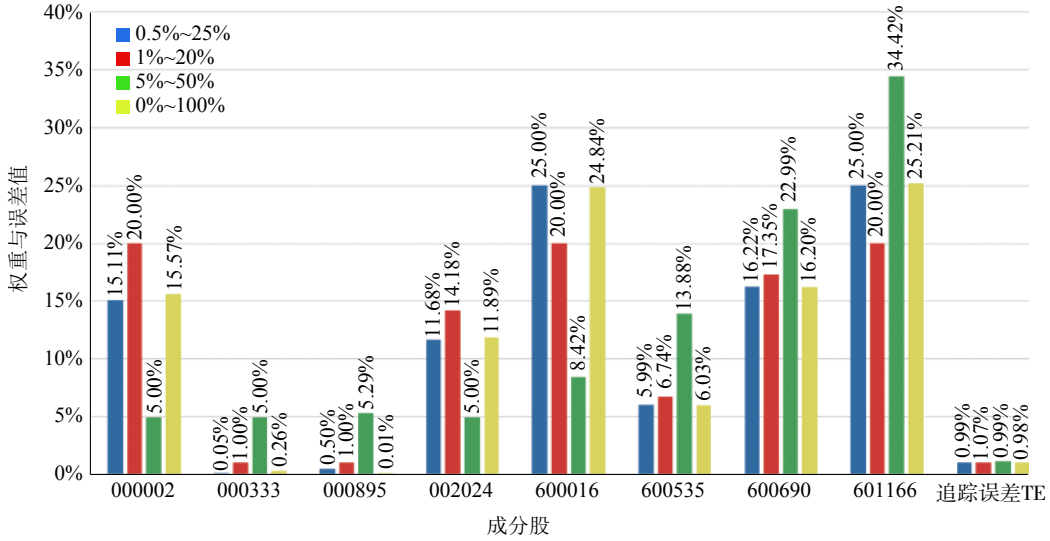


图 4　K-means 构建优化现货组合成分权重

Fig. 4　Constituent stock weight optimally created by K-means

表 2　两种方式构建现货组合的追踪误差

Table 2　Tracking error was optimally created by two methods　　%

约束条件	TCLP-DTW	K-means	random	$TER_{K\text{-means}}$	TER_{Random}
0.5~25	0.98	1.18	0.99	–16.95	–1.01
10~20	0.99	1.20	1.08	–17.50	–8.33
5~50	1.07	1.49	1.14	–28.19	–6.14
0~100	0.99	1.17	0.96	–20.51	–3.13

4　结束语

本文提出一种标签传播时间序列聚类方法，使用动态时间弯曲能够较好地度量时间序列之间的距离，结合距离阈值构建反映时间序列之间关联关系的网络，再利用标签传播来实现新方法。该方法能够较好地用于股票聚类，用来选择代表股票以确定投资组合，并通过最小误差追踪模型优化组合中的股票权重。实证分析得出，本文方法对追踪误差有一定优化，为进一步了解市场规律、提高投资效率提供一定的技术支撑。

参考文献：

[1] 柴尚蕾, 郭崇慧, 徐旭. 股指期货套利中的最优现货组合构建策略研究 [J]. 运筹与管理, 2012, 21(2): 154–161.
CHAI Shanglei, GUO Chonghui, XU Xu. Optimal spot portfolio construction strategy in stock index futures arbitrage[J]. Operations research and management science, 2012, 21(2): 154–161.

[2] 郑尊信. 股指期货持有成本模型的修正与比较 [J]. 哈尔滨工业大学学报, 2009, 41(2): 248–250.
ZHENG Zunxin. Modification and comparison of cost-of-carry model of index futures[J]. Journal of Harbin Institute of Technology, 2009, 41(2): 248–250.

[3] 韩立岩, 任光宇. 基于已实现二阶矩预测的期货套期保值策略及对股指期货的应用 [J]. 系统工程理论与实践, 2012, 32(12): 2629–2636.
HAN Liyan, REN Guangyu. Hedging strategy with futures based on prediction of realized second moment: An application to stock index futures[J]. Systems engineering—theory and practice, 2012, 32(12): 2629–2636.

[4] HOU Yang, LI S. Hedging performance of Chinese stock index futures: an empirical analysis using wavelet analysis and flexible bivariate GARCH approaches[J]. Pacific-basin finance journal, 2013, 24: 109–131.

[5] SU E D. Stock index hedging using a trend and volatility regime-switching model involving hedging cost[J]. International review of economics and finance, 2017, 47: 233–254.

[6] TRABELSI N, NAIFAR N. Are Islamic stock indexes exposed to systemic risk? Multivariate GARCH estimation of CoVaR[J]. Research in international business and finance, 2017, 42: 727–744.

[7] 苏治, 蔡腾腾, 马泽伟. 一种改进的不完全指数复制方法 [J]. 数量经济技术经济研究, 2013, 30(6): 149–160.
SU Zhi, CAI Tengteng, MA Zewei. An improved solution

for incomplete index tracking problem[J]. The journal of quantitative & technical economics, 2013, 30(6): 149–160.

[8] 倪禾. 基于启发式遗传算法的指数追踪组合构建策略[J]. 系统工程理论与实践, 2013, 33(10): 2645–2653.
NI He. Heuristic genetic algorithm for optimizing an index tracking portfolio[J]. Systems engineering—theory and practice, 2013, 33(10): 2645–2653.

[9] 胡春萍, 薛宏刚, 徐凤敏. 基于时间加权 SVM 的指数优化复制模型与实证分析 [J]. 系统工程理论与实践, 2014, 34(9): 2193–2201.
HU Chunping, XUE Honggang, XU Fengmin. An stock index replicating model based on time weighted SVM and it' s empirical analysis[J]. System engineering—theory and practice, 2014, 34(9): 2193–2201.

[10] 刘睿智, 周勇. 指数跟踪投资组合与多信息下指数可预测性——基于 Adaptive LASSO 和 ARIMA-ANN 方法[J]. 系统工程, 2015, 33(4): 1–7.
LIU Ruizhi, ZHOU Yong. The portfolio of index tracing and index predictability under multi-information—Based on adaptive LASSO and ARIMA-ANN method[J]. Systems engineering, 2015, 33(4): 1–7.

[11] CHEN Chen, KWON R H. Robust portfolio selection for index tracking[J]. Computers and operations research, 2012, 39(4): 829–837.

[12] GUASTAROBA G, SPERANZA M G. Kernel search: An application to the index tracking problem[J]. European journal of operational research, 2012, 217(1): 54–68.

[13] FILIPPI C, GUASTAROBA G, SPERANZA M G. A heuristic framework for the bi-objective enhanced index tracking problem[J]. Omega, 2016, 65: 122–137.

[14] DOSE C, CINCOTTI S. Clustering of financial time series with application to index and enhanced index tracking portfolio[J]. Physica A: statistical mechanics and its applications, 2005, 355(1): 145–151.

[15] NANDA S R, MAHANTY B, TIWARI M K. Clustering Indian stock market data for portfolio management[J]. Expert systems with applications, 2010, 37(12): 8793–8798.

[16] LEMIEUX V, RAHMDEL P S, WALKER R, et al. Clustering techniques and their effect on portfolio formation and risk analysis[C]//Proceedings of the International Workshop on Data Science for Macro-Modeling. Snowbird, UT, USA, 2014: 1–6.

[17] FERREIRA L N, ZHAO Liang. Time series clustering via community detection in networks[J]. Information sciences, 2016, 326: 227–242.

[18] RAGHAVAN U N, ALBERT R, KUMARA S. Near linear time algorithm to detect community structures in large-scale networks[J]. Physical review E: covering statistical, nonlinear, biological, and soft matter physics, 2007, 76(3): 036106.

[19] SAKOE H, CHIBA S. Dynamic programming algorithm optimization for spoken word recognition[J]. IEEE transactions on acoustics, speech, and signal processing, 1978, 26(1): 43–49.

[20] 李海林, 梁叶. 基于数值符号和形态特征的时间序列相似性度量方法 [J]. 控制与决策, 2017, 32(3): 451–458.
LI Hailin, LIANG Ye. Similarity measure based on numerical symbolic and shape feature for time series[J]. Control and decision, 2017, 32(3): 451–458.

[21] LI H. Accurate and efficient classification based on common principal components analysis for multivariate time series[J]. Neurocomputing, 2016, 171: 744–753.

[22] CHEN Yanping, KEOGH E, HU Bing, et al. The UCR time series classification archive[EB/OL]. 2015. (2015–07-01)[2015–12-01]. http://www.cs.ucr.edu/~eamonn/time_series_data/.

作者简介:

李海林, 男, 1982 年生, 教授, 博士, 主要研究方向为数据挖掘与决策支持。主持 2 项国家自然科学基金和 4 项省部级基金, 发表学术论文 50 余篇, 其中被 SCI 检索 11 篇, 被 EI 检索 20 余篇。

梁叶, 女, 1992 年生, 硕士研究生, 主要研究方向为金融时间序列数据挖掘。发表学术论文 5 篇。