

文章编号:1671-6833(2019)04-0080-06

基于随机森林 MOPSO 的城市最优资本结构分析

李燕燕¹, 杨昊天², 曾珣璠³

(1.郑州大学 商学院,河南 郑州 450001; 2.郑州大学 产业技术研究院 电气工程学院,河南 郑州 450001; 3.英国利物浦大学 数学科学系,利物浦)

摘 要:城市资本结构是一个受到多因素交互影响的复杂问题.试图基于随机森林多目标粒子群算法构建多目标多因素影响下的城市最优资本结构模型,对城市资本结构状况进行剖析.首先利用随机森林的拟合回归特性对历史数据进行拟合,从中找到历史数据特征之间的关系.随后采用多目标粒子群约束优化算法,根据已有的关系特征去寻找使目标同时达到最好效果的特征值,再根据这些效果最好的特征值从历史数据中寻找相关性最高的数据,从而分析出资本结构配置相对较优的城市以及年份.通过不断学习这些较优的结构配置,可以对各个城市的发展起到良好的借鉴作用.

关键词:随机森林;多目标粒子群约束优化算法;城市资本结构配置;拟合回归;相关性

中图分类号: TP393 **文献标志码:** A **doi:**10.13705/j.issn.1671-6833.2019.04.028

0 引言

城市作为国家经济增长最有力的支撑,也是各种资本最活跃的空间区域,各种经济问题在城市的平台上表现得最为直接.进入新常态,出现了动能困境、结构失衡、脱实就虚、公共服务产品短缺等一系列问题,这些问题的背后其实是资本规模大小的市场配置状况.随着城镇化进程的深入,城市的资本化程度也越来越高,所谓的城市资本化是使纳入城市的资产以货币化形式进行运作,是土地资本、金融资本、产业资本、科技资本(含人力资本)以及公共服务资本等高度耦合健康循环发展的模式.

目前理论界关于城市资本的研究多是从单个资本的视角来研究,诸如土地财政与城市发展、金融资本与经济增长、金融促进产业发展等,关于多重资本结构配置的研究相对较少.城市资本结构是一个受到多目标多因素交互影响的复杂问题,单一工具很难对此做出较为全面的解释.城市资本所涉及的土地资本、金融资本、产业资本、科技资本以及公共服务资本虽有各自的运行特征和规律,其作用机制和强度具有一定差别,但它们之间并非互相割

裂,而是互相影响、互动耦合发生作用.所以,笔者选取 30 个省级行政中心(拉萨除外)为代表,基于 2002~2016 年各城市地区生产总值、人均地区生产总值、全社会固定资产投资、土地出让金、第二产业增加值、规模以上工业企业利润总额、第三产业增加值、金融增加值、科教支出(科技支出和教育支出)、年末总人口数、城镇单位从业人员期末人数、城镇居民人均可支配收入、职工平均工资、当年实际利用外资金额等数据,透视土地资本、金融资本、产业资本、科技资本以及公共服务资本的作用特征,通过利用随机森林算法从多维角度考察城市资本结构的复杂关系,采用多目标粒子群约束优化算法构建多目标多因素影响下的城市最优资本结构模型,了解城市资本结构配置的状况以及相互作用的对冲效应,寻找最接近目标模式的年份和城市,为城市资本内部结构调整和功能提升提供参考价值.

1 相关算法

1.1 随机森林

随机森林回归(random forest regression, RFR)算法是 Breiman 等^[1]于 2001 年提出的.该算

收稿日期:2019-01-20;修订日期:2019-05-06
基金项目:国家社科基金资助项目(17BJY050)
作者简介:李燕燕(1968—),女,河南三门峡人,郑州大学教授,博导,研究方向为经济转型与金融投资,E-mail:liyiy@zzu.edu.cn.
通信作者:杨昊天(1994—),男,河南郑州人,郑州大学产业技术研究院硕士,研究生,主要研究方向为机器学习和进化计算,E-mail:yang.haotian@foxmail.com.

法将 Bagging 集成学习理论与随机子空间方法结合^[2],是一种基于决策树的集成学习算法。

随机算法不但能够处理连续的数据问题,而且也可以处理具有离散属性的数据,其运行速度快,且稳健性和抗噪性强,所以该算法被广泛应用于生物信息领域基因序列的分类和回归等^[3-4]、经济金融领域客户信用分析及反欺诈等^[5-6]以及数据挖掘领域异常检测和度量学习等^[7-8]方面。

为了能够提高模型精度并避免出现过拟合的情况,随机森林算法引入了 Bagging 和随机子空间思想^[9]。笔者通过评价模型的拟合效果和检验模型预测结果来评价模型的预测能力,通过计算 RFR 模型的决定系数(R^2)和均方根误差(RMSE)来评价模型的拟合效果^[9]。

1.2 粒子群优化算法

由于粒子群优化算法(particle swarm optimization)^[10]具有编程简单、直观易实现等优点,所以在工程上得到了广泛的应用。粒子群优化算法的思路:初始化种群规模设置随机为 N ,粒子群的维度设置为 D ,迭代次数为 k 时, $x_i = (x_{i,1}, x_{i,2}, \dots, x_{i,D})$ 作为第 i 个粒子的位置,此时这个粒子通过向它的个体极值 $pbest_i$ 和全局极值 $gbest$ 学习来更新它的位置,之后在迭代过程中不断计算每次粒子的个体极值和全局极值。如果此代极值比上一代极值好,那么就将这代极值更新,反之则保留上一代的个体极值和全局极值。接下来,判断粒子群算法是否达到设置的停止条件,若达到则将最优解输出,反之将继续迭代,更新粒子的速度和位置。经过不断地迭代更新后,最优解最终会出现。粒子群的速度更新如式(1)和位置更新如公式(2)。

$$v_{i,d}^{k+1} = wv_{i,d}^k + c_1r_1(pbest_{i,d}^k - x_{i,d}^k) + c_2r_2(gbest_d^k - x_{i,d}^k); \quad (1)$$

$$x_{i,d}^{k+1} = x_{i,d}^k + v_{i,d}^{k+1}, \quad (2)$$

式中: w 为惯性权重,取值范围为 $[0.4, 0.9]$,使其依据自身速度进行惯性运动; c_1 与 c_2 为学习因子,表示粒子受自身及全局的影响程度,调节向个体最优和全局最优方向飞行的最大移动步长; r_1 与 r_2 是 0 到 1 的随机数; $v_{i,d}^k$ 是第 k 次迭代的第 i 个粒子第 d 维的速度, $d = 1, 2, \dots, D$ 。

1.3 多目标粒子群约束算法

一般情况下,多目标优化问题的各个子目标是相互冲突的,一个子目标的改善有可能会引起另一个子目标的降低,同时使多个子目标一起达到最优值是不可能的,只能在它们中间进行协调和折中处理,使各个子目标都尽可能地达到最优

化。与标准的粒子群算法的流程大体一致^[11-12],流程参照文献[13]。

粒子群算法已经在多领域得到应用,如云计算领域的云计算消耗源调度问题^[14],电力系统领域用于电力系统的经济可靠运行^[15-16],解决电力系统经济调度等问题^[17],经济领域用于解决最大利益的股票交易决策多目标优化等问题^[18],数据挖掘领域应用于多目标数据分类规则挖掘以及对数据集进行分类规则挖掘等^[19]。

2 随机森林多目标粒子群算法

2.1 算法结构

随机森林算法是集成算法,能够对数据实现快速高度拟合。随机森林算法可以利用决定系数(R^2)和均方根误差(RMSE)得出数据特征与目标特征之间的关系。当然,此时特征与目标特征之间的关系并不能用具体的函数表达式表示,而是利用随机森林算法通过训练数据训练出来的模型来表示特征与目标特征之间的关系。通过这样的方式,就可以利用随机森林模型作为多目标粒子群算法的目标函数。因为利用多目标粒子群算法寻优需要使得多个目标同时达到最优,所以使用随机森林模型构建出各个目标特征与其他特征之间的关系作为目标函数。考虑到在实际应用中各个特征必然存在一定的约束条件,所以在使用多目标粒子群算法寻优之前,需要对各个特征添加约束条件。当粒子超过边界时,则重新随机进行初始化。用这样的结合方式,既可以利用随机森林的拟合特性将数据特征与所选目标特征进行高度拟合得出其关系,又可以根据多目标粒子群算法进行寻优,找到能够同时达到多目标的最优点。

2.2 算法应用

此次研究的目的是寻找城市最优资本结构。设定人均地区生产总值和城镇居民可支配收入同时达到最高时所对应的各个特征值为城市资本结构配置最优的条件。在进行多目标优化时,需要明确多目标的特征值与其他特征值之间的关系。基于所分析的数据逻辑关系复杂,计算量大,使用随机森林算法可以快速、准确、高效地从复杂的数据中提取出多目标特征与其他特征值之间的关系,而粒子群算法是目前传统又经典的寻优算法之一,将各个数据特征根据实际情况进行条件约束,随后利用多目标粒子群约束优化算法进行寻优。多目标粒子群约束优化算法与随机森林相结合的随机森林多目标粒子群算法,既能快速高效地得

出合理方案,又减少了算法的复杂性.利用该算法可以求得最优时各个特征的值即为城市最优资本结构,然后根据 30 个城市 2002~2016 年间数据与城市最优资本结构的相关性,得出资本结构配置相对较优的城市及年份.

3 实证检验

3.1 数据分析

表 1 中列出本次试验数据所选取的城市、城市类别、数据年份区间、数据所用到的特征以及目标特征.其中,城市类别 1 代表一线城市;类别 2 代表新一线城市;类别 3 代表二线城市;类别 4 代表三线城市;特征表示实验中选择城市的特征值;目标特征表示本次研究所选取的目标方向.

表 1 实验所选数据的结构

Tab.1 The structure of data selected for the experiment				
城市	城市类别	年份	特征	目标特征
北京	1	2002~2016	地区生产总值;全社会固定资产投资;土地出让金;第二产业增加值;规模以上工业企业利润总额;第三产业增加值;金融业增加值;科技支出;教育支出;年末总人口数;城镇单位从业人数;职工平均工资;当年实际利用外资金额	人均地区生产总值;城镇居民人均可支配收入
天津	2	2002~2016		
上海	1	2002~2016		
重庆	2	2002~2016		
石家庄	3	2002~2016		
太原	3	2002~2016		
呼和浩特	4	2002~2016		
沈阳	2	2002~2016		
长春	3	2002~2016		
哈尔滨	3	2002~2016		
南京	2	2002~2016		
杭州	2	2002~2016		
合肥	3	2002~2016		
福州	3	2002~2016		
南昌	3	2002~2016		
济南	3	2002~2016		
郑州	2	2002~2016		
武汉	2	2002~2016		
长沙	2	2002~2016		
广州	1	2002~2016		
南宁	3	2002~2016		
海口	3	2002~2016		
成都	2	2002~2016		
贵阳	3	2002~2016		
昆明	3	2002~2016		
西安	2	2002~2016		
兰州	3	2002~2016		
西宁	4	2002~2016		
银川	4	2002~2016		
乌鲁木齐	3	2002~2016		

由于需要对特征之间的关系进行研究,所以实验选用两个特征作为目标,用随机森林进行拟合求得这两个特征与其他特征之间的关系.

3.2 算法参数选择

目前面临的任务属于多特征的数据,因此为了进行复杂程度较高的拟合方式,参数需要进行一定程度的调整.随机森林中,选择随机个数为 10 个,选择迭代次数为 1 000 次,所选择算法的最大函数评价次数均设置为 240 000.粒子群优化算法的参数设置:种群规模为 50,惯性权重 w 随着迭代次数从 0.9 至 0.4 线性递减,学习因子 c_1 与 c_2 均为 1.494 45, r_1 与 r_2 每次迭代随机取 0~1 的任意数.

3.3 拟合效果评价

关于拟合效果如何,根据模型的决定系数(R^2)和均方根误差($RMSE$)来评价模型.利用随机森林模型拟合之后得出决定系数(R^2)和均方根误差($RMSE$)分别为 0.995 6 和 0.000 45.结果发现,随机森林模型可以实现高度的拟合.在对历史数据进行高度拟合之后,可以得出历史各个特征与目标特征之间的关系.

3.4 结果分析

通过拟合效果的评价,可知随机森林模型可以实现与历史数据的高度吻合,能够准确地代表历史中目标特征与各个特征之间的关系.在此基础上,利用多目标粒子群约束算法寻找多个目标特征最优化时所对应的各个特征之间的确定值.利用相关性函数式(2),求出最优的解集与历史数据的相关性系数,可以得出最优解,即设定的年份期间中最优的城市资本结构.

$$r(X,Y)=\frac{Cov(X,Y)}{\sqrt{Var[X]Var[Y]}}\quad (2)$$

式中: $Cov(X,Y)$ 为 X 与 Y 的协方差; $Var[X]$ 为 X 的方差; $Var[Y]$ 为 Y 的方差.

鉴于中国地区差异,依据第一财经新一线城市研究所对 338 个地级以上城市的排名,将城市分为一线城市、新一线城市、二线城市、三线城市.其中,笔者所选用的 30 个城市中,一线城市有北京、广州、上海;新一线城市有成都、杭州、济南、昆明、南京、沈阳、天津、武汉、西安、长沙、郑州、重庆;二线城市有福州、贵阳、哈尔滨、海口、合肥、兰州、南昌、南宁、石家庄、太原、乌鲁木齐、长春;三线城市有呼和浩特、西宁、银川^[20].

3.4.1 数据为真值的结果分析

当所有的数据使用真值进行实验时,依次

选用 2002~2016 年间全国的 30 个城市以及 30 个城市中的一线城市、新一线城市、二线城市、三线城市的数据进行实验.这里用热力图表示实验结果,颜色越深代表相关性越高,城市资本结构配置越好;颜色越浅则相关性越低,城市资本结构配置越差.

2002~2016 年间全国 30 个城市中,城市资本结构配置最好的是 2007 年西安市,相关系数为 0.994 787;其次是 2009 年的合肥市,相关系数为 0.993 822;再次为 2002 年的成都市,相关系数为 0.990 588.将各个城市 15 a 综合起来分析,全国 30 个城市中,城市资本结构配置最好的是西安市,相关系数之和为 14.812 322;其次是南宁市,相关系数之和为 14.765 423;再次为合肥市,相关系数之和为 14.761 251. 样本城市中的一线城市资本结构配置最好的是 2002 年的广州市,相关系数为 0.994 057;其次为 2002 年的上海市,相关系数为 0.993 213;再次为 2004 年的北京市,相关系数为 0.992 688.将各个城市 15 a 综合起来分析,样本城市中一线城市中城市资本结构配置最好的是广州市,相关系数之和为 14.840 293;其次为北京市,相关系数之和为 14.820 292;再次为上海市,相关系数之和为 14.799 533.

2002~2016 年间样本城市中的新一线城市中城市资本结构配置最好的是 2011 年的沈阳市,相关系数为 0.995 032;其次为 2012 年的西安市,相关系数为 0.994 559;再次为 2015 年的济南市,相关系数为 0.993 296.将各个城市在 2002~2016 年

共 15 a 综合起来分析,样本城市中新一线城市中城市资本结构配置最好的是济南市,相关系数之和为 14.771 292;其次为重庆市,相关系数之和为 14.763 693;再次为沈阳市,相关系数之和为 14.760 872.

2002~2016 年样本城市中的二线城市中城市资本结构配置最好的是 2002 年福州市,相关系数为 0.998 697;其次为 2015 年的太原市,相关系数为 0.998 690;再次为 2009 年的南昌市,相关系数为 0.998 664.将各个城市在 2002~2016 年 15 a 综合起来分析,二线城市中城市资本结构配置最好的是海口市,相关系数之和为 14.950 136;其次为南昌市,相关系数之和为 14.950 021;再次为南宁市,相关系数之和为 14.948 310.

2002~2016 年样本城市中的三线城市资本结构配置最好的是 2004 年银川市,相关系数为 0.999 594;其次为 2006 年西宁市,相关系数为 0.998 408;再次为 2006 年呼和浩特市,相关系数为 0.998 250.将各城市 15 a 综合起来分析,三线城市资本结构配置最好的是银川市,相关系数之和为 14.973 522;其次为西宁市,相关系数之和为 14.969 383;再次为呼和浩特市,相关系数之和为 14.968 272.

综上所述,通过多目标粒子群算法得出的最终结果作为最优城市资本结构配置.表 2 展示了所选数据为真值时,得出的与标准模式最接近的年份城市以及 15 a 综合分析最接近标准的城市.

表 2 数据为真值的实验结果
Tab.2 The experimental result of the data being true

实验选用城市	年份	最接近标准模式的年份城市	2002~2016 年 15 a 最接近标准模式的的城市
所有 30 个城市	2002~2016 年	2007 年的西安市	西安市
30 个城市中的一线城市	2002~2016 年	2002 年的广州市	广州市
30 个城市中的新一线城市	2002~2016 年	2011 年的沈阳市	济南市
30 个城市中的二线城市	2002~2016 年	2002 年的福州市	海口市
30 个城市中的三线城市	2002~2016 年	2004 年的银川市	银川市

3.4.2 数据为增长率的结果分析

将所有的数据使用增长率时再次进行实验,仍旧依次选用 2002~2016 年 15 a 全国 30 个城市和 30 个城市中的一线城市、新一线城市、二线城市、三线城市的数据进行实验.

2002~2016 年间样本城市中城市资本结构配置最好的是 2005 年的北京市,相关系数为 0.992 001;其次为 2011 年天津市,相关系数为

0.987 330;再次为 2002 年的济南市,相关系数为 0.982 865.将各个城市 15 a 综合起来分析,样本城市中城市资本结构配置最好的是上海市,相关系数之和为 14.553 823;其次为武汉市,相关系数之和为 14.538 332;再次为北京市,相关系数之和为 14.489 001.

2002~2016 年间一线城市资本结构配置最好的是 2016 年的上海市,相关系数为 0.675 601;其次为

2011 年的广州市,相关系数为 0.606 031;再次为 2006 年的北京市,相关系数为 0.536 633.将各个城市 15 a 综合起来分析,样本城市中一线城市资本结构配置最好的是北京市,相关系数之和为 3.205 879;其次为广州市,相关系数之和为2.792 489;再次为上海市,相关系数之和为 2.679 620.

2002~2016 年间样本城市中的新一线城市资本结构配置最好的是 2011 年的长沙市,相关系数为0.802 936;其次为 2006 年的重庆市,相关系数为 0.767 792;再次为 2004 年的西安市,相关系数为 0.749 976.将各个城市 15 a 综合起来分析,样本城市中新一线城市中城市资本结构配置最好的是长沙市,相关系数之和为 3.668 056;其次为西安市,相关系数之和为 2.916 951;再次为济南市,相关系数之和为 2.760 118.

2002~2016 年间样本城市中的二线城市资本结构配置最好的是 2003 年的长春市,相关系数为 0.785 886;其次为 2007 年的南昌市,相关系数为 0.677 504;再次为 2002 年的福州市,相关系数为

0.667 149.将各个城市 15 a 综合起来分析,样本城市中二线城市资本结构配置最好的是长春市,相关系数之和为 6.350 041;其次为福州市,相关系数之和为 4.654 140;再次为太原市,相关系数之和为 4.347 684.

2002~2016 年间样本城市中的三线城市资本结构配置最好的是 2011 年西宁市,相关系数为0.700 435;其次为 2010 年呼和浩特市,相关系数为 0.577 654;再次为 2006 年银川市,相关系数为 0.560 824.将各城市 15 a 综合起来分析,样本城市中三线城市资本结构配置最好的是西宁市,相关系数之和为 2.057 544;其次为呼和浩特市,相关系数之和为1.978 628;再次为银川市,相关系数之和为 1.302 793.

综上所述,通过多目标粒子群算法得出的最终结果作为最优城市资本结构配置.表 3 展示了当所选数据为增长率时,得出的与标准模式最接近的年份城市 and 15 a 综合分析最接近标准模式的城市.

表 3 数据为增长率的实验结果
Tab.3 The experimental result of the data being growth rate

实验选用城市	年份	最接近标准模式的年份城市	15 a 最接近标准模式的的城市
所有 30 个城市	2002~2016 年	2005 年的北京市	上海市
30 个城市中的一线城市	2002~2016 年	2016 年的上海市	北京市
30 个城市中的新一线城市	2002~2016 年	2011 年的长沙市	长沙市
30 个城市中的二线城市	2002~2016 年	2003 年的长春市	长春市
30 个城市中的三线城市	2002~2016 年	2011 年的西宁市	西宁市

4 结论

笔者将人均地区生产总值和城镇居民可支配收入同时达到最高时所对应的各个特征值为城市资本结构配置最优的条件,然后用真值数据和增长率数据分别将城市按照一线、新一线、二线、三线 4 类进行划分后,逐一分析计算出各个类型中最优资源配置的年份和城市.无论是用真值还是增长率实验的结果,除了 2016 年的上海市之外,其余得出的最优年份城市均在 2002~2011 年之间,这期间恰恰是中国经济持续快速发展的十年,2007 年中国 GDP 增速最高,达到 14.2%.用增长率实验的最优年份城市是 2005 年的北京市,用真值实验的最优年份城市是 2007 年的西安市,而西安市属于西北地区省会城市,相对发达地区发展要滞后一些,其城市资本结构在全国增速最高的年份趋于最优也是合理的.用真值实验的 15 a 最接近标准模式的城市仍是西安市,用增长率实验

的 15 a 最接近标准模式的城市则是上海市.此外,需要关注的城市是长沙市,不仅在 2011 年最接近标准模式,而且也是 15 a 最接近标准模式的城市.后续的研究需要进一步针对不同城市不同发展阶段深入分析其资本结构的变化,为城市资本结构优化配置提供更科学的参考依据.

参考文献:

[1] BREIMAN L. Random forests[J]. Machine learning, 2001, 45(1): 5-32.
[2] Ho T K. The random subspace method for constructing decision forests[J]. IEEE transactions on pattern analysis and machine intelligence, 1998, 20(8):832-844.
[3] ACHARJEE A, KLOOSTERMAN B, VISSER R G F, et al. Integration of multi-omics data for prediction of phenotypic traits using random forest[J]. BMC bioinformatics, 2016, 17(S):180.
[4] SVETNIK V, LIAW A, TONG C,et al.Random forest: a classification and regression tool for compound classi-

fication and QSAR modeling[J]. Journal of chemical information and computer sciences, 2003, 43(6): 1947-1958.

[5] PRASAD AM, IVERSON LR, LIAW A. Newer classification and regression tree techniques: bagging and random forests for ecological prediction [J]. Ecosystems, 2006, 9(2): 181-199.

[6] CUTLER DR, EDWARDSJR TC, BEARD KH, et al. Random forests for classification in ecology[J]. Ecology, 2007, 88(11): 2783-2792.

[7] XIONG CM, JOHNSON D, XU R, et al. Random forests for metric learning with implicit pairwise position dependence[C]//18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Beijing: ACM, 2012: 958-966.

[8] VERIKAS A, GELZINIS A, BACAUSKIENE M. Mining data with random forests: a survey and results of new tests[J]. Pattern recognition, 2011, 44(2): 330-349.

[9] 柴颖. 基于随机森林回归分析的径流预报模型[J]. 水利水电快报, 2018, 39(9): 36-38.

[10] LAGOS C, GUERRERO G, CABRERA E, et al. An improved particle swarm optimization algorithm for the VRP with simultaneous pickup and delivery and time windows[J]. IEEE latinamericatransactions, 2018, 16(6): 1732-1740.

[11] COELLO C A C, PULIDO G T, LECHUGA M S. Handling multiple objectives with particle swarm optimization[J]. IEEE transactions on evolutionary computation, 2004, 8(3): 256-279.

[12] COELLO C A C, LECHUGA M S. MOPSO: aproposal for multiple objective particle swarm optimization[C]//The 2002 Congress on Evolutionary Computation. Honolulu, HI: IEEE, 2002: 1051-1056.

[13] 丁知平. 一种多目标的粒子群算法的研究[J]. 科技通报, 2018, 34(7): 169-173.

[14] 贾嘉, 慕德俊. 基于粒子群优化的云计算低能耗资源调度算法[J]. 西北工业大学学报, 2018, 36(2): 339-344.

[15] AL-BETAR M A, AWADALLAH M A, KHADER A T, et al. Economic load dispatch problems with valve-point loading using natural updated harmony search[J]. Neural computing and applications, 2018, 29(10): 767-781.

[16] MEHMOOD K, AHMAD A. Improved grey wolf optimization for economic load dispatch problem considering valve point loading effect and prohibited operating zones[J]. The nucleus, 2018, 54(4): 250-257.

[17] 黄松, 王艳, 纪志成. 多目标粒子群算法的动态多燃料经济环境负荷分配[J]. 控制与决策, 2018, 33(7): 1255-1263.

[18] NENORTAITE J, SIMUTIS R. Stocks' trading system based on the particle swarm optimization algorithm[J]. Lecture Notes in Computer Science, 2004, 3039: 843-850.

[19] 段晓东, 王楠楠, 王存睿, 等. 一种基于粒子群算法的分类器设计[J]. 计算机工程, 2005, 31(20): 107-109.

[20] 中国经济网. 最新中国一二三线城市排名出炉! 快看你家排第几[EB/OL]. (2018-01-12). <https://baijiahao.baidu.com/s?id=1589345771536479546&wfr=spider&for=pc>.

Urban Optimal Capital Structure Analysis Based on Random Forest and MOPSO

LI Yanyan¹, YANG Haotian², ZENG Yufan³

(1. Business School Zhengzhou University, Zhengzhou 450001, China; 2. School of Electrical Engineering Zhengzhou University, Zhengzhou 450001, China; 3. Department of Mathematical Sciences, University of Liverpool, UK)

Abstract: Urban capital structure was a complex problem that was influenced by multiple factors. In this paper a city optimal capital structure model was constructed based on random forest and multi-objective particle swarm algorithm to analyze the state of capital structure. Firstly, historical data were analyzed by using the fitting regression characteristics of random forests to find the relationship among historical data features. According to the existing relationship characteristics, eigenvalues that could achieve the best results by using the multi-objective particle swarm optimization algorithm were identified. Then, according to these characteristic values with the best effect, the data with the highest correlation could be searched, so as to analyze the cities and years with relatively excellent capital structure allocation. By constantly learning these optimal structure allocation, it could serve as a good reference for the development of other cities.

Key words: random forest; multi-objective particle swarm optimization algorithm; urban capital structure allocation; fitting regression; correlation