

# 基于遗传算法优化神经网络的房地产评估模型及实证研究

吕 霁

(江西财经职业学院财税金融系 九江 332000)

**摘 要** 针对房地产价格和影响因素之间存在复杂且非线性的关系,用传统评估方法预测房地产价格精度不高的缺陷,提出一种基于遗传算法(Genetic Algorithm,GA)和 BP 神经网络(BP Neural Network,BPNN)的房地产评估预测模型。该模型利用 BPNN 确定房地产价格影响因子与评估价格之间的函数映射关系,并采用 GA 优化 BPNN 的权值和阈值来提高 BPNN 的收敛速度,解决了算法陷入局部极值点的问题。最后,通过对 100 组实际交易样本数据的计算机仿真实验,证明了该方法对房地产估价的有效性和准确性。

**关键词** 遗传算法,BP 神经网络,房地产估价

中图分类号 F270 文献标识码 A

## Real Estate Appraisal Model and Empirical Research Based on Genetic Algorithm to Optimize Neural Network

LV Ji

(Finance and Taxation Department, Jiangxi Vocational College of Finance and Economics, Jiujiang 332000, China)

**Abstract** Prediction accuracy of traditional assessment methods is not enough, because of the complicated nonlinear relationship between the real estate prices and influencing factors. This paper proposed an real estate appraisal predictive model based on genetic algorithms and BP neural network. It uses BPNN to determine the function mapping relationship impact factor and evaluated price of the real estate prices, which uses GA to optimize weights and thresholds of BPNN. The method improves the convergence rate of BPNN and solves the problem of getting into local extreme point. Finally, the computer simulation results of 100 sample data show that, the proposed method is valid and accurate for real estate appraisal.

**Keywords** Genetic algorithm, BP neural network, Real estate appraisal

## 1 引言

房地产市场的日趋活跃及交易额的日益增涨使房地产评估行业得以迅速发展。但当前房地产评估的理论研究和产业发展都尚未成熟。评估方法缺少准确、科学的依据而过于依赖评估人员的经验,这难以保证评估质量,影响了房地产评估行业的发展<sup>[1]</sup>。针对房地产价格与其影响因素之间复杂的非线性关系,本文采用神经网络来确定房地产价格和影响因素之间的映射关系,旨在有效降低由于评估人员主观因素所造成的评估结果的偏差。而针对 BP 算法存在收敛速度慢、易陷入局部极值点的问题,采用遗传算法来优化 BP 神经网络的权值和阈值,从而构建基于遗传算法和神经网络(GA-BPNN)的房地产评估模型。模型在实证中利用南昌市 2013 年房地产实际交易的 100 组数据,运用对模型进行编程,对样本中的 90 组数据进行训练,并对 10 组数据进行预测,将预测结果与 BP 神经网络和市场比较法进行了对比分析。

响因素之间存在高度的复杂性和非线性。房地产价格一般是指市场价格,房地产市场价格是在其交易时间内由众多因素综合作用的结果。其主要分为 3 大因素:一般因素、区域因素和个别因素。其中,一般因素在总体上影响某地区、某类房地产的价格水平,而个别因素和区域因素仅直接影响到某一具体房地产的价格。区域因素是指房地产所在地区的自然条件与社会、经济、行政等因素相结合所形成的地区特性对该区房地产价格水平的影响作用。个别因素是指构成房地产本身使用功能、质量优劣的因素<sup>[2]</sup>。具体评估指标体系如表 1 所列。

表 1 房地产价格评估指标体系

|                   |                   |                                                     |
|-------------------|-------------------|-----------------------------------------------------|
| 房地产<br>估价指<br>标体系 | 影响房地产价格<br>一般因素指标 | 供求状况、经济发展水平、物价水平、居民收入、人口因素、社会治安程度、城市化水平、政府对房地产市场的政策 |
|                   | 影响房地产价格<br>区域因素指标 | 地段繁华程度、交通条件、配套设施、区域环境                               |
|                   | 影响房地产价格<br>个别因素指标 | 建筑结构、工程质量、户型、楼层、朝向、装修情况、新旧程度、房产权、建筑面积、租金水平          |

## 2 房地产价格评估指标体系和方法

房地产价格是一个复杂的经济范畴,房地产价格与其影

房地产评估中常用的方法有市场比较法、收益法、成本法,本文采用最常用的市场比较法。市场比较法是根据经济

本文受江西省自然科学基金青年项目(20132BAB211038)资助。

吕 霁(1984—),女,博士生,讲师,主要研究方向为资产评估、工程概预算,E-mail:330529848@qq.com。

学中的替换原理,选择近期发生交易的实例,经过差别修正后,测算被评估房屋价格的方法。评估人员通过市场调查,选择几个同一供需圈、房屋用途一致、类型相似的交易案例作为参照物,然后经过评估人员实地查勘,对交易情况、交易日期、区域因素和个别因素等4项因素分别进行比较和修正,得出被评估对象最可能实现的合理价格<sup>[3]</sup>。

### 3 遗传算法优化神经网络的理论基础

#### 3.1 BP神经网络原理

BP神经网络是一种按误差逆传播算法训练的多层前馈网络,网络可分为输入层、隐含层和输出层,其中输入和输出都只有一层,隐藏层可有一层或多层,各层次之间结点全部互相连接,同层的网络结点之间没有连接。BP网络能学习和存储大量的输入-输出模式映射关系,而无需事前揭示描述这种映射关系的数学方程。每个结点表示一个神经元,传递函数通常采用 Sigmoid 函数,BP网络的训练其实是对网络输出结果与期望结果之间差值求最小值,BP网络的训练过程是采用梯度下降法来调整各个神经元之间的权值和阈值,直到误差函数值  $E$  达到最小<sup>[4]</sup>。如式(1)、式(2)所示:

$$\min E = \sum_{i=1}^n E_i \quad (1)$$

$$E_p = \frac{1}{2} \sum_{j=1}^k (O_j - P_j)^2 \quad (2)$$

其中, $E_p$  是训练样例的误差; $p$  为输出层的单元个数; $O_j$  为单元  $j$  对训练样例  $P$  的目标值; $P_j$  是给定训练样例  $p$  时单元  $k$  的输出值。但是 BP 算法的训练速度较慢且有可能陷入局部极小值。BP 神经网络的三层结构如图 1 所示。

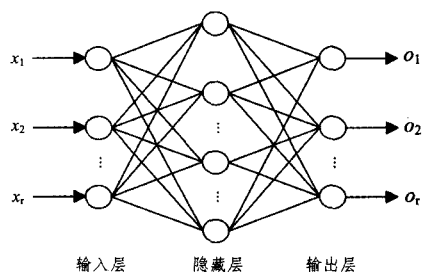


图1 三层BP神经网络的结构

#### 3.2 遗传算法优化原理

传统的 BP 网络训练需较长时间,学习性能不佳,采用遗传算法优化 BP 网络可改进神经网络的性能。遗传算法是借鉴生物界的适者生存,优胜劣汰遗传机制进化规律演化的随机化搜索方法。其主要特点是直接对结构对象进行操作,不求导和限定函数的连续性;具有内在的隐并行性和出色的全局寻优能力。将遗传算法应用于神经网络可以指导权重优化和拓扑选择<sup>[5]</sup>。遗传算法优化 BP 神经网络初始权值步骤如下:

①初始化种群,一个染色体对应一组网络权值,设定染色体种群数量为  $N$ ,遗传代数数为  $gen$ ,交叉概率为  $P_c$ ,突变概率为  $P_m$ 。

②如果达到最大遗传代数  $gen$ ,转至步骤⑧。

③计算每个染色体的适应性,染色体  $i$  的适应值  $f_i$  等于对应的网络误差  $E_i$  的倒数。

④采用轮盘赌方法选择群体中的一对染色体,以交叉概率  $P_c$  进行交叉操作,以  $P_m$  的概率对交叉后的染色体进行突变操作。

⑤将后代染色体放入新种群中。

⑥重复步骤④和⑤,直到新染色体种群的大小为  $N$  为止。

⑦用新染色体种群取代双亲染色体种群,转至步骤②。

⑧将适应性最高的染色体解码得到相应的网络初始权值。

在遗传算法全局寻优的基础上运用 BP 算法进行局部细致搜索,当达到要求的收敛精度时结束网络训练。

#### 3.3 遗传神经网络在房地产评估中应用的可行性分析

各种影响因素对房地产价格的影响方向、影响程度各有不同,它们之间是一种无法用具体的数学模型来表达的十分复杂的非线性关系。因此如何确定主要影响因素以及因素评估权重就成为房地产评估的一个难题。

神经网络具有极强的大规模并行处理能力和自组织、自适应、自学习的能力,这些特点使神经网络特别适用于处理需要同时考虑多个因素且条件模糊的信息处理问题。神经网络以影响房地产价格的多种因素量化值作为网络的输入数据  $x = (x_1, x_2, \dots, x_n)$ ,对应的房地产价格作为输出数据  $y$ ,利用足够的交易案例即训练样本对神经网络模型进行训练<sup>[6]</sup>。一旦网络训练成功,通过将所获取的数据输入到已经训练完成的网络,就可得到相应的房地产评估价格,其可以作为实际的房地产交易价格的参考。因此神经网络有能力发现房地产价格与其影响因素之间的客观规律,并且能够构建一个数学模型来表示这种映射关系。而遗传算法优化神经网络是一种有效的神经网络优化方法,能够提高神经网络的收敛速度和预测精度。

### 4 实证研究

#### 4.1 样本数据的选择

根据笔者在房地产估价工作中搜集的南昌市 2013 年度各类型房地产交易的详细资料(见表 2),选取其中 90 组完整样本数据作为遗传神经网络训练样本,10 组完整数据作为测试检验样本。房地产价格影响因素的评估指标体系应具有概括性和动态性,能够全面、客观地反映影响房地产价格的主要因素以及导致个体房地产价格差异的主要因素。由于 100 组住宅交易数据出自同一城市 and 同一年度,表 1 所述的房地产评估指标体系中的一般因素对 100 套住宅交易的价格影响大体相当,因此其无需列入评估指标。本文在全面参考房地产评估研究成果和神经网络模型的实际性能的基础之上,选定了一种房地产价格影响因素的评估指标体系,在区域因素和个别因素中选取 14 个指标体系成员:地段、交通条件、配套设施、区域环境、建筑结构、工程质量、户型、楼层、朝向、装修情况、新旧程度、房产产权、建筑面积、租金水平。

表2 2013年度南昌市100套普通住宅交易实例

| 实例编号              | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | ..... | 97   | 98   | 99   | 100  |
|-------------------|------|------|------|------|------|------|------|------|------|-------|------|------|------|------|
| 实际成交价格<br>(元/平方米) | 6680 | 6910 | 6665 | 6545 | 8480 | 8100 | 9030 | 4590 | 7650 | ..... | 7670 | 7755 | 5140 | 7645 |
| 地段                | 0.2  | 0.3  | 0.3  | 0.3  | 0.9  | 0.7  | 0.9  | 0.7  | 0.7  | ..... | 0.7  | 0.7  | 0.8  | 0.7  |
| 交通条件              | 0.2  | 0.2  | 0.2  | 0.2  | 0.5  | 0.5  | 0.5  | 0.2  | 0.2  | ..... | 0.2  | 0.2  | 0.2  | 0.2  |
| 配套设施              | 0.7  | 0.5  | 0.5  | 0.5  | 0.7  | 0.7  | 0.7  | 0.6  | 0.7  | ..... | 0.7  | 0.8  | 0.7  | 0.7  |
| 周边环境              | 0.5  | 0.5  | 0.5  | 0.5  | 0.9  | 0.9  | 0.9  | 0.7  | 0.7  | ..... | 0.9  | 0.7  | 0.9  | 0.7  |
| 建筑结构              | 0.3  | 0.9  | 0.9  | 0.9  | 0.9  | 0.9  | 0.9  | 0.7  | 0.9  | ..... | 0.9  | 0.9  | 0.9  | 0.9  |
| 工程质量              | 0.7  | 0.3  | 0.3  | 0.2  | 0.9  | 0.5  | 0.9  | 0.9  | 0.3  | ..... | 0.9  | 0.3  | 0.3  | 0.3  |
| 户型                | 0.9  | 0.5  | 0.3  | 0.3  | 0.7  | 0.9  | 0.7  | 0.7  | 0.7  | ..... | 0.7  | 0.9  | 0.9  | 0.7  |
| 楼层                | 0.5  | 0.3  | 0.9  | 0.3  | 0.9  | 0.7  | 0.9  | 0.9  | 0.3  | ..... | 0.9  | 0.7  | 0.5  | 0.5  |
| 朝向                | 0.5  | 0.3  | 0.9  | 0.3  | 0.9  | 0.7  | 0.9  | 0.9  | 0.3  | ..... | 0.9  | 0.7  | 0.5  | 0.5  |
| 新旧程度              | 0.7  | 0.7  | 0.3  | 0.3  | 0.7  | 0.9  | 0.7  | 0.7  | 0.7  | ..... | 0.6  | 0.7  | 0.9  | 0.9  |
| 房产产权              | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 0.5  | 1    | ..... | 1    | 1    | 0.5  | 1    |
| 装修情况              | 0.5  | 0.3  | 0.9  | 0.3  | 0.9  | 0.7  | 0.9  | 0.9  | 0.3  | ..... | 0.9  | 0.7  | 0.5  | 0.5  |
| 建筑面积              | 0.9  | 0.5  | 0.3  | 0.3  | 0.7  | 0.9  | 0.7  | 0.7  | 0.7  | ..... | 0.7  | 0.9  | 0.9  | 0.7  |
| 租金水平              | 0.5  | 0.5  | 0.5  | 0.5  | 0.9  | 0.9  | 0.9  | 0.7  | 0.7  | ..... | 0.9  | 0.7  | 0.9  | 0.7  |

## 4.2 样本数据的量化处理

在上述房地产价格影响因素评估指标体系中,部分指标是定性指标(如地段、环境、楼层),部分指标是定量指标(如租金水平)。为使系统中各指标具有可比性,应将定性指标进行量化处理,如果定性指标在一些通用的建筑物评分规则或国家标准中已有可行的量化方法,则采用已有的量化方法。如果定性指标还没有可行的量化方法,则采用五级打分制来区分其优劣。具体如表3所列。

表3 等级评分表

| 评定等级 | 优     | 良    | 中    | 差    | 极差   |
|------|-------|------|------|------|------|
| 评分标准 | 9~10分 | 7~8分 | 5~6分 | 3~4分 | 1~2分 |

由于各影响因素指标的表现形式不同,原始数据之间存在级差问题,个别输入分量差距悬殊,不能体现各分量的同等地位。且输入过大时,网络容易进入S型函数的包河区,导致网络无法收敛。因此在网络计算之前需要对样本数据进行标准化处理,以有利于提高网络训练效率。结合样本数据的特点及量化标准与房地产价格成正比的特性,本文采用了归一化的标准化方法。

$$P_i = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \quad i=1,2,\dots,n \quad (3)$$

式中, $P_i$ 表示目标值为 $X_i$ 的效用系数,表示第 $i$ 个评价指标值; $X_{\max}$ 、 $X_{\min}$ 分别表示样本数据中的最大值和最小值。

## 4.3 BP神经网络模型构建

### 1) 确定BP网络的层数

根据通用逼近定理等神经网络理论可知,具有偏差和至少一个S型隐含层加上一个线性输出层的网络能够逼近任何有理函数。增加隐含层层数在降低误差、提高精度的同时会使网络更加复杂,从而增加网络的训练时间。因此,本文中的BP神经网络确定选取3层网络模型,即输入层——隐含层——输出层结构,各层结点采用全相连方式。BP神经网络结构图如图2所示。

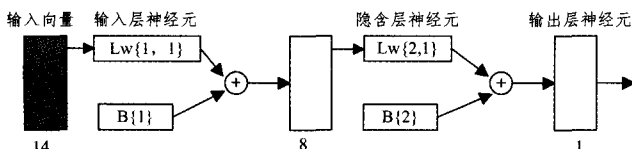


图2 BP神经网络结构图

### 2) 输入层与输出层神经元数目的确定

输入层节点的多少与评价指标个数相对应。本文采用14个指标对房地产价格进行评价,因此网络输入层神经元数目为14。对于输出层,只有1个神经元,即被评估对象的价格。

### 3) 隐含层的确定

确定隐含层神经元的数目是网络构建过程中非常关键但又很难解决的问题,目前精确预测隐含层神经元的数目仍没有广泛使用的明确理论。通过一些学者对此给出的经验性的公式,进行误差实验,本文确定隐含层节点数为8个。

## 4.4 遗传算法优化BP网络

BP神经网络隐含层的传递函数采用正切Sigmoid函数tansig(),输出层的传递函数采用线性函数purelin(),而网络训练函数采用traingdm()。设定遗传算法优化神经网络初始权值的种群数量 $N$ 为40,遗传代数 $gen$ 为900,交叉概率 $P_c$ 的值取0.9,突变概率 $P_m$ 的值取0.1。利用MATLAB R2011b对样本数据进行仿真,在GA进行了600代搜索后染色体的平均适应度趋于稳定。遗传算法优化BP神经网络的步骤如图3所示。

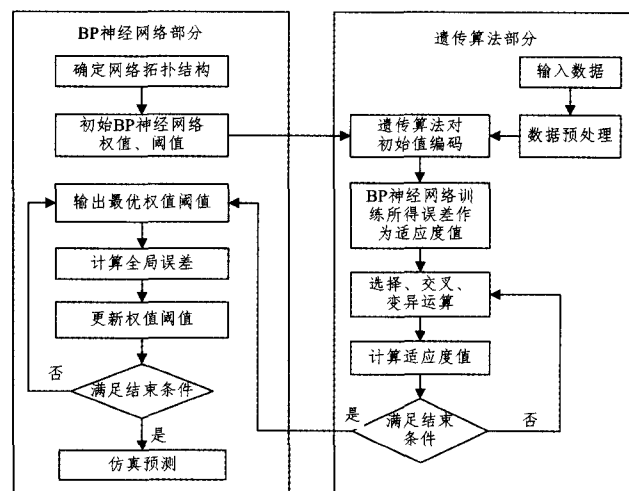


图3 遗传算法优化神经网络的步骤

## 4.5 模型训练与结果分析

将遗传算法的优化结果设置成BP网络的初始权值和阈值,训练BP网络,设定训练的误差精度为0.001,学习率为0.05。在100组样本中选取90组作为训练样本,选取10组作为预测样本。同时,为了比较遗传算法优化前后的区别,本文

(下转第87页)

[8] Zhang M, Hou Y, Hu Z. Accurate Object Tracking Based on Homography Matrix[C]//Computer Science & Service System (CSSS). 2012 IEEE International Conference, 2012:2310-2312

[9] McIntosh J H, Mutch K M. Matching straight lines[J]. CVGIP, 1988(43):386-408

[10] Schmid C, Zisserman A. The geometry and matching of lines and curves over multiple views[J]. International Journal of Computer Vision, 2000, 40(3):199-233

[11] Yang L, Wang J, Chao S, et al. A study on the correlation coefficient matching algorithm of general digital stereo under the approximate epipolar line constraints[C]//Multimedia Technology (ICMT). 2011 IEEE International Conference, 2011:954-957

[12] Liu Z, An J, Jing Y. A simple and robust feature point matching algorithm based on restricted spatial order constraints for aerial

image registration[J]. Geoscience and Remote Sensing, IEEE Transactions, 2012, 50(2):514-527

[13] Yang Y, Wang H, Liu B. A new stereo matching algorithm based on adaptive window[C]//Systems and Informatics (ICSAI). 2012 International IEEE Conference, 2012:1815-1819

[14] Smith S, Brady M. A new approach to low level image processing[J]. International Journal of Computer Vision, 1997, 23(1):45-78

[15] Zhang L. Automatic Digital Surface Model (DSM) Generation from Linear Array Images [D]. ETH Zurich, Switzerland; Institute of Geodesy and Photogrammetry, 2005

[16] Bay H, Ess A, Tuytelaars T, et al. Speeded-Up Robust Features (SURF)[J]. Computer Vision and Image Understanding, 2008, 110(3):346-359

(上接第 77 页)

也采用了相同参数对标准 BP 神经网络进行训练和预测。训练结束后,利用训练好的 3 层网络,分别输入测试的 10 组样本数据。计算结果分别如图 4 和图 5 所示。

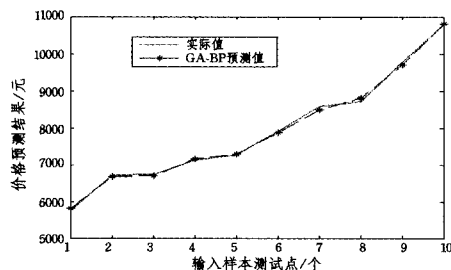


图 4 GA-BP 网络预测输出

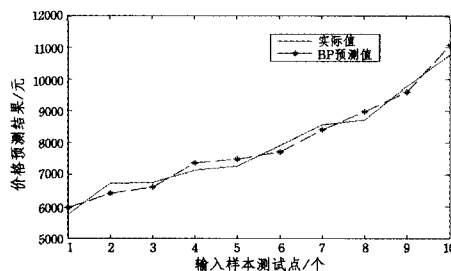


图 5 BP 网络预测输出

为了进一步说明 GA-BP 模型在房地产估价中的有效性,本文采用 BP 神经网络模型、市场比较法和 GA-BP 模型对房地产的估价进行了比较分析。3 种方法的预测结果如表 4 所列。

表 4 3 种方法预测结果相对误差表

| 样本<br>编号 | 实际值   | GA-BP<br>预测值 | 相对<br>误差<br>(%) | BP 神经<br>网络<br>预测值 | 相对<br>误差<br>(%) | 市场比较<br>法预测值 | 相对<br>误差<br>(%) |
|----------|-------|--------------|-----------------|--------------------|-----------------|--------------|-----------------|
| 1        | 5755  | 5811         | 0.97            | 5957               | 3.51            | 5987         | 4.03            |
| 2        | 6723  | 6675         | 0.71            | 6403               | 4.76            | 6418         | 4.54            |
| 3        | 6740  | 6711         | 0.43            | 6608               | 1.96            | 6638         | 1.51            |
| 4        | 7130  | 7164         | 0.48            | 7354               | 3.14            | 7304         | 2.72            |
| 5        | 7270  | 7298         | 0.39            | 7491               | 3.04            | 7491         | 3.04            |
| 6        | 7910  | 7875         | 0.44            | 7718               | 2.43            | 7709         | 2.54            |
| 7        | 8580  | 8498         | 0.96            | 8408               | 2               | 8389         | 2.23            |
| 8        | 8723  | 8809         | 0.99            | 8987               | 3.03            | 9056         | 3.82            |
| 9        | 9788  | 9693         | 0.97            | 9602               | 1.9             | 9543         | 2.5             |
| 10       | 10790 | 10822        | 0.3             | 11089              | 2.77            | 9989         | 7.42            |
| 平均相对误差   |       | 0.66         |                 | 2.85               |                 | 3.44         |                 |

从表 4 中可以看出,与 BP 神经网络和市场比较法相比,GA-BP 网络的预测平均相对误差是 0.66%,标准 BP 网络的预测平均相对误差是 2.85%,而传统市场比较法的平均相对误差是 3.44%。这说明运用 GA-BP 模型进行房地产价格的估价具有更好的预测精度。

**结束语** 考虑到房地产估价的复杂性,本文提出了一套房地产价格评价指标体系,将 BP 神经网络与遗传算法结合的混合启发式算法应用于房地产估价。采用遗传算法对 BP 神经网络的初始连接权值进行训练寻优,再用 BP 神经网络进行训练,从而获得较高精度的解。仿真计算表明,基于遗传算法优化神经网络的房地产评估模型训练时间短,收敛速度快,在学习过程中不易陷入局部极小点,具有良好的预测精度和应用前景。

## 参 考 文 献

[1] 赵财福,赵小红. 房地产估价[M]. 上海:同济大学出版社,2004

[2] 朱骏,张红. 基于时间序列与人工神经网络的房地产周期识别[J]. 清华大学学报:自然科学版,2006,46(6):781-784

[3] 杨励雅,邵春福. 基于 BP 神经网络与马尔可夫链的城市轨道交通周边房地产价格的组合预测方法[J]. 吉林大学学报:工学版,2008,38(3):514-519

[4] Liu X, Deng Z, Wang T. Real estate appraisal system based on GIS and BP neural network[J]. Transactions of Nonferrous Metals Society of China, 2011, 21:626-630

[5] 郭欣欣. 人工神经网络在住宅类房地产评估中的应用[D]. 北京:首都经济贸易大学,2012

[6] 李标. 基于神经网络的住宅房地产价格时间序列预测模型研究[D]. 浙江工业大学,2012

[7] 龙会典,严广乐. 基于实码加速遗传算法改进 TOPSIS 法在房地产预警体系中的应用[J]. 电子测试,2013(14):62-63,55

[8] 严蓓蓓,杨星光. 基于改进遗传算法的房地产投资项目风险评价模型[J]. 价值工程,2013(3):162-163

[9] 欧伟. 基于改进 PSO-BP 的房地产投资风险评价研究[D]. 邯郸:河北工程大学,2013

[10] 修雪嵩,徐爱农. 国外房地产评估方法总结与借鉴[J]. 中国资产评估,2013(10):41-46