

改进 TOPSIS 法用于出租房源综合评价研究

邓璐娟¹ 陈欣欣²

¹(郑州轻工业学院软件学院 河南 郑州 450002)

²(郑州轻工业学院计算机与通信工程学院 河南 郑州 450002)

摘要 综合目标地点和出租房源的面积、租金及间隔目标地点的距离等属性因素对出租房源进行综合排序研究。这对侧重点不同的租房者进行出租房源选择时具有较高的实用价值。通过对传统 TOPSIS 法缺陷的研究,提出一个基于最差理想解和马氏距离的改进 TOPSIS 法,用于对出租房源进行综合评价。该改进方法通过相关系数改进马氏距离,利用改进后的马氏距离替代欧式距离,解决传统 TOPSIS 法因出租房源的面积、租金等属性间相关性导致欧式距离失效的问题;根据传统 TOPSIS 法的正负理想解提取一个最差理想解,并用其代替一般 TOPSIS 法的负理想解,解决出租房源间隔正负理想解等同距离时无法准确确定位置的问题;通过租房者偏好系数将马氏距离和最差理想解所得距离尺度合成新的相对贴近度,对出租房源进行综合评价。实验证明利用该改进后的 TOPSIS 法排序的出租房源的推荐结果更加合理。

关键词 出租房源 TOPSIS 马氏距离 最差理想解

中图分类号 TP302

文献标识码 A

DOI:10.3969/j.issn.1000-386x.2019.02.013

COMPREHENSIVE EVALUATION OF RENTAL HOUSING SOURCE BASED ON IMPROVED TOPSIS

Deng Lujuan¹ Chen Xinxin²

¹(Software Engineering College, Zhengzhou University of Light Industry, Zhengzhou 450002, Henan, China)

²(School of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, Henan, China)

Abstract The comprehensive ranking of rental housing sources was studied by combining the attribute factors of target location and the area of rental housing, rent and distance between target locations. It is of great practical value to select rental housing sources for different renters. In this paper, an improved TOPSIS based on the worst ideal solution and Markov distance was proposed by studying the defects of the traditional TOPSIS, which could be used for comprehensive evaluation of rental housing resources. The method improved the Markov distance by correlation coefficient, and then replaced the Euclidean distance with the improved Markov distance. It solved the problem of Euclidean distance invalidation caused by the correlation between the area and rent of the rental housing resource in the traditional TOPSIS. A worst ideal solution was extracted from the positive and negative ideal solution of traditional TOPSIS. It replaced the negative ideal solution of the general TOPSIS, and solved the problem that the positive and negative ideal solution of the rental housing source spacing cannot accurately determine the location when the distance was equal. Through the renter preference coefficient, the distance scale obtained from the Markov distance and the worst ideal solution was combined into a new relative closeness degree, and the rental housing resources were evaluated comprehensively. Experiments show that the recommended results of the improved TOPSIS are more reasonable.

Keywords Rental housing resources TOPSIS Markov distance Worst ideal solution

0 引言

网上出租房源信息冗乱繁杂,很难进行比较和评价,研究一种改进的 TOPSIS 法对网络爬虫抓取的出租房源进行综合评价,评价结果以地图图标的形式推送给对出租房源属性侧重点需求不同的租房者,为亟待租房的对象提供了有价值的服务。

逼近于理想解排序 TOPSIS 法是按照各评估对象到理想化目标的接近程度对有限评估对象进行相对优劣选择的方法,常被用于多属性多目标排序的场景,其主要优势是简单、计算量小、几何意义直观、便于理解和运用等^[1-4]。然而,因传统 TOPSIS 方法中欧式距离本身的局限性,导致运用其进行综合评价时可能会出现两点不足:①当评估对象的属性线性相关时,欧式距离失效;②当评估对象距离正负理想解等同远近时,不能完全反映其对象的位置关系^[5]。近年来,纵观国内外对于传统 TOPSIS 方法的改进,一些专家专注于传统 TOPSIS 法的决策环境的拓展研究,一部分注重欧式距离权重确定方面的完善,还有一部分学者仅仅对传统 TOPSIS 方法的缺陷做了相应改进^[6],均没有同时解决利用传统 TOPSIS 法排序的两个缺陷。本文对出租房源评价应用中,其评估对象的面积、租金属性之间显然存在相关性及位置不确定性,一般 TOPSIS 法无法合理地对各个出租房源的优劣性进行判定。故而研究一种既能解决出租房源属性(面积、租金、间隔目标点距离)间线性相关问题,又能解决出租房源间隔正负理想解等同距离时无法准确定位的问题的改进 TOPSIS 法对出租房源进行合理的排序势在必行。

1 TOPSIS 方法

TOPSIS 法是通过计算待评估对象的各个属性指标与正反理想解的距离尺度,求得与理想解的相应贴近值,进而对评估对象进行评价的一种综合排序法^[7-8]。用 TOPSIS 法对出租房源评价的数学模型如下:

(1) 构建出租房源信息矩阵:

$$A = \begin{bmatrix} q_{11} & \cdots & q_{1j} & \cdots & q_{1n} \\ \vdots & & \vdots & & \vdots \\ q_{i1} & \cdots & q_{ij} & \cdots & q_{in} \\ \vdots & & \vdots & & \vdots \\ q_{m1} & \cdots & q_{mj} & \cdots & q_{mn} \end{bmatrix}$$

如矩阵 A 所示,表示出租房源的第 j 个属性指标值,本文具体指出租房源的面积、租金和间隔目标点的距离。

(2) 数据标准化处理:

$$p_{ij} = \begin{cases} \frac{\min p_{ij}}{p_{ij}} & p_{ij} \neq 0 \\ 1 & p_{ij} = 0 \end{cases} \quad i = 1, 2, \dots, m; j = 1, 2, \dots, n \quad (1)$$

$$p'_{ij} = \frac{P_{ij}}{\sqrt{\sum_{i=1}^m p'_{ij}{}^2}} \quad i = 1, 2, \dots, m; j = 1, 2, \dots, n \quad (2)$$

式(1)将成本型属性指标(出租房源的面积属性)归一化为取值范围在 $[0, 1]$ 之间效益型指标数据;式(2)对出租房源信息矩阵中所有属性数据进行标准化处理,如 A' 所示:

$$A' = \begin{bmatrix} p'_{11} & \cdots & p'_{1j} & \cdots & p'_{1n} \\ \vdots & & \vdots & & \vdots \\ p'_{i1} & \cdots & p'_{ij} & \cdots & p'_{in} \\ \vdots & & \vdots & & \vdots \\ p'_{m1} & \cdots & p'_{mj} & \cdots & p'_{mn} \end{bmatrix}$$

p'_{ij} 代表归一化后各出租房源 A_j 的 F_j 属性指标值, $\max_i p_{ij}$ 表示信息矩阵的一列 m 个出租房源针对第 j 个属性的最大指标值, $\min_i p_{ij}$ 表示 m 个出租房源信息中对第 j 个属性中的最小指标值。

(3) 特征矩阵权重规范化:

对出租房源进行综合评价时,鉴于不同用户对属性指标偏重不同,如何给各个属性指标赋予权重成为了利用 TOPSIS 法排序的关键问题之一。权重赋值方法很多,本文通过比较其优劣性,选择信息熵对各个属性进行权重赋值,信息熵越小,表示其透露出来的信息量较大;反之,表示其透露出来的信息量较少^[9]。对特征矩阵熵权重规范化的基本步骤如下:

(a) 计算出租房源各属性指标的熵值,其公式如下:

$$p_{ij} = \frac{p'_{ij}}{\sum_{i=1}^m p'_{ij}} \quad (3)$$

$$E_j = -\ln(m)^{-1} \sum_{i=1}^m p_{ij} \ln p_{ij} \quad (4)$$

$$w_j = \frac{E_j}{n - \sum_{j=1}^m E_j} \quad (5)$$

(b) 对各出租房源属性指标熵权重化,其公式如下所示,其中 $i = 1, 2, \dots, n$ 。权重规范化后的特征矩阵如 A'' 所示。

$$v_{ij} = w_j \times P_{ij} \quad (6)$$

$$A'' = \begin{bmatrix} v_{11} & \cdots & v_{1j} & \cdots & v_{1n} \\ \vdots & & \vdots & & \vdots \\ v_{i1} & \cdots & v_{ij} & \cdots & v_{in} \\ \vdots & & \vdots & & \vdots \\ v_{m1} & \cdots & v_{mj} & \cdots & v_{mn} \end{bmatrix}$$

(4) 确定正理想解和负理想解:

正理想解代表不同出租房源的同一属性指标中的最大值,而负理想解则代表不同出租房源中同一属性指标中的最小值。其公式如下:

$$v_j^+ = \max_i v_{ij} \quad i = 1, 2, \dots, m; j = 1, 2, \dots, n \quad (7)$$

$$v_j^- = \min_i v_{ij} \quad i = 1, 2, \dots, m; j = 1, 2, \dots, n \quad (8)$$

式中: v_j^+ 表示出租房源的第 j 个属性指标的正理想解, v_j^- 表示第 j 属性指标的负理想解。

(5) 计算距离尺度:

计算各出租房源属性指标值间隔正反理想解的距离尺度:

$$S_i^+ = \sqrt{\sum_{j=1}^n (v_{ij} - v_j^+)^2} \quad (9)$$

$$S_i^- = \sqrt{\sum_{j=1}^n (v_{ij} - v_j^-)^2} \quad (10)$$

式中: S_i^+ 表示第 i 个出租房源靠近正理想解的距离, S_i^- 表示第 i 个出租房源靠近反理想解的距离, 当 S_i^+ 越大, S_i^- 越小时, 则该候选出租房源综合性能越优。

(6) 计算与正反理想解的贴近值:

计算各出租房源与理想解的相应贴近程度:

$$C_i^* = \frac{S_i^-}{S_i^+ + S_i^-} \quad i = 1, 2, \dots, m \quad (11)$$

由公式可知, C_i^* 的取值范围为 $[0, 1]$, 当 $C_i^* = 0$ 时, 表示该出租房源最大地偏离了正理想解所对应的属性指标值, 该出租房源为最劣房源选择; 当 $C_i^* = 1$ 时, 表示该出租房源的面积、距离、租金等属性达到了正理想解所对应的属性指标值, 该出租房源为租房对象的最优选择。

(7) 对出租房源方案进行排序:

利用 TOPSIS 法对出租房源进行综合排序, 其最终评价结果由 C_i^* 的大小决定, C_i^* 越大, 其出租房源的排序越靠前; 反之, 该出租房源的排序靠后。

2 TOPSIS 方法的改进

2.1 马氏距离改进 TOPSIS 法

马氏距离具有不受坐标之间量纲的影响, 独立于测量尺度, 能够有效排除变量之间相关性的干扰优点^[9]。但它的协方差特性仅代表评估对象对各自平均值偏离的综合程度, 无法精确代表两属性之间的关联程度, 将它应用于本文对出租房源的综合评价中, 可能会导致计算结果不稳定的现象发生^[10-12]。通过研究

相关系数 $q_{xy} = \frac{\text{cov}(X, Y)}{\sqrt{DX} \sqrt{DY}}$ 与协方差 $\text{cov}(X, Y) = E$

$(X - EX) \times (Y - EY)$ 公式, 发现协方差 $\text{cov}(X, Y)$ 是依

赖于量纲的量, 而相关系数 q_{xy} 与它相反, 是不依赖量纲的量, 且 $|q_{xy}|$ 与 $\text{cov}(X, Y)$ 成正比, $|q_{xy}|$ 和 $\text{cov}(X, Y)$ 越大, X 与 Y 的相关程度均越高。因此, 利用相关系数矩阵代替协方差矩阵改进马氏距离, 可以有效解决协方差矩阵计算结果不稳定、易放大属性相关性、无法精确代表两属性间关联度的问题。利用改进后的马氏距离优化 TOPSIS 方法如下:

(1) 构建原始信息矩阵 A , 因马氏距离不受属性间相关性和量纲的影响, 故不需要对 A 进行规范化处理, 省略掉了传统方法模型中的第 2 步。马氏距离的权重是内化于公式之内的客观赋权, 故省掉了传统方法模型中的第 3 步。

(2) 确定正负理想解。假定 $B_i = (b_{i1}, b_{i2}, \dots, b_{in})$ 是第 i 个出租房源所对应属性的空间坐标, $S^+ = \{s^{+1}, s^{+2}, \dots, s^{+n}\}$, 表示负理想解所对应的空间坐标, $S^- = \{s^{-1}, s^{-2}, \dots, s^{-n}\}$ 对于本文爬取的郑州市的出租房源, B_i 的坐标是变化的, 正负理想解 S^+ 与 S^- 的坐标是固定的, Σ^{-1} 表示 n 个属性变量 b_1, b_2, \dots, b_n 的相关系数 Σ 的逆矩阵。则第 i 个评估对象 B_i 到 S^+ 与 S^- 的马氏距离, 如下式所示:

$$d(B_i, S^+) = \sqrt{(X_i - S^+)^T \Sigma^{-1} (X_i - S^+)} \quad (12)$$

$$d(B_i, S^-) = \sqrt{(X_i - S^-)^T \Sigma^{-1} (X_i - S^-)} \quad (13)$$

(3) 计算各出租房源的相对贴近度。计算爬取的郑州市的各出租房源到正负理想解的 S^+ 与 S^- 的贴近度。其贴近值计算方法如下:

$$C_i^* = \frac{d(B_i, S^-)}{d(B_i, S^+) + d(B_i, S^-)} \quad (14)$$

式中: $d(B_i, S^+)$, $d(B_i, S^-)$ 是求得的各评估对象到正负理想解 S^+ 与 S^- 的马氏距离, C_i^* 是各个出租房源距离理想解的贴近度, C_i^* 值大, 其出租房源排序越靠前。这种方式改进的 TOPSIS 法对出租房源进行综合排序时, 能够有效消除出租房源面积、租金等属性间相关性致使欧式距离失效的影响。

2.2 最差理想解改进 TOPSIS 方法

传统 TOPSIS 法排序无法准确确定两个距离目标点等同远近的出租房源的位置关系。为了便于说明, 如图 1 所示, 设点 $E(x_1^-, y_1^-)$ 对应传统 TOPSIS 法的负理想解向量 K^- 点, $B(x_1^+, y_1^+)$ 对应正理想解 K^+ 点, $C(x_1, y_1)$ 、 $D(x_2, y_2)$ 是 EB 中垂线上的点, E 、 B 两点距离 C 、 D 两点等同远近。显然, 这种特定情况下, 传统 TOPSIS 方法无法对 C 、 D 点所代表的出租房源的优劣性进行判定。

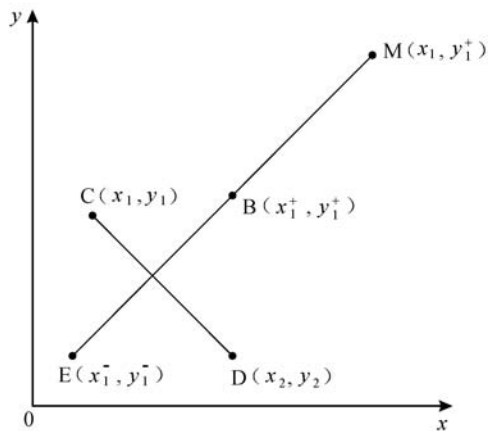


图 1 最差理想解图

为了解决这种特定现象,本文取 EB 延长线上一点 $M(x_1^*, y_1^*)$, 并使 $|EB| = |MB|$ (其中 $| \cdot |$ 表示欧式距离), 且记 M 点对应的向量为 $K^* = (K_1^*, K_2^*, \dots, K_m^*)$, 其中, $K_j^* = 2K_j^- - K_j^+$, $j = 1, 2, \dots, m$ 。定义 M 点为出租房源的最差理想解, 并用其代替传统 TOPSIS 法中的负理想解, 求解各评估对象到正理解的距离尺度 D_i^+ 、到最差理想解的距离尺度 D_i^* 、相应贴近度 C_i^* , 公式如下:

$$D_i^+ = \sqrt{\sum_{j=1}^m (k_{ij} - K_j^+)^2} \quad (15)$$

$$D_i^* = \sqrt{\sum_{j=1}^m (k_{ij} - K_j^*)^2} \quad (16)$$

$$C_i^* = \frac{D_i^*}{D_i^+ + D_i^*} \quad (17)$$

根据相对贴近度 C_i^* 对各评估对象进行综合评价, C_i^* 越大, 表明其对应的评估对象越接近最优选择。这种改进方式有效地解决了在正负理想解中垂线上的出租房源无法准确排序的问题。

2.3 结合马氏距离与最差理想解改进 TOPSIS 方法

分析最差理想解和马氏距离分别改进 TOPSIS 法的评价结果, 将马氏距离改进 TOPSIS 法求得的距离尺度 $d(B_i, S^+)$, $d(B_i, S^-)$ 与最差理想解改进 TOPSIS 法求得的距离尺度 D_i^+ 和 D_i^* 进行标准无量纲化处理之后, 根据租房者偏好系数程度定义出租房源与正负理想解 S^+ 、 S^- 的合成距离, 公式如下:

$$L_i^+ = \alpha \times d(A_i, S^+) + \beta \times D_i^+ \quad (18)$$

$$L_i^- = \alpha \times d(A_i, S^-) + \beta \times D_i^* \quad (19)$$

式中: α 、 β 代表租房者偏好系数, 且 $\alpha + \beta = 1$, 求解各出租房源的相对贴近度 C_i , 公式如下:

$$C_i = \frac{L_i^-}{L_i^+ + L_i^-} \quad i = 1, 2, \dots, m \quad (20)$$

式中: $C_i \in [0, 1]$, C_i 值越大, 代表出租房源排序结果性能越优, 排序顺序越靠前。将马氏距离和最差理想解相结合对传统 TOPSIS 法的改进, 涵盖了两者单独改进 TOPSIS 法的优良特性, 既解决了评估对象间隔正负理想解等同距离的问题, 又消除了欧氏距离本身局限性导致的出租房源属性之间相关性的影响。

3 出租房源综合评价

本文将爬虫系统爬取的某两大租房网站的郑州市出租房源信息作为评估对象, 以郑州大学为目的, 将出租房源的面积属性作为效益型指标, 租金和根据经纬度求得的距离属性作为成本型指标, 利用传统 TOPSIS 法和改进后 TOPSIS 法分别对其进行综合排序实现。

3.1 传统 TOPSIS 方法出租房源评价

利用传统 TOPSIS 法对爬虫系统爬取的某两大租房网站的 12 000 多条郑州市出租房源进行综合排序, 根据上述传统 TOPSIS 法模型中式(1) - 式(4)求得待评估出租房源的面积、租金、距离属性权重级 $w = \{0.060 7, 0.763 4, 0.483 3\}$ 。根据式(6)、式(7)求得各出租房源的正负理想解为:

$$v^+ = [0.008 5 \quad 0.195 2 \quad 0.030 0]$$

$$v^- = [0.002 0 \quad 0.050 0 \quad 0.011]$$

根据式(9) - 式(11)求得各出租房源的相应贴近度, 其综合排序结果的前 10 个最优出租房源及相对贴近度, 如表 1 所示, 其排序结果高德地图图标展示如图 2 所示。

表 1 传统 TOPSIS 法排序结果

排序 1	1	2	3	4	5
面积	35	35	80	37	125
租金	1 000	1 500	1 600	1 200	2 500
距离	38.77	123.1	75.52	76.07	132.7
贴近度	0.045	0.027	0.027	0.024	0.025
排序 1	6	7	8	9	10
面积	86	48	56	42	37
租金	1 600	1 700	1 600	1 700	1 500
距离	75.52	90.3	179	190	132
贴近度	0.020	0.019	0.017	0.016	0.015

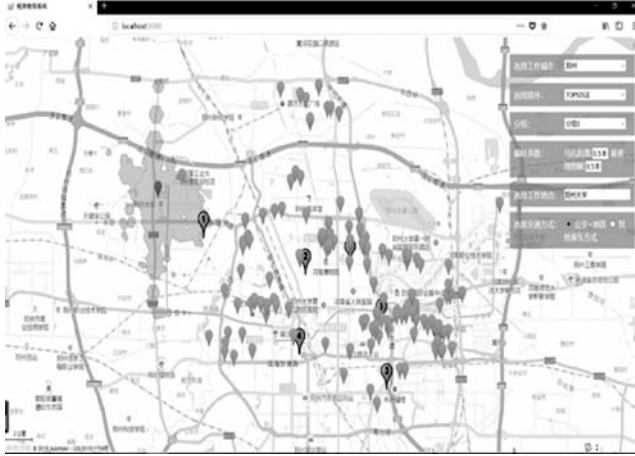


图2 传统 TOPSIS 排序效果图

图2为熵权重化的传统TOPSIS法对出租房源的排序效果展示图。图标1表示根据输入的目的地郑州大学和出租房源的面积、租金及经纬度求得的距离属性,利用传统TOPSIS法对爬虫系统爬取的郑州市的12000条出租房源排序的最终评价结果中贴进度最大的出租房源(最优出租房源),分组1表示其综合排序结果的前1~7个最优出租房源地图图标显示。

3.2 马氏距离优化 TOPSIS 出租房源评价

利用MATLAB对爬虫系统爬取的出租房源的面积、租金和通过经纬度信息求得的各出租房源间隔目标地点郑州大学的距离属性,求解相关系数矩阵,其结果如矩阵B所示:

$$B = \begin{bmatrix} 1 & 0.9018 & 0.1908 \\ 0.9018 & 1 & 0.5349 \\ 0.1908 & 0.5349 & 1 \end{bmatrix}$$

由B可知,出租房源属性间存在较大的相关性,面积与租金之间的相关系数甚至达到了0.9。租金、距离的相关性系数也达到了0.5,传统TOPSIS法对出租房源评价欧式距离失效,利用马氏距离替代欧式距离改进TOPSIS法对出租房源进行综合评价更具合理性。

根据马氏距离改进TOPSIS法模型中的式(12) - 式(14)分别求得各出租房源的相应贴进度,其综合评价的前10个最优出租房源与相应贴进度如表2所示,综合评价结果高德地图图标展示如图3所示。

表2 马氏距离改进 TOPSIS 法排序结果

排序2	1	2	3	4	5
面积	35	86	80	37	125
租金	1 000	1 600	1 600	1 200	2 500
距离	38.77	75.52	75.52	76.07	144.5
贴进度	0.985	0.984	0.978	0.974	0.972
排序2	6	7	8	9	10

续表2

排序2	1	2	3	4	5
面积	80	48	56	37	42
租金	1 600	1 700	1 600	1 500	1 700
距离	75.52	90.3	179	132	190
贴进度	0.971	0.9708	0.9698	0.9678	0.967

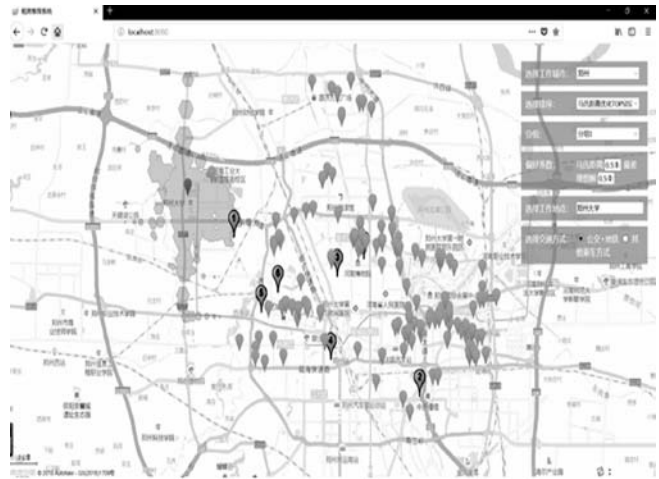


图3 马氏距离优化 TOPSIS 法效果图

图3为马氏距离优化TOPSIS法的排序效果展示图。图中水滴形图标表示依据输入的目的地地点郑州大学和出租房源的面积、租金及经纬度求得的距离属性,利用马氏距离改进的TOPSIS法对郑州市12000条出租房源进行综合排序的最终结果中的前200个出租房源,图标1表示根据上述排序法排序的最终综合评价结果中的最优出租房源选择,图中分组1表示综合评价结果的前1~7个最优出租房源展示。

3.3 最差理想解改进 TOPSIS 出租房源评价

通过爬虫系统爬取的原始出租房源数据和传统TOPSIS法求得的正、负理想解为:

$$v^+ = [0.0085 \quad 0.1952 \quad 0.0300]$$

$$v^- = [0.0020 \quad 0.0500 \quad 0.011]$$

根据最差理想解改进TOPSIS法模型求得最差理想解 $K^* = [-0.0045 \quad -0.0952 \quad -0.008]$,根据式(15) - 式(17)求得最差理想解下的距离尺度、贴进度。其排序结果的前10个最优出租房源选择、相应距离尺度和贴进度,如 D^-, D^* 和表3所示,综合排序结果地图图标展示如图4所示。

$$D^* = \begin{bmatrix} 3452984.548 & 7180342.15 & 3074082.861 \\ 2940061.853 & 6549101.029 & 4363638.394 \\ 13962.282 & 4699275.118 & 5427955.725 \end{bmatrix}$$

	21 950 748. 67		
D^+	=63 293 136. 32	51 792 762. 34	64 577 297. 98
	63 130 525. 87	53 363 598. 8	59 551 441. 88
	61 375 367. 95	58 610 154. 64	55 445 180. 23
	28 223 049. 91		

表 3 最差理想解优化 TOPSIS 排序结果

排序 3	1	2	3	4	5
面积	86	35	80	37	125
租金	1 600	1 000	1 600	1 200	2 500
距离	75. 5	38. 8	75. 5	76. 1	144. 5
贴近度	0. 929 0	0. 928 0	0. 927 0	0. 926 0	0. 924 5
排序 3	6	7	8	9	10
面积	35	48	56	42	37
租金	1 500	1 700	1 600	1 700	1 500
距离	123	90. 3	179.	190	132. 7
贴近度	0. 924 0	0. 921 80	0. 920 60	0. 919 9	0. 919 2

租房者对这两者的偏好度相同均为 0. 5,求得各出租房源的贴近度,其综合排序结果的前 10 个最优出租房源选择和相应贴近度,如表 4 所示,其最终评价结果高德地图图标展示如图 5 所示。

表 4 马氏距离综合最差理想解优化 TOPSIS 排序结果

排序 4	1	2	3	4	5
面积	35	86	80	37	125
租金	1 000	1 600	1 600	1 200	2 500
距离	38. 77	75. 52	75. 53	76. 07	144. 5
贴近度	0. 985 7	0. 984 2	0. 978 7	0. 972 1	0. 972 0
排序 3	6	7	8	9	10
面积	35	48	56	37	42
租金	1 500	1 700	1 600	1 500	1 700
距离	123. 1	90. 3	179	133	190. 2
贴近度	0. 971 7	0. 971 7	0. 970 4	0. 971 4	0. 969 3

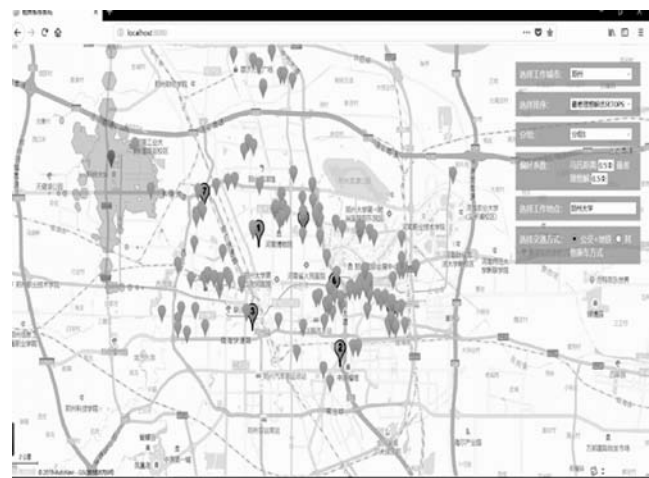


图 4 最差理想解优化 TOPSIS 效果图

图 4 为利用最差理解代替负理想解优化 TOPSIS 法的排序效果展示图。图标 1 表示根据输入的目的地郑州大学和出租房源的面积、租金及经纬度求得的距离属性,利用最差理想解改进的 TOPSIS 法对郑州市 12 000 条出租房源排序的最终综合评价结果中的最优出租房源,图中分组 1 表示其综合排序结果的前 1~7 个最优出租房源展示。

3.4 综合马氏距离和最差理想解优化 TOPSIS 评价

将马氏距离和最差理想解两种方式单独改进 TOPSIS 法所得距离尺度 $d(A_i, S^+)$ 、 $d(A_i, S^-)$ 、 D_i^* 、 D_i^- 进行无量纲化处理后代入式(18) - 式(20),假设

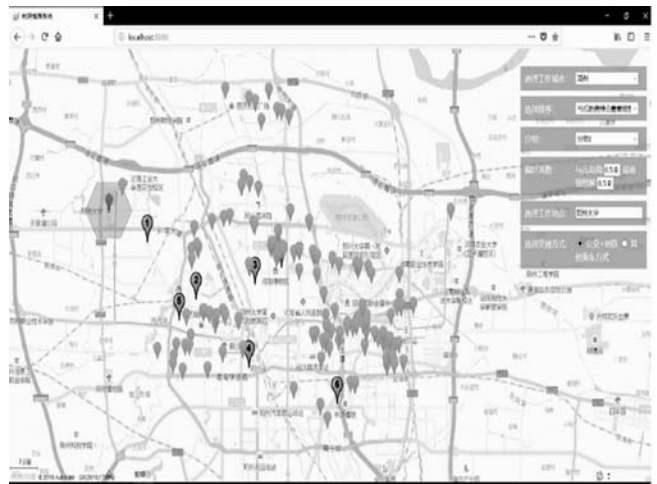


图 5 马氏距离综合最差理想解优化 TOPSIS 法效果

图 5 为马氏距离综合最差理想解优化 TOPSIS 法排序效果展示图。图标 1 表示根据输入的目的地郑州大学和出租房源的面积、租金及经纬度求得的距离属性,利用最差理想解综合马氏距离改进的 TOPSIS 法对郑州市出租房源进行综合评价的结果中的贴近度最大出租房源。偏好者系数表示当马氏距离结合最差理想解优化 TOPSIS 法时,租房者根据两者优化 TOPSIS 法产生的作用选择的的比例(总数是 1),它是可以改变的。如果租房者认为属性间的相关性对排序结果的影响较大,可以适当加大马氏距离优化 TOPSIS 法的比例,降低最差理解的优化 TOPSIS 法的比例,反之亦同。本文是以两者相同比例 0. 5 为例,对出租房源进行排序展示的。图中分组 1 表示其综合排序结果的前 1~7 个最优出租房源展示。

4 出租房源排序结果分析

4.1 排序结果误差差异值分析

验证本文改进的 TOPSIS 法综合评价的性能。假定 $RankResult = \{ Rank(1), \dots, Rank(j), \dots, Rank(m) \}$, 其中 $Rank(j)$ 表示排在第 j 个位置的评估对象。定义 $Diff_i = ((Rank(j), Rank(j'))$ 表示第 $Rank(j)$ 个评估对象和第 $Rank(j')$ 个评估对象在属性 i 上的评价结果与实际评价结果相违背的权重差异值, 其中 $j < j'$, 其计算公式如下:

$$Diff_i(Rank(j), Rank(j')) = \begin{cases} q_{Rank(j')i} - q_{Rank(j)i} & q_{Rank(j)i} - q_{Rank(i)j} > 0 \\ 0 & q_{Rank(j)i} < 0 \end{cases} \quad (21)$$

定义 $Error$ 为实际评价结果与分别针对单一属性值对评估对象评价的总违背差异值, 计算公式如下:

$$Error = \sum_{i=1}^n \sum_{j=1}^m \sum_{j'=j+1}^m Diff_i(Rank(j), Rank(j')) \quad (22)$$

本文将 $Error$ 作为评估标准, 验证不同改进方式对传统 TOPSIS 法综合评估性能的影响, 其总违背差异值 $Error$ 值越小, 评估性能越优。5 次评价实验结果如表 5 所示。

表 5 优化 TOPSIS 排序结果

出租房源排序性能比较	违背差异值
离差最大化权重化传统 TOPSIS 排序	129 423. 720 4
信息熵权重化传统 TOPSIS 排序	128 521. 747 5
最差理想解改进传统 TOPSIS 法排序	125 309. 185
改进的马氏距离优化 TOPSIS 排序	10 040. 835 5
马氏距离综合最差理想解优化 TOPSIS 排序	9 105. 942 8

本文进行了离差最大化权重化传统 TOPSIS 排序、信息熵权重化传统 TOPSIS 排序、最差理想解优化传统 TOPSIS 法排序、马氏距离优化 TOPSIS 排序、马氏距离结合最差理想解优化 TOPSIS 法排序共五种综合评估验证实验。根据实验结果, 显然, 利用离差最大化对传统 TOPSIS 法权重化求得的评价结果, 其总违背差异值是最大的。因此, 本文在对传统 TOPSIS 的改进中, 除马氏距离使用的是其内化权重, 其他方式的优化均是在熵权重化的条件下进行的, 由表 3 比较分析可知, 无论马氏距离和最差理想解两种方式单独对传统 TOPSIS 法的改进, 还是将两者结合对 TOPSIS 法的改进, 其评价结果的总违背差异值均有所降低。尤其是将马氏距离与最差理想解相结合优化 TOPSIS 法的评价结果其总违背差异值仅为 9 105. 942 8, 显然性能最优。

4.2 排序结果数据对比分析

4.2.1 改进前后的 TOPSIS 对比分析

将推荐结果表 1(传统 TOPSIS 方法)和表 3(最差理想解改进 TOPSIS 方法)进行对比, 其传统 TOPSIS 法排序 6 变成了最差理想解优化 TOPSIS 法的排序 1, 排序 1 变成了排序 2, 而排序 2 退到了排序 6 的位置。显然导致这种变化的根本因素是传统 TOPSIS 法对出租房源评价时, 存在出租房源间隔正负理想解等同距离导致无法被准确定位的问题。而本文利用最差理想解代替负理想解优化 TOPSIS 法对出租房源综合排序时, 很好地解决了这一问题。比较马氏距离优化 TOPSIS 方法所得排序结果(表 3)与传统 TOPSIS 法所得排序结果(表 1), 发现除排序 1、3、5、7、8 之外, 其他的出租房源排序均发生了较大偏移。这是因为出租房源的属性指标(面积, 租金)之间存在相关性, 导致传统的 TOPSIS 方法对其进行评价时欧式距离失效, 而马氏距离因不受量纲影响, 消除了属性指标间相关性影响。将马氏距离与最差理想解相结合优化 TOPSIS 方法的推荐结果(表 4)与传统 TOPSIS 法的推荐结果(表 1)进行对比。发现传统 TOPSIS 法排序 6 变成了改进 TOPSIS 法排序 2, 排序 4 变成了排序 6, 排序 2 则倒退到了排序 4 的位置, 排序 9 和 10 互换了位置, 这种差异主要是因为最差理想解综合马氏距离优化 TOPSIS 法不仅解决了出租房源面积、租金等属性间相关性, 又解决了出租房源距离正负理想解等同远近时无法准确定位的问题。

4.2.2 综合改进 TOPSIS 与单独改进 TOPSIS 对比分析

分析马氏距离综合最差理想解改进 TOPSIS 法的排序结果(表 4)与马氏距离改进 TOPSIS 法(表 3)可知, 其结果除 4 和 6 互换了位置, 其他均没太大变化。这是因为对出租房源进行综合排序时, 其面积、租金等属性间相关性对出租房源的排序结果影响较大, 而无论马氏距离综合最差理想解优化 TOPSIS 法还是马氏距离单独优化 TOPSIS 法对出租房源进行排序, 均克服了这一缺陷。这种情况体现了改进方法对出租房源评价的一致性, 表明了本文改进 TOPSIS 法排序的有效性。而细微的差异性则因为马氏距离改进传统 TOPSIS 法仅解决了出租房源面积、租金等属性之间相关性导致欧式距离失效的问题, 但可能存在因出租房源距离正负理想解相同距离, 导致排序无法确定位置的问题。

题。马氏距离与最差理想解相结合改进 TOPSIS 法同时解决了以上两个缺陷。将马氏距离结合最差理想解改进 TOPSIS 方法对出租房源排序的结果(表4)与最差理想解单独改进 TOPSIS 法的排序结果(表2)对比,其排序1变成了排序2,排序10变成了排序9,这种差异显然是因为最差理想解单独改进 TOPSIS 法只是解决了出租房源距离正负理想解等同远近的问题,没有解决出租房源属性(面积、租金、间隔目标点距离)间线性相关问题。

综上所述,本文提出的将马氏距离与最差理想解相结合优化的 TOPSIS 方法比马氏距离和最差理想解单独改进 TOPSIS 方法更具有科学性及其有效性,应用于对出租房源的评价也更合理。

4.3 排序结果实用价值分析

分析马氏距离综合最差理想解优化 TOPSIS 法对出租房源评价的结果(表4)的实用价值可知,排列在第一位的出租房源,其租金1000、面积35、距离38.77等指标方面分配较合理,性价比最高。对于排列在第2位和第3位的出租房源,在其租金、距离属性相同的条件下,第3位出租房源的面积却比第2位出租房源的面积少了6平方米,显然将第3位排到第2位之后也合理的。而对于第3位、第4位出租房源在其距离相差不大的情况下,第4位出租房源虽租金比第3位少了400元,但其面积减少了二分之一,故排序也是合理的,对于排列在第5位、第6位、第7位的出租房源,在其距离基本差异不大的基础上,第5位出租房源的租金虽比第6位多了1000元,而面积比第6位增加了四分之三,故其排序合理,排序在第7位的出租房源其面积虽比排列在第6位的出租房源多了10平方米,但租金却多出了200元,故将其排列在后面也是合理的。对于排列在第8位、第9位的出租房源,第9位的出租房源距离目的地点郑州大学的距离比排序在第8位少了50、但是面积却少了10平方米,故排序也是合理的。而对于排序在第10位的出租房源其面积是42平方米,距离是200,租金却高达1700元,可见性价比比前9个出租房源都低,故把它排最后一位也是合理的。

综上所述,本文将马氏距离和最差理想解相结合优化 TOPSIS 方法对出租房源进行综合排序的评价结果具有合理的现实意义和实用价值。

5 结 语

本文通过研究传统 TOPSIS 方法的缺陷,提出了一

个适用范围更广且更加合理的改进 TOPSIS 方法,即用马氏距离代替欧氏距离,用最差理想解代替负理想解优化传统的 TOPSIS 法,并将其应用到爬虫系统爬取的出租房源的多属性排序中。实验证明该方法有效地消除了传统 TOPSIS 法因出租房源面积和租金等属性间相关性导致的欧氏距离失效及距离正负理想解等同远近无法定位的缺陷,对出租房源的排序更具合理性及科学性,其排序结果更具实用价值和现实意义。本文的不足之处是该方法仅限于数值属性的决策分析,而对那些需要进行混合型属性分析的评估对象并不适用,故本文下一步将重点研究如何改进 TOPSIS 法,使其能够根据评估对象的混合属性进行综合评价。

参 考 文 献

- [1] 李华,何正柯,李群,等.改进的 TOPSIS 决策方法在供应商选择中的应用[J].数学的实践与认识,2016,46(16):93-101.
- [2] 胡永宏.对 TOPSIS 法用于综合评价的改进[J].数学的实践与认识,2002,32(4):572-575.
- [3] 李灿,张凤荣,朱泰峰,等.基于熵权 TOPSIS 模型的土地利用绩效评价及关联分析[J].农业工程学报,2013,29(5):217-227.
- [4] 谢小美.基于多信息融合的 Web 服务综合排序技术研究[D].南昌:江西财经大学,2015.
- [5] 孙若男,张斌,刘婷婷.一种使用改进熵权 TOPSIS 的 Web 服务质量评价排序方法[J].小型微型计算机系统,2017,38(6):1221-1226.
- [6] Hwang C L, Yoon K P. Multiple attribute decision making. Methods and applications. A state-of-the-art survey[J]. European Journal of Operational Research, 1981, 4(4):287-288.
- [7] 王先甲,汪磊.基于马氏距离的改进型 TOPSIS 在供应商选择中的应用[J].控制与决策,2012,27(10):1566-1570.
- [8] 刘晓峰,远亚丽.基于改进 TOPSIS 法的第三方物流供应商选择[J].物流科技,2011,34(3):102-105.
- [9] 信桂新,杨朝现,杨庆媛,等.用熵权法和改进 TOPSIS 模型评价高标准基本农田建设后效应[J].农业工程学报,2017,33(1):238-249.
- [10] 慈铁军.基于决策者偏好的区间数多属性决策方法研究[D].天津:河北工业大学,2014.
- [11] 何正柯. TOPSIS 多属性决策方法的改进研究[D].鞍山:辽宁科技大学,2017.
- [12] 余雁,梁樑,蒋跃进,等.一种新的基于模糊偏好的 TOPSIS 改进方法[J].系统工程,2004(8):87-90.