

个性化购房决策系统的设计与实现

邵小康[†], 王培进

(烟台大学 计算机与控制工程学院, 山东 烟台 264005)

摘 要: 研制个性化购房决策系统, 帮助购房者快速选到适合自己需求的住房显得十分必要。定义定量和定性购房指标, 利用网络爬虫技术获取房源信息; 将 GIS 和数字孪生技术结合并应用于购房决策; 采用模糊层次分析法和熵权法相结合的方式确定权重, 利用模糊 TOPSIS 决策方法优化购房方案, 为购房者优选出符合自己个性化需求的住房。个性化购房系统的研制可以很好地满足广大购房者购房需求, 实例分析证明, 有很好的应用价值。

关键词: 购房决策; 爬虫技术; 模糊层次分析法; 模糊 TOPSIS; 个性化; 数字孪生

中图分类号: TP311.1

文献标识码: A

Design and Implementation of Personalized House Purchase Decision System

SHAO Xiaokang[†], WANG Peijin

(School of Computer and Control Engineering, Yantai University, Yantai, Shandong 264005, China)

Abstract: It is necessary to develop a personalized purchase decision system to help homebuyers quickly choose the housing they need. Define quantitative and qualitative purchasing indicators, and use web crawler technology to obtain housing information; applying GIS and digital twin technology to assist decision-making; using a combination of fuzzy analytic hierarchy process and entropy weight method to determine the weights, and using the fuzzy TOPSIS decision-making method to optimize the purchase plan, buyers can select houses that meets their personalized needs. The development of a personalized home purchase system can effectively meet the needs of a large number of homebuyers, and case analysis has shown that it has good application value.

Key words: purchasing decision; crawler technology; fuzzy analytic hierarchy process; fuzzy TOPSIS; personalization; digital twin

在互联网及大数据时代背景下, 大数据技术不仅能够提高用户需求分析的精准性、全面性及有效性, 还能更好地满足用户多样化、个性化的需求^[1], 为设计实现个性化购房决策系统提供了技术上的保证。现有的房产信息筛选系统, 只是简单地将房产信息按照文本标签分类, 例如按照价格、户型、区域或者房屋类别等。分类后依然是繁杂的房产信

息, 用户很难筛选出满足自己个人实际需求的住房。文献[2]构建的购房平台系统结合百度地图应用程序编程接口(API), 但数据以文字形式表现且评价简单。文献[3]结合 ArcGIS 构建了选房系统, 但地图显示简单。综合各类购房系统, 传统的购房系统主要是依据房屋的定量指标进行筛选, 而在实际购房中, 用户的需求表达许多是定性的、模

糊的。将定性和定量指标结合,才能真正满足购房者的个性化需求。另一方面,现有的购房系统房源数据信息基于历史数据,不能及时更新,缺少实时性。

为解决目前购房决策系统存在的问题,研制了个性化、实时购房决策系统。系统的主要特征是:(1)设置较为全面的定性和定量化购房指标,满足购房者个性化购房需求;(2)采用网络爬虫技术,定时爬取网上房源信息,在线决策,满足实时性和有效性需求;(3)利用模糊多目标决策技术,为购房者优选住房。

数字孪生技术逐步应用与各行各业中,其可以通过虚拟模型实现物理产品的数字化表达^[4]。郑伟皓等^[5]以数字孪生五维模型为指导,以 GIS 为具体技术实现了交通基础设施的数字孪生系统。房产的地理周边信息可以辅助购房者进行决策,将

房产和周边信息以 3D 可视化形式展示易于进行主观辅助决策。本文基于 Unreal Engine^[6],设计了数字孪生体 3D 可视化购房系统,实现了系统设计的所有功能。

1 购房指标分析与处理

1.1 定量购房指标

购房指标中有确定数值的、可以用数值来衡量的即定量指标。购房定量指标是筛选分类房产的重要依据,用户可以直观地得到房产重要属性信息。定量购房指标又可分为正向购房指标、反向购房指标和中性购房指标,正向购房指标是指标的值越小越为最优;反向购房指标是指标的值越大越为最优;中性购房指标是根据用户的实际需求来判断。表 1 给出了相关定量购房指标。

表 1 定量购房指标	
指标	指标分析
价格	价格是购房的关键指标,是用户在筛选房产和决策时很看重的指标,属于反向购房指标。
用户评分	用户评分是用户在互联网上对该房产的打分,满分为 5 分,该指标可能存在误差,属于正向购房指标。
物业费	物业费的高低也对购房者选择房产有重要影响,属于反向购房指标。
容积率	容积率是指小区内总建筑面积与建筑用地面积的比值,这关系用户的居住舒适度,属于反向购房指标。
绿化率	绿化率是指小区内的绿化程度,以百分比描述,这体现了该小区的环境是否良好,属于正向购房指标。
户型面积	户型面积无法确定其好坏,因为有的用户想要大的;有的想要小的。用户面积属于中性购房指标。

1.2 定性购房指标

定性指标不能以具体数值所描述,表征购房者

的“软”需求,主要是交通、环境等因素影响购房者决策。表 2 给出了主要定性指标。

表 2 定性购房指标	
指标	指标分析
交通设施	房屋所在小区周围 500 m 内是否有公交站,或者距离地铁站是否远以及交通是否便捷等。
居住设施	房屋所在小区居住生活是否方便,周围 1000 m 内是否有利于生活便利的超市、门诊、快递站、洗衣店等。
环境设施	房屋所在小区周围 1000 m 内是否有大的化工厂污染空气,周围有没有公园、植物园、山、水系等。
地理位置	房屋所在小区周围是否有幼儿园、中小学等,周围 5000 m 内有没有大型医院,有没有购物中心或者图书馆等。

1.3 定性购房指标的量化

定性指标的量化处理可分为“直接量化”和“间接量化”两种方法,前者是对定性指标直接进行打分赋值,后者则是划分评价等级进行量化。我们采用的是“间接量化”法,购房者在系统中对某个定性购房指标进行细致了解后,选择对该指标的满意度,符合个性化服务的宗旨。根据评价标准将指标满意度划分为五个等级,对于该等级域上的任一元素 x ,都有一个 $f(x) \in (0,1]$ 与之对应,则称 $f(x)$ 为该域

上的模糊集, $f(x)$ 为 x 隶属度。随着该域上的 x 的变动, $f(x)$ 则为一个函数,称为隶属函数。因此,在划分完满意度等级后需要确定隶属函数^[7]。定性购房指标满意度映射等级如表 3 所示。

表 3 主观购房指标满意度映射等级				
很满意	满意	比较满意	一般	不满意
5	4	3	2	1

在对定性购房指标的满意度划分等级后,需要将划分的等级连续量化处理。根据购房者满意度的变化,在满意度到达极值时变化会变小,这里取偏大型柯西分布和对数函数作为隶属函数:

$$f(x)=\begin{cases} [1+\alpha(x-\beta)^{-2}]^{-1}, & 1\leq x\leq 3 \\ a\ln x+b, & 3<x\leq 5 \end{cases} \tag{1}$$

其中 α 、 β 、 a 、 b 为待定常数,值为:1.1、0.9、0.4、0.4。

当“很满意”时,隶属度为 1,则 $f(5)=1$;

当“不满意”时,隶属度为 0.01,则 $f(1)=0.01$ 。

满意度在 0—1 区间的变化曲线如图 1 所示。

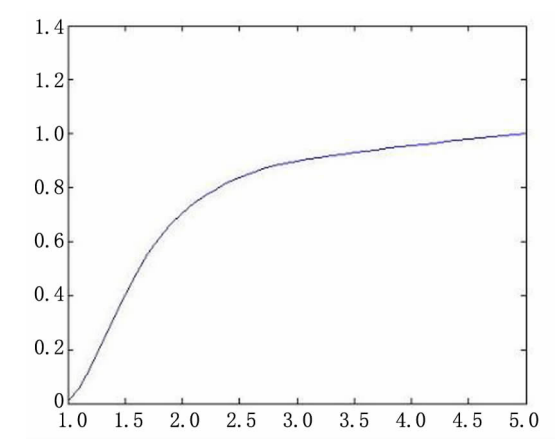


图 1 满意度在 0—1 区间的变化曲线

定性购房指标量化处理结果如表 4。

表 4 定性购房指标量化处理结果

很满意	满意	比较满意	一般	不满意
1	0.91	0.8	0.52	0.01

通过建立好的函数模型,用户选择的任何一个购房指标评价价值都可以得到一个合适的量化值,量化处理后的购房指标为综合评判奠定了基础。

1.4 定量购房指标的无量纲化

不同购房指标的单位是不同的,而且有的购房指标又有正向指标和反向指标之分。本文采用标准(0—1)区间变换的方法对定量购房指标进行无量纲化处理,该方法可以将原始购房指标数据进行线性变换,使其落到[0,1]区间内,公式如下:

正向购房指标:

$$r_{ij}=\frac{x_{ij}-x_{j_{\min}}}{x_{j_{\max}}-x_{j_{\min}}} \tag{2}$$

反向购房指标:

$$r_{ij}=\frac{x_{j_{\max}}-x_{ij}}{x_{j_{\max}}-x_{j_{\min}}} \tag{3}$$

2 房屋信息爬取技术

购房决策系统需要爬取网上大量有用的房源信息,为购房者决策提供数据支持。主题爬虫^[8]可以很好地爬取购房指标信息,但是爬取效率比较低,因为爬虫是一个需要访问网络的技术,这段时间内会有网络延迟,爬取数据需要一个时间过程,而爬虫速率关系到为用户提供的服务质量^[9]。为此我们采用了异步爬取技术策略,异步爬取是在执行一个爬取任务时,为了提高效率,由于网络延迟,让该爬取任务边执行边等待,然后去执行其他爬取任务,这样就保证了时间的利用率。

Aiohttp 是基于 Python 平台实现异步爬取的一个库,Aiohttp 需要先将爬取方法用 async 关键字标记,这样调用该函数时并不会执行该函数,只是返回一个协程对象。然后调用 Aiohttp 的 ClientSession 方法,可以自动打开和关闭 Aiohttp,使用 request 发送 get 请求爬取数据。由于需要执行多个爬取访问,爬取不同网址房产信息,因此需要将这些任务放到一个列表中,然后将其挂起,部分代码如表 5 所示。

通过 Request 库在百度地图开放平台上获取小区和周边信息的经纬度,获取的坐标系是经过处理的,所以存在偏差,先确定一个位置的经纬度及其偏差,然后在将经纬度转换为 Unreal Engine 坐标系时减去偏差。

3 基于主客观组合权重的模糊 TOPSIS

为了使决策结果更加符合用户的主观行为,同时决策的数据结果也具有客观性,本文采用模糊层次分析法和熵权法结合的方式来确定权重。这样既保证了决策结果贴合用户的行为,又具有客观可信度。

模糊 TOPSIS(逼近理想点法)是将传统 TOPSIS 与做模糊处理的购房指标矩阵相结合,使用处理后的购房指标组成模糊矩阵,利用 TOPSIS 进行决策。TOPSIS 法是一种接近理想解的排序方法^[10],详细算法过程见参考文献。

表 5 爬取购房指标信息的部分代码

Algorithm 1 asynchronous web crawler
Input: A list with an element of a URL
Output: String data
1: asyncio.run(); # Create loop and run
2: async with aiohttp.ClientSession() as session # Asynchronous context manager automatically aiohttp
3: async with session.get(url=url,) as response # Send get response
4: text = await response.text() #Pending and receive data
5: tasks.append(asyncio.create_task()) # Create tasks
6: await asyncio.wait(tasks) # Pending and waiting tasks

3.1 模糊层次分析法确定主观购房指标权重

模糊层次分析法是一种偏主观的方法,让决策过程贴合用户的主观行为,使用户在购房决策过程中参与购房指标权重的设定,让用户觉得哪一个购房指标更为重要,更符合个性化特征。

对于定性和定量指标,根据各个指标对购房者的购房决策影响力不同,赋予不同的权重,便于综合评判。中性购房指标是在用户做出筛选时所选择的,不给予权重,其他购房指标采用模糊层次分析法确定购房指标的权重。模糊层次分析法的结构模型分为目标层、指标层和方案层,目标层是购房的最优方案,指标层则是购房系统中所涉及的指标^[11]。购房决策层次结构模型如图 2 所示。

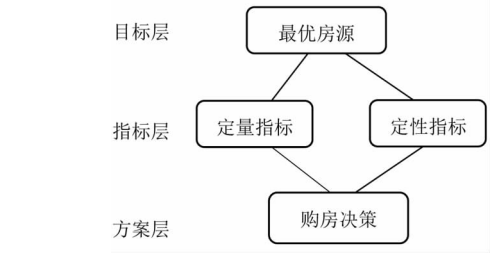


图 2 购房决策层次结构模型

使用模糊层次分析法确定指标的权重,本系统先要用户对购房指标的重要行进行排序,假设用户所选择的购房指标的排序如图 3 所示,数值越小说明越重要,同等数值说明一样重要。

根据购房指标之间的比较,1 表示购房指标 i 比购房指标 j 重要;0 表示购房指 j 比购房指标 i 重要;0.5 表示一样重要^[12]。根据这个准则构建模糊判断矩阵 $A_{n \times n}$,然后转化模糊一致矩阵 B_{ij} ,公式如下:

$$b_i = \sum_{j=1}^n A_{ij}$$

(4)

$$B_{ij} = \frac{b_i - b_j}{2n} + 0.5$$

(5)

模糊一致矩阵购房指标的数值的定性描述如表 6 所示。



图 3 用户所选择的购房指标的排序

表 6 模糊一致矩阵描述说明

指标 i 比指标 j	量化值
同等重要	0.5
稍微重要	0.6
重要	0.7
较强重要	0.8
非常重要	0.9
反比较	0.1、0.2、0.3、0.4
计算	$a_{ij} = 1 - a_{ji}$

然后以最优方案指标表标准对指标集进行层次单排序,计算出来的矩阵即购房指标的权重矩阵,如表 10 所示;再以每个购房指标为标准进行单排序,公式如下:

$$w_i^{\pm} = \frac{1}{n} \sum_{j=1}^n \frac{B_{ij}}{\sum_{k=1}^n B_{kj}}$$

(6)

得到购房指标的权重矩阵 $w_i^{\pm}$ 如表 7 所示。

表 7 购房指标的权重矩阵

价格	用户评分	容积率	绿化率	物业费	交通设施	居住设施	环境设施	地理位置
0.165360	0.0541875	0.1253758	0.0842399	0.069802	0.1520974	0.1253758	0.1253758	0.0981859

3.2 熵权法确定客观购房指标权重

购房决策的重点是确定购房指标权重,熵权法是一只客观的赋权方法。通过计算信息熵,根据购房指标的变化程度对整个购房决策影响确定权重^[13]。

本文还是选取融创·融公馆、新华万科·星耀天地、祥隆中心、源山郡和大栖地这几个购房方案作为例子进行决策。第一步进行概率矩阵 p 的计

算, h 为所有的购房指标矩阵:

$$p_{ij} = \frac{h_{ij}}{\sum_{i=1}^n h_{ij}}$$

(7)

第二步计算各个购房指标的信息熵:

$$e_i = -\frac{1}{\ln(n)} \cdot \sum_{i=1}^n [p_{ij} \ln(p_{ij})]$$

(8)

购房指标信息熵计算结果如表 8 所示。

表 8 购房指标的信息熵矩阵

价格	用户评分	容积率	绿化率	物业费	交通设施	居住设施	环境设施	地理位置
-0.130676	-0.124348	-0.146851	-0.130793	-0.134904	-0.133991	-0.155715	-0.154778	-0.135902

第三步根据所得的信息熵求取信息效用值,公式如下:

$$d_i = 1 - e_i$$

(9)

购房指标信息效用值计算结果如表 9。

表 9 客观购房指标的信息效用值矩阵

价格	用户评分	容积率	绿化率	物业费	交通设施	居住设施	环境设施	地理位置
1.130676	1.124348	1.146851	1.130793	1.134904	1.133991	1.155715	1.154778	1.135902

第四步计算购房指标的权值,公式如下:

购房指标权重计算结果如表 10。

$$w_i^{\text{客}} = \frac{d_j}{\sum_{j=1}^m d_j}$$

(10)

表 10 客观购房指标的权重矩阵

价格	用户评分	容积率	绿化率	物业费	交通设施	居住设施	环境设施	地理位置
0.110332	0.109714	0.11191	0.110343	0.110744	0.110655	0.112775	0.112684	0.110842

3.3 组合权重的确定

将所得的主观购房指标和客观购房指标利用公式(11)相结合, $w_i^{\text{总}}$ 是求得的组合权重; $w_i^{\text{主}}$ 是主观购房指标权重; $w_i^{\text{客}}$ 是客观购房指标权重; u 是比例系数,大于 0.5 就偏主观购房指标,小于 0.5

则偏客观购房指标,本文取 0.5,使得主观购房指标占比相同。最后求得的购房指标权重如表 11 所示。

$$w_i^{\text{总}} = u \cdot w_i^{\text{主}} + (1 - u) w_i^{\text{客}}$$

(11)

表 11 购房指标的权重矩阵

价格	用户评分	容积率	绿化率	物业费	交通设施	居住设施	环境设施	地理位置
0.137846	0.081951	0.118643	0.0972915	0.090273	0.131376	0.119075	0.119030	0.104514

3.4 基于 TOPSIS 的购房指标综合评价

第一步先要构造模糊决策矩阵 $E_{m \times n}$, m 为需要决策的房产的个数。模糊决策矩阵是由需要参加决策的数据和指标构成的,指标数据都是经过模糊处理的数值型数据。模糊决策矩阵和权重矩阵相乘得到模糊加权决策矩阵:

$$D_{m \times n} = E_{m \times n} \cdot w_i^{\text{总}} \tag{12}$$

基于模糊加权决策矩阵 $D_{m \times n}$,进行方案优选。在 TOPSIS 法中,优选方案称为理想方案。理想方案分为正、负理想方案。正理想方案是正向指标的最小值、反向指标的最大值;负理想方案是反向指标的最小值,正向指标的最大值。我们采用理想方案之间的贴进度表示优劣,TOPSIS 方法通过建立某种综合评价指标,即贴进度,来度量比较数列靠近正理想数列与远离负理想数列的程度,贴进度越高表示该方案越好^[14]。首先计算到理想方案的欧氏距离,然后计算理想方案的贴进度,最终计算的贴进度结果越接近 1,方案越好;越接近 0,表示方案越差。公式如下:

$$C = \frac{S^-}{S^+ + S^-} \tag{13}$$

其中 C 为贴进度, S^- 为到负理想方案的距离, S^+ 为到正理想方案的距离。

每个备选的购房方案的贴进度计算出来后,再将其排序,如表 12 所示。

表 12 购房方案的贴进度排序结果		
名称	相对贴进度	
祥隆中心	0.606025	
大栖地	0.431755	
融创·融公馆	0.430631	
源山郡	0.393054	
新华万科·星耀天地	0.386834	

4 应用

第一步:小区筛选。购房者首先根据自己的个性化需求选择定量指标,决策系统根据选择的指标列出满足需要的小区。筛选操作效果图如图 4 所示。

购房者在图 4 的“查询结果”模块中,可以点击相应的小区,画面就会转换到相应的小区的位置,并展示 3D 模型,购房者可以在右侧“周边信息”模块中先选择一级行业,然后选择二级行业,系统中就会显示周边信息的标签,效果图如图 5 所示。



图 4 筛选操作界面



图 5 周边信息展示效果图

第二步:“查询结果”模块的所有小区都会在确定模糊指标信息界面中,用户可以点击“删除”按钮,删除不想要参与决策的小区,效果图如图 6。



图 6 确定模糊指标信息效果图

第三步:用户选择结束后,点击“确定”按钮,显示决策结果信息。

5 结 论

个性化购房决策系统可以让购房者按照自己需求快速筛选房产信息,找到适合自己需求的最佳房源。所有房源信息和决策是定时更新、在线的,实现了房屋和周边信息的可视化,克服了现有一些购房系统功能上的不足,具有简洁、方便、快捷的特点。目前该系统正在进一步深化,以丰富、完善一些购房指标,达到推广、应用的目标。

参考文献

[1] 李春,朱珍民,叶剑,等. 个性化服务研究综述[J]. 计算机应用研究,2009,26(11):4001—4005.

[2] 程彬,段浩宇,董炜华. 基于百度地图 API 的购房平台系统开发[J]. 长春师范大学学报,2014,33(8):86—90.

[3] 刘雨婷,张超. 基于 ArcGIS Server 的在线选房系统的实现[J]. 城市勘测,2018,162(1):102—106.

[4] 刘芳,刘琪,黄美晨,等. 数字孪生:跨界赋能于多领域智能的新应用[J]. 计算机系统应用,2023,32(8):1—11.

[5] 郑伟皓,周星宇,吴虹坪,等. 基于三维 GIS 技术的公路交通数字孪生系统[J]. 计算机集成制造系统,2020,26(1):28—39.

[6] 董学宁. 基于虚幻引擎的地域文化展示设计研究[D]. 广州:华南农业大学,2020.

[7] 李金秋,于晶贤,苗晨. 基于模糊 AHP 与模糊 TOPSIS 的人才选拔方案[J]. 辽宁石油化工大学学报,2012,32(4):88—91.

[8] 赵康. 面向主题的网络爬虫系统的设计与实现[D]. 北京:北京邮电大学,2019.

[9] 李阳,范伊红,李彦蓉,等. 基于 aiohttp 的异步爬虫的设计与实现[J]. 无线互联科技,2021,18(17):56—57.

[10] 周亚. 多属性决策中的 TOPSIS 法研究[D]. 武汉:武汉理工大学,2009.

[11] 符学葳. 基于层次分析法的模糊综合评价研究和应用[D]. 哈尔滨:哈尔滨工业大学,2011.

[12] 李健平. 基于模糊层次分析法合肥市贾大郢安置房项目造价风险管理研究[D]. 合肥:安徽建筑大学,2022.

[13] 蒋毅,傅俊诚,李泽文,等. 基于改进 TOPSIS 的楼宇多指标综合评价[J]. 上海交通大学学报,2023,57(7):1—10.

[14] 王自成,陈华友,韩冰,等. 基于 TOPSIS 的最大贴近度的最优组合预测模型及其性质[J]. 中国科学技术大学学报,2019,49(9):751—761.

(上接第 45 页)

参考文献

[1] 陈磊,赵日升,矫镨达,等. 基于 ANFIS 和 SOM 的抽水蓄能电站主变压器健康状态监测[J]. 水电能源科学,2020,38(3):177—180.

[2] 张敏,朱杰兵,周黎明,等. 锚杆无损检测技术在抽水蓄能电站中的应用[J]. 地下空间与工程学报,2019,15(2):847—851.

[3] 蒋洪. 水电站机电设备的故障检修与运行维护[J]. 水利水电科技进展,2022,42(6):136.

[4] 赵洪山,高玉峰. 基于振动特征估计的 GIS 设备故障检测与分析[J]. 高压电器,2020,56(6):114—120.

[5] 刘晓璇,季怡,刘纯平. 基于 LSTM 神经网络的声纹识别[J]. 计算机科学,2021,48(2):270—274.

[6] 欧阳城添,袁瑾. 基于学习矢量量化的空调压缩机声纹诊断方法[J]. 计算机工程与设计,2021,42(9):2634—2641.

[7] 林春清,周颖,杨超,等. 面向电力变压器的声纹智能诊断装置设计与应用[J]. 传感技术学报,2021,34(10):1412—1420.

[8] 肖勇,赵云,涂治东,等. 基于改进的皮尔逊相关系数的低压配电网拓扑结构校验方法[J]. 电力系统保护与控制,2019,47(11):37—43.

[9] 纪德洋,金锋,冬雷,等. 基于皮尔逊相关系数的光伏电站数据修复[J]. 中国电机工程学报,2022,42(4):1514—1523.

[10] 张晓波,刘保华,于凯本,等. 基于灰度预测的变长度经验模态分解方法研究[J]. 海洋科学进展,2020,38(3):493—503.

[11] 杨鹤标,刘芳,胡惊涛. 基于 PSO 的小样本特征选择优化算法研究[J]. 江苏科技大学学报(自然科学版),2021,35(1):76—81+97.

[12] 张维,张浩晨. 一种基于最优集成随机森林的小样本数据特征提取方法[J]. 西北工业大学学报,2022,40(6):1261—1268.

[13] 耿苏杰,王秀丽. 基于模糊贝叶斯网络的电力设备故障诊断和状态评估[J]. 计算机集成制造系统,2021,27(1):63—71.

[14] 赵欢,阳浩,何亮,等. 高精度配电网电气设备故障识别检测方法[J]. 沈阳工业大学学报,2021,43(6):614—618.

[15] 李元,姚宗禹. 基于改进近邻标准化主多项式的故障检测研究[J]. 计算机仿真,2022,39(2):501—506+517.