

大数据下基于房屋交易网站的数据获取的二手房价格走势分析 ——以上海为例

张汉中¹ 张倩¹ 董起航² 周小平¹ 王斌¹

(1、上海师范大学 信息与机电工程学院, 上海 200234 2、青岛大学 数据科学与软件工程学院, 山东 青岛 266017)

摘要:随着城市化进程的不断加深, 互联网技术的飞速发展, 以及人们对于便于工作的住房需求, 致使房屋交易日趋活跃。房价是房地产市场健康稳定的重要指标。预测房价的变化有助于对房地产市场有一个清醒的认识。本文通过利用 Python 语言, 自动提取二手房交易信息, 并针对房屋成交量、成交户型、看房人数等多种因素进行分析, 通过统计图表分析上海市某一地区的房屋交易的价格变化以及相应地区的交易活跃度, 从而对二手房价走势作出预测。

关键词:房屋交易网站; 网络爬虫; 房价影响因素

Abstract: Housing transactions are getting more and more active resulting from the deepening urbanization, the rapid development of Internet technology, the demand for convenient traffic. Housing price is an important signal of the health and stability of the property market. Predicting the changes of the housing price will help people have a clear understanding of the property market. This paper uses Python language to extract second-hand housing transaction information automatically and analyzes a variety of data such as the housing trading volume, the housing type, the number of the people visited the house and so forth. By analyzing the changes of transaction price and activity for a certain area in Shanghai, we made the price tendency analysis for second-hand housing transaction in this paper.

Key words: Real estate transactional websites; Website spiders; Element of affecting housing prices

随着互联网技术的快速发展和房屋交易市场的日益活跃, 房地产市场也逐渐引起了人们的关注。近些年上海住宅价格水平在市场中变化很大。房态理论解释说, 价格取决于市场的特点^[1]。然而现有的文献大多数是通过对于一个或几个特定的常见因素(土地成交价格、公积金涨幅、人均收入等^[2])进行对房价的预测分析^[3], 现有的数据都倾向于分析房价的长期变化趋势, 这依然存在某方面的局限性^[4]。住房价格有很大的可预测性, 但考虑因素越全面, 则未来房价预测越准确^[5]。未来房价预测的准确性越高, 购买者在当前和未来消费中的分配就越理性^[6]。本文试图通过运用 python 语言构建爬虫算法对房屋交易的价格变动信息进行分析, 进而提出合理意见。

1 房屋交易网站网页结构分析

通过对目前主流的房屋交易网站的网页结构特征进行分析, 总结了此类网页的常见特征。房屋交易类网页通常含有房屋的报价, 大小, 结构类型, 地理位置, 交易次数, 周边配套设施, 相似类型房源推荐等几部分, 其中房屋的报价, 大小, 地理位置为我们需要的有用信息, 其他不相关信息需要在网页解析过程中过滤掉。一个典型网页结构如图 1 所示。

2 房屋交易价格信息获取方法的设计

从网页中实时获取房屋的信息的主要步骤为: 解析网页的地址(url), 获取网页数据, 进行数据筛选, 输出相应规范格式, 进行数据分析。(图 2)

3 房屋价格信息提取框架的实现

3.1 确定相应的 url

一般来说, 用户搜索租赁房产信息时, 都是首先进入交易房产信息的主页, 然后根据要交易的房产的地理位置, 以及自己所能承受的价格等因素进行进一步细化搜索, 本质来说就是对统一资源定位(url)的不断转变。主要分以下几步:

a. 通过判断用户输入的地区赋值给 part 变量传入值来返回



图 1 一般房屋交易网站网页结构

地区的 url。

b. 获得当前搜索结果总页数。

c. 利用 BeautifulSoup 解析 html 文档。

3.2 解析网页数据

通过 showDetail 函数对输入不同地区对应的 url 进行解析并抓取其中所需的有效信息, 并返回一个存有所需新的列表^[7], 主要步骤如下:

a. 输入 url。

b. `r=requests.get(url)# 获取网页请求。`

c. `soup=BeautifulSoup(r.text,"html.parser")# 解析网页。`

d. `tags=find_all(所需获取的标签名或属性, 要求能唯一标识一组或一个标签)# 如果不唯一则返回一个存有所需满足条件的列表。`

e. `print(tags.text)# 输出标签的内容。`

3.3 将网页数据储存并以 csv 格式输出

下面是对网页数据进行保存的部分代码:

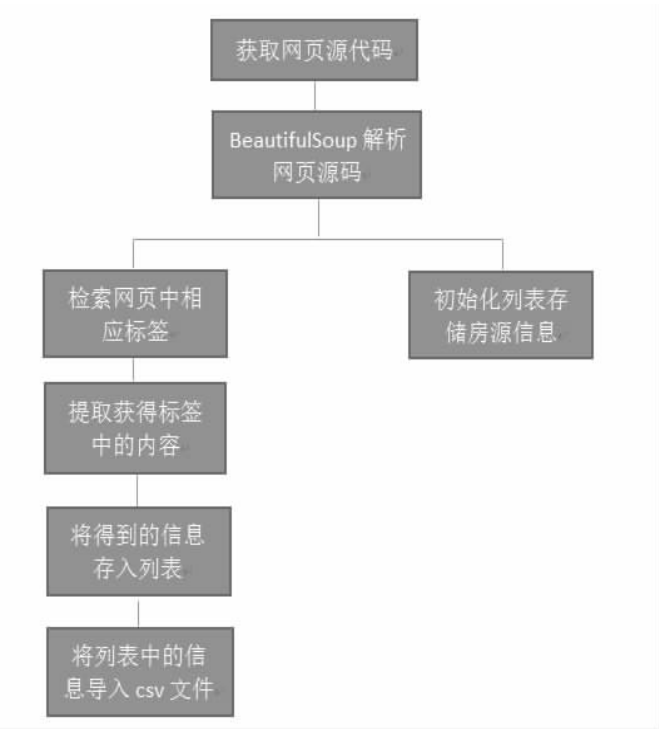


图 2 实时获取网页信息流程

	A	B	C	D	E
1	区域:	宝山			
2	位置	厅室	大小	成交日期	成交价格(万)
3	淞南二村	2室1厅	55.9平	2017.02.15	260
4	绿地领海(商业类)	1室1厅	51.38平	2017.02.15	120
5	月浦八村	2室1厅	51.52平	2017.02.15	190
6	旭辉城	2室1厅	73.5平	2017.02.15	206
7	四元路20弄	2室1厅	71.39平	2017.02.15	260
8	馨佳园(菊泉街1398弄)	2室2厅	76.42平	2017.02.15	300
9	枫桥湾名邸	3室3厅	93.09平	2017.02.14	620
10	通河八村	2室1厅	60.84平	2017.02.14	303
11	朗诗绿岛(公寓)	2室2厅	89.06平	2017.02.14	330
12	当代高邸	2室2厅	68.47平	2017.02.14	375
13	月浦九村	2室1厅	46.78平	2017.02.14	166
14	泗塘七村	2室1厅	65.11平	2017.02.13	281
15	绿地领海(商业类)	2室1厅	50.9平	2017.02.13	110

图 3 获取成交的二手房信息

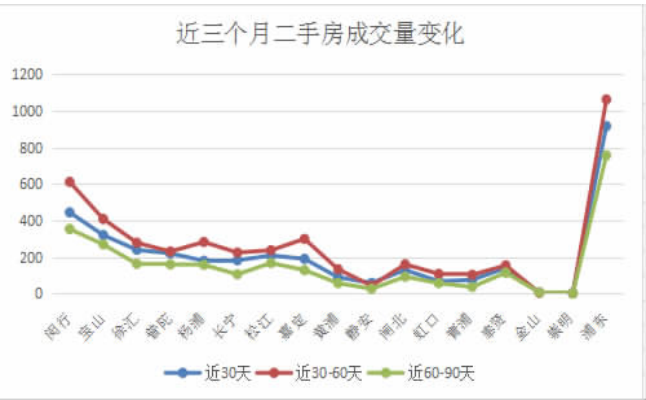


图 4 上海各区近三个月二手房成交量变化

```
# 将数据存储入 csv 文件
with open(csv 文件名, 'a+', newline='') as csvfile:
    writer = csv.writer(csvfile) # 初始化 writer 用来
    对 csv 文件进行写入
    writer.writerow(['区域:', part.__str__]) # 在 csv
```

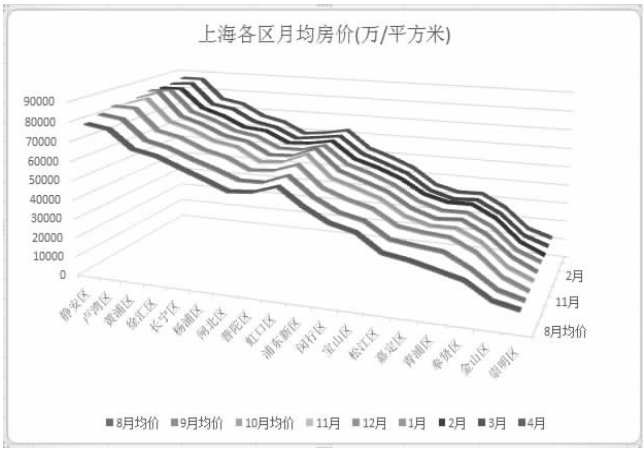


图 5 上海各区房价走势情况图



图 6 近一个月成交房中户型分配

```
文件中写入一行
writer.writerow(['位置', '厅室', '大小', '成交日期', '成交价格(万)'])
# 将 data 中的每一个元组写入 csv 文件
for item in data:
    writer.writerow(item)
```

3.4 输出数据格式

输出数据格式如图 3 所示。

4 实验结果分析

本文选取了链家网站的房屋交易数据作为实验样例。并对从 2016 年 8 月到 2017 年 4 月的数据进行统计(详见图 5),本文根据契税交纳比率的不同将户型分为大、中、小户型。90 平方米以内小户型契税交纳比率是 1%。90-144 平方米中户型契税交纳比率 1.5%。144 平方米以上大户型契税交纳比率是 3%。如图 6 所示为近 30 天来上海各区成交户型数据统计。由图 4 可以看出近 30 天来二手房成交数量下降 13.75%,与房价下跌的 2014 年同期成交量水平相近。由于政府对一手房开盘限价、去年 10 月“沪六条”的颁布以及上海市住建委发布了《关于开展商业办公项目清理整顿工作的意见》等多重因素作用下,降价预期扩大到二手房市场,买家出手趋缓,然而近期看房人数相较以往并没有太明显的下降趋势,说明二手房市场依旧活跃,大部分人依然持观望态度^[8]。

至 2016 年末,全市常住人口总数为 2419.70 万人。其中,户籍常住人口 1439.50 万人,外来常住人口 980.20 万人^[9]。2016 年,上海常住人口仅增加了 4.7 万人,其中,户籍常住 (转下页)

基于互联网技术的“e 家农”智能种菜系统设计

马玉志

(哈尔滨信息工程学院,黑龙江 哈尔滨 150000)

摘要:近年来随着媒体对各种食品的屡次曝光,食品的安全问题已昭然若揭。辐射、空气污浊、土壤腐蚀、居室甲醛超标等成为影响都市人健康的隐形杀手,然而楼宇林立与土色田园难以共存。都市人习惯了精致的自动化生活,田间的劳作早已离我们千里之遥。在享受都市繁华的同时,“健康”是都市人永恒不变的主题。因此 e 家农模块化智能种菜机应运而生。通过 e 家农模块化种菜机种植绿色蔬菜既可以自给自足又打破以往种植限制。培养花草可以丰富生活,种植植物可以通过光合作用吸收二氧化碳,净化我们空气环境。养殖花草可以让空气中的阴离子聚集,使得空气变得清新,有些植物还可以吸收空气中的有害气体,如今养殖绿色植物已被更多的人喜爱。城市中生活的人们没有充足的条件来种植蔬菜和花草,地点是一大限制,没有足够的空间来养殖植物,更没有充足的时间来照顾植物。为了解决在现有的基础设施上开发自家小菜园或小花园的问题,对此开发了 e 家农智能种植设备,能够充分的解决在楼房阳台及任何大小的空间上来开辟私人农场。实现自己种植自己食用,既安全又放心,为业余生活增加更多的色彩。e 家农模块化智能种菜机从始至终坚持科学健康的绿色种植方式。

关键词:e 家农;智能种菜;系统设计

1 设计背景及国内外研究现状

现如今已经衍生出智能花盆等种植技术,就目前来讲局限性较大,功能偏少,实用性小。本设计开发的 e 家农模块化智能种菜机可以通过连接手机 App(菜圈)终端来实现实时监测各项数据、达到远程监视、实时控制的效果。人们还可以通过数据库里提供的数据来进行科学的种植,使产量达到最大化。e 家农模块化智能种菜机采取模块化控制,需求者可根据空间的要求选装成模块。互联网上通过本产品设计的手机 App(菜圈)上实现附近的菜友交流,蔬菜的互换、买卖等功能,避免蔬菜的浪费和种植种类的局限性,智能室内种植机是当今世界家居和办公室种植花卉的新宠,主要通过传感器对植物的各生长环境参数进行了监测,例如,湿度、光照、肥料等,把采集到的生长环境数据和预先设定好的生长环境数据对比,若生长环境因素与设定值不同,系统会自动会定时定量的对花卉进行浇水,对其进行合理光照以及合理施肥等。随着科技的发展,智能家用市场已经做大做强,使得中国成为了全球各大生产智能家用设备厂商逐鹿

的大战场,家居设备的研发与测试技术也有了比较迅速的发展。智能种植机自然也包括在其中,近年来,国内外出现了许多种智能种植机,其功能与控制原理也有很多不同之处,但根本目的都是为了方便人们的日常生活和保证果蔬花草的正常生长。

2 “e 家农”智能种菜系统的设计思路

本产品理念基于简单生活、新鲜健康、追求品质生活的理念,因此设计该产品,解决社会需求,促进行业发展。本产品主要依靠传感器搭建、ARM 核心控制、语音控制系统、WIFI 透传、App 的上位机控制及大数据的应用,能够在网络平台建立起一个资源共享平台把使用此项目的消费者们联系起来,做到种植物的资源共享。通过 ARM 建立起中央控制系统,分别对温湿度、PH 浓度、电机控制、光照等参数,进行实时监测。当温度过高时,传感器会向 ARM 发送数据,中央控制系统作出控温操作,当 PH 浓度过低时,将会通过电机加压,抽取培养器皿中相应的培养液进行填补,当温度过低或光合作用时间 (转下页)

人口增加 6 万人,上海市的预期目标也是将人口控制在 2500 万以内^[9]。这说明未来房屋建设将会放缓,房价将会趋于稳定。

5 结论

通过上述整理分析,结果显示二手房价格总体呈上升减缓的趋势,近期二手房成交数量有明显的下降趋势,其中大部分交易来自于小户型。上海各区中整体二手房交易较为火爆的为闵行区 and 浦东新区,显然在地铁交通方便的条件下,人们会选择房价较为平均的地段买房。因此,随着政府宏观的房价调控政策的加强,买家会持更加谨慎的态度,未来上海房价不会有太大的变化波动,依然持稳中有升的走势。

参考文献

- [1]王萃梓,赖雯洁.基于房产交易网站的数据获取与在线工具开发[J].计算机技术与发展,2017(4).
- [2]Yu H. China's house price: Affected by economic fundamentals or real estate policy?[J]. Frontiers of Economics in China, 2010, 5(1):25-51.
- [3]张雨,顾俊,吉梨霞,刘岩,宋洁洁,杨静.关于二手房房价预测的分析研究[J].家饰,2016(5):188-190

- [4]Bork L, Moller S V. Housing Price Forecastability: A Factor Analysis[J]. Social Science Electronic Publishing, 2012.
- [5]Yu W, Yang C. Forecasting house prices using dynamic model averaging approach: Evidence from China [J].Economic Modelling, 2017, 61:147-155.
- [6]Paz P T D L. DETERMINANT OF HOUSING PRICES IN SPANISH'S URBAN AREAS [J].Paloma Taltavull De La Paz, 2000:1-19
- [7]嵩天,黄天羽,礼欣.程序设计基础:Python 语言[J].2014.
- [8]Mahalik M K. Factors determining regional housing prices: evidence from major cities in India [J].Journal of Property Research, 2015, 32(2):123-146.
- [9]上海市统计局,国家统计局上海调查总队.2016 年上海市国民经济和社会发展统计公报[J].统计科学与实践,2017-3-9(5).
- [10]上海市统计局.2015 年上海市国民经济和社会发展统计公报[J].统计科学与实践,2016(3):19-28.

通讯作者:张倩(1983-),女,博士,研究方向:视频、图像处理。