

基于AHP模糊分析法与随机森林的消费券满意度研究

蔡施宇, 许译之

杭州电子科技大学, 浙江 杭州

收稿日期: 2022年8月29日; 录用日期: 2022年9月19日; 发布日期: 2022年10月8日

摘要

为应对疫情冲击, 政府颁布了发放消费券的政策。本文正是围绕“消费券政策”展开, 通过线上线下发问卷以及网络爬虫收集数据做了初步分析, 利用随机森林对影响满意度的各项权重的重要性作综合分析。然后通过将AHP层次分析法和数学模糊分析法结合的AHP模糊分析法分析民众的整体满意度。通过随机森林可以得出优惠力度大、消费券宣传到位以及使用门槛低是影响满意度最重要的因素, 不同因素的影响力有着显著差别。进一步通过AHP模糊分析法得出结论: 目前过半杭州居民总体对消费券感到满意, 仅有极少数感到不满。

关键词

消费券, 民众满意度, AHP模糊分析法, 随机森林

Research on Consumer Coupon Satisfaction Based on AHP Fuzzy Analysis Method and Random Forest

Shiyu Cai, Yizhi Xu

Hangzhou Dianzi University, Hangzhou Zhejiang

Received: Aug. 29th, 2022; accepted: Sep. 19th, 2022; published: Oct. 8th, 2022

Abstract

In response to the impact of the epidemic, the government issued a policy of issuing consumer vouchers. This article is based on the “consumer coupon policy”, through the online and offline

distribution of questionnaires and the data collected by web crawlers to make a preliminary analysis, using random forest to make a comprehensive analysis of the importance of various weights affecting satisfaction. Then, the public's overall satisfaction is analyzed by AHP fuzzy analysis method which combines AHP analysis hierarchy process and mathematical fuzzy analysis method. Through random forest, it can be concluded that the most important factors affecting satisfaction are large discounts, well-publicized consumer coupons, and low thresholds for use, and the influence of different factors is significantly different. Further through the AHP fuzzy analysis method, it is concluded that more than half of Hangzhou residents are generally satisfied with consumer coupons, and only a few are dissatisfied.

Keywords

Consumer Coupons, Public Satisfaction, AHP Fuzzy Analysis Method, Random Forest

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

2020年初, 突如其来的新冠疫情使经济社会活动被迫陷入困境, 根据国家统计局初步核算, 我国一季度GDP约20万亿元, 同比下降了6.8%。今年一季度, 浙江省实现社会消费品零售总额5334亿元, 同比下降(下同)14.7%; 其中商品零售下降11.2%、餐饮收入下降42.3%。居家隔离制度使中小型企业亏损惨重[1]。

为激活消费市场发展, 加快提振消费信心, 据不完全统计, 从3月起有近50个城市推出消费券政策, 主要涉及餐饮和旅游业。消费券的发放充分刺激了消费, 起到了惠企惠民的作用。但消费券的发放也并不是十全十美的, 消费券目前仍具有有入门难、抢券难、用券难、难持久问题[2]。针对杭州市的消费券政策, 本文认为其刺激经济的作用虽然广度较大, 但是深度不足, 各方面也存在一定问题, 如消费券发放数量较少、使用有效期较短、面额较小难以刺激大额消费等, 这都是政府亟需解决的现实问题[3]。因此, 实现了解杭州市居民对消费券政策的认知度、参与度和满意度, 实现消费券刺激消费、消费带动生产、生产提高就业、就业促进消费的良性循环[4], 对于杭州政府以及全国各地的经济复苏都有着极其深远的现实意义。

在此背景下, 将实地调查所得到的问卷及访谈回馈作为基础, 通过随机森林与AHP模糊分析法对所获结果进行深度探讨, 据此得出影响消费券实施满意度各项因素之间的重要程度以及居民本身对消费券使用满意度的结论。

2. 数据获取与初步分析

2.1. 问卷调查内容

为了充分了解居民对于消费券的使用情况, 以及不同行业、不同社会背景的居民对杭州市消费券的满意度, 我们决定通过问卷的方式展开调查, 按比例分配原则确定了以西湖、余杭区为主的行政区域, 并在其中根据各行政区常住居民占总体比例抽取街道, 包括文新街道, 乔司街道、良渚街道、仁和街道等, 线下走访向当地居民等发放纸质及电子问卷, 同时也让居住于杭州的朋友帮助传播问卷, 让更多的

杭州居民参与问卷调查。通过分析通过问卷收集到的数据来分析影响居民消费券使用满意度的因素, 得出相应结论, 为杭州市消费券的改进提供一些参考意见。问卷内容提纲如表 1 所示:

Table 1. Outline of the content of the questionnaire

表 1. 问卷内容提纲

主维度	子维度
基本信息	性别
	年龄
	文化程度
	职业
	月收入
	婚姻状况
	是否为本地户籍
民众认知	是否了解过 了解渠道
参与情况	参加情况 领取情况 使用情况
民众意愿	是否增进消费欲望 理想申领方式
满意度	数量、有效期、门槛等

2.2. 访谈调查内容

为了更好地了解杭州市居民消费券领取和使用及满意度的情况, 我们还对走访街道的商家和消费者进行了一对一的访谈, 包括仁和街道某连锁餐饮店经理、乔司街道某大型家居商场经理等。访谈提纲如表 2 所示:

Table 2. Standard test system results data

表 2. 标准试验系统结果数据

访谈对象	访谈内容
商家	基本信息
	疫情对店内经营的影响
	面对疫情的影响做出的应对措施
	消费券的发放对生意的影响
	使用消费券的群体的情况
	使用消费券时遇到的问题
消费者	对于消费券发放的意见和建议
	基本信息
	领取消费券的情况
	使用消费券的情况
	身边的人使用消费券的情况
	消费券对于自身消费的影响
	在使用消费券时遇到的问题
对于消费券发放的意见和建议	

2.3. 网络数据爬取

我们通过网络爬虫爬取了有关消费券的网络评论。运用 Python 撰写爬虫程序, 在微博、知乎上根据正则表达式匹配消费券相关文章并爬取记录文章下的评论。然后进行数据预处理, 取出爬虫文件中的评论号, 评论内容和评论人, 进行精确分词后通过自定义的去词库去除掉无用、与消费券无关的评论, 过滤掉“可以”、“一个”、“就是”等无关的常见词, 合并含着相同或相近含义的词汇, 对处理后的网络评论进行词频分析, 得到排名前 45 的高频词汇, 结果如下(表 3):

Table 3. High-frequency vocabulary list

表 3. 高频词汇表

排序	词语	频数	排序	词语	频数	排序	词语	频数
1	消费	2430	16	领取	170	31	新闻	119
2	券	2141	17	双节	158	32	设计	119
3	发放	614	18	免费	156	33	亿元	117
4	2020	463	19	支付	150	34	商家	115
5	海报	394	20	平台	143	35	招聘	114
6	中秋	387	21	新	141	36	看中	114
7	电子	380	22	使用	137	37	电影	113
8	活动	369	23	在线	133	38	信息	112
9	万元	266	24	文化	130	39	小时	104
10	优惠	235	25	全国	130	40	餐饮	99
11	旅游	192	26	支付宝	128	41	惠民	96
12	国庆	186	27	本地	126	42	自愿	95
13	抢	183	28	新闻	119	43	数字	95
14	政府	178	29	使用	137	44	购物	95
15	寒露	171	30	生成	120	45	经济	94

“优惠”、“万张”、“活动”等词排在前列, 居民对于消费券活动兴趣高涨。在高频词汇中, “万元”、“活动”等词分别被提到了 266 次、369 次, 反映网友对于消费券发放活动的金额和活动内容非常感兴趣, 消费券的活动规模和活动金额得到了网友的广泛关注和讨论。“优惠”、“免费”、“惠民”等词分别被提到了 235 次、156 次、96 次。说明消费券的优惠力度具有非常高的讨论热度, 从“惠民”一词的热度来看, 网友对于消费券的优惠力度普遍持肯定态度, 认为消费券起到了惠民的作用。

“抢”、“领取”、“信息”等词被多次提到, 居民对于消费券领取方式十分关注。在高频词汇中, “抢”、“领取”、“信息”等词分别被提到 183 次、170 次、112 次, 体现网友对于消费券领取信息的关注度。“平台”、“支付宝”、“商家”等词分别被提到 143 次、128 次、115 次, 可以说明杭州消费券领取方式的广泛传播和网民对于消费券领取方式的热烈讨论。

多种多样的使用方式, 消费券促进多个产业的消费复苏。在高频词汇中, “旅游”、“电影”、“餐饮”、“购物”等词分别被提到 192 次、113 次、99 次、95 次, 排在高频词汇的前列。居民通过多种多样的方式使用消费券, 使消费券在不同领域的经济复苏中都起到了重要的作用。“餐饮”、“购物”等词被网友多次提到, 反映了消费券影响了很大部分网民的日常生活消费, 也反映了消费券对于餐饮行业

等的振兴作用, 也说明我们对消费券使用满意度的分析具有十分积极的现实意义。

2.4. 数据清洗

我们将所有获取的问卷数据等在整理后都进行了分析前的预处理, 运用 R 语言程序的数据清洗模块去除了内容有缺失的数据以及少量由于操作失误造成的重复数据。

3. 随机森林

这里选取了满意度量表中的消费券宣传到位、发放数量合理、单张金额设定合理、有效期合理、使用门槛低、使用商家多、优惠力度大、使用限制少、获取方式方便、使用方式简单共十个变量, 并将各个变量命名为 X1 至 X10, 进行随机森林回归分析。我们将每一个变量作为一个节点, 将这些节点串联在一起, 当 *Gini* 系数最小的时候, 就生成所需要的决策树[5], 树的数量为 1000 棵。随机森林输出结果如表 4 所示:

Table 4. Random forest result
表 4. 随机森林结果

变量	%IncMSE	IncNodePurity	Mean of squared residuals	%Var explained
消费券宣传到位 X1	26.39779	3.0423287		
发放数量合理 X2	22.43195	2.5155030		
单张金额设定合理 X3	19.83544	1.0836219		
有效期合理 X4	18.38317	1.6754199		
使用门槛低 X5	24.46095	3.0343412	0.02993734	96.79
使用商家多 X6	13.81305	0.8741188		
优惠力度大 X7	27.61496	3.1487245		
使用限制少 X8	20.85446	2.6479208		
获取方式方便 X9	17.86973	2.2267485		
使用方式简单 X10	18.17565	1.5335903		

由表 4 可知, 解释的方差为 96.79%, 说明该随机森林分析能够很好的解释各变量的重要性。IncMSE 代表的是 MSE 的增加值, 也被称为均方误差, 即对这十个变量中的任意一个变量随机给予一个值, 预测的结果相比之前结果的变化程度, 也就是说, 如果这个变量对结果的影响力越大, 那么预测结果的变化程度也就越大, 体现在均方误差的变大, 均方误差的值也就成了变量的重要程度的量化指标。IncNodePurity 是衡量自变量对因变量的重要程度, 该值越大表示此变量越重要[6]。由表 4 可知各个变量的 IncMSE 和 IncNodePurity 值不同, IncMSE 最大值为 27.615, 最小值为 13.813; IncNodePurity 最大值为 3.148, 最小值为 0.874, 说明不同指标的变量重要性差距较大。

这里各项指标的 IncMSE 和 IncNodePurity 值的趋势基本相同因此我们选取 IncMSE 的值作为标准。在表 4 中可以看到, X7 (优惠力度大)的 IncMSE 值为 27.61496、X1 (消费券宣传到位)的 IncMSE 值为 26.39779、X5 (使用门槛低)的 IncMSE 值为 24.46095, 这三个自变量是影响杭州市居民对消费券政策满意度比较重要的因素, 其中 X7 (优惠力度大)是最重要的因素。

4. AHP 模糊分析法

AHP 模糊分析法是将 AHP 层析分析法与数学的模糊分析法相结合的方法, 利用层析分析法得出权重, 在模糊分析法中利用上一步得出的权重, 运用模糊变化法得到最终的满意度综合评价向量结果。

4.1. 建立层次结构模型

在此模型中, 对消费券的满意度进行综合评价是总目标。我们将评价指标集分为了两个层次, 第一个层次是总目标集 $U = (U_1, U_2, U_3)$; 第二层是子目标因素集, $U_1 = (E_5, E_6, E_8, E_{10}, E_9)$, $U_2 = (E_2, E_3, E_4, E_7)$, $U_3 = (E_1)$ 。评价体系指标如表 5 所示:

Table 5. Evaluation index table

表 5. 评价指标表

总目标	一级指标	二级指标
消费券满意度综合评价 A	消费券使用便利性 U_1	门槛低 E_5
		使用商家多 E_6
		使用限制少 E_8
		获取方式方便 E_{10}
		使用方式简单 E_9
	消费券设置合理性 U_2	发放数量合理 E_2
		单张金额设定合理 E_3
		有效期合理 E_4
		优惠力度大 E_7
	消费券宣传力度 U_3	消费券宣传到位 E_1

4.2. 确定模糊评价集

基于商家和消费者的满意度, 我们决定将最终的评语分为五个级别。评价集为: $v = (v_1, v_2, v_3, v_4, v_5) = \{\text{很满意, 满意, 一般, 不满意, 很不满意}\}$ 。我们最终将通过计算得到一个向量, 将其对应于评价集则可判断出消费者们占不同满意程度的比例, 进而分析得出总体的满意程度[7]。

4.3. 建立评判矩阵及确定权重

这里参照之前随机森林分析所得因素重要性, 给出相应评分。可用评价模型上一层的一个因素(j)作为每一层所含因素(i)的比较准则, 相互之间比较重要程度, 因为将重要程度分为了 9 个层次, 从极端重要到极端不重要, 因此决定使用 1, 3, 5, 7, 9 以及它们的倒数的标度方法将重要性判断结果给量化以方便计算[8]。不同标度对应的含义如表 6 所示:

Table 6. Scale meaning table

表 6. 标度含义表

序号	重要性等级	C_{ij}
1	i, j 两元素同等重要	1
2	i 元素比 j 元素稍重要	3

Continued

3	i 元素比 j 因素明显重要	5
4	i 元素比 j 元素强烈重要	7
5	i 元素比 j 元素极端重要	9
6	i 元素比 j 元素稍不重要	1/3
7	i 元素比 j 元素明显不重要	1/5
8	i 元素比 j 元素强烈不重要	1/7
9	i 元素比 j 元素极端不重要	1/9

两两比较一级目标得出如下的判断矩阵(图 1):

A	U_1	U_2	U_3
U_1	1	$\frac{1}{3}$	3
U_2	3	1	5
U_3	$\frac{1}{3}$	$\frac{1}{5}$	1

Figure 1. Judgment Matrix
图 1. 判断矩阵

显然可以获得相应的特征方程 $A\omega = \lambda\omega$, 求解特征方程后得到解向量 ω , 权重向量则是要经过归一化后得出。同时对每一层作单排序时还需作一致性检验以防止出现计算错误。判断数据是否通过了一致性检验的方法是计算一致性比例, 其公式如下:

$$CR = CI/RI$$

$$CI = \frac{\lambda_{\max} - n}{n - 1}$$

其中, 判断矩阵 A 的阶数为 n , λ_{\max} 为特征方程的最大特征根, CR 是一致性比例, CI 是一致性指标, RI 则为平均随机一致性指标。因此, RI 的取值如表 7 所示:

Table 7. RI value

表 7. RI 值

n	1	2	3	4	5	6	7	8
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41

判断该层次单排序结果通过一致性检验的方式是 $CR < 0.10$, 若不然这需要修改判断矩阵的值。判断矩阵 A 的计算结果如表 8 所示:

Table 8. Matrix A calculation table

表 8. 矩阵 A 计算表

U	U_1	U_2	U_3	ω	一致性检验
U_1	1	1/3	3	0.26	
U_2	3	1	5	0.64	0.0216 < 0.10
U_3	1/3	1/5	1	0.11	

通过上表可以得到一级指标的一致性检验结果 $CR = 0.0216 < 0.10$, 证明判断矩阵是具有满意的一致性。我们进一步将特征向量归一化处理, 可以得到权重向量为 “ $\omega = (0.26, 0.63, 0.11)$ ”。通过同样的方法可得到二级指标的判断矩阵及其权重向量, 结果如表 9 和表 10 所示:

Table 9. Secondary Indicator Matrix

表 9. 二级指标矩阵

E	E_5	E_6	E_8	E_9	E_{10}	ω	一致性检验
E_5	1	7	3	5	5	0.50	
E_6	1/7	1	1/5	1/3	1/3	0.05	
E_8	1/3	5	1	3	3	0.25	0.036 < 0.10
E_9	1/5	3	1/3	1	1	0.10	
E_{10}	1/5	3	1/3	1	1	0.10	

Table 10. E_2, E_3, E_4, E_7 judgment matrix and its vector

表 10. E_2, E_3, E_4, E_7 判断矩阵及其向量

E	E_2	E_3	E_4	E_7	ω	一致性检验
E_2	1	3	3	1/3	0.29	
E_3	1/3	1	1	1/5	0.10	0.016 < 0.10
E_4	1/3	1	1	1/5	0.10	
E_7	3	5	5	1	0.56	

4.4. 评判矩阵

由表 11 构建模糊评判矩阵为:

Table 11. Evaluation result statistics table

表 11. 评价结果统计表

属性指标	非常满意 5 分	较满意 4 分	一般 3 分	较不满意 2 分	很不满意 1 分
门槛低 E_5	160	79	67	33	2
使用商家多 E_6	203	70	53	13	2
使用限制少 E_8	164	92	66	17	2
获取方式方便 E_{10}	177	96	48	15	5
使用方式简单 E_9	161	101	64	15	0
发放数量合理 E_2	55	59	150	52	25
单张金额设定合理 E_3	55	72	143	58	12
有效期合理 E_4	30	152	72	73	14
优惠力度大 E_7	99	49	120	70	3
消费券宣传到位 E_1	208	78	40	14	1

$$R_1 = \begin{bmatrix} 0.47 & 0.30 & 0.19 & 0.04 & 0 \\ 0.16 & 0.17 & 0.44 & 0.15 & 0.07 \\ 0.16 & 0.21 & 0.42 & 0.17 & 0.04 \\ 0.09 & 0.45 & 0.21 & 0.21 & 0.04 \\ 0.47 & 0.23 & 0.20 & 0.10 & 0.01 \end{bmatrix}$$

$$R_2 = \begin{bmatrix} 0.60 & 0.21 & 0.16 & 0.04 & 0.01 \\ 0.30 & 0.14 & 0.35 & 0.21 & 0.01 \\ 0.48 & 0.27 & 0.19 & 0.05 & 0.01 \\ 0.52 & 0.28 & 0.14 & 0.04 & 0.01 \end{bmatrix}$$

$$R_3 = [0.61 \quad 0.23 \quad 0.12 \quad 0.04 \quad 0]$$

接下来进行模糊变换, 就需要一个模糊算子“ \circ ”, 这里挑选“ $M(\cdot, \oplus)$ ”作为模糊变换的算子, 因为这是一种加权平均型算子, 有很强的综合性, 能充分利用判断矩阵的信息, 分析出足够全面的结果[9]。其对应公式为:

$$b_k = \min \left(1, \sum_{j=1}^m \omega_j R_j \right), k = 1, 2, \dots, n$$

将数据带入上式我们得到:

$$b_1 = \omega_j \circ R_j = \min \left(1, \sum_{j=1}^m \omega_j R_j \right)$$

$$= (0.50, 0.05, 0.25, 0.10, 0.10) \circ \begin{bmatrix} 0.47 & 0.30 & 0.19 & 0.04 & 0 \\ 0.16 & 0.17 & 0.44 & 0.15 & 0.07 \\ 0.16 & 0.21 & 0.42 & 0.17 & 0.04 \\ 0.09 & 0.45 & 0.21 & 0.21 & 0.04 \\ 0.47 & 0.23 & 0.20 & 0.10 & 0.01 \end{bmatrix}$$

$$= (0.34, 0.28, 0.26, 0.10, 0.02)$$

同理可得:

$$b_2 = (0.54, 0.26, 0.18, 0.06, 0.01)$$

$$b_3 = (0.61, 0.23, 0.12, 0.04, 0)$$

至此, 通过层次分析法求得的权重, 模糊分析法求得的评判矩阵以及一级指标的数据, 得到二级指标(b_1, b_2, b_3)的满意度综合评判。接下来依据二级指标的权重, 运用相同的模糊变化法便可得到最终的满意度综合评价向量。满意度综合评价向量:

$$A = \omega \circ R = \omega \circ \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = (0.50, 0.26, 0.20, 0.07, 0.01)$$

对原式进行归一化后得到: “ $A = (0.50, 0.25, 0.19, 0.07, 0.01)$ ”, 以上评价结果表明, 在被调查的顾客中, 有一半的杭州市居民都感到非常满意, 只有极少数感到不满。

5. 总结与分析

通过随机森林我们得到了优惠力度大、消费券宣传到位以及使用门槛低是影响满意度最重要的因素, 不同因素的影响力有着显著差别, 因此在改进消费券时应更加关注这三个方面, 做到合理规划金额, 扩大宣传影响力以及适当降低使用门槛。进一步通过 AHP 模糊分析法可以得到, 目前过半杭州居民总体对消费券感到满意, 仅有个别感到不满, 消费券的总体实施情况较为成功, 但仍需要进一步改进, 进行更严密审慎的规则设计, 进一步刺激居民消费, 实现经济复苏。

参考文献

- [1] 陈芳芳, 朱德超. 利用数字消费券迅速激活消费市场[J]. 政策瞭望, 2020(4): 27-28.
- [2] 刘颖, 马龙. 贵州贵阳消费券政策实施情况调查[J]. 中国国情国力, 2020(7): 65-67.
- [3] 高丽. 杭州消费券发放效果研究[J]. 黑龙江对外经贸, 2011(10): 66-68.
- [4] 多地发放消费券, 能否“券”动消费? [J]. 中国眼镜科技杂志, 2020(6): 70-74.
- [5] 唐大伟. 基于随机森林的日径流量预测模型及其 R 语言实现研究[J]. 黑龙江水利科技, 2019, 47(12): 6-9.
- [6] 赵北庚. 基于 R 语言 randomForest 包的随机森林建模研究[J]. 计算机光盘软件与应用, 2015, 18(2): 152-153
- [7] 曹铭, 张成科. 基于 AHP 模糊综合评价餐饮行业满意度研究[J]. 价值工程, 2013(30): 146-147.
- [8] 王蕾, 鲍新中. 基于模糊层次分析和实物期权法的专利组合价值评估研究[J]. 中国发明与专利, 2022, 19(8): 5-14.
- [9] 王龙帅, 吴洁, 王莹, 周思宇, 鞠伟轶, 蒋军成, 邢志祥. 基于 AHP-模糊综合评价法的化工园区应急能力评估[J]. 工业安全与环保, 2022, 48(6): 1-5.