



武汉大学学报(信息科学版)

Geomatics and Information Science of Wuhan University

ISSN 1671-8860, CN 42-1676/TN

《武汉大学学报(信息科学版)》网络首发论文

题目: 基于用户生成内容的大型商场情感地标提取方法
作者: 朱利涛, 沈婕, 王兴, 侯盈旭, 张成
DOI: 10.13203/j.whugis20210488
收稿日期: 2022-07-12
网络首发日期: 2022-07-27
引用格式: 朱利涛, 沈婕, 王兴, 侯盈旭, 张成. 基于用户生成内容的大型商场情感地标提取方法[J/OL]. 武汉大学学报(信息科学版).
<https://doi.org/10.13203/j.whugis20210488>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

DOI:10.13203/j.whugis20210488

引用格式:

朱利涛, 沈婕, 王兴, 等. 基于用户生成内容的大型商场情感地标提取方法[J]. 武汉大学学报·信息科学版, 2022, DOI: 10.13203/j.whugis20210488 (ZHU Litao, SHEN Jie, WANG Xing, et al. Extraction of Emotional Landmarks in large malls Based on User-Generated Content[J]. *Geomatics and Information Science of Wuhan University*, 2022, DOI: 10.13203/j.whugis20210488)

基于用户生成内容的大型商场情感地标提取方法

朱利涛^{1,2}, 沈婕^{1,2,3}, 王兴^{1,2}, 侯盈旭^{1,2}, 张成⁴

1 南京师范大学虚拟地理环境教育部重点实验室, 南京, 210023

2 南京师范大学地理科学学院, 南京, 210023

3 江苏省地理信息资源开发与利用协同创新中心, 南京, 210023

4 北京师范大学地理科学学部, 北京, 100875

摘要: 用户情感对空间注意力、决策和记忆力具有重要影响。情感与地标关联, 可在增强用户认知地图能力的同时, 提高导航效率。当前研究侧重于情感地标在导航中的作用, 对其在复杂室内环境中的提取方法关注甚少。以大型商场为研究对象, 提出基于用户生成内容的情感地标显著度定量评价模型, 进而实现室内情感地标的自动化提取。首先, 利用网络爬虫技术获取某大型商场的用户评论数据; 其次, 基于 SnowNLP 对用户评论进行情感分析, 并将分析结果扩展至认知显著性度量体系中; 然后, 利用层次分析法 (Analytic Hierarchy Process, AHP) 和指标相关性的权重确定法 (Criteria Importance Through Intercriteria Correlation, CRITIC) 的组合赋值法计算显著度指标的权重, 构建情感地标显著性综合评价模型; 最后, 利用层次聚类算法提取分级地标, 依据分级地标设计符合用户认知的多尺度室内导航地图, 通过用户实验验证地标提取方法的可用性。本研究推动室内导航地图设计的标准化, 为室内智能化导航服务提供有益的补充。

关键词: 用户生成内容 (UGC); 大型商场; 情感地标; 地标显著度; 情感分析

中图分类号: P208

文献标识码: A

移动计算技术和社交网络服务的快速发展推动了基于位置的社交网络 (LBSN) 的显著增长, 例如 Foursquare、Yelp 和大众点评^[1]。LBSN 会产生大量面向位置的用户生成内容 (user-generated content, UGC), 其蕴含人类感知位置的信息, 如空间、时间、情感、体验等^[2,3], 为研究用户行为提供了契机。借助海量的 UGC 可构建用户对地标认知的知识库^[4,5], 扩充传统地标提取方法的数据源^[6-12], 从而实现大规模自动化的地标提取。基于 UGC 的地标提取往往能更准确地反映用户情感与行为。目前基于 UGC 提取地标

的方法大多针对室外环境, 针对复杂室内环境却鲜有探讨, 如何提取室内地标成为亟需解决的问题。

地标是指环境中显著的对象^[13]。一个实体能成为地标, 应在结构 (如重要的空间位置)、视觉 (如颜色、大小、形状)、认知 (如文化和历史影响) 方面具有显著特征^[14]。基于地标显著性定量模型计算是常用的地标提取方法之一, 其基本思想是为人类感知地标的建模, 能从成因上对地标认知机理进行科学合理地解释。经典的地标显著度评价方法是由 Raubal & Winter (2002) 提出针对建筑物立面计算其地标性 (landmarkness)^[15]。室

收稿日期: 2022-07-12

项目资助: 国家重点研发计划(2021YFE0112300); 国家自然科学基金(41871371)。

第一作者: 朱利涛, 博士生, 主要从事空间认知、室内行人导航地图。181301028@njnu.edu.cn

通讯作者: 沈婕, 博士, 教授。shenjie@njnu.edu.cn

内地标作为室内智能导航的重要线索,一些学者提出了相关定量模型,如朱海红等(2016)从视觉、空间、认知角度分析影响大型商场内地标显著性的因素^[16];Zhu et al.(2019)提出一种基于AHP和实例计算的室内地标显著性评价体系^[17];Hu et al.(2019)提出一种基于遗传编程算法和数据驱动的室内地标显著度定量模型^[18];另外受室内空间认知的启发,Lyu et al.(2015)构建预定义权重的室内地标显著度计算框架^[19];Dubey et al.(2019)提出一种基于分层多标准决策模型的室内地标显著度计算框架^[20]。但上述研究都几乎忽视用户情感对地标认知的影响^[21]。

寻路感知(如空间决策、空间注意力和记忆力)受用户情感影响^[22],如情感积极的用户倾向于更全面地感知环境,而情感消极的用户关注局部和细节^[23]。另外,情感对地标的影响包含两个方面:地标可以触发用户情感如幸福或忧伤,地标本身存在情感负担如墓地^[24]。空间认知领域的寻路研究表明,

与中性地标相比,积极或消极地标更能提高导航效率和记忆能力^[25,26]。但是目前如何在复杂室内环境中提取有效情感地标的方法研究仍处于空白状态^[27]。

综上所述,针对目前室内地标显著度定量模型很少考虑用户情感维度的问题,本研究旨在提出一种基于UGC的大型商场情感地标自动提取方法。首先,基于SnowNLP的用户评论数据情感分析;然后,将情感分析结果拓展至室内地标认知显著性中,提出情感地标的定量评价模型;最后,提出基于层次聚类算法的分级地标提取方法,并应用在室内导航地图多尺度表达中,利用用户评价实验验证方法的有效性。

1 大型商场情感地标提取方法

本文研究框架如图1所示。研究方法包括三部分:(1)用户评论的情感挖掘;(2)情感地标显著度评价指标体系构建;(3)基于AHP-CRITIC的指标权重计算;(4)基于层次聚类算法的分级地标提取与表达。

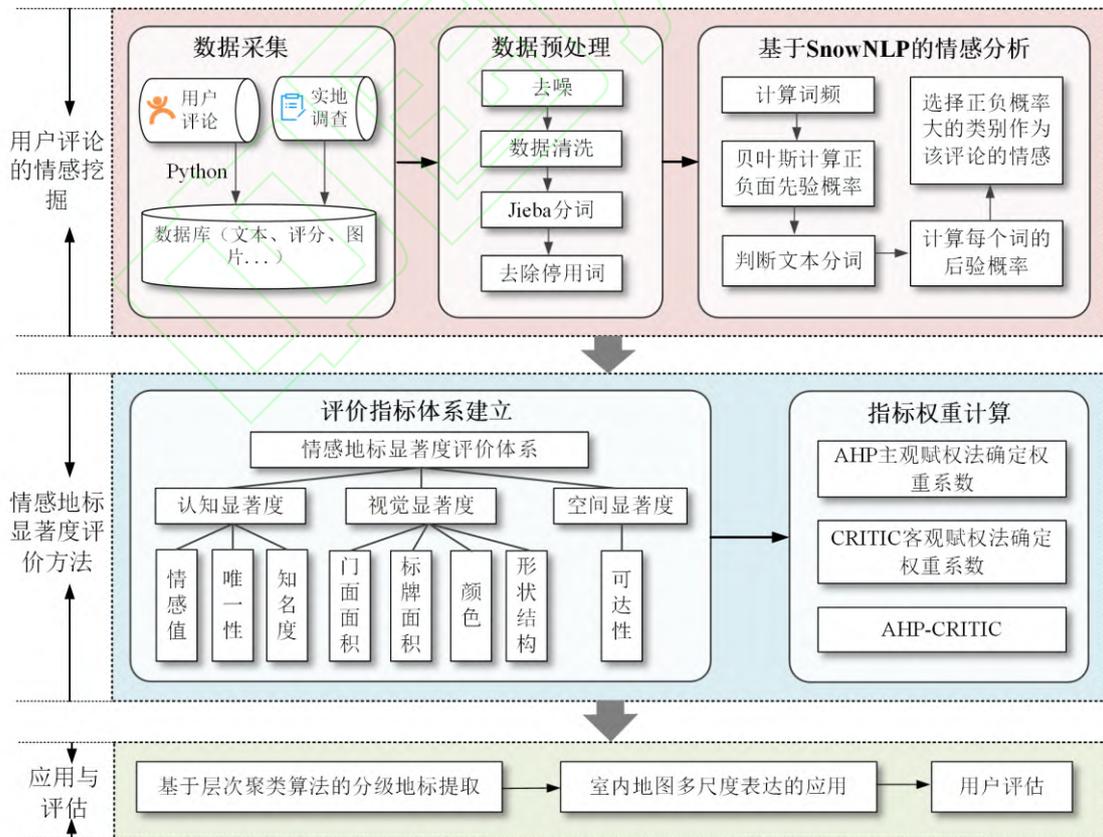


图1 基于UGC的大型商场用户情感地标提取方法

Fig.1 The extraction method for emotional landmarks for large malls based on UGC

1.1 用户评论的情感挖掘

1.1.1 数据获取与预处理

研究以“大众点评”(<https://www.dianping.com/>)用户在线评论为数据源,挖掘用户对大型商场室内感知。用户评论包含定量信息(如星级数、评分)和定性信息(如评论文本),是丰富的隐含信息来源^[27]。为保证所采集数据的现势性与完整度,在2022年某商场导览图的基础上,以商铺ID为基准,基于网络爬虫技术,筛选出2018到2022年的用户评论信息,共计3919条。用户评论的关键信息包括发表时间、评论数量、评论文本、评分等。

1.1.2 情感分析方法

$$P(\text{pos}|w_1, \dots, w_n) = P(w_1, \dots, w_n|\text{pos}) \times \frac{P(\text{pos})}{P(w_1, \dots, w_n)} \quad (1)$$

$$P(\text{neg}|w_1, \dots, w_n) = P(w_1, \dots, w_n|\text{neg}) \times \frac{P(\text{neg})}{P(w_1, \dots, w_n)} \quad (2)$$

$$P_{\text{sum}} = P(w_1, \dots, w_n|\text{pos}) \times P(\text{pos}) + P(w_1, \dots, w_n|\text{neg}) \times P(\text{neg}) \quad (3)$$

式中, $w = [w_1, w_2, \dots, w_n]$ 为特定单词序列,

P_{sum} 为该条评论情感值的计算结果, $P_{\text{sum}} \in [0,1]$; $P_{\text{sum}} < 0.5$ 为消极情感, $P_{\text{sum}} > 0.5$ 为积极情感。

1.2 情感地标显著度评价指标体系构建

本文从认知显著度,空间显著度,视觉显著度三个一级指标出发,构建基于UGC的大型商场情感地标显著性评价体系。

1.2.1 认知显著度

认知显著度 (S_c) 是指地标在社会文化的重要性的意义^[30]。衡量地标的认知吸引力包含三个二级指标:情感值、唯一性和知名度。

情感值 (X_1)。该指标表示用户对该地标的情感程度。为了更好地理解情感分析结果,将情感分数范围从 $[0,1]$ 转换为 $[-1,1]$ 。若情感值大于 0, 则认为该条评论的情感倾向

情感分析 (Sentiment Analysis) 是数据挖掘的重要领域之一,其定义是指利用自然语言处理(NLP)、文本分析、计算语言学和生物识别技术挖掘UGC中的情感倾向^[28]。SnowNLP是一个专门分析中文文本情感的Python库^[29],其建模原理是基于贝叶斯分类从0到1表达文本的情感程度,即判断正负句的概率。由于其包含购物类自然语言类型的情感字典,因此本文利用SnowNLP对大型购物中心的用户评论进行情感分析,其中SnowNLP的情感分析的步骤如下:计算词频,基于贝叶斯的正负先验概率计算,判断文本分词,计算每个词的后验概率和选择正负概率大的类别作为该评论的情感。其计算公式如下:

是积极的;反之则为消极。其计算公式如下:

$$X_1 = |P_{\text{sum}}| \quad (4)$$

式中,用户评论的情感得分 P_{sum} 的绝对值越大,表明情感越强烈, X_1 得分越大。

唯一性 (X_2)。该指标表示在区域范围内其名称不会被混淆或产生歧义。本研究利用布尔值对唯一性进行衡量, $X_2 \in \{T: 1, F: 0\}$ 。在数据库中检索地标名称,若同一商场内名称唯一为 True, 反之为 False。

知名度 (X_3)。该指标表示公众对其品牌认知和了解的程度、社会影响的广度和深度,根据用户评论数量进行衡量,用户评论数量越多,知名度数值越大。

1.2.2 视觉显著度

视觉显著度 (S_v) 表示地标在视觉上与邻近特征体对比的突出程度。衡量地标的视觉吸引力包含四个二级指标:门面面积、标牌面积、颜色和形状结构。

门面面积 (X_4)。该指标表示地标的立面可观察面积。门面面积明显大于其他周围

的地标更容易识别。本研究对门面面积的计算是通过门面的长度 (L) 乘以宽度 (W) 得到, 其计算公式如下:

$$\alpha = L * W \quad (5)$$

标牌面积 (X_5)。该指标表示地标上所粘附的标牌显目程度。人们很容易被具有宽阔入口或大标志的场所吸引^[15]。本研究对标牌面积的计算是通过标牌的长度 (l) 乘以宽度 (w) 得到, 其计算公式如下:

$$\beta = l * w \quad (6)$$

颜色 (X_6)。该指标是指物体与周围物体的色差。人们很容易被一个与周围环境的颜色形成鲜明对比的物体^[31]。基于 Hu 等人 (2019) 的方法^[18], 基于高维颜色变换方法将地标图像缩放为 256 级灰度, 每个像素分配到从 0 到 255 值。颜色越亮, 平均像素值 (γ) 越高。

形状结构 (X_7)。该指标是指物体与周围物体的形状规则对比。具有不规则形状的物体比其他规则形状的物体更突出。依据形状结构特征, 利用定序量表将其划分为三个等级: 封闭式是指独立存在于无密集地物遮挡的道路中间 ($X_7 = 3$); 半开放式是指商场内部用墙隔开的完整空间, 空间构造整齐有序, 处于主通道两边且容易识别 ($X_7 = 2$); 开放式大多位于次通道, 持久性较差, 且比较密集和混乱 ($X_7 = 1$)。

1.2.3 空间显著度

空间显著度 (S_s) 是指特定地标在空间环境中位置的重要程度, 包含可达性一个二级指标。

可达性 (X_8)。该指标表示了地标的容易到达的程度。依据空间可达性, 利用定序量表将其划分为三个等级: 位于出入口、自动扶梯、电梯附近 ($X_8 = 3$); 位于室内道路交汇口附近 ($X_8 = 2$); 位于室内中庭附近 ($X_8 = 1$)。

1.2.4 地标显著度的综合评价

为了消除量纲差异需对各指标进行标准化处理。情感地标显著度的三个一级指标认知显著度 (S_c)、视觉显著度 (S_v)、空间显著度 (S_s) 的计算公式如下:

$$\begin{cases} S_c = w_1 S_1 + w_2 S_2 + w_3 S_3 & (\\ S_v = w_4 S_4 + w_5 S_5 + & 7) \\ S_s = w_6 S_6 & \end{cases}$$

式中, S_1 、 S_2 、 S_3 、 S_4 、 S_5 、 S_6 、 S_7 、 S_8 分别表示情感值、唯一性、知名度、门面面积、标牌面积、颜色、形状结构、可达性二级指标得分的归一化值。 w_1 、 w_2 、 w_3 、 w_4 、 w_5 、 w_6 、 w_7 、 w_8 分别对应各三级指标的权重。

情感地标显著度的综合评价计算公式如下:

$$S_{sum} = w_v S_v + w_c S_c + w_s S_s \quad (8)$$

式中, w_c 、 w_v 、 w_s 分别表示认知显著度、视觉显著度、空间显著度三个一级指标的权重; 权重系数 $w_c, w_v, w_s > 0, \sum w = 1$ 。

1.3 基于 AHP-CRITIC 的指标权重计算

指标权重计算是评价指标不同重要程度的反映。指标权重值被用来反映指标的相对重要性, 指标越重要, 则赋予它的权重应越大, 反之则越小。确定指标权重的确定方法可以分为两类: 1) 主观赋权法, 它是根据决策者主观上对各指标的重要程度来确定权重的方法, 包含专家评判法、Delphi 法、层次分析法等; 2) 客观赋权法, 它根据指标值之间的相关关系或属性值的变异程度来确定权重, 包含熵值法、标准离差法、CRITIC 法等。主观赋权法和客观类赋权法各有优缺点^[32]。主观赋权法能直接反映评价目的, 将专家的知识 and 经验直接运用到指标赋权中, 但是也存在评价依赖专家知识体系以及专家偏好极端导致赋权偏离集体决策的不足, 评价结果的主观性较强; 客观赋权法具有赋权客观的优点, 但是该方法利用观察数据提供的信息来确定权重, 计算过程相对复杂, 个别情况下可能会得出与决策者的主观愿望相反的评价结果。为了改善主观赋权法和客观赋权法的不足, 本文提出一种 AHP 法和 CRITIC 法的组合赋权方法, 实现各指标权重系数的分配。

1.3.1 AHP 主观赋权法确定权重系数

层次分析法 (Analytic Hierarchy Process, AHP) 是一种主观赋权方法, 该方法将定性

与定量相结合的多目标决策分析方法。其赋权思路是对两两指标之间的重要程度做出比较判断，建立判断矩阵，通过计算判断矩阵的最大特征值以及对应特征向量，得出不同方案重要性程度的权重^[33]。其主要步骤如下。

步骤一，构建判断矩阵。依据所构建的评价指标体系，对各指标两两对比，建立判断矩阵 A 。

$$A = (a_{ij})_{n \times n} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \quad (9)$$

式中， a_{ij} 表示第 i 个因素与第 j 个因素的两两相对重要性量化数值。

本文采用专家调查问卷的形式，根据表1所示1-9标度进行各指标间重要性的判断。

表1 判断矩阵标度

Tab.1 Judgment matrix scale

标度	重要性	含义
1	同样重要	两因素对指定属性重要程度一样
3	略微重要	前因素比后因素略微重要
5	明显重要	前因素比后因素明显重要
7	强烈重要	前因素比后因素强烈重要
9	绝对重要	前因素比后因素绝对重要
标度 倒数	反向比较	因素 i 比因素 j 的重要性之比为 a_{ij} ，反向则因素 j 比因素 i 的重要性之比为 $1/a_{ij}$

步骤二，一致性检验。计算一致性比率 CR 值，若 $CR < 0.1$ ，则认为判断矩阵通过一致性检验。其计算公式如下：

$$CI = \frac{\lambda_{max} - n}{n - 1} \quad (10)$$

$$CR = CI/RI \quad (11)$$

式中， n 为判断矩阵的阶数， λ_{max} 为判断矩阵的最大特征根， CI 为一一致性指标， RI 为随机一致性指标。

步骤三，计算权重。对通过一致性检验的判断矩阵，以其归一化的特征向量作为权重向量 w_i ，计算公式如下：

$$w_i = \frac{1}{n} \sum_{j=1}^n \frac{a_{ij}}{\sum_{k=1}^n a_{kj}}, (i = 1, \dots, n) \quad (12)$$

1.3.2 CRITIC 客观赋权法确定权重系数

由于 AHP 主观赋权法存在一定主观性的弊端，本文引入一种客观赋权方法 CRITIC，该方法既考虑各指标自身信息量，还考虑指标之间相关性与变异性，使得计算结果更加地客观合理^[34]。其赋权思路是对比强度和指标间的冲突性两个方向为主。其主要步骤如下。

步骤一，指标标准化处理。由于指标量纲不同，需要对指标进行标准化处理。由于本研究所选取的指标均为正向指标，因此采用极大值法进行标准化处理，其计算公式如下：

$$Y_i = \frac{X_i - \min(X_i)}{\max(X_i) - \min(X_i)} \quad (13)$$

式中， Y_i 为指标 X_i 数据标准化后的量化值。

步骤二，计算指标 X_i 的信息量，其计算公式如下：

$$C_i = \theta_i \sum_{j=1}^n (1 - r_{ij}) \quad (14)$$

式中， C_i 为指标 X_i 的信息量， C_i 越大表明指标 X_i 所包含的信息量越大，指标 X_i 相对重要性也越大； θ_i 为指标 X_i 的标准差； n 为指标数； $\sum_{j=1}^n (1 - r_{ij})$ 为指标 X_i 和指标 X_j 的冲突性； r_{ij} 为评价指标 X_i 和指标 X_j 之间的相关系数。

步骤三，计算指标 X_i 的客观权重 w_i ，其计算公式如下：

$$w_i = \frac{C_i}{\sum_{i=1}^n C_i}, (i = 1, \dots, n) \quad (15)$$

1.3.3 AHP-CRITIC 指标综合权重的计算

AHP 法适用于处理决策者的主观信息，具有一定的主观局限性。CRITIC 法适用于挖掘样本数据中客观信息，但过于依赖原始数据质量的影响，且忽略专家经验和指标间实际的关系。因此需要进行组合赋权，本研究将基于 AHP 法和 CRITIC 的组合赋权法，其计算公式如下：

$$W_i = \rho W_{AHPi} + (1 - \rho) W_{CRITICi} \quad (16)$$

式中, ρ 为主观权重系数, 考虑主客观的均衡性, 本文将 ρ 取0.5。

1.4 基于层次聚类算法的分级地标提取方法

地标知识是空间知识的基础, 地标具有明显的区域指代特征, 其显著程度决定其指代的范围。基于显著度的地标分层建模, 其目的在于通过不同层次的地标获取不同层次的空间知识, 并将这种地标认知概念映射到智能机器领域。本文采用自底向上的凝聚层次聚类算法 (Agglomerative Nesting, AGENS) 定义层级与阈值, 基于地标显著度计算结果分级提取地标。地标层次聚类算法的伪代码如表 2 所示。首先将地标显著度数据集中的每个地标本看作一个初始聚类簇, 然后每次迭代时找出距离最近的两个簇进行合并, 迭代至簇类的个数达到指定个数为止。

表 2 地标的层次聚类算法

Tab.2 Hierarchical Clustering Algorithm for Landmarks

算法: 地标的层次聚类算法流程
输入: 地标集 $L = \{L_1, L_2, \dots, L_k\}$, 地标显著度值 $S = \{S_1, S_2, \dots, S_k\}$, 聚类距离函数 d , 聚类簇数 k
输出: 地标簇 $C = \{C_1, C_2, \dots, C_k\}$
1: for $j = 1, 2, \dots, m$ do
2: $C_j = \{x_j\}$
3: end for
4: for $i = 1, 2, \dots, m$ do
5: for $j = 1, 2, \dots, m$ do
6: $M(i, j) = d(c_i, x_j)$;
7: $M(j, i) = M(i, j)$;
8: end for
9: end for
10: $q = m$
11: while $q > k$ do
12: 找出距离最近的两个聚类簇 C_{i^*} 和 C_{j^*} .
13: $C_{i^*} = C_{i^*} \cup C_{j^*}$
14: for $j = j^* + 1, j^* + 2, \dots, q$ do
15: $C_j = C_{j-1}$
16: end for
17: 删除距离矩阵 M 的第 j^* 行与第 j^* 列

表 3 部分地标显著度评价的指标值

Tab.3 Index values for salience evaluation of selected landmarks

名称	S_c	S_v	S_s
----	-------	-------	-------

```

18:  for  $j = 1, 2, \dots, q - 1$  do
19:      $M(i^*, j) = d(C_{i^*}, C_j)$ ;
20:      $M(j, i^*) = M(i^*, j)$ ;
21:  end for
23:   $q = q - 1$ 
24: end while

```

2 实验结果与分析

2.1 基于 SnowNLP 的情感分析结果

基于 SnowNLP 对 3919 条用户评论数据进行情感分析, 结果表明, 呈消极情感的评论数为 1285, 占全部评论数的 32.79%; 呈积极情感的评论数为 2582, 占全部评论数的 65.88%; 呈中立情感的评论数为 52, 占全部评论数的 1.33%。该情感值计算结果将用于大型商场内情感地标的认知显著度计算。

2.2 大型商场情感地标显著度计算结果

本文基于用户评论数据及实地调查数据, 获取了某大型购物中心一楼商铺的属性数据。所构建的情感地标显著度评价指标体系如图 2 所示, 包含三个一级指标: 认知显著度 (S_c)、视觉显著度 (S_v)、空间显著度 (S_s) 和九个二级指标: 情感值 (X_1)、唯一性 (X_2)、知名度 (X_3)、门面面积 (X_4)、标牌面积 (X_5)、颜色 (X_6)、形状结构 (X_7)、和可达性 (X_8)。

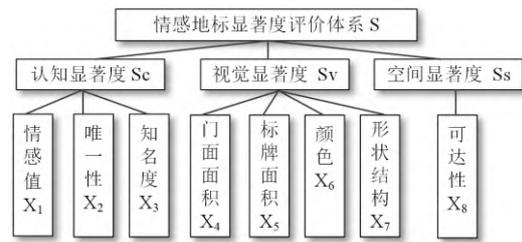


图 2 情感地标显著度评价指标体系

Fig.2 Emotional landmark salience evaluation index system

根据第 1.2 节各二级指标的计算方法, 表 3 为部分地标的显著度各指标计算评价结果。

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
LOUISVUITTON	0.63	0	609	143	4.1	213	3	3
CHANEL	0.81	1	528	90	3.5	187	3	3
GUCCI	0.86	1	403	82	3	160	3	3
LOEWE	0.74	0	76	20	2	68	2	2
MiuMiu	0.65	1	103	24	3.1	164	3	1
Celine	0.79	1	89	16	2.4	179	2	1
BLANCPAIN	0.87	1	8	4	0.5	36	1	2
AudemarsPiguet	0.78	1	7	4	0.3	64	1	2

依据第 1.3 节的指标权重计算方法，分别采用经 AHP 法、CRITIC 法及 CRITIC-AHP 复合加权法得到各指标权重系数，结果如表 4 所示。首先，根据 AHP 法的步骤，首先构造判断矩阵；然后由 5 名专家协商一致后进行打分，得到最大特征根 $\lambda_{max} = 8.740$ 、一致性指标 $CI = 0.106$ ，随机一致性

指标 $RI = 1.410$ ，一致性比率 $CR = 0.075 < 0.1$ ，通过一致性检验。然后，根据 CRITIC 法的步骤，利用公式 (13) 对情感地标显著度评价指标数据进行标准化处理，利用公式 (14)、(15) 得出客观权重。最后，利用公式 (16) 得到情感地标显著度评价指标的综合权重。

表 4 情感地标显著度评价指标权重

Tab.4 Weights of indicators for evaluating the emotional landmark salience

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
AHP (%)	6.20	2.76	32.06	16.92	17.65	15.82	3.82	4.77
CRITIC (%)	10.52	13.34	19.05	9.67	10.30	12.10	12.88	12.14
综合(%)	8.36	8.05	25.56	13.30	13.97	13.96	8.35	8.45

结合 CRITIC-AHP 的综合权重计算结果，情感地标显著度定量模型计算如下：

$$S_{sum} = 0.42S_c + 0.5S_v + 0.08S_s \quad (17)$$

将各指标的综合权重系数与其相对应的地标显著度指标得分归一化结果相乘，可得到地标的显著度评价结果，如表 5 所示。

表 5 部分地标显著度的综合评价结果

Tab.5 Comprehensive evaluation results of the salience of selected landmarks

名称	w_1S_1	w_2S_2	w_3S_3	w_4S_4	w_5S_5	w_6S_6	w_7S_7	w_8S_8	S_{sum}
LOUISVUITTON	0.053	0.000	0.256	0.133	0.140	0.140	0.084	0.085	0.889
CHANEL	0.068	0.081	0.256	0.133	0.140	0.140	0.084	0.085	0.984
GUCCI	0.072	0.081	0.256	0.133	0.140	0.140	0.084	0.085	0.988
LOEWE	0.062	0.000	0.128	0.067	0.070	0.070	0.042	0.042	0.480
MiuMiu	0.054	0.081	0.256	0.067	0.140	0.140	0.084	0.000	0.820
Celine	0.066	0.081	0.128	0.067	0.070	0.140	0.042	0.000	0.592
BLANCPAIN	0.073	0.081	0.000	0.000	0.000	0.000	0.000	0.042	0.195
AudemarsPiguet	0.065	0.081	0.000	0.000	0.000	0.070	0.000	0.042	0.258

2.3 情感地标的分级提取与表达

利用 SPSS 软件对某大型商场一楼的地标显著度进行层次聚类分析，根据聚类结果

分别提取显著度大于 0.85、0.72、0.35、0.20 作为四个层次地标的阈值，地标分级提取结果如表 6 所示。

表 6 地标分层的结果

Tab.6 Results of stratified landmarks

比例尺	显示地标级别	地标个数
1:50	一级	4
1:25	一级, 二级	9
1:10	一级, 二级, 三级	33
1:5	一级, 二级, 三级, 四级	37

为了在导航系统中引入地标概念来模拟人类的交流行为, 在人机交互过程中生成具有认知效率的表达式, 即需要提供放大或缩小两种层次结构的表达式, 进而本文将层

次地标应用在不同尺度的室内地图中进行要素表达。该商场的室内导航地图比例尺范围主要有 1:50、1:25、1:10 和 1:5, 不同尺度对应表达地标数量分别为 4、9、33、37 (如图 3 所示)。高层级的地标可作为子集包含在低层级地标内。当用户处于陌生的环境中, 利用显著度高的地标能进行初步的全局定位, 确定当前所处的位置和方向; 然后进一步通过层级较低、显著度较小的地标进行局部定位。



图 3 情感地标在室内导航地图的多尺度表达。(a) 地标在 1:50 的室内地图表达; (b) 地标在 1:25 的室内地图表达; (c) 地标在 1:10 的室内地图表达; (d) 地标在 1:5 的室内地图表达

Fig.3 Multi-scale representation of emotional landmarks in indoor navigation maps. (a) Landmark representation in the indoor map with 1:50. (b) Landmark representation in the indoor map with 1:25. (c) Landmark representation in the indoor map with 1:10. (d) Landmark representation in the indoor map with 1:5.

2.4 情感地标提取方法的可用性评价

为验证本文提出的情感地标显著度评价方法的有效性, 本文在该商场的 1F 实施用户实验。参与者共有 30 人 (15 男 15 女, 平均年龄 33 岁)。具体流程为被试在跟随实验员沿固定的路径步行, 然后在显示的 37 个地标图片的手机屏幕上进行李克特五级量表评分, 其中“1”代表非常不显著, “5”代表非常显著。

为了验证情感地标显著度计算结果和用户评分结果之间的相关性, 本文进行了皮尔逊相关系数(Pearson correlation coefficient, PCC)检验^[35], 其计算如式 (18)。

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (18)$$

式中, \bar{x} , \bar{y} 分别为第 n 个地标显著度计算结果的均值; r 为皮尔逊相关系数, 其结果介于 $-1 \sim 1$, $r > 0$ 表示正相关, $r < 0$ 表示负相关, $r = 0$ 表示零相关。| r | 值越大, 表明变量之间的相关性越强; 检测水准以 $P < 0.05$ 为差异有统计学意义。

结果分析表明, 通过本文所提出方法计算的情感地标显著度与用户评价结果呈高度线性相关 ($r = 0.601, P < 0.001$), 表明基于本文所提出的方法符合用户认知。

3 结论与展望

本文提出一种基于 UGC 的大型商场情感地标自动提取方法, 基于 SnowNLP 的用户评论情感分析, 并将分析结果拓展至室内地标认知显著度的计算, 基于 AHP 计算地标显著度的权重因子, 利用层次聚类算法提取层次地标并应用到实际大型商场中, 取得很好的表达效果。针对从室内复杂环境中提取地标的的需求, 本研究结合现有的室内外地标提取方法, 从认知显著度、视觉显著度、空间显著度三个方面对情感地标显著度评价模型进行构建, 其创新之处在于构建了情感地标显著性, 将用户情感嵌入到室内空间认知领域的理论框架。但是研究存在一些局限性, 首先由于用户评论数据的局限性, 本文关注讨论大型商场中功能类地标(主要为商铺)作为情感地标, 而对于交通类地标(入口、门厅、中庭、楼梯、电梯等)、构建类地标(屋顶、墙面、地面、立柱、门窗等)和服务类地标(洗手间、客服中心、自助机、停车场等)的讨论并未有涉及。其次在构建情感地标显著度定量评价模型过程中, 利用 AHP 确定指标权重, 其计算结果会由于专家认知差异而具有一定的主观性。因此, 未来研究可借鉴客观认知实验(如眼动)和权重计算方法(如机器学习), 以提取更符合用户认知的室内地标并在多尺度室内地图中应用表达。

参考文献

- [1] Chen Y, Hu J Y, Xiao Y, et al. Understanding the User Behavior of Foursquare: A Data-Driven Study on a Global Scale[J]. *IEEE Transactions on Computational Social Systems*, 2020, 7(4): 1019-1032
- [2] Marchiori E, Cantoni L. The Role of Prior Experience in the Perception of a Tourism Destination in User-Generated Content[J]. *Journal of Destination Marketing & Management*, 2015, 4(3): 194-201
- [3] Feick R, Robertson C. Identifying Locally- and Globally-Distinctive Urban Place Descriptors from Heterogeneous User-Generated Content[M]//*Advances in Geographic Information Science*. Cham: Springer International Publishing, 2015: 51-63
- [4] Richter K F. Prospects and Challenges of Landmarks in Navigation Services[M]//*Lecture Notes in Geoinformation and Cartography*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013: 83-97
- [5] Chen H, Vasardani M, Winter S, et al. A Graph Database Model for Knowledge Extracted from Place Descriptions[J]. *ISPRS International Journal of Geo-Information*, 2018, 7(6): 221
- [6] Kim J, Vasardani M, Winter S. Landmark Extraction from Web-Harvested Place Descriptions[J]. *KI - künstliche Intelligenz*, 2017, 31(2): 151-159
- [7] Tezuka T, Tanaka K. Landmark Extraction: A Web Mining Approach[M]//*Spatial Information Theory*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005: 379-396
- [8] Sadeghian P, Kantardzic M. The New Generation of Automatic Landmark Detection Systems: Challenges and Guidelines[J]. *Spatial Cognition & Computation*, 2008, 8(3): 252-287
- [9] Zhao Weifeng, Li Qingquan, Li Bijun. Extracting Hierarchical Landmarks from Urban POI Data[J]. *Journal of Remote Sensing*, 2011, 15(5): 973-988 (赵卫锋, 李清泉, 李必军. 利用城市 POI 数据提取分层地标[J]. *遥感学报*, 2011, 15(5): 973-988)
- [10] Wang Ming, Hu Qingwu, Li Qingquan, et al. Extracting Hierarchical Landmark from Check-in Data[J]. *Chinese Journal of Computers*, 2016, 39(2): 405-413 (王明, 胡庆武, 李清泉, 等. 基于位置签到数据的城市分层地标提取[J]. *计算机学报*, 2016, 39(2): 405-413)
- [11] Weng M, Xiong Q, Kang M J. Saliency Indicators for Landmark Extraction at Large Spatial Scales Based on Spatial Analysis Methods[J]. *ISPRS International Journal of Geo-Information*, 2017, 6(3): 72
- [12] Samany N N. Automatic Landmark Extraction from Geo-Tagged Social Media Photos Using Deep Neural Network[J]. *Cities*, 2019, 93: 1-12
- [13] Caduff D, Timpf S. On the Assessment of Landmark Saliency for Human Navigation[J]. *Cognitive Processing*, 2008, 9(4): 249-267
- [14] Sorrows M E, Hirtle S C. The nature of landmarks for real and electronic spaces[C]//*International conference on spatial information theory*. Springer, Berlin, Heidelberg, 1999: 37-50.

- [15] Raubal M, Winter S. Enriching Wayfinding Instructions with Local Landmarks[M]//Geographic Information Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002: 243-259
- [16] Zhu Haihong, Wen Ya, Mao Kai, et al. A Quantitative POI Salience Model for Indoor Landmark Extraction[J]. *Geomatics and Information Science of Wuhan University*, 2018, 43(3): 336-341 (朱海红, 温雅, 毛凯, 等. 室内地标提取的 POI 显著度定量评价模型[J]. *武汉大学学报·信息科学版*, 2018, 43(3): 336-341)
- [17] Zhu L T, Švedová H, Shen J, et al. An Instance-Based Scoring System for Indoor Landmark Salience Evaluation[J]. *Geografie*, 2019, 124(2): 103-131
- [18] Hu X K, Ding L, Shang J G, et al. Data-Driven Approach to Learning Salience Models of Indoor Landmarks by Using Genetic Programming[J]. *International Journal of Digital Earth*, 2020, 13(11): 1230-1257
- [19] Lyu H, Yu Z H, Meng L Q. A Computational Method for Indoor Landmark Extraction[M]//Lecture Notes in Geoinformation and Cartography. Cham: Springer International Publishing, 2014: 45-59
- [20] Dubey R K, Sohn S S, Thrash T, et al. Identifying Indoor Navigation Landmarks Using a Hierarchical Multi-Criteria Decision Framework[C]//MIG '19: Motion, Interaction and Games. Newcastle upon Tyne, United Kingdom. 2019: 1-11
- [21] Balaban C Z, Karimpur H, Röser F, et al. Turn Left where You Felt Unhappy: How Affect Influences Landmark-Based Wayfinding[J]. *Cognitive Processing*, 2017, 18(2): 135-144
- [22] Mendez-Lopez M, Fidalgo C, Osma J, et al. Wayfinding Strategy and Gender - Testing the Mediating Effects of Wayfinding Experience, Personality and Emotions[J]. *Psychology Research and Behavior Management*, 2020, 13: 119-131
- [23] Gasper K, Clore G L. Attending to the Big Picture: Mood and Global Versus Local Processing of Visual Information[J]. *Psychological Science*, 2002, 13(1): 34-40
- [24] Ruotolo F, Claessen M H G, van der Ham I J M. Putting Emotions in Routes: The Influence of Emotionally Laden Landmarks on Spatial Memory[J]. *Psychological Research*, 2019, 83(5): 1083-1095
- [25] Palmiero M, Piccardi L. The role of emotional Landmarks on Topographical Memory[J]. *Frontiers in Psychology*, 2017, 8: 763
- [26] Balaban C Z, Florian R, Kai H. The Effect of Emotions and Emotionally Laden Landmarks on Wayfinding[J]. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 2014, 36(36)
- [27] Kim J Y, Choi J K, Han W H, et al. The Influence of Users' Spatial Familiarity on Their Emotional Perception of Space and Wayfinding Movement Patterns[J]. *Sensors (Basel, Switzerland)*, 2021, 21(8): 2583
- [28] Liu B. Sentiment analysis and subjectivity[J]. *Handbook of natural language processing*, 2010, 2(2010): 627-666.
- [29] SnowNLP: Simplified Chinese Text Processing, Oct. 2017, [online] Available: <https://github.com/isnowfy/snownlp>
- [30] Nothegger C, Winter S, Raubal M. Selection of Salient Features for Route Directions[J]. *Spatial Cognition & Computation*, 2004, 4(2): 113-136
- [31] Kim J, Han D, Tai Y W, et al. Salient Region Detection via High-Dimensional Color Transform and Local Spatial Support[J]. *IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society*, 2016, 25(1): 9-23
- [32] Tai Huixin, Yang Yulan, Shen Li, et al. Study on weighting indicators of historic building value assessment[J]. *Journal of Zhejiang University of Technology*, 2021, 49(3) (邵惠鑫, 杨玉兰, 沈黎, 等. 历史建筑价值评估中的指标权重研究[J]. *浙江工业大学学报*, 2021, 49(3): 300-307)
- [33] Saaty T L. How to Make a Decision: The Analytic Hierarchy Process[J]. *European Journal of Operational Research*, 1990, 48(1): 9-26
- [34] Diakoulaki D, Mavrotas G, Papayannakis L. Determining Objective Weights in Multiple Criteria Problems: The Critic Method[J]. *Computers & Operations Research*, 1995, 22(7): 763-770
- [35] Lee Rodgers J, Nicewander W A. Thirteen Ways to Look at the Correlation Coefficient[J]. *The American Statistician*, 1988, 42(1): 59-66

Extraction of Emotional Landmarks in Large Malls Based on User-Generated Content

ZHU Litao^{1,2}, SHEN Jie^{1,2,3}, WANG Xing^{1,2}, HOU Yingxu^{1,2} and ZHANG Cheng⁴

1. Key Laboratory of Virtual Geographic Environment (Nanjing Normal University), Ministry of Education, Nanjing, 210023, China
2. School of Geography, Nanjing Normal University, Nanjing, 210023, China;
3. Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing, 210023, China;
4. Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China.

Abstract: User emotions have a significant impact on spatial attention, spatial decision-making, and spatial memory. Emotions associated with landmarks can improve navigation efficiency while enhancing users' cognitive map abilities. Previous studies focused on the role of emotional landmarks in navigation, but few studies paid attention to landmark extraction methods in complex indoor environments. In this paper, we propose a quantitative model of salience to automate the extraction of emotional landmarks in large shopping malls based on user-generated content (UGC). Firstly, we obtain user comment data of a large shopping mall using web crawler technology. Second, we conduct the sentiment analysis of user comments based on SnowNLP and extend the results to landmark cognitive salience. Then, we combine the Analytic Hierarchy Process (AHP) and Criteria Importance Through Intercriteria Correlation (CRITIC) to calculate the weights of indoor landmark salience indicators and construct a quantitative evaluation model of emotional landmark salience. Finally, we extract hierarchical landmarks using hierarchical clustering algorithms, design multi-scale indoor navigation maps based on hierarchical landmarks to meet user cognition, and verify the usability of the landmark extraction method through user experiments. This work can promote indoor navigation map design standardization and provide a valuable complement to intelligent indoor navigation services.

Keywords: user-generated content (UGC); large shopping mall; emotional landmarks; landmark salience; sentiment analysis

First Author: ZHU Litao, PhD candidates, specializes in spatial cognition and indoor pedestrian navigation map.
E-mail: 181301028@njnu.edu.cn

Corresponding author: SHEN Jie, PhD, professor. E-mail: shenjie@njnu.edu.cn

Foundation Support: The National Key Research and Development Program of China (2021YFE0112300); the National Natural Science Foundation of China (41871371).

网络首发:

标题: 基于用户生成内容的大型商场情感地标提取方法

作者: 朱利涛, 沈婕, 王兴, 侯盈旭, 张成

DOI: 10.13203/j.whugis20210488

收稿日期: 2022-07-12

引用格式:

朱利涛, 沈婕, 王兴, 等. 基于用户生成内容的大型商场情感地标提取方法[J]. 武汉大学学报·信息科学版, 2022, DOI: 10.13203/j.whugis20210488 (ZHU Litao, SHEN Jie, WANG Xing, et al. Extraction of Emotional Landmarks in large malls Based on User-Generated Content[J]. *Geomatics and Information Science of Wuhan University*, 2022, DOI: 10.13203/j.whugis20210488)

