

華南農業大學

课程论文

基于熵权法和因子分析法的城市宜居性评价模型

张恒鑫

201512010223

学 院 名 称	<u>数学与信息学院</u>
专 业 班 级	<u>信息与计算科学 2 班</u>
提 交 日 期	<u>2018 年 6 月</u>

评阅分数 _____

评阅人 _____

《数学建模》课程论文任务书/承诺书

姓 名	班 级	学 号																																																													
张恒鑫	信息与计算科学 2 班	201512010223																																																													
课题题目 <div style="text-align: center;">宜居城市评价</div>																																																															
课题来源 <div style="display: flex; justify-content: space-around;"> 课程论文指导书 <input checked="" type="checkbox"/> 自选题 <input type="checkbox"/> </div>																																																															
课题设想 <p>宜居城市是城市发展的最终产物，也是最终形态。宜居城市是我们开启美好生活的基础，请完成下列任务：</p> <p>(1) 定义衡量宜居城市的指标；</p> <p>(2) 建立合适的评价宜居城市的模型；</p> <p>(3) 选择一个城市，就你的模型而言，谈谈如何提高该城市的宜居水平。</p>																																																															
<p>我承诺我的论文的主要内容来自于本人的独立思考，引用别人的成果都在文中加以注明。</p> <p>(请在下面的田字格中抄写以上文字)</p> <table border="1" style="width: 100%; height: 150px; border-collapse: collapse;"> <tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> </table>																																																															
学生（签名）		指导教师（签名）																																																													

填写日期：2018 年 月 日

基于熵权法和因子分析法的城市宜居性评价模型

摘 要

为了量化城市宜居水平，从其概念出发，我们构建了一个三级综合评价指标系统。整个模型的流程见图 1。

首先，我们定义了一个综合指数，称为城市宜居指数（CLI），它包括社会、经济、政治和气候四个方面。在四个选择准则的指导下，我们选择了 15 个指标，具体见表 2。城市宜居指数是 4 个原则和 15 个指标的线性组合。所以我们采用熵值法和因子分析法分别确定了它们的权重。此外，我们划分出不同的城市宜居水平等级，得分 0-3 为差、3-7 为中、7-10 为优。

然后，应用我们构造的模型，计算了北京、上海、广州、杭州、重庆五个城市城市宜居指数。我们发现，杭州和重庆的城市宜居水平最高，其次是广州。根据模型划分，这三者的城市宜居水平都可以归为优。而北京的宜居水平最低，属于差的一类。

目 录

一、引言.....	3
1.1 问题提出.....	3
1.2 问题分析.....	3
1.3 我们的工作.....	4
二、模型假设和符号说明.....	5
2.1 模型假设.....	5
2.2 符号说明.....	5
三、数据处理.....	6
3.1 数据收集.....	6
3.2 数据填补和归一化.....	6
四、城市宜居指数模型.....	7
4.1 指标体系的建立.....	7
4.2 城市宜居指数基线模型.....	9
五、模型求解与分析.....	11
六、模型的评价与改进.....	12
6.1 模型优势.....	12
6.2 模型不足.....	12
参 考 文 献.....	13

一、引言

1.1 问题提出

许多国家在城市规划和管理过程中都遇到了不少问题，特别是当今世界城市化趋势正在加速。新兴经济使城市以前所未有的速度发展，也导致了城市空间的无序扩张，给环境和资源带来了巨大压力，使人地关系更加紧张。许多人提出了“逃离北上广”的口号，原因在于他们认为这些城市的不适合居住。

城市是人类文明的标志，特别是我国目前的城镇化正处于由快速大规模发展向绿色、智慧可持续发展转变的关键时期。量化城市宜居性对政府制定政策有着重大的意义，尤其是随着大数据时代的到来，这使其变成了当务之急的工作。城市宜居性是当前城市科学研究领域的热点议题之一，也是政府和城市居民密切关注的焦点。我国宜居城市的排名每年都是热门话题，不同机构对宜居城市的排名结果也不尽相同。中科院在 2016 的报告说中国宜居城市排名广州倒数第二，而美世人力资源咨询公司（William Mercer）公布的 2016 全球宜居城市中，广州位列中国城市第三名。由此可见，宜居城市评价指标体系不同，宜居城市排名结果也会发生变化。

因此我们需要建立起一套完备的合理科学的城市宜居度评价体系，以更容易观察城市宜居度指数所揭示的危险性并制定相应的政策。

1.2 问题分析

针对问题，我们必须首先理解城市宜居性的定义。这个术语--城市宜居性，被各种学术团体以不同概念所使用，因而对此产生分歧十分常见。通过查阅相关资料，我们将城市宜居性定义为经济、社会、政治和气候条件综合反映的客观指标。

这是一个典型的综合评价问题，其难点在于建立起合理科学的指标体系和收集相关数据。典型的综合评价方法有层次分析法、模糊评价方法、综合指数法和熵权法等。我们要结合城市宜居的特征从中选择合适的模型，从而定量计算出城市宜居指数。

1.3 我们的工作

我们需要提出一系列任务的解决方案。我们的工作可以列举如下：

- 首先，我们基于这个问题重新描述了城市宜居性的定义；
- 然后，考虑到获取指标的基本思想和四个准则（城市、经济、政策、气候）等方面，我们设计了一个综合指标体系；
- 借助因子分析法和熵权法，我们得到了计算城市宜居性的客观模型。

整个模型的流程图如下图所示：

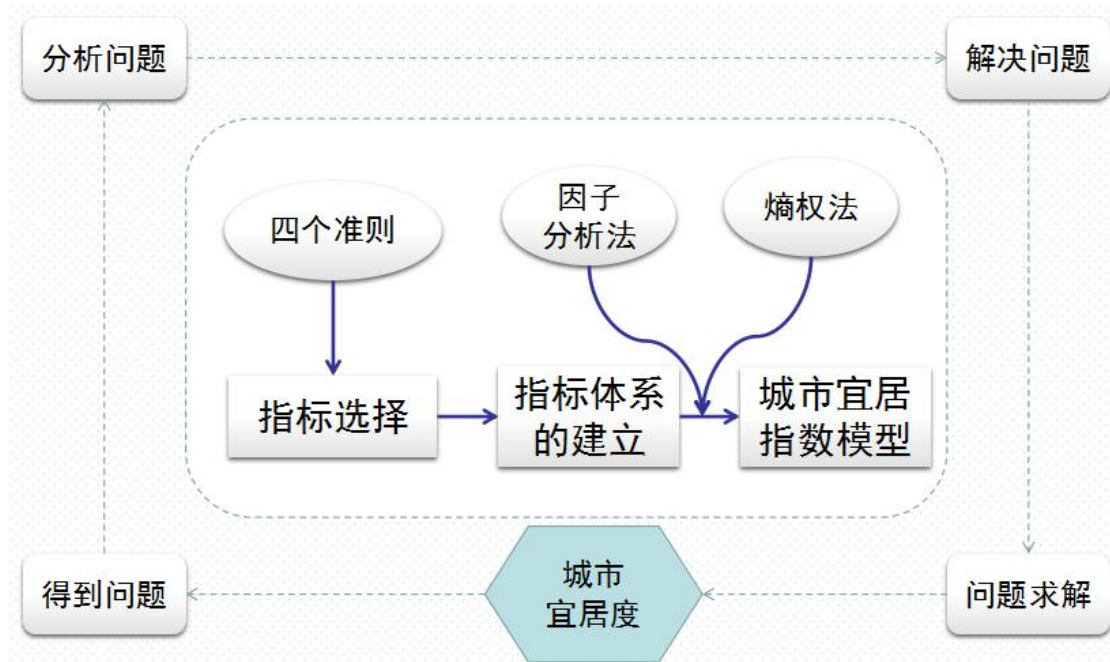


图 1 流程图

二、模型假设和符号说明

2.1 模型假设

我们提出以下假设，以简化复杂的问题，因为本文中讨论的问题是庞大且抽象的。

- ❖ 我们的数据来自于各省市统计年鉴，假定其数据是准确、可靠且相互一致；
- ❖ 城市的宜居水平可以通过所选指标衡量，未经选择的指标数据对城市宜居性没有影响或影响极低；
- ❖ 地理条件对城市宜居水平的影响可以完全转化为其他因素的影响；
- ❖ 我们假设在因子分析法中，共同因子是不相关的，特殊因子是不相关的，同时共同因子和特殊因子间也不相关。

2.2 符号说明

表 1 符号说明表

符号	意义
x_i	第 i 个指标的时间序列矩阵
p_{ij}	城市宜居指数的第 j 个性能指标下的第 i 个样本的比例
e_j	第 j 个指标的熵
d_j	信息熵冗余（差异）
W_j	第 j 个指标的权重
$CLI(i)$	City Livable Index, 第 i 年的城市宜居指数

三、数据处理

3.1 数据收集

收集足够的数据是建立完整的指标体系的基础。

首先，我们检索了大多数关于城市宜居度指标的文献，并参考选择现有模型的指标，以筛选出我们需要的指标。

然后，我们查阅了各省市的统计年鉴，找到了 5 个城市的相关指标。这 5 个城市分别为北京、上海、广州、杭州、重庆。

3.2 数据填补和归一化

数据丢失在统计中是十分常见的。当值得丢失阻止我们成功地观察变量和规律时，我们需要考虑到数据的可用性是一个重要的问题。如果数据不可靠，我们就不能提供准确的评估。

因此，我们应采取一定的措施，确保数据的真实性和连续性。

- ✧ 指标缺失率超过 60%：我们寻找足够的和等效的指标来替换它，或者丢弃它；
- ✧ 指标缺失率不超过 30%：我们寻找其他具有相同变化特征的国家的指标来填充它；
- ✧ 指标缺失率低于 30%：采用回归、预测或模拟的方法来填充它。

在处理缺失数据之后，我们对数据进行归一化处理，将其转换为无量纲表达式，并将数据映射到 0 到 1 的范围。本文使用以下公式进行归一化：

$$x_{ij} = \frac{\max(x_j) - x_{ij}}{\max(x_j) - \min(x_i)}$$

四、城市宜居指数模型

在这一节中，我们将介绍如何选择城市宜居指数指标和如何应用因子分析法和熵权法定量计算城市宜居指数。

4.1 指标体系的建立

毫无疑问，量化一个城市的宜居性是非常重要的复杂和困难，因为其中包含了大量的指标。

基于可用性、可比性和可操作性的总体思路，我们选择了四个准则层，包括社会、政治、经济和气候。事实上，许多机构在评估的时候也会关注这些核心准则。接着，我们选择与这些核心准则相关的性能指标。在选择指标选择过程中，我们试着去关注**指标的典型性和代表性**，这会让后面的工作变得简单容易（指标间不存在共线）。

因此，我们构建了一个基于上述原则的包含 15 个客观指标的评价指标体系，并以此建立起模型。然后我们试着分析了它们的优点和局限性。当然，为了简化计算，我们将所有指标归一化以确保它们在相同的范围内（从 0 到 1）。

具体的选择思想如下图所示：

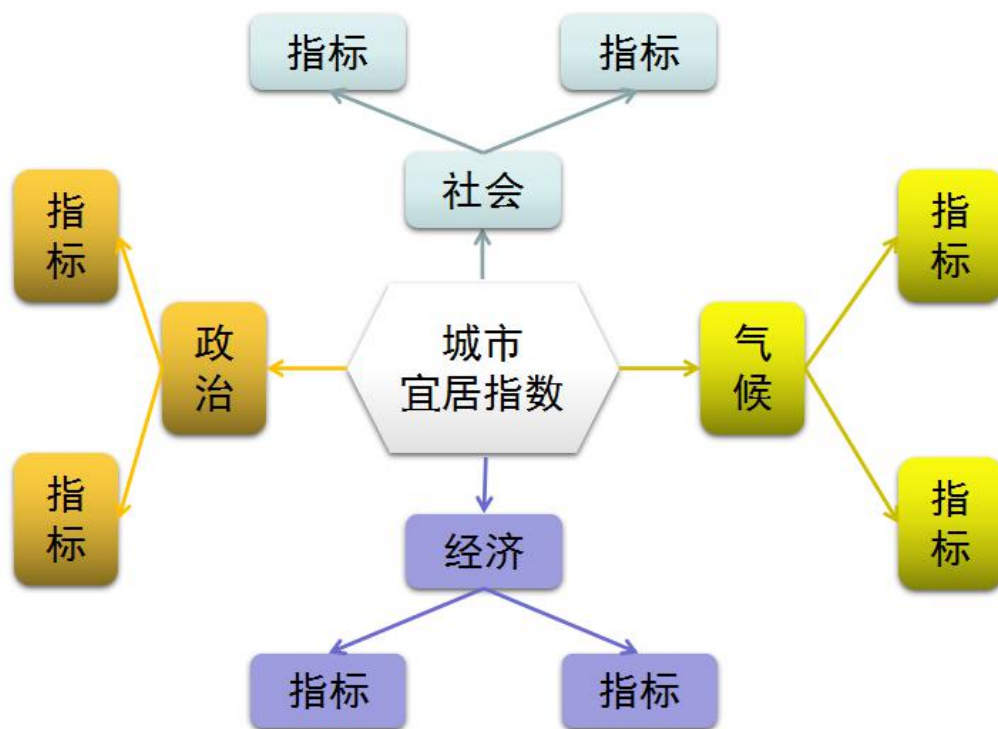


图 2 如何选择指标示意图

按照这个思想，我们挑选出以下 15 个指标：

表 2 具体指标选择表

指标	
经济	生产总值
	生产总值年均增长率
	人均生产总值
社会	人口密度
	人口增长率
	失业率
	人均资源拥有率
	民主满意度
政治	廉政指数
	法律完善程度
	公共服务支出
	每万人犯罪率
	军事支出
气候	年平均温度
	年平均降雨

4.2 城市宜居指数基线模型

基于上述潜在指标的分析,我们开发了一个名为城市宜居指数基线模型(City Livable Index Baseline Model,CLIBM)来量化城市宜居性。

我们所要做的就是确定这些变量的权重,因为我们假设城市宜居指数是所有绩效指标的线性组合。在这里,我们使用熵权法(Entropy Weight Method,EWM)来确定各指标的权重。

熵权法是一种客观的加权方法,因为它只依赖于数据本身的离散性。在信息论中,熵是不确定度的量度。这意味着信息量越大,熵越小,不确定性越小,熵越大,信息量越小,不确定性越大。

根据熵的特点,熵值可以用来判断事件的随机性和无序度。熵值也可以用来判断指标的分散程度。指标离散程度越大,指标对综合评价(权重)的影响越大。

假设我们有一个国家 m 个指标的 n 年数据。我们将其表示为

$$x_i = (x_{i1}, x_{i2}, \dots, x_{in}) \quad i = 1, 2, \dots, m$$

代表第 i 个指标的时间序列矩阵。所以全部指标的序列矩阵可以表示为

$$X = (x_1, x_2, \dots, x_m)$$

然后,我们可以用下面的公式计算模型中的第 j 个性能指标下的第 i 个样本的比例:

$$P_{ij} = x_{ij} / \sum_{i=1}^n x_{ij}$$

接着计算第 j 个指标的熵值:

$$e_j = -k \sum_{i=1}^n p_{ij} \ln(p_{ij})$$

上式中, $k = 1/\ln(n) > 0$ 。

因此我们可以得到各个指标的权重:

$$w_j = d_j / \sum_{j=1}^m d_j = (1 - e_j) / \sum_{j=1}^m (1 - e_j)$$

最后，我们应用因子分析法（Factor Analysis Method,FAMM）去确定四个准则（经济、社会、政治和气候）的贡献率（即权重） c 。那么，城市宜居指数可以表示为

$$CLI = \sum c_i w_{ij} x_{ij}$$

计算中所有指标均为正向指标，模型综合得分越低，城市宜居性越差。本文将城市宜居性等级定义如下表 2。

表 2 城市宜居性等级定义表

定义	得分
优	7-10
中	3-7
差	0-3

五、模型求解与分析

应用上述方法，经过计算得到北京、上海、广州、杭州、重庆五个城市 2017 年宜居水平的排名（见表 3）

表 3 城市宜居水平排名表

宜居水平排名	城市	综合得分 CLI
1	杭州	8.086
2	重庆	7.516
3	广州	7.023
4	上海	3.578
5	北京	2.896

由表 3 可以看出，杭州和重庆的城市宜居水平最高，其次是广州。根据模型，这三者的城市宜居水平都可以归为优。而北京的宜居水平最低，属于差的一类。排名落后的城市应加强社会、经济、环境和政治各方面的发展，提高城市宜居水平，促进可持续发展。

当然，从模型中我们可以看出哪些指标对城市的宜居水平权重影响较大，政策制定者可以尝试从这些指标下手以改善城市宜居水平。比如说北京，在气候和社会两方面得分十分低，政府应该努力改善气候环境，并且提高社会福利、减缓人民在北京的生活压力，这样才可以提高北京的宜居水平。

六、模型的评价与改进

6.1 模型优势

我们的模型在统计和评价方面具有许多优点。

- ✓ 评价模型完全基于数据，没有主观或任意的决策；
- ✓ 我们的模型成功地确定了气候（代表环境）对城市宜居水平有潜在影响的机制，而不是仅仅关注对一个国家的肤浅评价；
- ✓ 我们的模型充分利用时间序列数据以产生合理的结果。这充分反映了真实世界，并可以帮助我们充分理解真实世界。

6.2 模型不足

虽然我们的模型成功地计算了城市宜居水平，但有几个方面有待改进：

- ❖ 由于部分数据的缺失，一些指标的值只能是使用数学方法计算得到，这可能是不准确的。我们致力于对一些数据进行预测以填充表格或矩阵。但是不准确的数据往往会引起错误。
- ❖ 在计算城市宜居指数之前，我们仅找出两个三级指标来衡量气候指标。同样，一些重要指标在我们首次建立模型时放弃，因为数据的难收集或丢失。为了模型的适应能力，以及进行更充分的分析，我们需要在更广泛的范围内收集更多的研究数据。

参 考 文 献

- [1] 张文忠. 宜居城市的内涵及评价指标体系探讨[J]. 城市规划学刊, 2007.
- [2] 杨静怡, 赵平, 马履一. 宜居城市绿化评价指标体系研究——以北京市为例[J]. 西北林学院学报, 2012.
- [3] 易红, 王植, 王志章. 生态文明视域下宜居城市构建路径探析: 基于重庆的实证研究[J]. 城市观察, 2014.
- [4] 宁艳杰, 刘远军, 张志强. 宜居城市生态住区评价模型研究[J]. 北京林业大学学报(社会科学版), 2008.