

基于 PCA 和 SOFM 的农村宅院综合分类及优化利用

—以东营市初家村为例

王瑗玲¹,李方方²,初康杰³,石宁⁴

1. 山东农业大学 资源与环境学院, 山东 泰安 271018

2. 聊城市国土资源局东昌府分局, 山东 聊城 252000

3. 新泰市国土资源局, 山东 新泰 271000

4. 龙口市国土资源局, 山东 龙口 265700

摘要: 以东营市初家村为研究区, 基于遥感技术解译和实地调查数据, 采用主成分分析 (PCA) 和自组织特征映射网络模型 (SOFM) 进行农村宅院综合分类, 并提出各类宅院利用建议, 以为各类宅院优化利用、村庄整治提供科学依据。结果表明: 初家村宅院类型综合影响因素包括宅院经济、人口、建筑、结构性及占地等 5 个方面, 宅院综合类型有非农高收入、非农中收入、中收入高利用、低收入高利用和低收入老龄等 5 类宅院, 提出村庄保留情景下非农低效利用住户到城镇落户、有偿退出宅院、宅院继续利用, 兼业、农业住户宅院整治、改造利用, 老龄宅院改造为绿地或复垦为耕地, 以及村庄整体迁入新型社区情景下较高质量宅院继续利用、老龄宅院复垦为耕地等。

关键词: PCA; SOFM; 农村宅院; 综合分类; 优化利用

中图分类号: K901

文献标识码: A

文章编号: 1000-2324(2017)01-0121-05

Comprehensive Classification and Optimal Utilization of Rural House Based on PCA and SOFM

—A case of Chujiacun village, Dongying City

WANG Ai-ling¹, LI Fang-fang², CHU Kang-jie³, SHI Ning⁴

1. College of Resources and Environment/Shandong Agricultural University, Tai'an 271018, China

2. Dongchangfu Substation of Bureau of Land and Resource of Liaocheng City, Zhucheng 262200, China

3. Bureau of Land and Resource of Xintai City, Xintai 271000, China

4. Bureau of Land and Resource of Longkou City, Longkou 265700, China

Abstract: Taking Chujiacun village, Dongying city as a research region, based on remote sensing interpretation and field investigation data, using principal component analysis (PCA) and self-organizing feature map (SOFM) network model, the paper carried through comprehensive farmers house classification and put forward the suggestion of all kinds of house use, so that the paper can give scientific suggestions for all kinds of comprehensive house optimal utilization and village renovation. Results showed that the influence factors of rural house in Chujiacun village included 5 aspects, which were house economy, population, architecture, constitutive property and land area, and there were 5 comprehensive house types, which were high-income and non-agricultural house, median-income and non-agricultural house, median-income and high-use house, low-income and high-use house, and low-income and elderly house. Based on that, the paper put forward targeted suggestions of low-use and non-agricultural resident move to the town, give up house for compensation and the house continue to be used, multiple-occupations and agricultural resident house are renovated and reconstructed, old house are transformed into green space or reclaimed into cultivated land under the scenario of the whole village reservation, and the high quality home continue to use, the old house reclaimed into cultivated land under the scenario of the village integral moving.

Keywords: PCA; SOFM; rural house; comprehensive classification; optimal utilization

农村宅院是农村居民点的基本组成单位和主要组成部分, 也是中国农民传统的居住形式、农村重要的人文资源, 在满足村民居住、增加经济收入、美化生活环境、传承乡村文明等方面发挥了重要作用。随着中国社会发展和农村经济条件改善, 农村宅院也发生了变化和分化: 有些宅院建筑面积增加、质量提高, 居住功能强化, 城镇化特征明显; 有些宅院依然兼具生活居住、生产经营等功能; 有些宅院由于部分家庭成员常年外出打工, 利用效率低下; 还有些宅院被废弃、闲置, 造成村庄“空心化”^[1], 这与我国土地资源短缺的国情极不相称, 影响了经济社会发展及新农村建设。因此, 需加强农村宅院综合分类研究, 探讨不同类型宅院优化利用对策, 这对促进城乡一体化发展、农村

收稿日期: 2015-10-18

修回日期: 2015-11-15

基金项目: 山东省自然科学基金(ZR2013DM006); 山东农业大学现代农业发展研究院第三批研究课题(14xsk2-03)

作者简介: 王瑗玲(1972-), 女, 博士, 教授, 主要从事土地利用、农村居民点用地变化研究. E-mail: ailingwang@sdau.edu.cn

新型社区和新农村建设及土地集约节约利用意义重大。

近年来与农村宅院相关的农村居民点及宅基地等研究成为热点,学者分析了农村居民点用地结构^[2]、时空特征^[3]和集约利用水平^[4],研究了农村宅基地转型^[5]、退出机制^[6]等。上世纪 80 年代,农村宅院生态系统^[7]、庭院经济^[8]引起学者关注,认为庭院经济对提高农户家庭收入等发挥了巨大作用^[9]。新农村建设时期,庭院经济应拓展经营领域,重点发展乡村旅游业和乡村庭院文化产业^[10]。目前,农村宅院研究取得了较多研究成果,但针对宅院类型分化,尤其是采用定量方法进行宅院类型综合划分研究薄弱。主成分分析法(PCA)是利用降维思想对高维变量系统进行最佳综合和简化,在多因素综合评价中应用广泛^[11]。在分类模型方面,基于自组织特征映射网络模型(SOFM)的综合分类研究增多,如乡村地域功能分区研究^[12],显示其极大优势。本文借鉴相关研究,以东营市初家村为例,采用 PCA 和 SOFM 模型,对农村宅院进行综合分类,并提出不同类型宅院优化利用建议,从而为农村新型社区和新农村建设提供理论支撑。

1 研究区概况及数据来源

1.1 研究区概况

初家村位于山东省东营市东营区史口镇南部,距镇政府驻地约 4.5 km,距东营区约 29 km。2012 年,全村共 116 户,总人口 356 人,常住 136 人,耕地面积 333335 m²。村内无小学、幼儿园及诊所等公共设施,有 2 处小型超市,村内主街道实现了硬化,建有垃圾池 1 处,全村绿化、亮化,排水设施较齐全。居民点总面积 62238 m²,农户宅院(宅基地)43933 m²,道路用地 11382 m²,村委用地 188 m²,空闲地及其他土地 6735 m²。

初家村农户宅院共 127 处,最大面积 1695 m²,最小 52 m²,处均 346 m²,人均 124 m²,宅基地超标比较严重;建筑容积率平均 0.43,土地利用强度较低;建筑结构为砖混或砖木,一成至全新不等。宅院平均人口 2.8 人,实际常住人口 1.07 人,有效居住率较低。住户从事农业、打工、兼业及承包工程等不同职业,收入差距较大。无收入宅院 10 处,均 1 人 1 院,多为老龄人口。

1.2 数据来源

从天地图·山东网站得到研究区高分辨率遥感影像,进行人机解译,并实地校核,得到宅院分布图及宅院面积等数据,共 127 处农户宅院。以宅院为单位,开展微观调研,获得宅院土地利用、人口及经济等信息。127 处宅院中,闲置、常年无人居住及无收入的宅院 18 处,实际利用 109 处。走访村委会,获取村庄历史、人口变化、宅基地增减等情况。收集研究区统计年鉴、有关政策文件等,得到研究区经济发展、自然资源及相关政策制度等。

2 基于 PCA 的农村宅院综合分类影响因素分析

2.1 农村宅院综合分类影响因子初步选择

本文所指农村宅院是农村集体经济组织为满足本集体组织成员的生活生产需要而划拨给其建造住房、辅助用房的土地及庭院用地,即宅基地。随社会经济发展,农户从业类别、收入水平等发生分化,宅院建设、居住时间等发生分化,因此农村宅院出现了农业型、兼业型、非农型等类型,也出现了闲置、高效利用等分化,形成农业高利用、非农闲置等综合类型。

农村宅院类型分化是自然环境、社会经济等因子综合作用的结果。根据农户生活生产需求及受到的影响,可将宅院类型影响因子分为外部因子和内部因子两大类。外部因子即宏观性因子,包括村庄自然环境及国家政策,作用范围大、时间长,对各宅院影响基本相同。内部因子则主要指宅院自身条件,是宅院利用行为的原动力,也是区域内农户个体差异的主要因素,包括家庭人口状况、资源禀赋、经济状况等。考虑初家村特点,基于农户视角,重点分析影响宅院利用的内部因子。征求相关专家与村民意见,初步选择 15 个宅院综合类型影响因子,包括农户人口规模、有效居住率(实际居住人口占家庭总人口的比重)、宅院占地面积、人均建筑面积等(表 1)。

2.2 基于 PCA 的初家村宅院综合分类影响因素分析

初家村宅院综合分类初步选择了 15 个影响因子,对其直接分析工作量较大。PCA 方法基于数

据资料,在信息损失较少的前提下把多个指标转化为几个综合指标即主成分,同时客观给出各主成分权重(方差贡献率)。本文采用 PCA 进行宅院分类影响因素综合分析。

2.2.1 评价宅院确定及因子指标标准化 剔除闲置 18 处宅院,对使用中的 109 处宅院进行分析。实地调查各宅院各影响因子状况,并计算其原始值。采用 Z-score 方法对原始指标值进行标准化处理,得到数据标准化表。公式为:

$$X_{ij} = \frac{Y_{ij} - \bar{Y}_j}{S_j}$$

式中: X_{ij} 表示标准化值, Y_{ij} 表示第 i 宅院第 j 个指标值, \bar{Y}_j 表示第 j 个指标的样本均值, S_j 表示第 j 个指标的样本方差。

2.2.2 方差分解及主成分提取 基于初始因子标准化表、利用 SPSS19.0 进行主成分分析,得到各初始因子相关系数矩阵、方差分解主成分提取分析表及初始因子载荷矩阵。根据方差分解主成分提取分析表,选取特征值 > 1、累积贡献率大于 80% 的为主成分,得到 5 个主成分因子载荷矩阵(表 1)。

从表 1 可以看出,主成分 F1 在家庭总收入、非农收入、家庭支出、家庭人均收入等方面占有较大载荷,均与经济状况有关。主成分 F2 在家庭人口规模、人均居住建筑面积、人均宅基地面积及耕地面积等方面占有较大载荷,而初家村最初耕地分配按人均 1.2 亩分配,因此均与人口规模有关;主成分 F3 主要在有效居住率、宅院容积率及房屋成新度等宅院建筑方面有较大载荷;主成分 F4 主要在劳动力比重、农业收入等宅院结构性方面有较大载荷;主成分 F5 主要在宅院占地面积占有较大载荷。因此,初家村宅院类型的影响因素可综合为经济、人口、建筑、结构性及占地等 5 个方面。

表 1 初始因子载荷矩阵
Table 1 The initial factor loading matrix

初始因子 Initial factor	主成分 Principal component				
	F1	F2	F3	F4	F5
家庭总人口	0.6447	-0.7047	0.0861	-0.0423	0.1515
有效居住率	-0.1612	-0.0962	0.5268	0.2899	-0.1373
宅院占地面积	0.2330	-0.2008	0.1161	0.4325	0.7764
人均建筑面积	-0.3015	0.7254	0.3919	-0.0217	0.1773
人均宅基地面积	-0.5205	0.6047	-0.0527	0.3433	0.4244
劳动力比重	-0.1646	-0.2093	0.4826	0.5343	-0.2479
家庭总收入	0.8765	0.3521	-0.1537	0.2287	-0.1430
农业收入比重	-0.2556	-0.2320	0.3868	0.5701	-0.3905
非农收入	0.8754	0.3598	-0.1678	0.2118	-0.1317
家庭支出	0.9366	0.1162	-0.0552	0.2321	-0.0080
耕地面积	0.6128	-0.6968	0.0919	-0.0383	0.1665
建筑结构	0.5004	0.1809	0.4776	-0.2761	0.3140
容积率	0.3366	0.2047	0.5464	-0.4705	-0.3107
房屋成新度	0.3168	0.1829	0.7524	-0.1924	0.1701
家庭人均收入	0.7634	0.5272	-0.1479	0.2053	-0.1668

2.2.3 主成分表达式及宅院主成分分值计算 基于主成分载荷矩阵及相对应的特征值,在 SPSS19.0 中利用“转换→计算变量”功能,计算主成分中每个影响因子对应的系数,将得到的特征向量与标准化后的数据相乘,得到各主成分表达式。利用主成分表达式,计算各宅院各主成分得分,并将其作为宅院综合分类的指标值。

3 基于 SOFM 的初家村宅院综合分类

3.1 SOFM 原理

自组织特征映射网络(Self-Organizing feature Map, SOFM)是一种人工神经网络,由芬兰 Kohonen T 教授提出^[13]。它可以自动、同时实现模式识别和数据分类,是一种较好的非监督分类,可以对多维数据进行平行高速处理,适于多维数据集分类,当前已逐渐应用于地理学的分区与分类研究。

SOFM 的自组织功能通过竞争学习实现,即根据有无导师组织或信号实现数据的分类和聚类两个功能。其竞争学习规则为:网络输出神经元之间相互竞争以求被激活或点火,结果在每一时刻只有一个输出神经元被激活或点火。这个被激活的神经元称为竞争获胜神经元,而其它神经元的状态被抑制。SOFM 训练过程为:

3.1.1 向量归一化 首先将当前输入模式向量 X 和竞争层中各神经元对应的内星向量 W_j 全部进行归一化处理,归一化公式为:

$$\hat{\mathbf{X}} = \frac{\mathbf{X}}{\|\mathbf{X}\|} = \left[\begin{matrix} x_1 / \sqrt{\sum_{j=1}^m x_j^2} & \dots & x_m / \sqrt{\sum_{j=1}^m x_j^2} \end{matrix} \right]^T$$

式中： $\hat{\mathbf{X}}$ 为归一化后矩阵， \mathbf{X} 为输入向量矩阵， x_j 为第 j 项原始数据， $j=1,2,\dots,m$ 。

3.1.2 寻找获胜神经元 当网络得到一个输入模式向量时，竞争层的所有神经元对应的内星权向量均与其进行相似性比较，并将最相似的内星权向量判为竞争获胜神经元。

3.1.3 权值调整、训练 对竞争层领域内所有神经元与输入层神经元之间的连接权值进行修正，并进行训练，直至结果满意。

3.2 基于 SOFM 的初家村宅院综合分类

初家村宅院综合分类采用 SOFM 网络，具体过程在 matlab 软件中的神经网络工具箱进行。基于主成分分析确定的 5 个主成分得分为指标，确定输入层为 109 个宅院单元的 5 项指标导入宅院综合分类 SOFM 网络模型。训练分类过程：测试网络训练总步数为 5000，并将输出层神经元数量 S 依次设置为 3、4、5、6、7、8，分别训练分类；分析不同分类结果下各单元的分布特点与归类情况，并对个别宅院进行微调，最终将初家村宅院分为非农高收入、非农中收入、中收入、高利用、低收入高利用和低收入老龄宅院 5 类（图 2）。

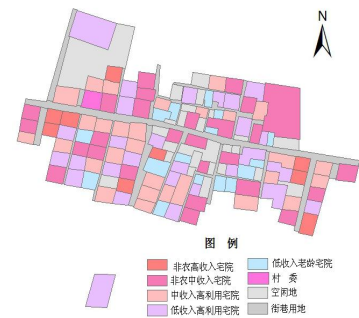


图 1 基于 PCA 和 SOFM 的初家村宅院分类结果
Fig.1 Rural houses classification results based on PCA and SOFM

4 初家村各类宅院特征分析及优化利用建议

由图 1 可以看出，初家村各类宅院空间分布特征较明显。低收入老龄宅院及空闲地主要集中在村中心，非农高收入宅院主要分布于村外围，其他 3 类分散分布。分析初家村各类宅院特征（表 2），考虑我国社会经济发展、农村新型社区建设及初家村社会经济条件，提出各类宅院优化利用建议。

表 2 初家村各类宅院特征

Table 2 Characters of each kinds of rural house in Chujiacun Village

宅院类型 House type	户数 Household	人口 Population	职业 Profession	家庭收入 Income	建筑结构 Building structure	成新度 Old grade	容积率 Plot ratio	有效居住率 Effective occupancy
非农高收入宅院	9	3.56	承揽工程等	47.56 万	砖混	7.4	0.54	0.28
非农中收入宅院	27	5.26	二、三产业、兼业	8.27 万	砖混	6.19	0.43	0.56
中收入高利用宅院	29	3.45	兼业、农业	5.77 万	砖混、砖木	7.17	0.45	0.84
低收入高利用宅院	28	2.14	农业、兼业	2.58 万	砖木、砖混	5.85	0.37	0.78
低收入老龄宅院	16	1.25	无业	0.67 万	砖木	4.94	0.41	0.50

4.1 非农高收入宅院

非农高收入宅院共 9 处，占总处数的 8.26%。住户从事承揽工程、销售等非农业，家庭收入高。房屋建筑结构基本为砖混，建筑成新度 7.4，平均容积率 0.54，房屋成新度和土地利用强度较高。宅院有效居住率为 0.28，人口利用率很低。住户已脱离农业，经济收入高，已完全有能力在城镇居住，为城镇化宅院。此类家庭在城镇已有住房，长期在城镇生活，农村宅院利用率很低。因此，应鼓励此类住户退出农村宅院，并给予经济补偿，在城镇落户，享有城镇居民各项权利。鉴于宅院建筑结构良好、成新度高，坚持低碳、节约原则，应继续利用，如村庄继续保留，可允许其有偿转让、作为老年房或公共用房等；如村庄整体迁入农村新型社区，可作为农业休息房、农具间或乡村旅馆等。

4.2 非农中收入宅院

非农中收入宅院共 27 处，占总处数的 24.77%。住户从事二三产业或兼业，家庭收入较高，平均 8.27 万元。房屋建筑结构主要为砖混，平均成新度 6.19，平均容积率 0.43，房屋成新度和土地利用强度一般。宅院有效居住率为 0.56，人口利用率一般。住户从事非农或兼业，经济收入较高，和第一类宅院相似，在村内居住时间较少，宅院利用率较低。因此，也应鼓励此类非农家庭有偿退出宅院，到城镇落户，享有城镇居民各项权利。闲置宅院 3~5 成新，成新度较低，不适合再利用，可整治、复垦或作为养殖用房。对于成新度较高的宅院，可作为老年房、公共用房或农业休息用房等。

4.3 中收入高利用宅院

中收入高利用宅院共 29 处，占总处数的 26.61%。住户务工、兼业或承包了较多耕地，家庭收入中等。房屋平均成新度和土地利用强度一般。宅院有效居住率为 0.85，人口有效利用率高。此类

住户兼业或农业,基本生活在农村,宅院利用率高。因此,村庄继续保留情况下,该类宅院可进行整治、改造,改变建筑结构、增加建筑面积,提高建筑质量和土地利用率,尤其是农业生产大户。如村庄整体搬入新社区,5成新以上、连片房屋可作为农业生产的休息用房、农具间等,而5成新以下的房屋可拆除,复垦为耕地、增加耕地面积。

4.4 低收入高利用宅院

低收入高利用宅院共 28 处,占总处数的 25.69%。住户务农(规模小)或兼业,家庭收入较低。房屋平均成新度 5.85,平均容积率 0.37,土地利用强度低。户人口较少,但有效居住率 0.78,人口有效利用率较高。此类住户从事农业或兼业,也基本生活在农村,收入较低。因此,村庄继续保留情况下,该类宅院可进行维修、改造,重点是增加建筑安全度和成新度,改善居住条件。如果村庄整体迁入新社区,则拆除宅院、复垦为耕地。

4.5 低收入老龄宅院

低收入老龄宅院共 16 处,住户均不从事农业生产。大部分宅院收入很低,其中 12 户无业,收入主要来源于承包地的流转收入。宅基地获取时间早,房屋结构多为砖木,成新度低。家庭总人口较少,其中 10 户为 1 人。该类宅院总体老龄化、利用率低,主要为无收入老年人居住,房屋老化严重,不适宜继续居住,且居住人口较少。因此,村庄继续保留情况下,可直接拆除村边的老龄宅院、复垦为耕地,村内的老龄宅院可改造为绿地、休闲广场。而原住户可无偿居住或有偿购买、租住城镇化家庭退出的宅院。如村庄整体迁入新社区,则直接拆除宅院,复垦为耕地。

5 结论与讨论

5.1 结论

(1) 农村宅院类型划分 以东营市史口镇初家村为研究区,基于宅院遥感影像解译及实地入户调查数据,综合采用 PCA 和 SOFM 模型,将初家村农户宅院划分非农高收入、非农中收入、中收入高利用、低收入高利用和低收入老龄宅院等 5 种综合类型。

(2) 宅院优化利用建议 村庄保留情景下非农低利用宅院有偿退出、住户到城镇落户、宅院继续利用,兼业、农业住户宅院整治、改造利用,老龄宅院改造为绿地或复垦为耕地。村庄迁入新社区情景下较高质量宅院继续利用、老龄宅院复垦为耕地。

5.2 讨论

综合运用 PCA 和 SOFM 划分宅院类型,得出的综合分类结果基本符合研究区情况,说明研究方法可行。今后可扩大研究区,探索研究方法在农村宅院类型划分的适用性。另外,目前我国正处于城镇化快速发展、新型户籍制度建立时期,大量农村人口转移至城镇居住、就业,但却没有落户城镇、农村宅院依然保留,农村宅基地制度与新型户籍制度不衔接。应尽快改革现有农村宅基地制度,与城镇化发展、新型户籍制度协调。

参考文献

- [1] 刘彦随,刘玉,翟荣新.中国农村空心化的地理学研究及整治实践[J].地理学报,2009,64(10):1193-1202
- [2] 姜广辉,张凤荣,周丁扬,等.北京市农村居民点用地内部结构特征的区位分析[J].资源科学,2007,29(2):109-116
- [3] 田光进,刘纪远,庄大方.近 10 年来中国农村居民点用地时空特征[J].地理学报,2003,58(5):651-658
- [4] 刘洁,王瑗玲,姜曙千,等.胶州市胶北镇 10 个村庄土地集约利用水平评价[J].农业工程学报,2012,28(25):244-249
- [5] 龙花楼.中国农村宅基地转型的理论与证实[J].地理学报,2006,61(10):1093-1100
- [6] 欧阳安蛟,蔡锋铭,陈立定.农村宅基地退出机制建立探讨[J].中国土地科学,2009,23(10):26-30
- [7] 云正明.农村庭院生态系统研究[J].农业现代化研究,1987(3):12-16
- [8] 陆志刚.发展农村庭院经济广辟农民致富新路[J].中国农业资源与区划,1988(5):51-56
- [9] 孙好勤.新农村建设背景下的庭院经济战略转型[J].农业经济问题,2006(7):54-56
- [10] 虞志淳,刘加平,雷振林.从户型到宅院组合—陕西关中地区农村住宅研究[J].建筑学报,2010(8):10-13
- [11] 傅湘,纪昌明.区域水资源承载力综合评价—主成分分析法的应用[J].长江流域资源与环境,1999,8(2):168-173
- [12] 刘玉,刘彦随,郭丽英.基于 SOFM 的环渤海地区乡村地域功能分区[J].人文地理,2013(3):114-120
- [13] Kohonen T. Self-Organizing Maps[M]. Berlin:Springer,1995