

# Analysis of Arab Israeli's Sense of National Belonging Based on K-Means Cluster Algorithm

Wenjia Liu<sup>1,\*</sup>

<sup>1</sup>School of Middle Eastern Study, Beijing International Studies University, Beijing 100024, China

\*Corresponding author. Email: liuwenjia126@126.com

## ABSTRACT

The sense of national belonging is an important feature of national identity and the core factor of the establishment of a nation state. Based on the cluster analysis method in social statistics, a quantitative analysis on the sense of belonging of Arab-Israelis is conducted in this paper. First of all, the Arab-Israelis can be divided into two groups according to age structure and educational background. The new generation of Arab-Israelis is about 30 years old and has a good education background. The older ones are 50 years old on average and with a poor education less than 10 years. Secondly, the state of national belonging of Arab-Israelis is analyzed. The results show that the state belonging of Arab-Israelis is generally good, but there is a significant difference between the old and the new generations. The sense of national belonging of the new generation of Arab-Israelis is significantly higher than that of the older generation.

**Keywords:** national belonging, K-means cluster analysis, Arabs, Israel

## 基于K-means聚类算法的阿拉伯裔以色列人国家归属感分析

柳文佳<sup>1,\*</sup>

<sup>1</sup>北京第二外国语学院, 中东学院, 中国北京, 100024

\*通信作者. Email: liuwenjia126@126.com

## 摘要

国家归属感是国家认同的重要表现形式, 是一个民族国家得以成立的核心因素。本文基于国际社会调查项目国家归属感调查数据, 采用K-means聚类分析方法, 对阿拉伯裔以色列人的国家归属感进行定量分析。首先, 从年龄结构上对阿拉伯裔进行划分, 新一代以色列阿拉伯人年龄在30岁左右, 有着良好的教育背景; 而老一代阿拉伯人年龄在平均值在50岁, 受教育程度一般, 其平均受教育年限不到10年。其次, 分析了以色列裔阿拉伯人国家归属感情况, 结果显示: 阿拉伯裔以色列人的国家归属感总体状况良好, 但新老两代阿拉伯裔以色列人的国家归属程度存在明显差异。新一代阿拉伯裔以色列人的国家归属感明显高于老一代阿拉伯裔以色列人。

**关键字:** 国家归属感; K-means 聚类分析; 阿拉伯; 以色列

## 1. 引言

归属感是佛洛姆理论中的术语, 意指心理上的安全感与落实感。归属感是与生俱来的, 个体作为社会群体的一员也会有归属感, 以其文化和价值归属感为核心<sup>[1]</sup>。国家归属感指的是源于对国家的政治、经济、文化、传统等的认可和接受, 而产生的对国家的

归属认知和感情依附。国家归属感是国家认同的重要表现形式, 是一个民族国家得以成立的核心因素。就全球范围来看, 现代民族国家的稳定与否, 与其国家内部各个族群国家认同的强度有着紧密的关联性。以色列建国以来, 阿拉伯国家和阿拉伯人与以色列的关系一直是中东地区的核心矛盾之一, 阿以关系在中东研究中的重要地位, 是观察和理解中东国际政治的一个重要维度。

目前,国内外在国家民族认同研究领域多采用简单的定量评价方式或描述性分析,不能对民族归属感进行科学、系统的分析。个人的国家归属感是其国家认同的一项重要表征,是基于社会背景及身份作出的个人主观认同,但同时也受到相同或相似族群间认同的影响。由于国家归属感是多种因素相互作用的结果,研究中也存在不少使用相关性分析的研究,但该类研究仍以定性相关性分析为主,以简单统计为辅,缺乏数据和量化指标支持。

随着计算技术的快速发展,多元统计分析方法已被广泛应用于多个领域,如医学、金融、电商等。但该类统计方法在国家归属研究中尚未得到充分利用,聚类分析是一种用于分析多变量的技术。其根据研究对象多样性的相似程度将其聚合在一起,相似度大的优先聚合,最终按照类别的综合性质将多个品种聚合,进而完成聚类分析的过程<sup>[1]</sup>。

为揭示以色列阿拉伯人国家归属的关联性,为后续阿拉伯人国家认同研究提供直观量化的依据,结合已有研究和数据分析,本文基于聚类分析方法提出了以色列阿拉伯人国家归属感分析模型,以此阿拉伯人国家归属感现状进行了定量探讨和研究。

首先,本文基于国际社会调查项目(International Social Survey Programme, ISSP)所提供的国家归属调查结果数据库进行了数据筛选,获得以色列裔阿拉伯人国家归属调查问卷结果样本174条;然后,对数据集进行预处理,重点进行了评价数据属性的有效性分析与过滤;最后,通过K-means 聚类分析算法,基于年龄结构及教育背景对阿拉伯裔进行层次划分,分析了不同层面以色列裔阿拉伯人国家归属感的差异性,为以色列裔阿拉伯人国家认同研究奠定一定的理论基础。

## 2. 样本数据集

### 2.1. ISSP简介

国际社会调查项目<sup>[2]</sup>是一个跨国合作调查项目,ISSP成立于1984年,是自发的非政府非盈利学术组织,旨在推动社会科学研究的标准化和国际化,协调跨国跨地区的联合项目,建立共享的社科数据库。目前会员国已经扩大到48个国家和地区,中国自2007年成为会员国。该项目影响广泛,对跨国比较研究具有重要价值。其中ISSP在1995、2003、2013年先后三次开展了关于国家认同的相关调查,以色列分别在2003年和2013年参加ISSP关于国家认同的调查研究,由于以色列社会构成的复杂性,阿拉伯裔和犹太裔分别作为不同的调查对象展开调查。本文采用了ISSP的2013年国家认同模块收集的数据,该数据集来自2013年冬季在以色列收集的数据。采用多阶段随机抽样的方法从以色列人口(居住在以色列至少一年)中选择18岁以上的受

访者。面对面的采访在他们的家中进行,采访用希伯来语、阿拉伯语和俄语进行,时长30-45分钟。调查表主要包括关于身份、民族自豪感和对文化多样性、移民等的态度等项目。

### 2.2. 国家归属感调查问题

2013年ISSP以色列国家认同数据集共有368个变量,共为174名阿拉伯裔受访者,受访者年龄分布如下图所示。

表1 受访者年龄分布

	个案数	范围	最小值	最大值	平均值	标准差
阿拉伯受访者年龄	174	67	18	85	37.67	16.154

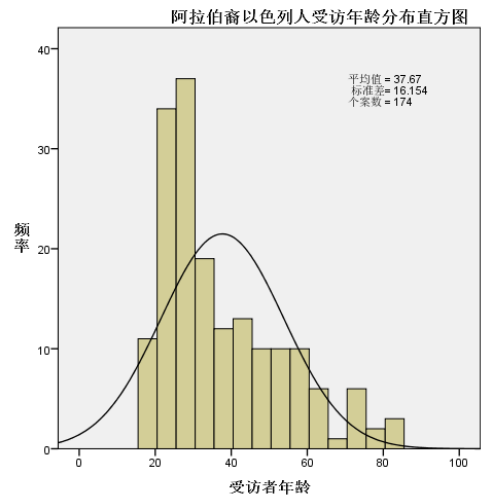


图1 阿拉伯受访者年龄分布直方图

本文提取了其有以色列阿拉伯人涉及的国家归属的调查的相关问题,得出对应聚类属性:受访者的年龄、受教育情况以及有关国家归属的六个问题,问题内容如下表所示。

表2 聚类分析国家归属感涉及问题

问题编号	问题内容
a	How important: born in [Country]?
b	How important: have [Country Nationality] citizenship?
c	How important: living most of life in [Country]?
d	How important: able to speak [Country Language]?
e	How important: to feel [Country Nationality]
f	How important: to have [Country Nationality] ancestry

每个问题对应选项分别为: 1) 非常重要; 2) 一般重要; 3) 不是非常重要; 4) 一点也不重要。

### 2.3. 数据与编码

本文以ISSP数据作为研究基础，包含受访者的年龄、受教育情况、国内出生重要性、国籍重要性、国内居住重要性、能过说本国语言重要性、感觉到国籍的重要性、拥有本国先祖的重要性。为便于后续分析，将调查的数据结果进行预处理，，可得样本集为：

$$U_i = \{u_{i1}, u_{i2}, \dots, u_{ik}\} \quad (1)$$

其中， $i \in [1, N]$ ， $N$  为描述集合大小，即受访者人数， $N = 174$ ； $u_{ik}$  为描述集合中第  $i$  个样本  $u_i$  的第  $k$  个属性值。 $k = 1$  表示受访者年龄； $k = 2$  表示受访者受教育年限； $k = 3 \sim 8$  分别表示受访者国内出生重要性、国籍重要性、国内居住重要性、能过说本国语言重要性、感觉到国籍的重要性、拥有本国先祖的重要性。

### 2.3. 数据预处理

原始数据中，国内出生重要性、国籍重要性、国内居住重要性、能过说本国语言重要性、感觉到国籍的重要性、拥有本国先祖的重要性均为属于文本类型，不能被多因子分析方法<sup>[4]</sup>直接使用。因此，本文选用 One hot 编码<sup>[5]</sup>将文本数据转换成数值型，如表3所示

表 3 代码属性转换表（以 k=3 为例）

文本类型	表格属性	数值
国内出生重要性	非常重要	0
	一般重要	1
	不是非常重要	2
	一点也不重要	3

本文采用人工标定方法对样本集进行标定，形成标定的数据集，来更好地评价分析模型。

## 3. 基于K-means聚类分析算法的国家归属感分析

### 3.1. K-means聚类分析算法

在上述数据预处理的基础上，将原始数据集数据采用K-means算法进行聚类分析，其算法流程如下：

1) 将收集好的以色列阿拉伯裔国家归属特征数据转化为向量预先保存到文本中。

2) 构建随机质心O（中心点），并设置k=2，即将数据集分为“老一代以色列阿拉伯人”、“新一代以色列阿拉伯人”2个簇。随机质心O在整个数据集的边界之内，这可以通过找到数据集每一维的最小和最大值来

完成，然后生成 0~1.0之间的随机数并通过取值范围和最小值，以便确保随机点在数据的边界之内。

3) 基于上述初始化过程，构建K-means算法。创建K个质心，然后将每个点分配到最近的质心，再重新计算质心。这个过程重复数次，直到数据点的簇分配结果不再改变位置。返回类质心与点分配结果。停止条件为没有需要分配的任务到不同的簇，质心不再发生变化，或者均方误差E的值下降幅度最小。E的计算公式为：

$$E = \sum_{k=1}^K \sum_{x \in c_k} d(x, m_k)^2 \quad (2)$$

其中， $c_k$  为第k个簇； $m_k$  为簇  $c_k$  的质心； $d(x, m_k)$  为x和质心  $m_k$  之间的距离。欧氏距离是每个目标点到簇中心的距离：

$$d(x, m_k) = \|x, m_k\| = \sqrt{(x_1 - m_k)^2 + (x_2 - m_k)^2 + \dots + (x_n - m_k)^2} \quad (3)$$

### 3.2. K-means聚类分析结果

采用K-means聚类算法基于年龄大小，受教育年限对受访者进行年龄结构划分，其结果如下：

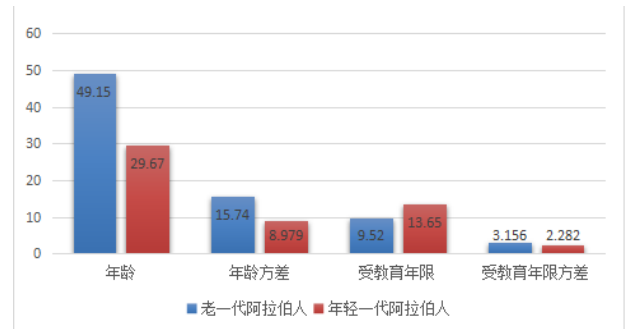


图 2 新老两代阿拉伯裔以色列人分布对比图

以受访者的年龄和受教育年限作为分类的质心，第一类属于新一代的阿拉伯裔，年龄在平均值在29.67岁，标准差为8.979岁，有着良好的教育背景，其平均受教育年限为13.65年；第二类属于老一代的阿拉伯裔，其年龄在平均值在49.15岁，标准差为15.740岁，受教育程度一般，其平均受教育年限不到10年。

关于新老阿拉伯人对于以色列国家归属感聚类分析结果如下图所示：

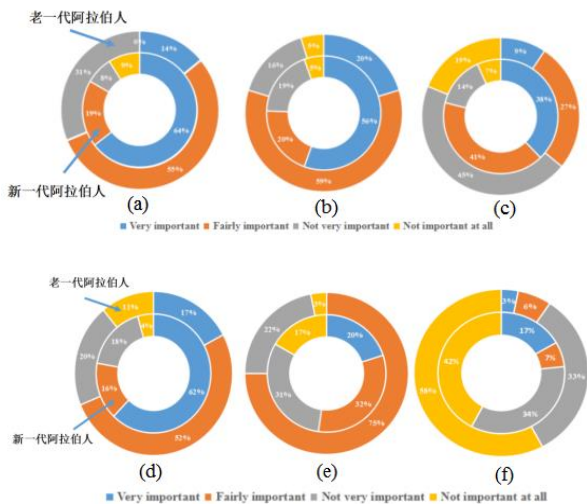


图3 新老阿拉伯裔以色列人对社会认同的对比

可以看出，新老阿拉伯裔以色列人在关于国家归属感的问题的具有一定的差异性，观念理解略有不同。例如，调查项a：在以色列出生的重要性时，无论是老一代阿拉伯裔以色列人还是新一代的阿拉伯都认为在以色列出生是很重要的。但是，具体上还有很大的差异性，在老一代的阿拉伯裔以色列人中只有14%的人认为该项非常重要，但是在新一代阿拉伯裔以色列人中，这比例高达62%。这是由于早期以色列对老一代阿拉伯裔以色列人采取的政策缺乏公平性，导致其遭受了一定的不平等待遇。随着以色列对阿拉伯裔以色列人的政策缓和，新一代阿拉伯裔以色列人的国家归属感也随之增加。对于问题c：在以色列居住时间的重要性是，新老阿拉伯裔以色列人的对于此反馈的差异性显著。在老一代阿拉伯裔以色列人中，36%的受访者认为在一生的大部分时间生活在以色列是重要的，45%的受访者认为其不是那么重要，此比例占老一代阿拉伯裔以色列人的绝大部分。而在新一代阿拉伯裔以色列人中，79%的新一代阿拉伯受访者认为在一生的大部分时间生活在以色列是重要的，认为不那么重要的比例只有可怜的14%。这一组数字的对比充分表现出以色列阿拉人更加倾向于在以色列生活居住，也说明了阿拉伯裔以色列人的国家归属感的显著提升。调查问题e：感觉到以色列国籍的重要性时，在老一代阿拉伯裔以色列人有75%的老一代受访者认为该项目是比较重要的，而在新一代阿拉伯裔以色列人中这一比例降到了52%，另外有31%的新一代阿拉伯裔以色列人认为感受到以色列国籍不是很重要。值得注意的是，对于问题f：拥有以色列祖先的重要性时，无论是老一代还是新一代阿拉伯裔以色列人绝大多数受访者都认为拥有以色列祖先是不重要的，这也说明阿拉伯裔以色列人即使生活在以色列境内，仍然保持着对自己的阿拉伯祖先的强烈的认同感。

#### 4. 结论

本文基于聚类分析对阿拉伯裔以色列人的国家归属感进行了定量分析，首先，从年龄结构上对阿拉伯裔进行划分，新一代以色列阿拉伯人年龄在30岁左右，有着良好的教育背景，而老一代阿拉伯人年龄在平均值在49.15岁受教育程度一般，其平均受教育年限不到10年。其次分析了以色列裔阿拉伯人国家归属感情况，结果显示：阿拉伯裔以色列人的国家归属感总体状况良好，但新老两代阿拉伯裔以色列人的国家认同程度存在明显差异。新一代阿拉伯裔以色列人的国家认同明显高于老一代阿拉伯裔以色列人，这表明以色列国家认同建设取得了一定成效。但鉴于阿拉伯裔以色列人身份的特殊性，这种国家认同将来还会随着以色列内部条件和外部环境的变化而改变，从而影响以色列的国家安全和社会稳定，乃至影响整个中东地区的和平与发展。

#### REFERENCES

[1] Shangtao Gao. A Study of Arab-Israeli Relations from the Perspective of Power Constructivism [M]. Beijing: World Affairs Press, 2018

[2] Sarbu C, Nascu-Briciu R D, Kot-Wasik A, et al. Classification and fingerprinting of kiwi and pomelo fruits by multivariate analysis of chromatographic and spectroscopic data[J]. Food Chemistry, 2012, 130(4):994-1002.

[3] <http://www.issp.org/menu-top/home/>

[4] Xiaoyang Chen. Semantic Similarity Calculation of Short Text Based on Deep Learning[D]. Beijing: Beijing University of Technology, 2015.

[5] RODRIGUEZ P, BAUTISTA M A, GONZALEZ J, et al. Beyond one-hot encoding: lower dimensional target embedding[J]. Image and vision computing, 2018, 75:21-31.

[6] Polykovskiy, D. and Novikov, A., Bayesian Methods for Machine Learning[J]. Coursera and National Research University Higher School of Economics, 2018.

[7] The Arab Population in Israel[J]. Israel Central Bureau of Statistics, Retrieved 9 July 2016.

[8] Jacob M. Landau. The Arab minority in Israel, 1967-1991: political aspects [M]. London: Oxford University Press, 1993.

[9] Wenjia Liu. A Study on The National Identity of Arab Israelis [D]. Beijing: Beijing Language and Culture University, 2019