

# 基于因子-聚类分析的大学生就业 满意度统计及预测

吴亚娟<sup>1</sup>

## 摘要

引用满意度指数的概念,根据实际满意度调查的数据,利用多元统计分析的方法,对2010届毕业生选择的第一份工作的满意度进行了实证分析,得出影响毕业生就业满意度的主要因子分别为“单位内部因子”、“个体发展因子”、“外显所得因子”,其中起主要影响的因素是“单位因子”,三者的影响程度依次递减,同时对2010届大学毕业生对工作的满意程度进行了预测。

## 关键词

大学生;就业;满意度;因子分析;聚类分析

中图分类号 O212.4

文献标志码 A

## 0 引言

### Introduction

用户满意度指数(Customer Satisfaction Index, CSI)是目前许多国家积极研究和使用的—种新型宏观经济指标,是专门用来测量用户对产品和服务满意程度的质量指标,用来反映产品和服务的质量状况<sup>[1]</sup>,目前美国已经完成了行业覆盖面最广泛的用户满意度指数体系。

就业是民生问题,它关系到社会的稳定 and 经济发展。近年来,大学生就业观念和就业环境正日益发生变化,毕业生择业过程中也呈现出许多新的特点,影响因素也越来越多<sup>[2]</sup>。为进一步了解当代大学生的就业观念,清楚的反映当代大学生就业方面的要求,协调好社会、学校、学生三者间的关系,引用用户满意度指数的理论和方法,研究大学生对就业的满意度是可行的,也是必要的。

## 1 因子分析与聚类分析

### Factor analysis and cluster analysis

### 1.1 因子分析

因子分析是主成分分析的推广和发展,它将具有错综复杂关系的变量(或样品)综合为数量较少的几个因子,再现原始变量与因子之间的相互关系,同时根据因子的不同还可以对变量进行分类的一种多元统计分析方法<sup>[3]</sup>。

因子分析的基本过程:

1) 因子分析的前提条件鉴定。考察原始变量之间是否存在较强的相关性,是否适合进行因子分析。因为因子分析的主要任务之一就是対原有变量中信息重叠的部分提取和综合成因子,最终实现减少变量个数的目的,所以要求原有变量之间应存在较强的相关关系。否则,如果原有变量相互独立,不存在信息重叠,也就无需进行综合和因子分析。

2) 因子提取。研究如何在样本数据的基础上提取综合因子。

3) 因子旋转。通过正交旋转或斜交旋转提取出的因子具有可解释性。

4) 计算因子得分。通过各种方法求解各样本在各因子上的得分,为进一步分析奠定基础。

收稿日期 2010-06-12

作者简介

吴亚娟,女,硕士,讲师,研究方向为数学教学论。wuyajuan2002@sohu.com

<sup>1</sup> 南京信息工程大学 数理学院,南京,210044

## 1.2 聚类分析

聚类分析(Cluster Analysis)是根据事物本身的特性研究个体的一种方法,目的在于将相似的事物归类.它的原则是同一类中的个体有较大的相似性,不属于同一类的个体之间差异性很大.聚类分析的大部分应用都属于探测性研究,最终结果是产生研究对象的分类,通过对数据的分类研究还能产生假设.聚类分析也能用于证实性目的,对于通过其他方法确定的数据分类,可以应用聚类分析进行检验<sup>[4]</sup>.

聚类分析有利于处理随意和主观性的不确定问题,并达到归类效果,而因子分析法是用于将诸多影响因素归为几个主要指标的统计方法,故将其相结合用于大学生就业满意度的研究与分析.

本文采用了K-均值聚类法<sup>[5,6]</sup>,这种算法的基本思想是将每一个样品分配给最近中心(均值)的类中,具体算法至少包括以下3个步骤:

1) 从 $n$ 个数据样本中任意选择 $k$ 个对象作为初始聚类中心,对于剩下的样本,根据其到聚类中心的距离,分配给距离最近的类;

2) 计算获得每个新类的中心,根据各个样本与聚类中心的距离,重新分配各个样本归属类别;

3) 不断重复上述过程,直到样本所属类别达到稳定,或者达到指定的循环次数.

衡量样本间或者类与样本间距离的方法常用的是欧氏距离.假设有 $p$ 个变量,则样本 $i(x_{i1}, x_{i2}, x_{i3}, \dots, x_{ip})$ 与样本 $j(x_{j1}, x_{j2}, x_{j3}, \dots, x_{jp})$ 间的距离 $d$ 计算公式如下:

$$d(x, y) = \sqrt{(x_{i1} - x_{j1})^2 + \dots + (x_{ip} - x_{jp})^2}.$$

## 2 实例分析

### Case study

### 2.1 数据来源

本研究中就业满意度评价体系列举了现实工作中的18个评价指标,具体为签约单位的性质 $X_1$ ;单位在同行业的竞争力及实力 $X_2$ ;工作地点及周围环境 $X_3$ ;单位人员结构及用人机制 $X_4$ ;规章制度 $X_5$ ;单位在这次聘用的新人情况(学历、毕业学校、户籍等) $X_6$ ;与专业的对口 $X_7$ ;工作强度 $X_8$ ;硬件设施 $X_9$ ;实习期间人际关系 $X_{10}$ ;薪酬及福利待遇 $X_{11}$ ;在单位里拥有的个人发展空间 $X_{12}$ ;对所在单位工作领域获得的社会声望 $X_{13}$ ;单位能够提供的培训计划或进修机会 $X_{14}$ ;能够对社会做出的贡献 $X_{15}$ ;实现自我 $X_{16}$ ;工作的挑战性 $X_{17}$ ;工作的创造性 $X_{18}$ .每个评价指标

分别按照非常满意、满意、基本满意、不满意和非常不满意5个等级要求毕业生评判,评价结果量化分别对应5分、3分、1分、1/3分和1/5分.

在满意度调查中,通过校内纸质调查问卷和校外网络调查,随机抽取了不同地区、不同学校、不同专业的2010届毕业生进行工作满意度调查评价.为保证评价信息的效度和信度,校外网络调查在校园和高校论坛里进行,评价时间选择安排在毕业生签约后实习期间,这样保证对用人单位基本情况有所了解.本研究采用随机抽样的方法,以100名毕业生的评价数据为样本.

### 2.2 因子分析结果

用SPSS软件进行数据处理.

#### 1) 可行性检验

首先对原始数据进行主成分分析可行性检验,由KMO和Bartlett检验结果知,KMO值为 $0.905 > 0.6$ ,同时Bartlett检验的伴随概率为0,小于显著水平0.05,以上检验结果都说明原始数据能够进行主成分分析.

#### 2) 方差分解

接下来对原始数据进行总方差分解,在此设置提取的因子对应的特征值界限为1,因子提取和因子旋转的分析结果表明,采用主成分分析提取公因子后,前3个变量特征值的累积贡献率为79.193%,这说明前3个特征值包含要调查信息的79.193%.所以可以选择前3个特征值对应的向量为公共因子.

#### 3) 因子旋转

将评价指标体系中的18个评价指标划分为3类,再对样本进行分析.为了清楚解释该3类因素的构成,可对因子载荷进行方差最大化旋转,得到旋转前后矩阵.从旋转前的因子载荷矩阵中,每个因子在不同原始变量上的载荷没有明显的差别.为了便于对因子进行命名,还需对因子载荷阵进行旋转,而后得出旋转后的因子载荷阵,经过旋转后的载荷系数已近明显地两级分化了.根据因子载荷的大小(在此取载荷因子 $>0.5$ ),第1个公共因子( $F_1$ )在指标 $X_1, X_4, X_5, X_6, X_8, X_9, X_{10}$ 上有较大载荷,说明这7个指标有较强的相关性,可以归为一类,从整体看这7个指标属于单位内部工作实际环境的指标,因此可以把第1个因子命名为“单位内部因子”;第2个公共因子( $F_2$ ) 在指标 $X_7, X_{12}, X_{14}, X_{16}, X_{17}, X_{18}$ 上有较大载荷,同样可以归为一类,这6个指标多数属于毕业生后续发展的指标,因此可以把第2个因子命名为

“个体发展因子”；同样  $X_2$ 、 $X_3$ 、 $X_{11}$ 、 $X_{13}$ 、 $X_{15}$  归为第 3 类，与工作所带来的经济、地位等变化有关，因此把第 3 个因子 ( $F_3$ ) 命名为“外显所得因子”。

从以上可以得出这样的结论，即影响毕业生就业满意度受多方面因素影响，但可将若干因素归结为 3 个方面，按影响力的大小顺序依次为  $F_1$ “单位内部因子”、 $F_2$ “个体发展因子”、 $F_3$ “外显所得因子”。

#### 4) 因子分值计算

通过以上研究可以发现影响毕业生就业质量的主要因素，接下去计算这 3 个因子的分值。首先计算这 3 个因素的因子得分系数矩阵，可以得出因子的得分表达式如下：

$$F_2 = 0.129X_1 - 0.041X_2 + 0.086X_3 + 0.272X_4 + 0.322X_5 + 0.240X_6 - 0.148X_7 + 0.189X_8 + 0.214X_9 + 0.289X_{10} + 0.024X_{11} - 0.051X_{12} - 0.205X_{13} - 0.108X_{14} - 0.169X_{15} - 0.051X_{16} - 0.150X_{17} - 0.062X_{18},$$

$$F_2 = -0.084X_1 - 0.116X_2 - 0.160X_3 - 0.048X_4 - 0.109X_5 - 0.181X_6 - 0.162X_7 - 0.049X_8 - 0.057X_9 + 0.073X_{10} - 0.012X_{11} + 0.236X_{12} + 0.048X_{13} + 0.198X_{14} + 0.140X_{15} + 0.314X_{16} + 0.325X_{17} + 0.400X_{18},$$

$$F_3 = 0.084X_1 + 0.319X_2 + 0.201X_3 - 0.130X_4 - 0.129X_5 + 0.049X_6 - 0.480X_7 - 0.035X_8 - 0.046X_9 - 0.296X_{10} + 0.134X_{11} - 0.056X_{12} + 0.335X_{13} + 0.047X_{14} + 0.186X_{15} - 0.145X_{16} - 0.033X_{17} - 0.226X_{18}.$$

其中  $X_i$  ( $1 \leq i \leq 18$ ) 也表示第  $X_i$  个就业满意指标的得分值。取前 10 位毕业生就业满意指标的得分代入表达式，计算各因子得分值。以每个公共因子所对应的特征值占所提取公共因子总的特征值之和的比例作为权重计算综合得分模型：

$$F = \frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3} F_1 + \frac{\lambda_2}{\lambda_1 + \lambda_2 + \lambda_3} F_2 + \frac{\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3} F_3.$$

其中  $\lambda_1$ 、 $\lambda_2$ 、 $\lambda_3$  分别为各对应因子的特征值。依此式计算 10 名毕业生综合得分值，并对其按综合主成分值进行排序，即可对每个毕业生进行满意度综合评价。

由表 1 可以看出，满意度最高的是学生 4，最低的是学生 6，这符合毕业生实际情况。通过这个实例证明采取因子分析法对毕业生就业满意度的探讨是可行的。

表 1 毕业生满意综合评价表

Table 1 Graduate satisfaction evaluation

学生	因子综合评价值 $F$	排名
1	0.118 6	7
2	0.412 6	5
3	-0.399 6	8
4	1.361 1	1
5	0.160 7	6
6	-0.562 0	10
7	0.737 2	3
8	0.621 5	4
9	1.270 8	2
10	-0.501 7	9

## 2.3 聚类分析结果

利用因子分析法把 18 个影响因素降维到 3 个主要因子，其累积贡献率为 79.193%，在聚类过程中只取这 3 个因子进行聚类分析。

由 SPSS 自动给定 K-均值迭代的初始类中心坐标，从 5 个初始类的中心开始进行 K-均值迭代，每次迭代都将导致类中心的变化，同时每次迭代导致的类中心变化量在逐渐减少，第 9 次迭代导致的类中心的变化量已经为 0，达到了收敛。至此 100 名毕业生被分成 5 类。

第 1 类包括 15 名学生，其类中心 3 个主因子分值分别为 -0.771 61, 1.251 25, 0.703 57。依大小顺序排列可知，主因子  $F_2$  分值最大，可以理解为这类毕业生中该因子满意度较高，其次是因子  $F_3$ ，满意度最低的是  $F_1$ ，从这个数据可以看出，这一类的毕业生对单位实际工作环境满意程度较弱，应值得用人单位和有关部门采取措施，提高这类毕业生的满意度。

第 2 类包括 42 名学生，其类中心 3 个主因子分值分别为 -0.233 60, -0.117 53, -0.930 75。由于 3 个因子分值都是负数，这一类毕业生在这 3 个因子方面满意度普遍不高，且比重较大，学校应该加强就业方面的指导工作。

第 3 类包括 28 名学生，其类中心 3 个主因子分值分别为 -0.299 67, -0.796 14, 0.564 41，满意的程度依大小分别为  $F_3$ 、 $F_1$ 、 $F_2$ ，这类毕业生对单位情况及自我的后续发展空间不够满意。

第 4 类包括 2 名毕业生，其类中心 3 个主因子分值分别为 3.219 09, -2.339 71, 1.328 72。数据很明显， $F_1$  非常高，而因子  $F_2$  却很低，可知这一类毕业生对自己的评价和预期可能有些误差，应积极调

整发展方向,改变对已有工作的某些看法和认识。

第5类包括13名学生,其类中心3个主因子分值分别为1.116 00,1.011 67,0.775 16。从这组数据可知,这类毕业生各项因子满意程度比较均衡,应当做好追踪毕业生满意度工作,维持好毕业生的平均满意度。

### 3 结语

#### Conclusion

因子分析法结果表明,高校2010届毕业生就业满意度影响因素可分为3大类,它们分别是“单位内部因子”、“个体发展因子”、“外显所得因子”,其中起主要影响的因素是“单位因子”,依次影响程度递减。可见,当高等教育从精英教育向大众教育的转型过程中,大学生择业观也逐渐趋于现实和理性,将“个体的可持续发展”放到了重要位置,这点在当代中国社会转型期——农业社会向工业社会、知识经济社会转变中显得尤为重要。

结合聚类分析后的结果数据可以看到,目前大学生对就业的满意度普遍不高,有42/100的学生对上述3个指标都未能表示满意,这就要求各就业主管部门和高校把毕业生就业工作摆在重要议事日程,继续开辟新的就业渠道,进一步完善促进毕业生就业的政策,努力形成长效机制。

该调查的数据主要来源于网络调查,有来自江苏、山东、四川、广东、北京、上海、陕西等省份的毕业

生参与调查,区域分布较为广泛,但是忽略了高校的类型、学校所处地理位置、学生的专业以及性别差异等影响因素,这些问题有待进一步研究和讨论。

### 参考文献

#### References

- [1] 梁燕. 关于顾客满意度指数的若干问题研究[J]. 统计研究, 2003(11): 52-56  
LIANG Yan. Some aspects of customer satisfaction index[J]. Statistical Research, 2003(11): 52-56
- [2] 武毅英. 高校毕业生就业问题的教育学审视[M]. 厦门: 厦门大学出版社, 2006  
WU Yiyang. The education review about employment of graduates [M]. Xiamen: Xiamen University Press, 2006
- [3] 李静萍, 谢邦昌. 多元统计分析方法与应用[M]. 北京: 中国人民大学出版社, 2008: 21-60  
LI Jingpin, XIE Bangchang. Multivariate statistical analysis method and application [M]. Beijing: China Renmin University Press, 2008: 21-27, 51-60
- [4] 肖宜滨. 聚类分析的理论及其应用[J]. 江苏统计, 2001(11): 13-15  
XIAO Yibin. Theory and application of cluster analysis[J]. Jiangsu Statistics, 2001(11): 13-15
- [5] 李柏年. 模糊数学及其应用[M]. 合肥: 合肥工业大学出版社, 2007: 54-60  
LI Bonian. Fuzzy mathematics and its application[M]. Hefei University of Technology Publishing House, 2007: 54-60
- [6] 勺海(北京)市场研究公司. 主成分分析和聚类分析在满意度研究中的应用[EB/OL]. (2009-12-02) [2010-05-23]. <http://www.datasea.com.cn/article/detail.jsp?aid=79>  
Datasea Marketing Research. Application of principal component analysis and cluster analysis in satisfaction research [EB/OL]. (2009-12-02) [2010-05-23]. <http://www.datasea.com.cn/article/detail.jsp?aid=79>

## Undergraduate employment satisfaction statistics & evaluation based on factor analysis and cluster analysis

WU Yajuan<sup>1</sup>

1 School of Mathematics and Physics, Nanjing University of Information Science & Technology, Nanjing 210044

**Abstract** Finding satisfactory jobs is pursued by undergraduates, the university and the whole community, but what is the real picture of the graduate employment? Using the concept of satisfaction index and method of multivariate statistical analysis, this paper analysed the 2010 graduates' satisfaction on their first job based on the investigative statistics. Results showed that the influence factors of employment satisfaction mainly include company environment, individual promotion and income level, and the three factors exert depressive influence on respondents' satisfaction.

**Key words** undergraduate; employment; satisfaction; factor analysis; cluster analysis