

基于主成分 logistic 回归模型的餐饮商务信息意识研究^{*}

李红霞^{a, b}

(重庆工商大学 a.管理学院; b.电子商务及供应链重庆市重点实验室 重庆 400067)

摘 要: 数据挖掘技术日趋成熟,广泛应用于金融、地产、投资、评估等各个领域。数据挖掘技术亦可应用于餐饮业,为其经营决策做出分析。其他领域的数据挖掘的成功案例也可以引用到餐饮行业中。本文对餐饮人士在餐饮商务中是否具备信息意识、对信息意识的重视程度进行调查,利用数据挖掘中主成分分析和 logistic 回归处理和分析收集来的数据,从而说明数据挖掘在餐饮业管理中的重要意义,反映餐饮人士信息意识状况。利用数据挖掘技术对餐饮商务资讯进行管理势必提高企业效率和盈利水平,促进餐饮业健康稳定的发展。

关键词: 数据挖掘; 餐饮商务; 主成分分析; logistic 回归; 信息意识

中图分类号: F719.3 **文献标志码:** A **文章编号:** 1672-0598(2016)04-0056-06

引言

近年来,数据挖掘引起了信息产业界的极大关注,其主要原因是存在大量数据,可以广泛使用,并且迫切需要将数据转换成有用的信息和知识。数据挖掘是从大量数据中挖掘出隐含的、先前未知、对决策有价值并能被人们所理解的模式(知识)的高级数据处理过程(胡家声,2012)^[1]。数据挖掘的任务主要是关联分析、聚类分析、分类、预测、时序模式和偏差分析等。数据挖掘技术日趋成熟,广泛应用于金融、地产、投资、评估等各个领域。数据挖掘技术亦可应用于餐饮业,为其经营决策做出分析。其他领域的数据挖掘的成功案例也可以引用到餐饮行业中。

我国现有各类饭店、餐馆数量庞大,可是,据相

关调查显示,大部分餐馆对于信息化管理水平的运用仍处在初级摸索阶段。许多餐饮管理者,由于缺乏对先进管理理念的理解,以及对信息化的认识 and 意识不够,再加上目光短浅舍不得投入,致使餐馆的管理,至今仍停留在一个比较原始和落后的水平上,餐馆的发展因此很受限。在科技飞速发展的今天,餐饮管理者也应当与时俱进,不断提高自己对信息化管理的先进意识,只有这样,才能使餐馆在竞争中不断发展和前进。本文对大中型餐饮企业进行了餐饮商务问卷调查,并进行数据挖掘,以研究餐饮老板以及餐饮职业经理人等餐饮人士在餐饮商务中是否具备信息意识,对信息意识的重视程度以及信息意识在企业战略角度上的影响。

数据挖掘技术日趋成熟,广泛应用于金融、地

^{*} [收稿日期] 2016-01-21

[基金项目] 重庆市教委科学技术项目(KJ1400631); 重庆市教育规划项目(2014-GX-092); 重庆市企业管理研究中心开放基金项目; 重庆市教委人文社会科学项目(14SKH07)。

[作者简介] 李红霞(1976—),女,研究生,重庆工商大学管理学院副教授,电子商务及供应链系统重庆市重点实验室研究员,主要从事信息管理研究。

产、投资、评估等各个领域,尤其在国内外,数据挖掘理论被运用到多个领域进行了研究,如国内学者黄解军^[2](2004)基于数据挖掘进行了电子商务的策略研究,他将数据挖掘运用在企业资源计划、客户关系管理、产品数据管理、商业信用评估等方面,有利于促进企业开拓市场,优化企业资源,提高企业的经营效率和管理水平。而后姚淼^[3](2009)介绍了数据挖掘在电子商务中的运用,将关联分析、序列模型分析、分类分析、聚类分析应用在电子商务营销、电子商务网站系统和安全、客户管理、网络广告等方面,通过对电子商务过程中的各种数据和信息的挖掘,为商务活动的具体实施提供决策基础,使得电子商务能够真正的更快、更好、更高效地发展。刘红光^[4](2011)则定性描述了数据挖掘在网络信息资源管理中的应用,主要包括优化组训方法及教学训练手段、合理设置教学训练内容、转变学院学习行为模式、科学评价教学训练效果、改进信息化教育训练系统,提出让技术真正与教育训练整合在一起,对提高军事教育训练的质量和效益意义深远。不仅在国内,国外同样有学者将数据挖掘运用在各种商业行为中,取得了丰硕的成果。滕广青、毛英爽^[5](2005)从商业机构和非营利组织两大领域出发,总结和分析了国外近年来数据挖掘技术在零售、金融、电信、电子商务、工业生产和生物与医学、天文与气象、地理与环保等领域的应用,并发现随着数据挖掘的迅猛发展,其被运用的领域数量和在各种领域的运用次数逐年上升。可见随着数据挖掘技术应用范围的不断扩展,人类社会的方方面面几乎都会涉及数据挖掘。尽管数据挖掘原本是作为一项技术出现的,但由于数据挖掘本身独有的理念给人们处理解决各类问题都提供了一个新的思路和方法,在这一点上,数据挖掘一定程度上等同于一种方法论,在未来的一段时期里必将对人类生产生活产生重大影响。

数据挖掘技术已经发展到了一定的程度,其不仅运用在金融、电子商务、地产、投资等热门行业,还被运用在餐饮业的各个方面。周大兴^[6](2012)将数据分析应用在餐饮企业经营中,以小南国和至味馆两个餐饮公司为实例,建立成本回归模型,对餐饮企业的场地、能源成本、人力资源进行了研究。数据挖掘运用在餐饮企业的经营策略中,运用关联

规则和聚类分析,研究了影响顾客点菜的因素,并且对顾客进行了分类,从而有利于企业决策者作决策(孔令娟,2012;付婷,2015)^[7,8]。本文通过调查问卷,收集餐饮企业的管理数据,涉及菜品的形式、服务、采购、经营,等等。为了从海量数据中挖掘信息,本文基于数据挖掘中的主成分分析的 logistic 回归模型,定性研究餐饮商务信息意识,分析了餐饮的各种商务信息对决策者进行决策的影响。因此,餐饮数据挖掘技术就是在这些信息中查找到有效的信息用来帮助监管部门及投资部门进行有效监督和投资管理,可以帮助餐饮管理者分析客户以往的需求趋势并预测未来。

一、研究方法

(一) 主成分分析

设模型样本数 n , $X = (X_1, X_2, \dots, X_m)^T$ 是一个 m 维随机向量,是问卷中每个问题调查者的选项, m 是选取的题目的个数。主成分可以从 $n \times m$ 矩阵的特征向量来推算,令 λ 是相应的特征值, Π 是相应的特征向量,则第 k 个主成分等于

$$Y_k = \prod_k^T X = \pi_{1k}X_1 + \pi_{2k}X_2 + \dots + \pi_{mk}X_m \quad (1)$$

Y_k 的方差以及 Y_k 和 Y_l 之间的协方差等于

$$\text{Var}(Y_k) = \prod_k^T \sum \prod_k (k=1, 2, \dots, m) \quad (2)$$

$$\text{Cov}(Y_k, Y_l) = \prod_k^T \sum \prod_l (k, l=1, 2, \dots, m \text{ 且 } k \neq l) \quad (3)$$

要使得式(2)最小,求解可得到其特征根。由于特征值 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k \dots \geq \lambda_m$ 都是非负数,因此求得其特征向量。再将其代入式(1)就可以求解主成分。如果在模型自变量过多的情况下,假如较少的 k 个主成分就能够解释原始变量的大部分方差,就说明这 k 个主成分可以在原始信息丢失很少的情况下对原有数据进行替换,并且这 k 个主成分之间是不相关的。

(二) 二项 logistic 回归模型

在餐饮企业是否重视商务信息的研究中,因变量是二分类的,且其取值是在 0—1 之间的,违背了一般线性回归的基本前提假设。因此,本文采取 logit 转换,使其取值在整个实数集内,采用二项 logistic 回归方法研究餐饮企业是否重视商务信息

与各主成分之间的关系。因此,本文建立的二项 logistic 回归方程为:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 F_1 + \beta_2 F_2 + \cdots + \beta_k F_k \quad (4)$$

其中 P 是餐饮企业重视餐饮商务信息的概率,上式右边表示餐饮企业重视餐饮商务信息的分类值。其中 F_1, F_2, \cdots, F_k 是餐饮企业重视餐饮商务信息的解释变量(本文中即每个主成分), β_0 为常数项, $\beta_1, \beta_2, \cdots, \beta_k$ 是解释变量的 logit 系数。

二、实证分析

(一) 数据来源及其描述

1. 数据来源

调查的方式主要是实体问卷调查和网络问卷调查。实体问卷调查对象主要是重庆市江北区和渝北区域内的大中型的中餐店、西餐厅、咖啡茶艺厅和快餐店美食街;网络调查主要在认识的餐饮老板和一些餐饮人士交流群里面,请大家参与调查。本文采用随机抽样方法发放问卷 116 份,其中有效问卷 100 份,有效百分比达 86.21%。

2. 调查数据的描述统计分析

(1) 调查用户特征

在 100 份有效样本中,有 45%的中餐厅、20%的茶艺厅和咖啡厅、13%的西餐厅以及 22%的其他类型。其中,所调查的管理者大多数都从事了该行业 5 年以上,占总调查人数的 74%,仅有 20%的管理者从事该行业还不到 1 年,这也可以充分说明此次调

查的可信度。

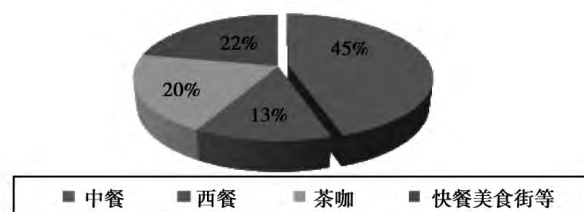


图1 调查餐厅分布饼图

(2) 调查对象对餐饮商务信息重要度认识的基本态度分布

在调查中,有 66%的餐饮企业应用了餐饮管理软件,另外 34%的餐饮企业并未采用餐饮管理软件。我们还发现,所调查的餐饮企业管理人员中,有 55%的人经常关注政府发布的与餐饮相关的政策,40%的人偶尔关注,仍然有 5%的人根本不关注政府发布的与餐饮相关的政策。除此之外,我们还发现仍然有 10%的餐饮企业不知道采用信息化技术管理餐馆的作用。

(二) 主成分分析

问卷调查数据各问题从不同方面说明了餐饮企业对商务信息的重视程度,由于问题数目过多,用一般的二项 logistic 回归模型会产生指标的多重共线性问题,因此,本文应用主成分分析法,将各个指标的影响分配到 5 个主成分上,不仅达到了降维的效果,还很好地解决多重共线性的问题。

表1 主成分分析各因子解释原有变量总方差的情况

成分	初始特征值			提取平方和载入		
	合计	方差的 %	累积 %	合计	方差的 %	累积 %
1	6.717	41.982	41.982	6.717	41.982	41.982
2	2.772	17.323	59.305	2.772	17.323	59.305
3	1.447	9.042	68.347	1.447	9.042	68.347
4	1.146	7.165	75.512	1.146	7.165	75.512
5	1.037	6.484	81.995	1.037	6.484	81.995
6	0.823	5.144	87.139			
7	0.598	3.739	90.878			
8	0.442	2.765	93.643			
9	0.376	2.352	95.995			

续表

成分	初始特征值			提取平方和载入		
	合计	方差的 %	累积 %	合计	方差的 %	累积 %
10	0.294	1.840	97.835			
11	0.173	1.079	98.914			
12	0.070	0.437	99.351			
13	0.047	0.294	99.646			
14	0.031	0.196	99.842			
15	0.014	0.090	99.932			
16	0.011	0.068	100.000			

提取方法: 主成分分析

从表 1 可以看出,前五个主成分对模型的解释程度就超过了 80%,这意味着它们包含了企业是否重视商务信息指标中的大部分信息。在综合考虑各主成分对餐饮企业是否重视商务信息的经济意

义以及各种统计量的基础上,本文将五个主成分作为二项 logistic 回归模型的自变量。从表 2 可以得到各个变量在 5 个因子上的载荷。

表 2 因子载荷矩阵

	成分				
	1	2	3	4	5
1	0.618	-0.573	0.011	0.381	-0.027
5	-0.779	0.285	0.368	-0.076	0.177
6	0.672	-0.136	0.151	-0.264	0.410
7	0.348	0.824	-0.265	-0.013	0.143
9	0.308	0.528	0.590	0.306	-0.031
10	0.602	-0.210	-0.111	0.473	-0.087
11	0.802	0.238	0.043	-0.201	-0.359
12	0.849	-0.057	-0.127	0.163	0.317
13	0.674	-0.262	0.008	-0.280	0.180
14	0.622	-0.127	0.587	-0.108	0.450
15	-0.090	0.762	-0.308	0.347	0.349
16	0.561	0.373	-0.428	-0.460	0.051
17	0.570	0.553	0.145	0.233	-0.123
18	0.704	-0.299	-0.274	0.186	0.071
19	0.913	0.008	-0.069	-0.088	-0.195
20	0.718	0.286	0.341	-0.150	-0.402

提取方法: 主成分

a. 已提取了 5 个成分

(三) 结合主成分分析的 logistic 回归分析

从上述主成分分析过程得到五个主成分因子得分值,并将其作为本文二项 logistic 回归模型的因变量,问卷调查中问题 2 的统计数据作为因变量进行模型拟合,得到拟合结果如下表 3 和表 4(在这里由于篇幅的原因仅选取了步骤三的拟合数据)。

表 3 二项 logistic 回归模型拟合优度分析

步骤	-2 对数 似然值	Cox & Snell R 方	Nagelkerke R 方
3	60.027	.494	.684

表 4 二项 logistic 分析结果

		B	S.E.	Wals	df	Sig.	Exp (B)
步骤 3 ^a	FAC1_1	3.522	1.107	10.115	1	0.001	33.836
	FAC4_1	-0.756	0.396	3.644	1	0.056	0.470
	FAC5_1	1.016	0.412	6.086	1	0.014	2.762
	常量	-0.651	0.386	2.837	1	0.092	0.522

a. 在步骤 1 中输入的变量: FAC1_1, FAC2_1, FAC3_1, FAC4_1, FAC5_1

从表 4 可以得出第一个主成分和第五个主成分对企业是否关注商务信息有正影响,而第四个主成分对其产生负的影响。从影响大小上看,第一主成分的影响力最大,第五主成分次之,而第四主成分最小。

其 logistic 表达式可表示为:

$$\ln\left(\frac{p}{1-p}\right) = -0.651 + 3.522F_1 - 0.756F_4 + 1.016F_5 \quad (5)$$

其中 p 代表餐饮企业注重商务信息的概率,而 $1-p$ 代表其不注重商务信息的概率。

三、结论及建议

(一) 二项 logistic 回归模型在餐饮商务信息意识研究方面的意见和展望

本文选用 logistic 回归模型,将其应用在餐饮商务信息意识的影响因素分析上。选用该模型有两大优势:一是 Logistic 回归模型能够很好地解决非线性问题,且可以解决多个自变量的问题,确定各自变量对因变量的影响强度,模型相对较成熟,准确度较高。二是该模型不要求变量服从多元正态分布,适用范围更广。另外,对于决策者是否有餐饮

从拟合优度上看,表 3 步骤三剔除变量的 Cox 的 R 方值以及 Nagelkerke 的 R 方值分别为 0.494 和 0.684,一定程度上说明模型拟合效果不错;表 4 中步骤三剔除变量后保留了第一、第四和第五个主成分,从表中可以看出各变量的显著性 p 值均很小,可以达到拟合要求。根据表 2 将第一、第四以及第五个主成分分别定义为:餐饮行业信息的了解程度、培训和相关政策的了解以及企业内部战略发展分析工作。

商务信息意识这一定性因变量,用二项 logistic 模型恰好得以运用,从文章的结论也可见其拟合效果较好。

(二) 实证结论和意见

从二项 logistic 模型分析的结论来看,第一和第五个主成分,即餐饮行业信息的了解程度和企业内部战略发展分析工作两个方面,其参数分别是 3.522 和 1.016,相对其他因素来说,其值较大,说明二者对餐饮商务信息意识会产生较大的影响。同时,这两个参数对应的 Wals 统计量的值较大,其 p 值较小,则应该拒绝原假设,即认为该因素的作用较为显著,说明餐饮行业信息的了解程度和企业内部战略发展分析工作两个要素对餐饮商务信息意识的影响大且显著。而第四个因素,即培训和相关政策的了解方面,其参数数值较小且不显著,说明该要素对餐饮商务信息意识的影响小且不显著。因此,本文建议餐饮业管理者或者决策者:首先,应该重视搜集并了解餐饮行业信息,作为该行业的一分子,了解行业宏观发展状况和发展趋势,有利于企业自身定位,一方面可以以其他发展较好的企业为参考,另一方面可以看准形势发展,提高商务信息

意识,将有利于本企业的未来发展;其次,企业内部战略发展分析也是餐饮商务信息的一个重要方面,了解企业自身的经营状况,将有助于企业在行业中的定位,提高商务信息意识,从主观方面找到发展中存在的根本问题;再次,建议充分利用企业经营过程中的信息数据,实现餐饮管理信息化,使用餐饮管理软件,不但可以提高餐饮企业日常管理的工作效率,还可以将现有的数据进行自动整合,并以此为依据,帮助企业管理者作出科学、严密、精细化的战略决策,其优点在于错误少且信息快、准而全。也就是说,餐饮管理软件不单单是操作层面上的作用,应慢慢地转向战略层面的管理作用;最后,应该重视对企业雇员的培养,包括素质的提高和技能的提升等,可以通过各种形式的员工培训来实现,更可以在招聘雇员时以多种方式招贤纳士,如招聘海报、报纸、人才市场、网络招聘,等等,这是餐饮商务信息意识提高的一个重要体现。

[参考文献]

- [1] 古家声.数据挖掘在寿险客户分析中的应用[D].华南理工大学,2012.
- [2] 黄解军,万幼川.基于数据挖掘的电子商务策略[J].计算机应用与软件,2004,21(7):12-13.
- [3] 姚淼.数据挖掘在电子商务中的应用[J].图书情报论坛,2009(1):56-58.
- [4] 刘红光.数据挖掘在网络信息资源管理中的应用[J].中国信息技术教育,2011(18):55-57.
- [5] 滕广青,毛英爽.国外数据挖掘应用研究与发展分析[J].统计研究,2005(12):68-70.
- [6] 周大兴.数据分析在餐饮企业经营中的应用[D].上海交通大学硕士论文,2012.
- [7] 孔令娟.数据控制在连锁餐饮业中的研究和应用[D].南京理工大学,2012.
- [8] 付婷.大数据在餐饮行业的应用研究[J].中国商贸,2015(15):147-149.

(责任编辑:朱德东)

Research on the Catering Business Information Consciousness Based on Logistic Regression Model of Principal Component Analysis

LI Hong-xia^{a, b}

(a. Chongqing Key Laboratory of Electronic Commerce and Supply Chain System; b. School of Management, Chongqing Technology and Business University, Chongqing 400067, China)

Abstract: Data mining technology has become increasingly mature and has been widely used in financial, real estate, investment, evaluation, banks and other fields. Data mining techniques can be applied to the catering industry to make analysis of its business decisions. Data mining in the field of other successful cases can also refer to the catering industry. This article has carried out the survey on whether the dining people in the restaurant business had information awareness and the extent the people emphasized information awareness. Principal component analysis and logistic regression in data mining process are used to analyze the collected data so as to illustrate the important value of data mining in the catering industry management and to reflect the restaurant information awareness of the dining people. Using data mining technology on food and beverage business information management is bound to improve business efficiency and profit level and to promote the healthy and stable development of the catering industry.

Key words: data mining; catering business; principal component analysis; logistic regression; information awareness