

大数据环境下老年人失能失智因素关联

李 涵 李慧嘉 张林姿 黄钰瑛

(中央财经大学管理科学与工程学院 北京 100081)

摘 要 近年来,老年人失能失智问题愈发普遍,不断增加的老年失能失智人口在为家庭和政府带来沉重养老负担的同时,也引发了社会公众的广泛关注。老年人年龄增长、状态和控制能力调查(ASOC)是通过 1995 年—2001 年每隔 3 年连续调查美国老年人健康方面的问题而得到的数据。利用 ASOC 的 3 轮调查数据,通过跟踪研究和大数据分析,尝试将能力、智力衰退与生活习惯、情绪发泄渠道等诸多因素相关联,全面考察老年人失能失智的影响因素。在描述性统计的基础上,通过 Logistic 回归并去除关联性强的因素发现,年龄、性别、抽烟、饮酒、运动、心脏病、关节炎/风湿病、参与社区服务 8 个因素与老年人失能有较强的相关性,年龄、饮酒、高血压、参与社区服务、婚姻状况 5 个因素与老年人失智显著相关。借助本研究成果,可以对老年人的日常行为、生活习惯、健康状况等进行评价,找出其失能失智的潜在风险,从而指导家庭和养老服务机构等采取相应的措施以规避风险,为改善老年人的健康状况、提高生活质量和延长生命周期提供参考依据。

关键词 老年人,失能,失智,大数据,关联

中图法分类号 C931 **文献标识码** A

Correlative Factors about Elderly Disabled and Dementia in Big Data Environments

LI Han LI Hui-jia ZHANG Lin-zi HUANG Yu-ying

(School of Management Science and Engineering, Central University of Finance and Economics, Beijing 100081, China)

Abstract Recently, with the deepening of China's population aging problem, the rising family pension burden, government pressure and the decline of demographic dividend have gained wide attention. Aging, status, and sense of control (ASOC) is a continuous survey data about elderly's health in American which was conducted every three years from 1995 to 2001. Using three waves of ASOC, we have found influential factors related to disability and dementia by logistic regression based on the descriptive statistical analysis. On the basis of the descriptive statistics, by Logistic regression and removing strong related factors, we discovered that age, gender, smoking, drinking, exercise, heart disease, arthritis/rheumatism, and participation in community service have strong correlation with the elderly disability. Age, drinking, hypertension, participation in community service and marital status are significantly related to dementia in the elderly.

Keywords Elderly, Disability, Dementia, Big data, Correlation

1 引言

1.1 人口老龄化趋势不断加剧,社会问题层出不穷

近年来,中国老年人口比例不断攀升。国家统计局数据显示,2011 年以来,中国老年人口比率达到 9%,2013 年更是达到了 9.7%。根据联合国对于老龄化社会的新标准,65 岁以上的老人占总人口的 7%的地区视为进入老龄化社会,中国早在 2000 年就已经步入了老龄化社会,并且老龄化危机逐年加剧,随之而来的老年人的失能失智问题日益突出。在不久的将来,中国人口老龄化将达到高峰。有专家预测中国在 2050 年,老龄人口将超过 5 亿。在过去的 20 年中,中国的财富增长速度已是一个非凡奇迹,但人口老龄化比人们的财富增长速率更快,因此养老问题异常严峻。

1.2 老年人健康状况不容乐观,失能失智问题严峻

随着时代的进步,人们的生活在各方面都得到发展,但仍

有严峻的问题等待大家共同解决,老龄化及老年人的健康问题即为其中一个。因此,国家与相关机构越来越重视这类问题,并展开了多个方向的深入调查。

全国老龄办^[1]发布的调查结果显示,根据 2015 年 1%人口的抽样调查,60 岁以上人口占总人口的比例比 5 年前提升了 2.89 个百分点,达到 16.15%,而且老年人的健康状况不容乐观,失能和半失能老年人口的数量较大。2015 年,自评健康状况“好”的老年人只有 32.8%,全国城乡失能、半失能老年人口占老年人口总量的 18.3%,总量约为 4063 万。

中国健康与养老追踪调查(CHARLS)发布追踪中国人口结构转变的结果:老年人中有 40%具有明显的抑郁症状,超过 50%的老年人患有高血压。

我国卫生部“中国居民健康素养调查”表明,我国城乡居民健康素养总体水平为 6.48%,即每 100 人不到 7 人的身体状况为健康,而 55~64 岁年龄组的为 4.69%,65~69 岁年龄

组的最低为 3.1%。根据世界卫生组织的统计,全球 49 个高收入国家的人均健康寿命为 70 岁,而中国为 66 岁;发达国家 60 岁以上老年人身体状况为健康的比例超过 60%,而中国仅为 43%。

上述数据表明,中国老年人口数量持续增加,人口老龄化程度持续加深,必须高度重视老龄化持续加深的人口现状及影响。

1.3 针对失能失智老人的服务和医疗保障严重滞后

我国作为人口大国,老年人口规模大,失智老人的照顾不能得到保障,如以北京为例,即使已规划好了养老模式,即“9064”模型,但因为长期居住的老人有 300 万,且进行特殊看护的高龄失智失能老人约达到 60 万,而为失智失能老人提供的床位供给不足,大约为 5 万张,这使得失能失智老人不能得到床位和相应的照料,需求大大超过供给。目前,我国的失能失智老人多数由家属看护,失能失智老人因记忆模糊和行为异常往往会给家人带来不安和担忧,而照顾老人需要花费的人力和物力也给家庭带来负担。失能失智老人比健康老人需要更高水平的医疗服务和照料看护,但疗养院提供该服务的数量少,且费用昂贵。

1.4 政府积极应对人口老龄化

随着我国人口老龄化形势日趋严峻,近年来人口老龄化频频成为两会的热点议题。2013 年《政府工作报告》提出积极应对人口老龄化,大力发展老龄服务事业和产业。2017 年,“两会”就应对人口老龄化上升为国家战略,应对人口老龄化已刻不容缓。

人口老龄化问题无疑给医疗卫生行业带来了巨大冲击,加快医疗服务发展及技术改革刻不容缓。2016 年 12 月 27 日,国务院发布《“十三五”深化医药卫生体制改革规划》,提出应继续组织国家科技重大专项和重点研发计划项目,提升科技创新能力。近年来,政府不断出台相关扶持政策,鼓励医疗研究,推动医疗技术发展。

1.5 我国老年人失能失智问题研究匮乏

关于我国老年人健康问题的研究已经受到以曾毅等^[2]为代表的学者们的广泛关注。而张向明^[3]则是简单地分析了城市老年人健康的影响因素,未做深入具体病情影响因素分析。由于老年人的健康状况具有多样性和难以刻画等特点,当前对影响老年人健康因素的研究,尤其对失能和失智两个方面的分析研究较少。因此,在利用大量数据进行系统分析的基础上,提取影响老年人健康的相关因素并对其健康状况进行分级预测,以更好地理解老年人健康问题的内在机理。

2 研究目的和意义

2.1 研究目的

在我国老龄化趋势不断加剧、失能失智老年人问题严峻的大背景下,对老年人失能失智的预防和治疗的研究迫在眉睫。本文的研究目的是通过跟踪研究和大数据分析揭示老年人失能失智的关联因素,尝试将能力衰退与生活习惯、情绪发泄渠道等诸多因素相关联,例如抽烟是否会导致老年人肺功能下降,是否有渠道发泄消极情绪影响老年人的心理健康水平等来全面考察老年人失能失智的影响因素。

2.2 研究意义

2.2.1 理论意义

收集老年人健康状况的数据,研究老年人的失能失智的关联因素,有助于从理论上构建老年人健康状况的指标体系和实施健康状况影响因素的实证研究。运用定性和定量的方法检验和充实关于老年人失能失智现有的研究成果,完善老年人健康问题研究的理论体系将构成老年服务研发的核心知识基础。

2.2.2 现实意义

通过研究结果,可以对老年人的生活习惯、身体心理健康状况进行评价和指导,找出老年人失能失智的原因,采取合理的方式管理和规避风险,改善和提高老年人的健康状况、生活质量并延长生命周期。本项目成果可以作为保险公司设计养老保险的依据,有利于养老保险的完善和成熟。

2.2.3 研究创新

本文使用“老年人年龄增长,状态和控制能力调查”(ASOC)数据进行了实证分析,该调查分别于 1995 年、1998 年和 2001 年进行,调查了受访者在生活中的控制感和年龄变化之间的关系。利用 logistic 回归分析提取老年人失能失智的影响因素,根据该结果可以在一定程度上降低老年人失智失能的比。

本研究既有挖掘影响因子的回归分析又有健康状况的预测分析,既有研究分析又有验证分析。而从研究结果来看,除年龄、烟酒等普遍认知的影响因子外,还发现了家务、社交活动等对失能失智的强影响。

3 相关工作

老年人的健康问题被社会广泛关注,许多学者对其进行了不同层次和角度的研究。Rong 等^[4]特别观察了中国女性老龄人口的健康状况,其研究表明,在老年人口中女性的医疗支出逐年递增。

在影响老年人健康状态的因素方面,国内外学者已经取得了一定的成果。翟德华^[5]对东西部各省市老年人健康状况进行了分析,考虑了性别、年龄、居住地区、经济条件等因素的影响。谷琳等^[6]在研究上述因素的基础上,还考虑了婚姻、文化程度、就业类型以及心理和疾病等因素与老年人健康自评结果的关系。

在这类研究中,周琼等^[7]对老年痴呆即老年人失智进行了分析,研究发现老年痴呆的患病率与年龄、性别、教育程度、婚姻状况等有关。性别方面,女性的患病率高于男性;年龄方面,随着年龄的增大,患病率不断攀升;教育程度方面,教育程度低者患病率高;婚姻状态方面,没有或失去伴侣的老年人的患病率明显高于有伴侣的老人。Dean 等^[8]对影响老年人心理健康的因素进行了研究,他们认为相较于年龄等因素,生活功能障碍、经济拮据以及生活中不如意的事件等,更能直接使老年人产生抑郁症状,同时朋友对老年人心理健康的影响比子女更加明显。Lum 等^[9]在论文中指出,适当参与老年志愿活动更有利于老年人自身的心理健康。

在对老年人健康做因素关联研究时,很多学者都采用了 Logistic 回归的方法。傅东波等^[10]应用 Logistic 回归模型对 1453 名社区老年人近 3 年的纵向观察资料进行了分析,得出

如下结论:社区保健的重点对象为高龄老人、无职业者、患严重疾病尤其是患心脑血管疾病者;社区老年保健的首要任务是预防和阻止躯体健康和日常活动能力的损害,加强老年人心血管疾病的防治。Merril 等^[11]也运用 Logistic 回归模型对老年人健康因素关联进行研究,他们认为子女对父母健康的促进作用只有在老年人丧偶后才能得到体现。Schone 等^[12]则发现,一段新的婚姻更可能帮助丧偶老人积极面对生活,从而得出良好的婚姻有益于老年人身心健康的结论。

4 数据及其描述性统计

4.1 数据来源

本文使用的数据来自 ICPSR(Interuniversity Consortium for Political and Social Research)数据库上“老年人年龄增长,状态和控制能力调查”。第一部分的研究包含数据的 Wave1(2593 人受访)和 Wave2(1378 人受访)。第二部分的研究包含数据的 Wave3(1444 人的回访并加入新的老年人样本),其中研究的问题与第一部分相同。

由于前期调查的部分老年人去世以及新的老年人的加入,3 轮调查人群并不完全相同。为了研究老年人连续时间内的状态转移情况,本文选取了 3 轮重叠的调研对象,综合考虑了老人的身心健康状况。

4.2 描述性统计

本文使用的数据涵盖了与老年人相关的 250 多个变量,本文根据现有文献研究成果和日常生活认知提取了 22 个可能与失能或失智相关的变量,并将其作为 Logistic 回归的自变量,具体变量信息如表 1 所列。

表 1 自变量名称及其含义

变量名称	变量含义
年龄	受访老年人的年龄
性别	老人性别(1 为男性,0 为女性)
锻炼	锻炼频率
抽烟	一周抽烟是否超过 7 支
吃肉	吃肉的频率
吃蔬菜水果	吃蔬菜或水果的频率
饮酒	饮酒的频率
心脏疾病	是否患心脏疾病
高血压	是否患高血压
肺部疾病	是否患肺部疾病
癌症	是否患癌症
变量名称	变量含义
乳腺癌	是否患乳腺癌
糖尿病	是否患糖尿病
关节炎/风湿病	是否患关节炎/风湿病
骨质疏松症	是否患骨质疏松症
过敏/哮喘	是否患过敏/哮喘
消化问题	是否有消化问题
社区服务	参与社区服务的频率
教堂志愿	参与教堂志愿的频率
婚姻	是否已婚且与伴侣同住
做家务	做家务的频率
居住情况	居住地为乡村或城镇

因变量方面,将失能变量(爬楼梯、跪或弯腰、携重、做家务、购物、视力、听力 7 类)的综合指标作为是否失能(基于 ADL 量表选取低于 40% 的为显著样本)的判断标准。求出变量的和 $X(7-21)$,数值从小到大代表失能程度依次递增。使用计算式 $Y=100-(X-7)/(21-7) \times 100$ 进行标准化, Y 定义为能力值。根据日常生活活动能力量表(ADL),将 $Y < 60$

视作失能。在失智回归中,主要采取老年人记忆能力和健忘天数作为指标,设 X 为一周中健忘的天数,根据简易智能精神状态检查量表(MMSE)使用计算式 $Y=30-X/7 \times 30$ 进行标准化,定义 Y 为智力值,将 $Y < 15$ 视作失智。

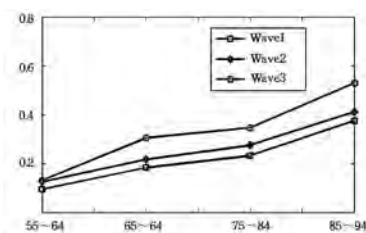


图 1 Wave1—Wave3 各个年龄段失能和失智老人的比例

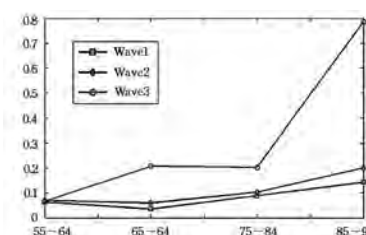


图 2 Wave1—Wave3 各个年龄段失智和不失智老人的比例

从图 1 和图 2 可以看出,随着年龄的增长,老年人失能的比例逐渐增大,虽然失智比例在部分年龄段略有下降,但 85 岁以上老年人失智比例大幅度提高。因此,年龄与老年人的失能可能存在较大的相关性。另外,随着年龄的增长,老年人会出现身体机能衰退、疾病增加等症状,从而导致老年人失能。在后文的 Logistics 回归分析中,我们也将进一步验证上述结论的正确性。

图 3 中的 3 年统计数据分析表明,从 Wave1 到 Wave3 中的男、女比例的变化情况来看,男女比例呈递增趋势。并且,随着年龄的增长,男女比例差距增大。以 Wave3 为例,55~64 岁年龄段老人男/女比例约 0.8,但到 85 岁以上,男/女比例便降到 0.4 左右。据此可以推断,女性老年人较男性老年人可能具有一定的生存优势。

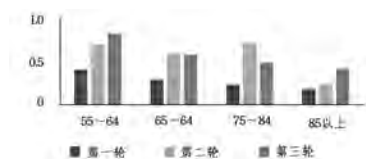


图 3 Wave1—Wave3 调查中,各个年龄段老人的男/女比例

5 回归分析

5.1 Logistic 回归模型

在老年人 wende 失能失智因素关联部分,本文使用二元 Logistic 回归模型分别考虑失能和失智的影响因素。因变量为是否失能或是否失智,自变量则选取可能对失能失智产生影响的因素。本文从老年人基本信息(如年龄、性别)、生活方式(如饮食与锻炼情况)、疾病情况(如高血压、癌症)、婚姻家庭情况(如婚姻状况)4 个方面选取了 22 个指标并在下文列示,最终形成如下 Logistic 回归模型:

$$\ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k \quad (1)$$

在模型中, β_i 作为回归系数,反映了在其他变量固定后,

$X_i=1$ 与 $X_i=0$ 相比发生失能或失智的对数优势比,一般认为 β 越大,该因素越有可能对老年人失能或失智产生影响。 P 表示失能或失智的概率, P 的计算式如下:

$$P = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k)} \quad (2)$$

5.2 回归结果

根据上文所述处理方法,将提取的 22 个因素作为自变量。将是否失能和是否失智作为因变量。利用 Logistic 模型对 Wave1—Wave3 分别进行回归,得到以下结果,表中加粗的数据为 P 值小于 0.05 的因素。

表 2 Wave1 失能因素回归结果

因素	<i>B</i>	<i>P-value</i>	<i>OR</i>
年龄	0.742	0.016	2.102
性别	-0.423	0.093	0.655
锻炼	0.280	0.004	1.323
抽烟	-0.645	0.587	0.525
吃肉	0.005	0.014	1.005
吃蔬菜水果	-0.002	0.045	0.998
饮酒	1.230	0.034	3.431
心脏疾病	0.024	0.015	1.024
高血压	-0.285	0.761	0.752
肺部疾病	-1.084	0.684	0.338
癌症	0.870	0.001	2.387
乳腺癌	-0.273	0.824	0.761
糖尿病	17.228	0.098	0.432
关节炎/风湿病	-2.324	0.048	0.098
骨质疏松症	18.209	0.199	0.433
过敏/哮喘	0.106	0.938	1.112
消化问题	-1.192	0.313	0.304
参与社区服务	0.014	0.273	1.015
参与教堂志愿活动	-0.100	0.259	0.905
婚姻状况	0.387	0.305	1.472
做家务	0.025	0.009	1.253
居住情况	0.626	0.202	1.869

表 3 Wave1 失智因素回归结果

因素	<i>B</i>	<i>P-value</i>	<i>OR</i>
年龄	0.557	0.042	1.745
性别	2.186	0.244	8.900
锻炼	-0.003	0.529	0.997
抽烟	-1.167	0.358	0.311
吃肉	0.002	0.503	1.002
吃蔬菜水果	0.000	0.786	1.000
饮酒	1.029	0.048	2.799
心脏疾病	-2.642	0.027	0.071
高血压	1.008	0.343	2.740
肺部疾病	20.901	0.999	0.822
癌症	0.662	0.637	1.938
乳腺癌	0.853	0.614	2.346
糖尿病	-0.649	0.708	0.523
关节炎/风湿病	-1.304	0.211	0.272
骨质疏松症	-1.593	0.203	0.203
过敏/哮喘	0.167	0.872	1.182
消化问题	0.203	0.770	1.224
参与社区服务	0.592	0.035	1.808
参与教堂志愿活动	-0.012	0.592	0.988
婚姻状况	0.431	0.014	1.539
做家务	0.002	0.028	1.002
居住情况	-16.679	0.998	0

由回归结果可知,Wave1 中与失能显著性相关的因素有年龄、锻炼、吃肉、吃蔬菜水果、饮酒、心脏疾病、癌症、关节炎/风湿病、做家务。与失智显著性相关的因素有年龄、饮酒、心脏疾病、参与社区服务、婚姻状况、做家务。

同理,从回归的结果可以得到,Wave2 中与失能显著性相关的因素有年龄、锻炼、吃肉、心脏疾病、高血压、癌症、糖尿病、消化病、做家务。与失智显著性相关的因素有年龄、饮酒、肺部疾病、糖尿病、骨质疏松症、参与社区服务、做家务。Wave3 中与失能显著性相关的因素有年龄、性别、抽烟、饮酒、癌症、关节炎/风湿病、骨质疏松症、参与社区服务。与失智显著性相关的因素有年龄、锻炼、饮酒、心脏疾病、关节炎/风湿病、骨质疏松症、婚姻状况、做家务。

取 3 轮调查数据回归结果的均值,得到失能和失智的回归结果,如表 4 所列。

表 4 失能和失智回归结果

失能因素	<i>OR</i>	失智因素	<i>OR</i>
年龄	1.40	年龄	2.02
性别	2.46	锻炼	1.10
锻炼	1.32	喝酒	1.18
抽烟	0.18	高血压	1.99
喝酒	1.34	参与社区服务	2.56
心脏疾病	1.55	婚姻状况	1.54
癌症	2.39	做家务	1.11
关节炎/风湿病	0.33		
参与社区服务	1.07		
做家务	1.25		

从上述回归结果可以看出,年龄、性别、抽烟、饮酒、运动、心脏病、癌症、关节炎/风湿病骨质疏松症、做家务、参与社区服务这 10 个因素与老年人失能显著相关。其中,年龄在 OR 值为 1.4 时代表年龄每增加 1 岁,失能的风险增加为原来的 1.4 倍,这与前文描述性统计分析结果相吻合——随着年龄的增长,失能失智的老年人的比重也增大。同时可以看到,性别作为显著变量,男性失能风险为女性的 2.46 倍,同样与前文分析相符。癌症的 OR 值最高, OR 为 1.72 时代表着患癌症老人失能风险是未患癌老人的 1.72 倍。患心脏疾病和关节炎/风湿病的老年人分别是未患病老年人失能风险的 1.02 倍和 0.31 倍。另外,结果显示锻炼频率增加会增大失能风险,其原因可能是因锻炼次数多更易导致老人意外伤害,具体原因还有待进一步研究。

年龄、锻炼、饮酒、高血压、参与社区服务、婚姻状况、做家务 7 个因素与老年人失智显著相关。其中,年龄每增加 1 岁,对应失智风险变为原来的 2.02 倍,而饮酒频率每增加一个单位,对应失智风险变为原来的 1.79 倍。

由上述结果和分析可知:年龄在失能和失智两回归结果中均为显著的变量,其与因变量有着很强的相关性,这要求我们给予高龄老人更多的关注,以规避更可能出现的失能或失智风险,同时也要重点关注男性的失能风险;对于患有心脏疾病、癌症、关节炎或风湿病的老人,要特别防范失能情况的出现;对于患骨质疏松症的老年人,其可能更容易出现失智,同时也要控制老年人的饮酒量以避免失智的发生。

5.3 关联性分析

从上述回归结果中可以看出,年龄、性别、抽烟、饮酒、运动、心脏病、癌症、关节炎/风湿病骨质疏松症、做家务、参与社区服务 10 个因素与失能显著相关;年龄、锻炼、饮酒、高血压、参与社区服务、婚姻状况、做家务 7 个因素与老年人失智显著相关。为进一步验证因素选择的准确性与可靠性,我们对上述因素之间的关联性进行 SPSS 分析,以考察影响因素之间的关联性,结果如表 5 所列。

表 5 因素关联性分析的结果

因素	<i>B</i>	<i>P-value</i>
年龄 & 性别	0.008	0.781
吸烟 & 饮酒	0.046	0.118
饮酒 & 锻炼	0.183	0.002
心脏病 & 饮酒	-0.038	0.202
癌症 & 饮酒	0.068	0.021
锻炼 & 心脏病	-0.031	0.301
婚姻 & 癌症	0.009	0.723
癌症 & 吸烟	0.088	0.003
婚姻 & 做家务活	0.239	0.001
参加社交活动 & 做家务活	0.564	0.000
锻炼 & 做家务活	0.003	0.923
锻炼 & 癌症	-0.022	0.453
婚姻 & 参加社交活动	-0.043	0.147

由上述信息可知,因素之间的相关性存在差异,饮酒与锻炼、抽烟(曾经),癌症与吸烟,婚姻与做家务活,参加社交活动与做家务活具有很强的相关性。为了保证 Logistic 回归函数的准确性,对因素进行去除关联性处理,结合表 5 中的数据,选择性删除做家务活、癌症和锻炼这 3 个因素。

因此,最终失能的影响因素为年龄、性别、吸烟、饮酒、心脏病、关节炎、参加社交活动这 7 个因素,失智的影响因素为年龄、饮酒、高血压、参加社交活动、婚姻状况 5 个因素。

6 结果验证分析

6.1 数据介绍与处理

为验证回归分析结论具有广泛的适用性与实际意义,我们另选取了中国老年健康影响因素跟踪调查(CLHLS 2008-2011)进行验证分析。该数据收集的信息包括受访者的基本信息(性别、年龄、同居人口等)、生活方式(饮食、抽烟、饮酒、社交活动等)、心理状态(害怕、孤独、无力感等)、疾病情况(心脏病、高血压等疾病)4 个方面的多个变量指标。利用该新数据进行的第二轮验证对保证以上结论的可靠性具有相当重要的意义。

在将 2008 年和 2011 年案例中的部分变量导入 SPSS 中并去除含无效值过多的变量与个案后,本部分最终采用的自变量包括上述 4 个方面共 37 个变量。同时,添加“年龄”变量,并将年龄分段为 50~60,60~70,70~80,80~90,90~100,100 以上 6 类,分别标号为 1—6。因变量方面,本文选取了 10 个失智变量(分别考察了受访者的算数能力、语言能力、记忆力等)与 8 个失能变量(分别考察了受访者在日常生活中的自理能力),变量处理方式与分级标准同前文。

6.2 描述性统计分析

在本文选取的数据中,按照上文的划分方法,2008 年与 2011 年受访者的失智比例均达 5 成,失能比例也都在 40%左右。在自变量(在此仅阐述部分经后文回归分析验证对因变量影响较大的自变量)方面,受访者绝大多数为男性;受访者饮酒、抽烟的比例均在 30%左右;两次调查的受访者主要收入来源为退休金或子女赡养的均在 35%以上,同时两次调查各有一部分老年人自给自足(2008 年 18.54%,2011 年 14.57%)。婚姻状况方面,已婚且与配偶生活的接近 70%,丧偶的老年人为 30%左右,而这两种情况的比例从 2008 年到 2011 年分别有所下降和上升;在年龄结构方面,受访者的年龄多为 60~90 岁,其中以 70~80 岁居多。

研究发现,男性的失能比例与女性相比更低,而失智比例在性别上的差异并不明显;抽烟者比不抽烟者的失能比例更低,同样失智比例在两类人群间的差异不明显;收入充足的老年人失能和失智比例更低;离婚或从未结婚者的高失能和高失智比例十分突出,但 2011 年离婚者的失智比例和从未结婚者的失能比例都很低;随着受访者年龄的上升,失能和失智比例呈现非常明显的上升趋势。

以是否失能(“是”=1,“否”=0)、是否失智(“是”=1,“否”=0)分别作为因变量对两年数据进行回归,结果如表 6 所列,打勾选项为回归结果中 *P* 值小于 0.05 的变量。从该验证数据的回归结果来看,年龄、现在是否抽烟、中风 3 个指标对两轮数据失能和失智 2 个方面均有较强影响。包括婚姻次数、无力感等在内的 11 个指标在两轮调查均与失能有很强的相关性,其中 5 个指标与 ASOC 数据的回归结果一致。而在一轮或两轮调查中表现出与失智具有强相关性的因素中,年龄和婚姻状况 2 个指标与 ASOC 数据的回归结果保持一致。

表 6 验证数据回归结果

因素	08 失能	08 失智	11 失能	11 失智
年龄	✓	✓	✓	✓
性别			✓	
同居人口		✓		
房屋产权		✓		✓
婚姻状态	✓	✓		
婚姻次数	✓		✓	
孤独		✓		
害怕	✓			✓
无力感	✓	✓	✓	
现在抽烟	✓	✓	✓	✓
过去饮酒			✓	
现在饮酒			✓	
参加社交	✓		✓	
主要收入	✓	✓	✓	
收入足够	✓			
60 岁吃肉		✓		
高血压	✓			
糖尿病	✓			
心脏病	✓			
中风	✓	✓	✓	✓
胃病		✓		
肺部疾病	✓		✓	✓
帕金森	✓			
关节炎	✓		✓	

结束语 利用“老年人年龄增长,状态和控制能力调查”(ASOC)的数据,本文挑选了变量来刻画老年人失能失智状况并进行相应的分类,通过 Logistic 回归发现了年龄、性别、抽烟、饮酒、运动、心脏病、关节炎/风湿病、参与社区服务 8 个因素与老年人失能有较强的相关性,年龄、饮酒、高血压、参与社区服务、婚姻状况 5 个因素与老年人失智显著相关。

本文主要存在以下两点不足:1)由于分析的因素较多,未就提取的相关因素对失能或失智的具体影响过程或原理进行理论分析;2)虽然考虑了因素关联,但未得出各关联因素所起到的作用。

我国严峻的老龄化形势对各行各业和家庭都提出了挑战。从政府的角度来看,建立科学的养老制度,推行有效的惠老措施都是积极应对社会老龄化的表现。而在家庭的日常生活中,也应该着重关注影响老年人健康的关键因素以及老年

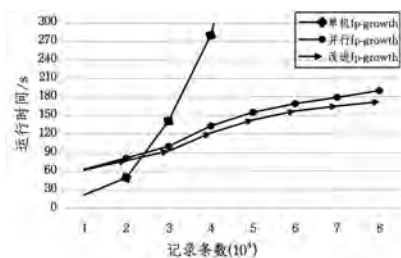


图6 算法的运行时间对比

由图6可得,传统单机 fp-growth 在数据量较少的情况下具有绝对优势,到达一定数量级时计算时间会大幅上涨直到内存不足而无法计算;改进后的 fp-growth 由于基于 hadoop 的并行运算,在处理大数据量时具有一定优势,同时基于兴趣列的选择后减少了频繁集输出的数量,从而较并行 fp-growth 的计算效率有了一定的提升。

结束语 本文在传统 FP-growth 算法的基础上提出了一种改进算法,基于 Hadoop 框架,采用 MapReduce 编程解决了内存溢出的可能,同时实现了对兴趣属性列的筛选。实验结果表明,海量大数据下,减少冗余信息的挖掘,提高了对有效信息的挖掘效率;同时改进后的算法在工业环境设备下的应用有效地解决了传统工业设备故障诊断的难题。

改进的算法虽然解决了大数据故障设备诊断过程中的瓶颈,兴趣度的选择也提高了挖掘效率,但还存在一些问题,fp-growth 算法本身两次访问数据降低了算法效率,针对 hadoop 集群而言,并行计算过程中可能存在负载不均衡问题,而且在工业设备的应用过程中,前期需要足够的监控数据,否则无法构造较好的模型。未来研究将减少算法的访问次数,同时实现 MapReduce 过程中负载均衡的优化,从而将其更好、更准确地应用于工业设备检测。

参考文献

- [1] WU L, YAO B, PENG Z, et al. Fault Diagnosis of Roller Bearings Based on a Wavelet Neural Network and Manifold Learning[J]. Applied Sciences, 2017, 7(2): 158.
- [2] CERRADA M, ZURITA G, CABRERA D, et al. Fault diagnosis

in spur gears based on genetic algorithm and random forest[J]. Mechanical Systems & Signal Processing, 2016, 70-71: 87-103.

- [3] 胡辉. 基于 Hadoop 的动车组故障数据关联规则挖掘研究与实现[D]. 北京: 北京交通大学, 2015.
- [4] WHITE T, CUTTING D. Hadoop: the definitive guide[J]. O'reilly Media Inc Gravenstein Highway North, 2009, 215(11): 1-4.
- [5] SHVACHKO K, KUANG H, RADIA S, et al. The Hadoop Distributed File System[C]// MASS Storage Systems and Technologies. IEEE, 2010: 1-10.
- [6] DEAN J, GHEMAWAT S. MapReduce: A Flexible Data Processing Tool[J]. Communications of the Acm, 2010, 53(1): 72-77.
- [7] AGRAWAL R, SHAFER J C. Parallel Mining of Association Rules[J]. Journal of Supercomputing, 2007, 39(3): 273-299.
- [8] LEI Y, JIA F, LIN J, et al. An Intelligent Fault Diagnosis Method Using Unsupervised Feature Learning Towards Mechanical Big Data[J]. IEEE Transactions on Industrial Electronics, 2016, 63(5): 3137-3147.
- [9] WANG B, CHEN D, SHI B, et al. Comprehensive Association Rules Mining of Health Examination Data with an Extended FP-Growth Method[J]. Mobile Networks & Applications, 2017, 22(2): 1-8.
- [10] 刘鑫, 贾云献, 孙磊, 等. 基于 BP 神经网络的变速箱故障诊断方法研究[J]. 计算机测量与控制, 2017, 25(1): 12-15.
- [11] 谢宏, 程浩忠, 牛东晓. 基于信息熵的粗糙集连续属性离散化算法[J]. 计算机学报, 2005, 28(9): 1570-1574.
- [12] 杨世海, 李涛, 陈铭明, 等. 基于数据挖掘的智能电网在线故障诊断与分析[J]. 电子设计工程, 2017, 25(1): 136-139.
- [13] LIU D, LIN Y, HUANG P C, et al. Durable and Energy Efficient In-Memory Frequent Pattern Mining[J]. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2017, PP(99): 1-1.
- [14] ZENG Y, YIN S, LIU J, et al. Research of improved FP-Growth algorithm in association rules mining [J]. Scientific Programming, 2015, 2015: 6.
- [15] 杨鹏坤, 彭慧, 周晓峰, 等. 改进的基于频繁模式树的最大频繁项集挖掘算法——FP-MFIA[J]. 计算机应用, 2015, 35(3): 775-778.

(上接第 501 页)

人健康状态的发展动向。改善老年人的生活质量与身体状况需要有针对性的工作,本研究对具体的家庭护理提出了针对性建议。

参考文献

- [1] 中国老龄科学研究中心课题组. 全国城乡失能老年人状况研究[J]. 残疾人研究, 2011(2): 11-16.
- [2] 曾毅, 顾大男, PURSER J, 等. 社会、经济与环境因素对老年健康和死亡的影响——基于中国 22 省份的抽样调查[J]. 中国卫生政策研究, 2014, 7(6): 53-62.
- [3] 张向明. 城市老年人健康状况相关因素初步探讨[J]. 山东医药, 2008, 48(19): 18.
- [4] RONG P, LI L, QUN H. Self-rated health status transition and long-term care need, of the oldest Chinese[J]. Health Policy, 2010, 5(9): 259-266.
- [5] 翟德华, 陶立群. 高龄老人性格心理特征、饮食习惯与健康长寿关系研究[J]. 中国人口科学, 2004, S1(1): 83-87, 177.

- [6] 谷琳, 乔晓春. 我国老年人健康自评影响因素分析[J]. 人口学刊, 2006, 6(6): 25-29.
- [7] 周琼, 纪颖, 张蕾, 等. 高龄老人痴呆发生的相关因素浅析[J]. 中国人口科学, 2004, S1(1): 126-129, 178.
- [8] DEAN A, KOLODY B, WOOD P. Effects of social from various sources on depression in elderly persons[J]. Journal of Health and Social Behavior, 1990, 31(2): 148-161.
- [9] LUMV T Y, LIGHTFOOT E. The Effects of Volunteering on the Physical and Mental Health of Older People[J]. Research on Aging, 2005, 27(1): 31-55.
- [10] 傅东波, 傅华, 李洋. 上海市高桥社区糖尿病新型管理模式的效果评估[J]. 中国社会医学杂志, 2011, 28(6): 412-414.
- [11] MERRIL S, BENGTON V L. Do close parent-child relations reduce the mortality risk of older parents? [J]. Journal of Health and Social Behavior, 1991, 32(4): 382-395.
- [12] SCHONE B S, WEINICK R M. Health-related behaviors and the benefits of marriage for the elderly persons[J]. The Gerontologist, 1998, 38(5): 618-627.