

基于科学数据的用户相关性判断实证研究^{*}

刘建平¹, 王健¹, 周国民¹, 石蕾², 张贵兰¹, 高飞¹, 韦草原¹

(1.中国农业科学院农业信息研究所, 北京 100081; 2.国家科技基础条件平台, 北京 100862)

摘要: 为探索科学数据用户相关性判断中线索、标准与感知价值关系, 发现科学数据用户相关性判断的认知机理, 本研究对影响科学数据用户相关性判断的线索、标准和感知价值因素进行测量, 以问卷调查形式对“共享杯”参赛人员进行调研, 获得相关数据。使用探索性因子对线索集、标准集和感知价值集进行探索性分析, 基于线索、标准、感知价值间已有理论关系, 通过结构方程方法对三者关系进行验证分析。研究发现, 影响科学数据用户相关性判断的主要线索、标准和感知价值因素, 通过对线索、标准和感知价值关系的实证分析, 探究科学数据用户相关性判断的认知机理。

关键词: 科学数据; 标准; 线索; 感知价值; 结构方程; 探索性因子分析

中图分类号: G202

DOI: 10.3772/j.issn.1673-2286.2017.04.004

相关性判断研究是信息查询与检索的主题, 根据研究的侧重点划分为面向系统相关性判断研究和面向用户相关性判断研究。面向用户相关性判断研究重视用户信息查询与检索的主体地位, 主要根据信息行为学、认知心理学等理论与方法对影响用户相关性判断的各维度因素进行相关研究。影响用户相关性判断的因素主要包括相关性线索、相关性标准及感知价值。相关性线索是信息的客观表达, 通常表现为用户能直接看到的元信息, 如标题、作者、发表时间等信息; 相关性标准是用户进行相关性判断的依据和工具性概念, 如用户通过关注作者的知名程度形成相关性标准的判断; 感知价值是从消费心理学引入的概念, 因为用户对文档相关性的判断和商品价值的感知具有相似性, 都是依据一定心理感知判断其相关性或价值(如用户基于文档能够弥补知识上的空白感知文档相关性, 则用户是根据文档的认知价值进行相关性判断的)。研究者基于以上研究理论, 对不同信息载体用户相关性判断进行研究, 如以文档类信息、网页类信息、图像类信息及音乐类信息为载体进行相关性判断研究。科学数据作为重要的信息载体之一, 用户在科学数据相关性判断过程中关注的主要线索、使用标准、感知价值及三者间关系将成为本研究重点。

1 相关性判断的研究综述

1.1 相关性判断研究中的线索、标准及其关系研究

Saracevic认为线索是使信息或信息实体相关的因素, 具体指在用户信息查询过程中能帮助用户进行相关性判断的客观因素^[1]; Wang等将文档元信息定义为影响用户文档相关性判断的线索^[2]; Barry等将标准定义为为用户判断信息相关或决策使用信息的依据^[3]; 成颖将标准定义为判据, 即相关性判断依据^[4]。国内外学者对于相关性判断研究大多基于相关性线索与标准关系开展, Barry等针对18位学者在其工作情境中, 利用访谈法和内容分析法研究用户判断文档相关的线索, 并从中识别出7大类标准^[5]; Cool等对300名大学新生在完成课程论文撰写中查找文献使用的标准进行研究, 利用内容分析法等从文档线索中识别出影响用户文档相关性判断的6大类标准^[6]。有研究者借鉴Barry和Cool等研究成果, 基于不同目标信息类型、信息用户, 从用户关注的线索中总结归纳出相关性判断的标准集, 对Barry等的研究成果进行扩展与延伸。Savolainen^[7]和Choi^[8]

^{*} 本研究得到中国农业科学院协同创新工程重点任务“多源异构农业大数据关联发现与计算挖掘”和国家社会科学基金“科学数据用户相关性标准与使用模式实证研究”(编号: 14BTQ056)资助。

等分别以网页和图像为信息载体, 研究用户在信息查询过程中关注的线索和标准使用; Laplante以音频为信息载体, 研究用户音乐类信息查询过程中关注的线索和标准使用^[9]。

1.2 相关性判断中的感知价值及其与线索和标准关系研究

Wang等提出文档选择模型, 将消费心理学中核心概念感知价值引入用户相关性判断研究, 利用内容分析法从用户访谈记录中总结出影响用户相关性判断的5类感知价值(认知价值、功能价值、条件价值、社会价值及情感价值), 同时指出感知价值与标准的关系(相关性标准的组合使用形成感知价值)^[2,10]; 后续研究将消费心理学的感知价值观引入信息查询领域, Gasper等研究情绪和行为的观念如何改变信息的感知价值和信息搜索, 研究结果表明, 积极情绪能帮助个人获取解决当前问题的信息, 而消极情绪则能帮助个人获取还未出现问题的信息^[11]; Choi等研究网络环境信息的感知价值^[12], 但关于价值与线索关系、价值与标准关系的后续研究较少, 这也是本文需要进行探索研究的一个方面。

综上所述, 研究者基于影响用户相关性判断的线索、标准、感知价值及三者关系开展相关研究, 主要存在两方面不足: 第一, 目前针对科学数据用户相关性判断的研究较缺乏; 第二, 缺乏基于线索、标准与感知价值三者间关系的实证研究。

本文针对相关性判断研究的不足, 以科学数据用户为研究对象, 构建影响科学数据用户相关性判断的线索集、标准集、感知价值集, 以问卷调查的形式获得用户对不同集合测量指标的评价数据。结合探索性因子法发现影响科学数据用户的主要线索、标准和感知价值因素, 在因子分析的基础上, 结合已有研究理论, 提出这3类因素间的假设关系, 利用结构方程方法验证该关系的合理性, 以期发现科学数据用户相关性判断的认知机理。

2 科学数据用户相关性判断的研究设计

2.1 研究目的

本研究的主要目的包括探索影响科学数据用户相

关性判断的线索、标准、感知价值, 识别科学数据相关性线索、标准、感知价值间的关系, 探索科学数据用户相关性判断的认知机理。

2.2 调查对象

本文以科学数据用户为研究对象。全国“共享杯”大学生科技资源共享服务创新大赛是我国首个以推动科技资源共享和利用为主题的大学生竞赛平台, 通过国家科技基础条件平台中心联合各地方科技平台, 为参赛人员共享各领域的科学数据集。参赛者通过对相关数据进行查询与分析形成参赛作品, 从而激发大学生的科研创新与自主创业能力。因此, 本研究以“共享杯”参赛者为主要研究对象, 调查其科学数据查询与使用中线索、标准和感知价值的理解情况。

2.3 问卷设计

本研究的问卷设计分3个阶段进行, 即前期调研、量表设计、样本测试。前期调研阶段主要通过调研与访谈的形式, 对科学数据用户在数据查询中关注的线索、所用标准及感知价值情况进行记录; 量表设计借鉴相关性判断中已有研究和前期调研阶段采集的数据, 对量表进行设计; 样本测试阶段通过分析测试样本, 修正和改进问卷。

2.4 问卷发放与回收

通过网络问卷形式, 借助全国“共享杯”大学生科技资源共享服务创新大赛平台, 对问卷进行有偿发放与回收。为控制网络问卷反馈的真实性与准确性, 本文采取的过程控制措施主要有4项: 第一, 在“共享杯”官方网站以正式通知的形式邀请来自全国的参赛者进行问卷填写; 第二, 问卷采取有偿填写, 要求问卷填写客观真实, 不可重复填写; 第三, 样本对象控制, 将目标对象定为硕士及以上学历和使用数据较频繁的本科生; 第四, 对问卷填写者进行随机抽样电话访谈, 要求重新填写问卷, 测量填写的一致性。网络问卷发放时间为2016年12月22日—2017年2月7日, 累计收到问卷673份, 来自全国21个省份、4个直辖市、3个自治区。在673份问卷样本中, 男生350人, 女生323人; 年龄段23—30岁的人数总计569人, 占总问卷样本数的84.5%; 硕士研究生

524人、博士研究生115人,共占总人数的94.9%。通过过程控制环节和问卷填写时间限制(低于500秒不纳入研究范畴),最终筛选出有效问卷542份。

2.5 测量指标

2.5.1 线索集的选取与得分

在以文档为目标研究类型的相关性判断研究中,Wang等将文档元信息定义为影响用户文档相关性判断的线索^[2],当研究对象转为科学数据时,线索的表现形式与文档不尽相同,具体表现为数据元信息(如数据覆盖时间、数据格式、数据名称等)。本研究对国内各大数据共享平台(包括地球系统科学数据共享平台、气象科学数据共享中心、国家标准物质资源共享平台、国家标准物质资源共享平台等23个数据共享平台)的元数据信息进行调研,最终总结出19条线索。调研对象根据数据相关性线索对其数据查询的影响程度进行打分,采用5分制李克特量表,量表中变量命名及通过SPSS20.0对

各线索均值与方差得分进行统计(见表1)。

2.5.2 标准集的选取与得分

本文相关性标准集的选取借鉴Schamber^[13]和Barry^[5]等的研究成果,综合前人研究和实验总结共得出12个标准,调研对象根据标准对其数据查询的影响程度进行打分,量表采用李克特5分制,量表中变量命名及通过SPSS20.0对各标准均值与方差得分进行统计(见表2)。

2.5.3 感知价值集的选取与得分

Sheth等提出的消费者价值理论指出,用户感知价值指用户在使用或选择某商品时,对该商品满足用户某方面需求的感知^[14]。本研究根据前期调研将用户在数据查询和使用过程中,感知数据满足用户的需求分为9种情况,采用5分制李克特量表,量表中变量命名及通过SPSS20.0对各感知价值均值与方差得分进行统计(见表3)。

表1 线索集均值与方差得分

线索集	均值	方差	线索集	均值	方差
是否支持下载(X1)	4.88	1.72	数据访问量(X11)	4.11	2.14
关键词(X2)	4.75	1.42	检索结果的排列顺序(X12)	4.02	2.00
数据质量说明(X3)	4.63	1.75	数据提供平台(X13)	4.00	1.93
数据摘要(X4)	4.58	1.65	数据共享级别(X14)	3.99	2.05
数据覆盖时间(X5)	4.52	1.72	数据格式(X15)	3.80	1.89
数据名称(X6)	4.41	1.87	数据生产者(X16)	3.76	1.85
数据获取渠道(X7)	4.27	1.64	空间位置(X17)	3.73	1.97
数据评级(X8)	4.23	1.78	数据提交者所在单位(X18)	3.67	1.99
数据全文(X9)	4.23	1.93	数据大小(X19)	3.40	1.80
数据生产和加工方法(X10)	4.22	1.84	-	-	-

表2 标准集方差与均值得分

标准集	均值	方差	标准集	均值	方差
质量(X20)	5.18	1.03	可用性(X26)	4.78	1.38
主题性(X21)	5.12	1.39	规范性(X27)	4.75	1.31
权威性(X22)	5.06	1.19	可获得性(X28)	4.70	1.33
时效性(X23)	4.84	1.38	新颖性(X29)	4.49	1.42
个人需求(X24)	4.82	1.47	可理解性(X30)	4.49	1.32
全面性(X25)	4.81	1.50	便利性(X31)	4.48	1.51

表 3 数据感知价值集均值与方差得分

感知价值集	均 值	方 差	感知价值集	均 值	方 差
获取合适的数据作为研究对象 (Y1)	3.69	0.89	使用数据建立模型 (Y6)	3.15	1.25
参考相关数据以指导后续研究 (Y2)	3.67	0.65	导师或长辈让查找数据 (Y7)	2.96	0.85
辅助说明研究理论 (Y3)	3.61	0.87	期望获得灵感或新思路 (Y8)	2.85	0.83
验证研究模型或假设 (Y4)	3.40	1.05	没有明确目的随意浏览 (Y9)	1.79	0.57
弥补知识上的空白 (Y5)	3.29	0.97	-	-	-

2.6 测量指标的信度与效度分析

本研究对测量指标信度的检验采用Cronbach’s α 系数, 利用SPSS20.0进行信度与效度分析, 测得各潜变量对应测量指标的 α 系数得分。经计算19项线索集、12项标准集、9项感知价值集对应的 α 系数分别为0.91、0.91、0.63, 且均在可接受范围; 研究主要通过结构效度检验数据效度, 使用KMO和巴特利特球形检验判别数据是否适合因子分析, 即测量变量间是否存在结构关系。经计算, 集合线索集、标准集、感知价值集KMO值得分分别为0.91、0.93、0.64, 巴特利特球形检验的显著性均为0, 即样本数据能够反映测量变量间的结构关系, 结构效度显著。

表明样本充足; 巴特利特球形检验的近似卡方值为4 352.11, 对应概率值为0 ($p<0.01$), 即相关系数矩阵与单位矩阵有显著差异, 测量指标间有强相关关系, 说明线索指标集适合做因子分析。该研究根据累计方差贡献率超过70%确定公因子个数为7, 利用主成分提取前7个因子作为公因子; 因子负荷矩阵说明公因子与原始测量指标间的关联程度, 由于未旋转的因子负荷矩阵结构不够简明, 因子解释力较弱。本文使用正交旋转法对因子进行旋转, 其优点是因子间提供的信息不会重叠, 使公共因子的负荷系数趋向0或1。本研究中根据因子负荷矩阵中, 在测量指标因子负荷系数大于0.5为依据, 选择表示公因子的测量指标, 标准集和感知价值集因子分析也采用该标准值。

3 测量指标的探索性因子分析

3.1.2 因子解释与命名

因子分析是将描述某一事物或概念的多个观测变量简化为少数几个潜在变量的多元统计分析方法^[15]。其前提是观测变量间存在相关关系, 即测量变量间存在公因子; 公因子是不可直接测量的潜变量, 能更具概括性地反映事物特征。因子分析根据是否使用先验条件, 分为探索性因子分析和验证性因子分析。探索性因子分析是在事先不知道影响因素的基础上, 利用测量数据和统计软件, 根据一定原则确定公因子的方法^[16]。本研究对影响科学数据用户相关性判断的3类因素集合进行探索性因子分析。

公因子命名根据以下两条原则: 第一, 公因子方差贡献率从大到小的顺序对7个公因子依次命名; 第二, 选取因子负荷矩阵中测量指标负荷系数大于0.5的测量指标对因子进行解释。(1) 第一公因子方差贡献率为12.75%, 该因子在原始测量变量X13、X14、X16、X18上有较大负荷值, 其共同反映数据的来源信息, 因此将其定义为数据来源信息; (2) 第二公因子方差贡献率为12.22%, 其在X9、X15上有较大负荷值, 将其命名为数据详细内容信息; (3) 第三公因子方差贡献率为11.88%, 其在X2、X4上有较大负荷值, 将其命名为主题详细信息; (4) 第四公因子方差贡献率为11.42%, 其在X3、X7、X10上有较大负荷值, 将其命名为获取与质量信息; (5) 第五公因子方差贡献率为9.55%, 其在X5、X17上有较大负荷值, 反映与数据时空相关的信息, 将其命名为时空信息; (6) 第六公因子方差贡献率为7.68%, 其在X8、X19上有较大负荷值, 反映与数据

3.1 线索集的因子分析

3.1.1 确定公因子与因子旋转

在测量指标的效度分析中, 线索集的KMO值为0.91,

评价相关的信息,将其命名为数据评价信息;(7)第七公因子方差贡献率为6.63%,其在X6、X16上有较大负荷值,但是这两个测量指标所表达的信息差异较大,X16在第一公共因子中也有较大的负荷值且与第一因子中其他测量指标表达的信息较一致,所以第七公共因子以X6所反映的信息命名,将该因子命名为主题信息。

3.2 标准集的因子分析

3.2.1 确定公因子与因子旋转

在测量指标的效度分析中,标准集KMO值为0.93,巴特利特球形检验的近似卡方值为3 231.81,对应概率值为0 ($p < 0.01$),标准指标集适合做因子分析。在标准集因子分析中,确定公因子与因子旋转和线索集分析过程一样,此处不再赘述,最终根据累计方差贡献率大于70%确定4个公因子。

3.2.2 因子解释与命名

(1)第一公因子方差贡献率为24.91%,原始测量变量X25、X26、X30、X31上有较大负荷值,共同反映用户对数据内容的理解,将其定义为数据内容相关标准;(2)第二公因子方差贡献率为19.90%,其在X20、X22、X27上有较大负荷值,反映用户对数据质量的理解决,将其命名为数据质量相关标准;(3)第三公因子方差贡献率为15.13%,其在X21、X28上有较大负荷值,而主题性与可获得性两个测量指标理论关系不大,故根据因子解释力较强的Y2对该因子命名,将其命名为数据获取相关标准;(4)第四公因子方差贡献率为12.40%,其在X23、X29上有较大的负荷值,反映用户对数据内容新与时间新的理解,将其命名为数据新颖度相关标准。

3.3 感知价值集的因子分析

3.3.1 确定公因子与因子旋转

在测量指标的效度分析中,感知价值集的KMO值为0.64 ($KMO > 0.5$),表明样本充足;巴特利特球形检验的近似卡方值为823.80,对应的概率值为0 ($p < 0.01$),说明

感知价值指标集适合做因子分析。在感知价值集因子分析中,确定公因子和因子旋转分析与线索集因子分析方法类似,最终根据累计方差贡献率大于70%确定4个公因子。

3.3.2 因子解释与命名

(1)第一公因子方差贡献率为20.56%,在原始测量变量Y4、Y6上有较大负荷值,反映用户对数据能进行建模和验证模型的感知,即用户对数据功能价值的感知,将其命名为功能价值;(2)第二公因子方差贡献率为17.12%,其在Y3、Y5上有较大负荷值,反映数据对用户知识的弥补和理论知识提升,即数据对用户认知的提升作用,将其命名为认知价值;(3)第三公因子方差贡献率为15.99%,在Y1、Y2、Y8上有较大的负荷值,反映用户对数据未来某一时刻价值的期望,即当满足一定条件时,数据能够带来的价值,将其命名为条件价值;(4)第四公因子方差贡献率为15.32%,其在Y7、Y9上有较大的负荷值,反映用户感知数据价值受外界影响较大的情况,将其命名为社会价值。

通过探索性因子分析对影响科学数据用户相关性判断主要线索、标准和感知价值因素进行分析,发现影响科学数据用户的7类线索及其重要程度、4类标准及其重要程度、4类感知价值及其重要程度。为更好地探究科学数据用户相关性判断的认知机理,对各因素关系的研究尤为重要。

4 科学数据用户相关性判断的关系模型构建与验证

基于以上研究现状,本研究从科学数据用户相关性判断的认知机理出发,结合结构方程法,拟通过以下假设对关系模型进行验证。

H1: 相关性线索使用对相关性标准使用有正向影响作用。

H2: 相关性线索使用对数据感知价值有正向影响作用。

H3: 相关性标准使用对感知价值有正向影响作用。

H4-1—H4-7: 线索集对应的公因子对线索使用有正向影响。

H5-1—H5-4: 标准集对应的公因子对标准使用有正向影响。

H6-1—H6-4: 数据感知价值集对应的公因子对数据价值感知有正向影响。

4.1 关系模型构建

利用结构方程法, 使用AMOS21.0, 结合探索性因子分析的结果, 对影响科学数据用户相关性判断的线索-标准-感知价值关系进行二阶验证性因子分析。构建

科学数据相关性判断过程中线索-标准-感知价值关系模型(见图1), 在探索性因子分析中因子作为一阶潜变量, 其对应的测量指标变量为旋转后因子负荷值大于0.5的指标项。线索集探索性因子分析得出的7类一阶潜变量对应的二阶潜变量为线索使用; 标准集探索性因子分析得出的4类一阶潜变量对应的二阶潜变量为标准使用; 感知价值集探索性因子分析得出的4类一阶潜变量对应的二阶潜变量为数据感知价值。

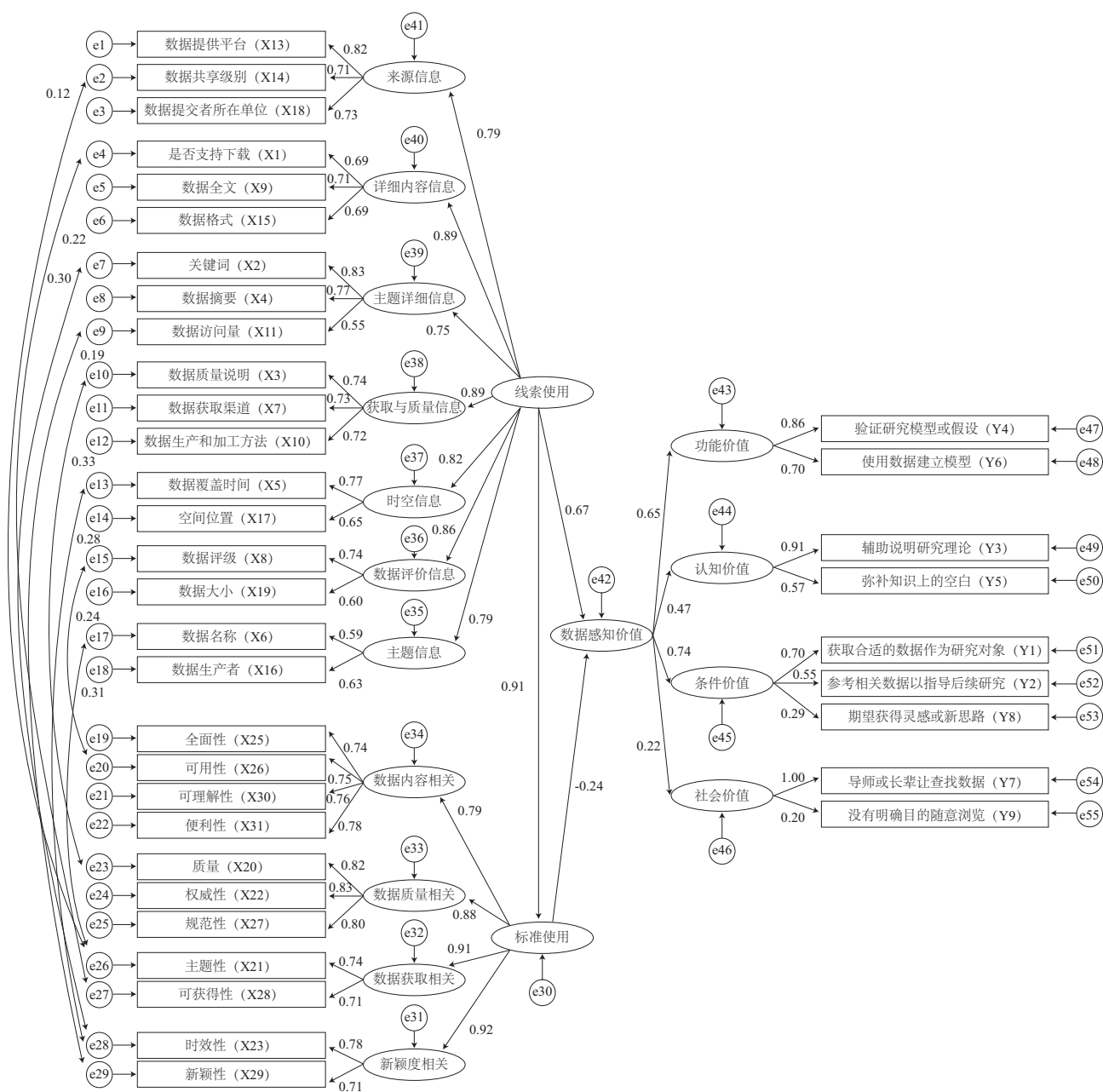


图1 科学数据相关性判断中线索-标准-感知价值关系模型

4.2 关系模型验证

4.2.1 模型拟合检验

结构方程通过构建一系列模型拟合指数,测量模型的可靠性与真实性。本研究根据侯杰泰等的研究^[17],选取拟合指数: χ^2/df (不受模型的复杂度影响)、 GFI (模型解释力)、 $RMSEA$ (比较理论模型与饱和模型的差距)、 NFI (比较假设模型与独立模型的卡方差异)、 CFI (假设模型与独立模型的非中央性差异)、 RMR (未标准化假设模型整体残差)。经计算本研究模型中各拟合指数 χ^2/df 、 GFI 、 $RMSEA$ 、 NFI 、 CFI 、 RMR 得分分别为2.60、0.90、0.05、0.90、0.92、0.06,各项拟合指数均在判断值范围,该模型总体拟合较理想,样本数据能较充分反映理论模型。 $RMSEA$ 值在0.05—0.08,表示模型适配度尚可,小于0.05表示模型适配度佳^[18];本研究 $RMSEA$ 为0.05,在可接受范围,模型适配度较好。

4.2.2 模型参数估计与假设检验

模型参数主要是对路径系数值和残差项的估计,本研究采用极大似然估计,路径系数的检验采用 t 检验。

(1) 模型回归路径系数估计与假设检验。如表4所

示,模型的回归路径系数与假设检验结论,标准化路径系数说明了研究假设中原因变量对结果变量的解释力。其中标准化路径系数均较高,仅个别路径系数较低;在路径系数 t 检验中,只有H3不显著,其他均为显著,且大部分是在0.01水平下显著,这充分说明研究假设的合理性。表4的研究结果验证了该研究提出的H1、H2、H4-1—H4-7、H5-1—H5-4、H6-1—H6-4的合理性,同时根据H3对应的标准化路径系数得分与 p 值判断应拒绝研究假设,即标准使用对数据感知价值有正向影响不成立。

(2) 模型相关性路径系数估计与假设检验。模型中相关关系说明变量间互相影响程度,该研究理论假设没有对测量变量间关系进行相应假设,表5列出测量变量间相关关系通过结构方程修正系数 MI 建立。修正指数是优化结构方程模型的主要方法,通过建立变量间关系以减少卡方统计量,进而提升模型整体拟合效果,同时通过建立变量关系发现有意义的信息。表5中研究假设关系建立遵循两条原则:第一,变量间关系的建立必须有理论解释力;第二,变量间关系建立能够较明显减少卡方统计量,提升模型整体拟合度。表5中 t 检验结果显示,通过修正指数建立的测量指标变量误差项间的相关关系均显著,验证了H7—H14的合理性。探索性发现测量变量相关关系信息。

表 4 模型回归路径系数估计与假设检验

研究假设	标准化路径系数	t 值	p 值	结 论
H1: 线索使用对标准使用有正向影响	0.91	11.96	***	支持
H2: 线索使用对数据感知价值有正向影响	0.67	3.07	**	支持
H3: 标准使用对数据感知价值有正向影响	-0.24	-1.16	0.25	不支持
H4-1: 线索使用对来源信息线索有正向影响	0.79	-	-	支持
H4-2: 线索使用对详细内容信息线索有正向影响	0.87	11.94	***	支持
H4-3: 线索使用对主题详细信息线索有正向影响	0.75	9.55	***	支持
H4-4: 线索使用对获取与质量信息线索有正向影响	0.89	12.45	***	支持
H4-5: 线索使用对时空信息线索有正向影响	0.82	12.18	***	支持
H4-6: 线索使用对数据评价信息线索有正向影响	0.79	11.57	***	支持
H4-7: 线索使用对主题信息线索有正向影响	0.86	10.20	***	支持
H5-1: 标准使用对数据内容相关标准有正向影响	0.79	12.98	***	支持
H5-2: 标准使用对数据质量相关标准有正向影响	0.88	14.23	***	支持
H5-3: 标准使用对数据获取相关标准有正向影响	0.91	13.04	***	支持
H5-4: 标准使用对新颖度相关标准有正向影响	0.92	-	-	支持

续表

研究假设	标准化路径系数	t值	p值	结 论
H6-1: 数据感知价值对功能价值有正向影响	0.65	-	-	支持
H6-2: 数据感知价值对认知价值有正向影响	0.47	6.07	***	支持
H6-3: 数据感知价值对条件价值有正向影响	0.74	4.30	***	支持
H6-4: 数据感知价值对社会价值有正向影响	0.22	3.67	***	支持

注: “***”表示0.01水平上显著, “**”表示0.05水平上显著。

表 5 模型相关性路径系数估计与假设检验

研究假设	标准化路径系数	t值	p值	结 论
H7: 数据共享级别与可获得性互相影响	0.12	2.39	**	支持
H8: 是否支持下载与可获得性互相影响	0.22	4.18	***	支持
H9: 关键词与主题性互相影响	0.30	4.93	***	支持
H10: 数据访问量与新颖性互相影响	0.19	3.75	***	支持
H11: 数据质量说明与质量互相影响	0.33	5.91	***	支持
H12: 数据覆盖时间与时效性互相影响	0.28	4.65	***	支持
H13: 数据评级与可用性互相影响	0.24	4.21	***	支持
H14: 数据名称与主题性互相影响	0.31	5.87	***	支持

注: “***”表示0.01水平上显著, “**”表示0.05水平上显著。

5 结论

本研究从影响科学数据用户相关性判断的线索、标准、感知价值及三者关系出发,通过探索性因子分析发现影响科学用户相关性判断的主要线索、标准以及感知价值因子;在已有研究基础上,通过结构方程法对三者关系进行验证性分析,构建科学数据相关性判断线索-标准-感知价值关系模型。

用户相关性判断以用户信息查询与检索中的主体地位为研究出发点,探究影响用户相关性判断的主要影响因素:线索是用户形成相关性判断的刺激因素,标准是线索刺激而产生或者激活的用户相关性判断的工具性概念,感知价值是用户基于信息需求对信息作用的感知。线索是对信息的客观表达,标准与感知价值是用户对信息的主观理解。探究影响用户相关性判断的主要因素及三者关系,能够揭示用户相关性判断的认知机理。本文探索性研究了影响科学数据用户相关性判断的线索、标准、感知价值三者之间的影响关系,从以下三方面内容揭示科学数据用户相关性判断过程中的认知机理特征。

5.1 线索使用对标准使用的正向影响作用

通过科学数据相关性判断中线索-标准-感知价值关系模型显示,线索使用对标准使用正向影响力达0.91,这充分说明科学数据用户通过关注客观线索形成相关性标准进行相关性判断的理论假设。通过探索性因子分析发现,科学数据用户关注的线索及根据探索性因子分析中各因子方差解释率由高到低依次为数据来源信息、数据详细内容信息、主题详细信息、获取与质量信息、时空信息、数据评价信息、主题信息;使用的标准及根据探索性因子分析中各因子方差解释率由高到低依次为与数据内容相关标准、数据质量相关标准、数据获取相关标准、新颖度相关标准。

同时,线索使用与标准使用的关系还体现在具体测量指标间的关系,本文通过结构方程修正系数M.I建立测量指标误差项间相关关系,对应研究假设H7—H14,结果表明研究假设H7—H14是合理的。该结论探索性揭示了线索与标准生成的对应关系:数据共享级别与可获得性、是否支持下载与可获得性、关键词与主题性、数据访问量与新颖性、数据质量说明与质量、数

据覆盖时间与时效性、数据评级与可用性、数据名称与主题性。如用户通过关注数据共享级别形成相关性判断标准可获得性进行判断,同时用户如果在判断数据相关性时对标准可获得性较重视,他会首先关注数据共享级别,其他相关关系解释类似。该结论从科学数据用户关注的线索、用到的标准及三者间关系的角度,揭示科学数据用户认知机理的主要内容。

5.2 线索使用对数据价值感知正向影响作用

模型验证分析中,线索使用对数据价值感知正向的标准回归路径系数为0.67,说明用户对数据价值的感知同样是通过关注数据的客观线索形成的,科学数据用户主要的数据感知价值及根据探索性因子分析中各因子方差解释率由高到低依次为功能价值、认知价值、条件价值、社会价值,该结论探索性验证了消费心理学中价值理论在用户相关性判断研究中的适用性。

5.3 标准使用对数据价值感知的影响不明显

在模型验证分析中,标准使用对数据价值感知的标准回归路径系数为-0.24,显著性检验 p 值为0.25,拒绝原假设,即相关性标准使用对数据价值感知正向影响不明显。因为从用户认知的角度出发,相关性标准和感知价值是用户基于客观线索的主观感知,二者影响关系不明显。这一结论对Wang等提出的文档选择模型中的结论(相关性标准的组合使用形成感知价值的理论假设)进行了验证,从实证分析的角度证明该关系不成立。

在大数据时代,科学数据作为一种非常重要的信息表现形式,科学数据相关性判断研究显得尤为重要。该研究结论为科学数据查询与检索理论研究提供一定的理论支持和方法借鉴,同时为开发更加高效、便捷、用户友好型的新一代智能数据搜索引擎提供一定的理论支持。

参考文献

- [1] SARACEVIC T.Relevance:a review of the literature and a framework for thinking on the notion in information science.Part III:behavior and effects of relevance[J].Journal of the Association for Information Science and Technology,2007,58(13):2126-2144.
- [2] WANG P,SOERGER D.A cognitive model of document use during a research project.Study I.Document selection[J].Journal of the Association for Information Science and Technology,1998,49(2):115-133.
- [3] BARRY C L,SCHAMBER L.Users' criteria for relevance evaluation:a cross-situational comparison[J].Information Processing & Management an International Journal,1998,34(2/3):219-236.
- [4] 成颖.信息检索相关性判断及应用研究[M].北京:科学出版社,2011.
- [5] BARRY C L,CAROL L.User-defined relevance criteria:an exploratory study[J].Journal of the Association for Information Science and Technology, 1994,45(3):149-159.
- [6] COOL C,BELKIN N J,FRIEDER O,et al.Characteristics of text affecting relevance judgments[J].Automotive News,1993,17(4):77-84.
- [7] SAVOLAINEN R,KARI J.User-defined relevance criteria in web searching[J].Journal of Documentation,2006,62(6):685-707.
- [8] CHOI Y,RASMUSSEN E M.User's relevance criteria in image retrieval in American history[J].Information Processing & Management,2002,38(5): 695-726.
- [9] LAPLANTE A.Users' relevance criteria in music retrieval in everyday life:an exploratory study[C]//International Society for Music Information Retrieval Conference, August 2010,Utrecht.[S.l.]:[s.n.],2012.
- [10] WANG P L,WHITE M D.A cognitive model of document use during a research project.Study II.Decisions at the reading and citing stages[J]. Journal of the Association for Information Science and Technology, 1999,50(2):98-114.
- [11] GASPER K,ZAWADZKI M J.Want information?How mood and performance perceptions alter the perceived value of information and influence information-seeking behaviors[J].Motivation and Emotion,2013, 37(2):308-322.
- [12] CHOI D,SHAH C.Perceived value of information sharing in online environments:user engagement and social reputation[J].Proceedings of the American Society for Information Science & Technology,2015,51(1):1-10.
- [13] SCHAMBER L.Users' criteria for evaluation in a multimedia environment[J].American Society for Information Science & Technology, 1991,28(6):126-133.
- [14] SHETH J N,NEWMAN B I,GROSS B L.Why we buy what we buy:a theory of consumption values[J].Journal of Business Research,1991,22(2):159-170.
- [15] 杜智敏,樊文强.SPSS在社会调查中的应用[M].北京:电子工业出版社,2015.
- [16] 王松涛.探索性因子分析与验证性因子分析比较研究[J].兰州学刊, 2006(5):155-156.
- [17] 侯杰泰,温忠麟,成子娟.结构方程模型及其应用[M].北京:教育科学出版社,2004:70-74.

[18] 吴明隆. 结构方程模型——AMOS的操作与应用: 第2版[M]. 重庆: 重庆大学出版社, 2010.

作者简介

刘建平, 男, 1989年生, 硕士研究生, 研究方向: 信息查询与检索, E-mail: 1592126095@qq.com。
王健, 男, 1971年生, 博士, 通讯作者, 研究员, 研究方向: 信息处理, E-mail: wangjian01@caas.cn。
周国民, 男, 1969年生, 研究员, 博士生导师, 研究方向: 农业科学数据共享技术、农业网络信息智能搜索技术、果园信息化技术。
石蕾, 女, 1982年生, 副研究员, 研究方向: 科学数据管理与共享。
张贵兰, 女, 1993年生, 博士研究生, 研究方向: 信息查询与检索。
高飞, 女, 1987年生, 博士研究生, 研究方向: 信息查询与检索。
韦草原, 女, 1993年生, 硕士研究生, 研究方向: 信息查询与检索。

An Empirical Study on User's Relevance Judgment Based on Scientific Data

LIU JianPing¹, WANG Jian¹, ZHOU GuoMin¹, SHI Lei², ZHANG GuiLan¹, GAO Fei¹, WEI CaoYuan¹

(1. Agricultural Information Institute, CAAC, Beijing 100081, China; 2. National Science & Technology Infrastructure, Beijing 100862, China)

Abstract: In order to explore the relationship among clues, criteria and perceived value in the process of scientific data relevance judgement and find the cognitive mechanism of user's relevance judgment in scientific data. This study measures the variables of clues, criteria and perceived value that affect scientific data relevance judgement. By investigating the participants who took part in "The Shared Cup" through questionnaire survey in China, this study obtains relevant data. The exploratory factor method is used to explore the clue set, the criteria set and the perceived value set. Based on the existing theoretical relationship between the clues, the criteria and the perceived value, and the relationship are verified by the method of structural equation modeling. The study finds the main clues, criteria and perceived value that affect the user's relevance judgement in scientific data. Through the empirical analysis of the relationship between the cues, the criteria and the perceived value, the cognitive mechanism of the relevance judgment in the scientific data is explored.

Keywords: Scientific-Data; Criteria; Clue; Perceived-Value; Structural Equation Modeling; Exploratory Factor Analysis

(收稿日期: 2017-03-21)

■ 书 讯 ■

《网络环境下叙词表编制与发展》

叙词表作为一种有效的知识组织工具, 在网络环境下继续发挥着重要作用。中国科学技术信息研究所常春研究员及其项目研究团队, 依托国家社科基金项目“网络环境下叙词表的编制模式与应用方式研究”(10BTQ048), 对这一课题开展研究, 并于完成《网络环境下叙词表编制与发展》。

该著作主要论述网络环境下叙词表的编制、维护与应用的理论和方法。编制方法包括网络环境下总体策略、总体形态、选词方法、词间关系建立方法、编制管理机制、维护方式方法等; 应用研究包括网络环境下相关技术的突破给叙词表带来的各类新的应用方式, 从术语服务、多语种翻译、概念组配、知识单元、概念映射、国外应用等多个方面, 阐述了网络环境下叙词表的发展方向。最后按年代顺序介绍了国内历年编制的、可查阅的重要中文叙词表, 理、工、农、医四大领域20多个可从网络上在线获取的英文叙词表。该书可供图书馆学、情报学专业相关专业人员参考使用。

《网络环境下叙词表编制与发展》于2015年4月由科学技术文献出版社出版, 定价38.00元。