

元分析方法在信息分析中的应用*

陈鹤阳¹² 朝乐门¹ (1 中国人民大学信息资源管理学院 北京 100872;

² 天津外国语大学图书馆 天津 300204)

摘要 文章通过对元分析方法的由来、概念以及特点的介绍,总结了元分析方法适用于信息分析的三个基本属性,梳理了元分析方法应用于信息分析时的主要流程。

关键词 信息分析 情报分析 元分析

Application of Element Analysis Method in Information Analysis

Chen Heyang¹² Chao Lemen¹

(¹School of Information Resources Management, Renmin University of China, Beijing, 100872;

²Tianjin Foreign Studies University Library, Tianjin, 300204)

Abstract Through the introduction of the origin, concepts and characteristics of Meta-analysis, the paper summed up three basic properties and the main flow of Meta-analysis for application in intelligence analysis.

Keywords information analysis, intelligence analysis, Meta-analysis

1 引言

情报分析(信息分析)是对所有来源的资料进行综合、评估、分析和整理记忆对情报结果的准备,将处理过的信息转换成情报,以支援已知或预期用户的需求^[1]。对已有信息的情报化、知识化是信息分析的目的。通常情况下,对已有研究的综合分析多采用文献综述的方法,它能够对源数据的综合和评估并对现有的研究总结出符合其未来发展的科学规律和现实实践的结论,但传统的方法是一种一般性叙述方法,文献的来源不全面,仅对能够检索到的文献进行综合分析,这样存在潜在的偏倚性,忽略了未发表且有价值的文献,对检索到的原始文献也没有统一的评价方法或根本未评价,这就会导致“垃圾进,垃圾出”的情况。同时,不同的研究者对同一主题的综合研究得出的结论可能有不一致或者相悖的情况,从而使同一主题分析的结论产生了多种可能性,导致这种情况出现的原因主要是研究者主观判断和分析上的差异,在一定程度上存在“评

述偏见”的问题^[2],很难对同一主题的研究得到全面、客观、明确的结论。同时,这种一般性的描述多以定性分析单个研究的结论为主,无法回答结果变量是否对总体存在一定的影响效应、效应的大小、效应的显著性以及独立研究结果存在差异性的原因等问题^[3]。综上所述,传统的信息分析方法在原始文献的来源、选择、评价、综合上都存在一定的局限性,随着信息的指数级增长也给应用传统信息分析方法带来了巨大的挑战,因此,我们需要引入一种新的方法来解决传统信息分析过程中存在的偏见和局限。元分析作为一种定性和定量分析相结合的方法,能够综合分析同一主题研究下研究结果不一致的独立研究之间的共性,并通过整合个体研究最终形成一致性、普适性的研究结论,从而得到比任何独立研究都更为精准的研究结论^[4]。元分析能够避免传统方法上的种种不足,鉴于此,将具有比较和综合分析的元分析方法引入到信息分析是十分必要的,能最大限度地降低信息分析中的各种偏倚,能够识别出潜在的变量,挖掘个体研究的关联性并发现新观点,为

* 本文系国家社会科学基金重大项目“云计算环境下的信息资源集成与服务研究”(编号:12&ZD220)、中国人民大学科学研究基金(中央高校基本科研业务费专项资金资助)项目(编号:14XNQ022)的研究成果之一。

信息分析的研究提供新的方法。

2 元分析方法

“Meta”来自希腊语的“After”,是指对至少两篇原始文献进行的分析,并且在每篇文献发表或交流之后进行的分析^[5-6]。元分析(Meta-analysis)是对所有筛选后符合标准的研究发现进行整合的分析,国内也将 Meta-analysis 译为“荟萃分析”、“总体分析”、“分析的分析”、“整合分析”等。元分析思想最早可以追溯到 20 世纪 30 年代, Tippett 等人^[7-10]都在寻找将不同农业研究结论整合的方法。Fisher 最先^[11]提出“合并 P 值”,后由 Beecher^[12]进一步改进,并提出了元分析的初步概念,1974 年美国教育心理学家 Glass 在综合前人研究的基础上,对 Primary analysis 和 Secondary analysis 进行了区分并提出了元分析概念,指出元分析的涵义:“就是对大量个体研究结果的集合进行统计学的分析,用严格的统计分析方法代替随意性的描述”^[13]。换句话说,元分析就是对个体研究结果的量化分析。Glass 运用元分析发展了心理学研究效度的方法,使得元分析方法受到很多研究学者的关注,尤其是医学领域的随机对照试验中。因此,早期的元分析也被称为典型定量元分析(classic meta-analysis or quantitative meta-analysis)^[14]。20 世纪 60 年代,随着社会科学研究的迅速发展,元分析方法逐渐引起学者们的关注,20 世纪 70 年代中期,美国的社会科学家才真正将元分析方法应用于实际研究中。随着应用领域的扩大以及方法在使用过程中的不断改进和发展,许多研究者也逐步将定性分析融入到定量元分析中,元分析已不仅仅是一种统计分析,而是对研究效应综合定量合并的系统分析方法。如 Eilenberg 提出定性元分析的思想,对个体研究在原有定量分析的基础上进行合并分析,并对特征的差异性进行分析,综合评价研究结果^[15]。Sacks 等提出运用系统论述的方法对统计分析结果进行特征分析^[16]。因此,元分析不同于传统文献综述描述性的介绍,是以定量和定性相结合的方法对以往不同时期、不同阶段的研究成果进行全面系统综合的分析。

虽然表述各异,但总体上元分析方法就是运用统计学的方法和特定模型对以往同一主题的独立研究进行完整的、系统的定量与定性分析,整合了以往同一主题下因研究设计、研究时期不同而产生的不同结果,是一种回顾性和验证性的方法,它的应用已经被扩大到非实验性研究中。

元分析方法与传统的描述性综合方法相比,有其独特的优势,为了更清晰的认识元分析方法的特点,笔者拟从研究对象、研究方法、研究目的、研究效果、适用范围五个方面总结元分析方法的特点。

(1)研究对象。传统的描述性综合方法的研究对象来源于正式发表的期刊论文,一些会议论文、重要报纸等其他来源中同一主题的研究成果都不在其研究讨论的范围内,而元分析方法则不同,它是要对所有来源的

资料进行搜集,得到同一主题研究下相关的所有文献,包括会议论文、研究报告、报纸、一次文献、未公开发表的研究等都包含在其检索范围内,然后经过严格的筛选机制和标准,确定最终的研究对象以保证原始资料的高质量。

(2)研究方法。从方法的角度看元分析,它是一种定性方法与定量方法相结合的一种研究方法,传统的描述性综合方法多是从定性的角度对以往同类研究中的文献进行描述性的分析,并没有对这些文献的研究结果进行有效的综合,而元分析方法有效避免了这种局限。它运用统计学方法从同类研究中组织和抽取信息,整合研究结果,并以平均效果量的形式表示整合后的结果,效果量的生成是经过加权平均处理得到的,受样本量大小的影响较小。传统的描述性综合方法应用定量分析时,通常采用频次分析的方法进行简单统计,但频次的统计与样本量的大小有直接关系,样本量较大则出现的频次相对较高,反之,频次的统计数量则较小。相比两种综合分析方法,元分析的研究结果更具有完整性、客观性和科学性。

(3)研究目的。元分析的目的是为了得到普适性、规律性的研究结论。社会科学领域不同学者有不同的研究视角和方法,对同类主题的研究往往会出现结论不一致的情况,研究主体的知识背景、研究设计的设置、样本的选择、研究方法的不同等原因都有可能造成结论的偏差,而元分析的目的就是要对这些同一主题下不一致的研究结果进行再分析,挖掘出这些研究成果所反映的共同效应,揭示各因素变量之间的相互关系,理清大量研究成果所代表的涵义。

(4)研究效果。元分析方法得出的结论是系统的并可重复的,并对研究结果的稳定性和是否存在偏倚进行检验,如果得到的结果变化较大说明稳定性不高,各研究结果之间的差异比较显著,因此需要解释造成这种差异的原因。如果得到结果的稳定性较高,说明各研究之间的差异不显著,比较平均,进行整合后可以得到一个普适性、规律性的结论,并且具有验证性,可以接受不同研究者的验证,避免了传统描述性统计方法的主观性问题。

(5)适用范围。元分析是定量与定性相结合的分析方法,就定量而言,它是通过平均效果量对研究对象与影响因素之间关系进行量化分析,那么进行分析的逻辑起点就要保证原始文献中必须含有所需的数据,能够计算出平均效果量以供进行综合统计分析时使用。但对于原始文献中未能提供可计算的数据或者研究设计无法满足定量分析的原始文献则可利用定性的方法加以总结分析。

3 元分析方法在信息分析中的意义

我国开展的信息分析活动是以文献信息为基础,并伴随大量实证研究来验证理论研究的科学性,通过对某一领域的文献研究进行分析、综合、归纳,为用户

提供专题情报、解决方案和实施办法。元分析方法是在满足传统文献综述的基础上,融合定量的研究方法,将局部的、片面的、分散的研究结果再分析,从而对研究主题有全面、系统的认识,对研究结论形成普遍性的认识。具体来说,元分析方法体现了信息分析以下三点属性:

3.1 评估信息分析的信度和效度

信度和效度是衡量信息分析结论一致性、稳定性、可靠性、准确性的标准。个体研究由于受到样本采集范围、大小以及研究方法的局限,使得各个独立研究的研究结论不一致,存在较大的差异或矛盾。传统的信息分析方法无法解决和调和研究结论的不一致或矛盾,导致同一研究问题存在多种不同的研究结论,其可靠性和准确性会遭到质疑,从而造成信息分析结果信度和效度不高,不利于信息分析活动的开展。而元分析方法在整合个体研究结论基础上,形成了相关研究的大样本,避免了个体研究的缺陷,利用定量综合评价效应大小的方法,估计研究效应的平均水平来解决个体研究中研究结论不一致或矛盾的情况。通过对同一个研究问题增加同质的题目,样本量增大,加大了分布的范围,那么形成的误差就会降低,从而改进了项目取样的代表性,更好地反映了样本的真实水平,同时也改进和提高了检验功效,使每个问题的随机误差相互抵消,提高研究结果的信度和效度,一些在个体研究中没有体现出来的相对较弱的效应通过元分析也能够被显现出来,提高了对结论的论证强度和效应的分析评估力度。元分析方法能够揭示和分析同一主题研究下多个独立研究之间的分歧,发现独立研究中的不确定性,并通过异质性(齐性检验)判断分析多个独立研究间异质性的原因,估计可能存在的各种偏倚(bias)。Ali Reza Montazemi 和 Hamed Qahri-Saremi^[7]对影响用户采用在线支付方法的因素进行分析,不同的研究者对在线支付的影响因素得出的结论具有异质性,他们对搜集实证研究得到的 25 265 个案例进行元分析,最终得出影响采用在线支付方法的 10 个因素,并分析了 10 个因素之间的相互关系和各个因素对采纳在线支付行为的不同重要性进行了对比。类似的研究还有 Eldon^[8]对不同组织信息系统成功因素的元分析,在已有 8 个维度分析的基础上添加了几个额外的因素,为信息系统的管理实践提供了依据。Ankem^[9]从人口统计学和情境变量对癌症患者的信息需求进行综合分析,结果表明癌症患者的性别、教育程度、癌症诊断时间、癌症阶段与他们的信息需求不相关,并得出老年癌症患者信息需求较低是因为他们对情感的压抑,根据此结论,医护人员应为老年癌症患者积极传递健康信息促进他们的信息需求。因此,在信息分析中应用元分析方法能够有效的排除个体研究中统计学的误差,通过平均效应量的计算,得到同一主题研究下的平均效应水平,使效应估计的有效范围更精确,大大提高研究结果的信度和效度。

3.2 实现研究问题的集成分析

各个领域的研究者对同一主题的研究可能会产生

大量的研究文献,这些不同领域研究证据的积累和评价对合并结论的客观性有重要作用。在物理学中,元分析被看作是批判性评价方法,强调发现单个研究本身所存在的偏见,并在尽可能降低偏见的情况下挖掘有价值的相关物理属性,实现对单个研究的集成分析^[20]。早在 1978 年 Rosenthal 和 Rubin^[21]将人际关系期望对行为影响的 345 个单个研究进行合并、集成分析,从最初实验主体的自我实现到老师对学生、雇主对雇员、治疗师对患者的行为影响,扩大了研究样本,并将这些单个研究整合分成了 8 个维度即反应时间、墨迹测试、实验室动物学习、访谈、心理判断、学习能力、感知能力和日常生活情景进行再分析,确定了人际关系期望对行为影响发生的概率,估计了 8 个维度人际关系期望对行为影响的平均量,经过合并、集成分析使研究结果更加清晰化,对各个分类维度的影响因素更加具体化。Glass 和 Smith^[22]对班级规模与学习成绩之间影响关系的 725 个单个研究进行合并分析,在单个研究中,得出的结论并不一致存在正相关、负相关和不相关几种可能性,将这些单个研究整合分析,最终得出班级规模与学生成绩有密切关系,并确定班级规模在某一范围内与学生成绩的显著影响关系。Guo 和 Gifford^[23]通过 74 个个体研究分析了土地使用情况的改变对土壤碳储量的影响,最终得出当土地使用从牧草地到种植园、原始森林到种植园、原始森林到农作耕地、牧草地到农作耕地,土壤的碳储量会下降,下降的比例分别是 10%、13%、42%、59%;而从原始森林到牧草地、农作耕地到牧草地、农作耕地到种植园、农作耕地到次生林,碳储量会增加,分别为 8%、19%、18%、53%,对各种土地使用变化对碳储量的影响得出了精确的百分比,对实际工作有重要的指导意义。元分析在综合大量研究文献的基础上,注重独立研究之间的差异,利用元分析将集成化单个研究,并对同一主题研究中存在的分歧运用定量的手段判断差异对研究结果的影响程度,使对主题的研究更加精确全面,便于得出一般性的研究结论。

3.3 提出新的分析假设

信息分析的目的就是为了发现新问题,预测未来,并对未来的问题研究提供帮助。相比传统的信息分析方法,元分析通过对以往文献的反复调查和纵向调查,能发现单个研究中尚未提及或不能回答的问题,并据此提出新的研究假说及时发现新的有价值的研究问题,了解研究主题的发展趋势和方向,并能够为未来的研究精确计算出研究主题所需要的样本量^[24]。Brewin 等^[25]用元分析的方法整合单个研究中对创伤后应激障碍(PTSD)的影响因素,并对有影响的 14 个不同危险因素进行分析,从人口学特征(性别、年龄、种族)、成长背景(教育程度、创伤因素、童年灾难)、病史(精神病例、虐待儿童、精神病家族史)三方面预测 PTSD。Lazarou 等^[26]利用元分析对 1966~1999 年间美国 39 家医院住院患者中由严重或者致命的不良药物反应的发生率进行分析,估计出严重或者不良药物反应的发生率是 6.7%和

0.32%, 并指出未来不良药物反应的发生率是极高的。哈佛大学的 Ambady 和 Rosenthal^[27]从社会和临床心理学的角度用元分析方法对表现性行为进行短期观察(小于5分钟),精确地预测所能出现的各种客观结果。Hwang 和 Lin^[28]探讨了信息过载对决策质量的影响,从信息过载的两个维度(信息多样性和信息重复性)对破产预测进行测试,结果表明,无论信息维度是多样性还是重复性对决策质量都有负面影响。利用元分析方法也能够了解一个领域的发展趋势,Alexander Serenko^[29]对108篇知识管理科学计量学的文献进行元分析,指出知识管理高产的前6个国家分别是美国、英国、加拿大、德国、澳大利亚、西班牙,并提出知识管理已是一个学术相对成熟的领域,有很深的历史根基,研究者在未来应更加注重科学计量的方法进行知识管理领域的实证研究,尤其是知识管理新框架的研究。

4 基于元分析方法的信息分析流程

首先,元分析方法是建立在原始分析基础上的一种分析方法,原始分析按照信息分析的一般流程^[30],应用定性和定量的各种分析方法对研究主题进行分析,最终得出研究结论。而元分析的信息分析流程是对原始分析得出的研究结论再分析的过程,对原始分析中文献来源、选择、评价、综合等方面存在潜在偏倚的修正,如图1所示。

从图1中可以看出,元分析的信息分析流程分为两个主要部分,即原始分析和元分析。首先,某一主题研究的大量原始分析是元分析的必要条件,这些原始分析的研究结论可能会存在不一致性。通过对原始分析所得结论的分析提出研究问题或研究假设,制定研究设计方案,包括研究对象、研究路径、效应量设计;然后确定检索策略,进行全面、系统的搜集文献资料,在实际应用中可以先进行预检索,确定大概范围,根据检索得到的结果适当的修改检索策略;提取符合标准的文献资料,根据资料类型确定反映研究效应的指标;确定多个个体研究的统计量,整合汇总成单一的效应量;对纳入进行异质性(齐性)检验,剔除低质量的研究,否则将不同背景的研究结果

合并不但毫无意义,也容易导致错误的结论,对符合要求的文献选择恰当的统计模型进行统计分析;对得出的估计值进行统计学检验,看是否达到显著性水平;挖掘因变量和结果变量之间的关系强度,进行灵敏度分析和偏倚分析,验证研究结果的稳定性;最后得出研究结果,若验证结果的稳定性不高,则需解释造成这种差异的原因,若稳定性较高,则可以认为元分析的研究结论具有一般性。

从图1中可以看出元分析的信息分析过程具有以下三个特征,首先,通过对原始文献的标准化筛选、评价、整合,扩大了样本范围,降低了个体取样的误差,对统计偏误进行了修正,实现量化的信息分析;第二,根据异质性检验的结果,调节纳入文献的数量,提高研究结论的准确性;第三,发现个体研究中未阐明的问题,提出新的假说和研究思路。

5 结语

从分析任务类型看,信息分析可以分为原始分析和元分析两类,前者主要基于原始数据进行挖掘和创

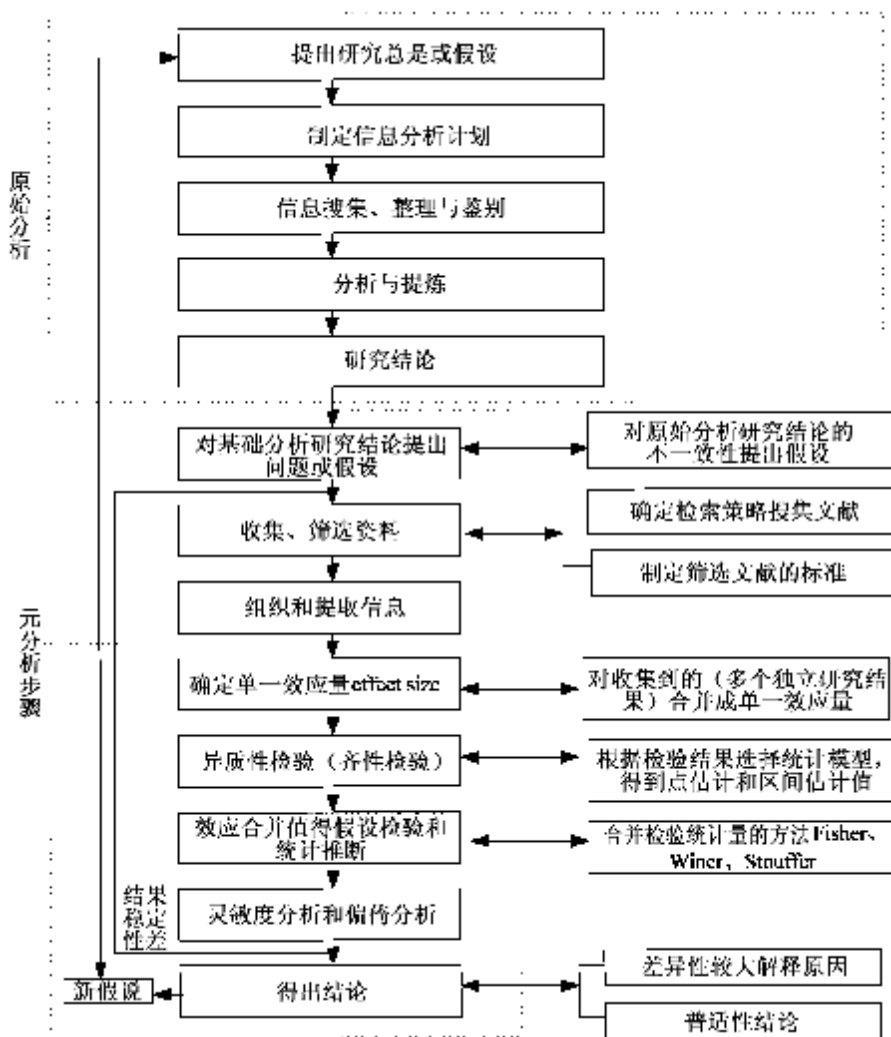


图1 基于元分析方法的信息分析流程

新的一次分析方法;后者主要侧重于在已有研究结果的基础上进行验证和集成的二次分析方法。目前,信息分析及其相关研究对原始分析的重视较多,而对元分析方法的研究很少,元分析方法作为基于原始分析结果的二次分析方法将广泛应用于信息分析工作中。元分析方法作为定量综合为主的分析方法,在许多其他领域的应用取得了很好的研究效果,能够有效克服传统文献综述中存在的不足,合理、清晰地解释独立研究中存在的模糊或冲突性的结论;能够解决信息分析过程中精度不高、受主观因素影响严重的问题;能够对某一主题未来研究的变量关系实现量化预测,并为未来的研究方向提供参考意见,为信息分析研究工作者提供了新的方法。目前,元分析方法在信息分析领域的研究还处于起步阶段,但随着信息分析实证研究的日益增多,利用元分析方法实现结果的整合分析将是未来的研究趋势,将发挥知识创造的价值。

参考文献

- [1] 程勇,程石.新编美国军事术语词典[M].北京:国防工业出版社,2008:46-47.
- [2] Wolf F M. Meta-analysis: Quantitative Methods for Research Synthesis[M].Newbury Park: Sage Publications,Inc.,1986.
- [3] Borenstein M, Rothstein H, Cohen J. Comprehensive Meta-analysis: A Computer Program for Research Synthesis[M].Englewood:Biostat,1999.
- [4] Hunter J E, Schmidt F L. Methods of Meta-analysis: Correcting Error and Bias in Research Findings [M].Newbury Park: Sage Publications,Inc.,2004.
- [5] John H,Noble J R. Meta-analysis: Methods, strengths, weaknesses, and political uses[J].Journal of Laboratory and Clinical Medicine ,2006(1):7-14.
- [6] Wolf F M. Meta-analysis: Quantitative Methods for Research Synthesis[M]. Beverley Hills, Calif.: Sage, 1986:13.
- [7] Tippett L H C. The Methods of Statistics[M]. London: Willam and Norgate, 1931.
- [8] Fisher R A. Statistical Methods for Research Workers[M]. 4th ed. London: Oliver and Boyd, 1932.
- [9] Pearson K. On a method of determining whether a sample of size n supposed to have been drawn from a parent population having a known probability integral has probably been drawn at random [J] . Biometrika, 1933,25(1):379-410.
- [10] Cochran W G. Problems arising in the analysis of a series of similar experiments [J].Journal of the Royal Statistical Society, 1937,100(Suppl 4):102-118.
- [11] Fisher R. A mathematical examination of the methods of determining the accuracy of an observation by the mean error, and by the mean square error [J].Monthly Notices of the Royal Astronomical Society,1920,80(3):758-770.
- [12] Beecher H. Appraisal of drugs intended to alter subjective responses, symptoms[J].Journal of the American Medical Association,1955,158(5):399-401.
- [13] Glass G. Primary, secondary, and meta-analysis of research[J]. Educational Researcher,1976,5(10):3-8.
- [14] Milos J. Meta-analysis in medicine where we are and where we want to go[J].Journal of Clinical Epidemiology,1989,42(1):35-44.
- [15] Ellenberg S S. Meta-analysis: The quantitative approach to research review[J].Seminars Oncol, 1988,15(5):427.
- [16] Sacks H S,Berrier J,Reitan D,et al. Meta-analysis of randomized controlled trials [J].The New England Journal Medicine, 1987,316(8):450-455.
- [17] Ali Reza Montazemi, Hamed Qahri-Saremi. Factors affecting adoption of online banking A meta-analytic structural equation modeling study [J].Information & Management,2015, 5 (2)210-226.
- [18] Eldon Y Li. Perceived importance of information system success factors: A meta analysis of group differences[J].Information & Management,1997,3(2):15-28.
- [19] Ankem k. Factors influencing information needs among cancer patients: A meta-analysis [J]. Library & Information Science Research,2006,28(1): 7-23.
- [20] Schoolman H M. Anatomy, physiology and pathology of biomedical information [J] .Western Medical Journal, 1982,137(5):460-466.
- [21] Robert Rosenthal, Donald B Rubin. Interpersonal expectancy effects: The first 345 studies [J].The Behavioral and Brain Sciences,1978,1(3):377-415.
- [22] Glass G V, Smith M L. Meta-analysis of the relationship between class size and achievement [J]. Educational Evaluation and Policy Analysis, 1979,1(1): 2-16.
- [23] Guo L B,Gifford R M. Soil carbon stocks and land use change: A meta analysis[J].Global Change Biology, 2002,8(4):3345-360.
- [24] Egger M, Smith G D. Meta-analysis: Potentials and promise[J]. British Medical Journal,1997,315(7119):1371-1374.
- [25] Brewin C R, Andrews B,Valentine J D. Meta-analysis of risk factors for posttraumatic stress disorder in trauma-exposed adults [J]. Journal of Consulting and Clinical Psychology, 2000,68(5):748-766.
- [26] Lazarou J, Pomeranz B H, Corey P N. Incidence of adverse drug reactions in hospitalized patients - A meta-analysis of prospective studies [J]. Jama-Journal of The American Medical Association,1998,279(15):1200-1205.
- [27] Ambadv Nalini, Rosenthal Robert. Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis[J]. Psychological Bulletin,1992,111(2):256-274.
- [28] Hwang M I,Lin J W. Information dimension, information overload and decision quality [J]. Journal of Informaiton Science, 1999,25(3): 213-218.
- [29] Alexander Serenko. Meta-analysis of scientometric research of knowledge management: Discovering the identity of the discipline[J]. Journal of Knowledge Management,2013,17(5):773-812.
- [30] 卢小宾.信息分析概论[M].北京:电子工业出版社,2014:20-39.

[作者简介]陈鹤阳,女,1983年生,天津外国语学院图书馆馆员,中国人民大学信息资源管理学院博士生。

朝乐门,男,1978年生,中国人民大学信息资源管理学院副教授,硕士生导师。

收稿日期:2015-06-10