

doi:10.3772/j.issn.2095-915x.2015.01.012

情报方法库研究

王强^{1,2,3}, 陈安琪^{2,3}

(1 北京大学新闻与传播学院 北京 100871; 2 北京市科学技术情报研究所 北京 100048; 3 北京市科学技术研究院 竞争情报与创新评估重点实验室 北京 100048)

摘要: 以情报方法论相关研究为理论支撑, 提取情报方法的属性及相互关系, 分析情报方法应用的一般过程, 在此基础上构建情报方法的概念模型, 设计情报方法库。

关键词: 情报方法, 方法库, 概念模型

The Study on the Intelligence Method Base

WANG Qiang^{1,2,3}, CHEN Anqi^{2,3}

(1 School of Journalism & Communication, Peking University, Beijing, 100871; 2 Beijing Institute of Science and Technology Information, Beijing, 100048; 3 Key Laboratory of Competitive Intelligence and Innovation Evaluation, Beijing Academy of Science and Technology, Beijing, 100048)

Abstract: The attributes and relationship between different methods are extracted and the general process of the application of intelligence method is analyzed on the theoretical support of intelligence research methodology, based on which the conceptual model describing intelligence method is constructed and the method base is designed.

Keywords: Intelligence method, method base, conceptual model

为不断适应情报工作的新变化, 情报方法也在不断发展。一方面, 文献计量法等情报学专门的研究方法的外延在不断拓展, 方法实施的客体由文献数据向网络数据拓展; 另一方面, 情景分析、社会网络分析等其他学科的方法也被不断引入情报工作中, 充实了情报方法体系。随着情报方法

体系的不断扩充和方法数量的增加, 在情报研究和服务中, 哪些方法可以达到研究和服务的目标, 这些方法有哪些功能, 如何实施, 对数据和实施人员等外部资源有怎样的要求, 有哪些局限性, 怎么弥补这些局限性等, 是情报工作人员在选择方法中常常遇到的问题。除了方法的功能外, 方

法对数据、方法之间的兼容或者互斥关系也分别提出了不同的要求。目前,情报方法选用更多地依赖于情报人员的经验和研究习惯,方法是否适用只能通过经验判断或者不断尝试,造成了时间的延误和资源的浪费。

本文以情报方法论相关研究成果为理论支撑,提取了情报方法的属性和相互之间的关系,分析了情报方法应用的一般过程,构建了描述情报方法的概念模型,在此基础上设计了情报方法库,用于对情报方法进行组织、存储和检索。

1 相关研究

方法库是按照一定的组织原则将方法汇集起来,以便于方法的组织、检索和调用。在决策支持系统中,方法库建设的目的是为了提高方法的运行效率,实现决策过程中的资源共享,同时方便进行方法扩展,从而使计算过程可以根据决策需要从方法库中选择方法,从数据库中选择数据,然后将结果输出,方法库因此成为决策支持系统的重要组成部分^[1],并已经在成本估算、多目标优化、工业设计等领域得到广泛应用^[2,3]。方法库一般由方法库管理系统、方法库和方法字典库三部分构成。方法库管理系统的负责对库中管理的各种方法进行集中控制,功能包括方法命名、分类、合成、执行等;方法库负责对方法数据进行存储和调用;方法库字典则集中管理方法信息,包括对方法类别、方法功能、方法使用范围、调用形式、参数等内容的描述^[4]。

方法库作为方法数据存储和组织的一种形式,在数据挖掘领域得到的进一步发展。随着数据挖掘方法的不断丰富、发展和广泛应用,国内外学者尝试使用本体对数据挖掘方法进行组织,其中有代表性的有 OntoDM、EXPO、DAMON、KDDONTO、KD ontology 等^[5,6]。数据挖掘本体在复用 Basic Formal Ontology (BFO)、OBO

Relation Ontology (RO) 等成熟本体基础上将数据挖掘涉及的数据类型、算法、执行过程等要素及其相互关系进行抽象和组织,为数据挖掘方法的组织、查询、使用、扩展奠定了基础^[7,8]。

决策支持系统中的方法库和数据挖掘方法本体等相关研究为情报方法库的研究和设计提供了很好的借鉴。然而与决策支持系统和数据挖掘系统中的方法不同,情报方法来源更加广泛,不仅有定量方法,同时也有大量定性方法,这些方法同样需要有序的组织 and 存储。与此同时,由于情报需求的多样性、情报方法来源的广泛性及情报研究人员自身的研究水平和研究习惯等原因,造成当前情报方法的选择和使用存在一定的随意性和不规范性,情报方法库的建设将对以上现象的改善发挥积极作用。

2 情报方法的属性及相互关系

方法论研究作为情报基本理论的重要组成部分,一直以来是情报理论研究的主要方向之一,内容集中在情报方法研究的范围、情报方法的形成和发展过程、情报方法的来源、情报方法体系的结构、情报方法的表现形式等方面。情报方法由哪些属性构成,方法之间存在怎样的关系,是对情报方法进行数据建模必须要回答的两个关键问题,本文试图从这些研究成果中,提取情报方法的构成要素和影响因素。

2.1 情报方法的属性

属性是指事物本身所固有的性质及与其他事物之间的关系^[9]。情报研究方法论是关于情报研究方法的知識体系,其研究内容主要围绕方法品质、性能、特点、适用范围、实施过程、注意事项、优缺点、来源及方法之间的关系展开^[10-12]。以上内容作为情报方法的固有性质,体现了方法之间的主要区别,在具体研究中,也成为方法分类或

层次划分的主要标准以及情报方法选择的主要依据^[13-16]，因而这些内容构成了情报方法的基本属性。

2.2 情报方法之间的关系

情报方法体系对情报方法的研究范围和内容进行了界定，探讨了不同来源、不同应用的研究方法。体系是指“若干有关事物或某些意识互相联系而构成的一个整体”¹，不少学者对方法之间的联系进行了探讨，然而更多关于联系的研究焦点在情报方法之间的类目关系方面，混合方法的引入丰富了情报方法体系研究中方法之间联系的类型和内容。除类目关系外，从混合方法研究目的中可以看出，方法之间的联系至少还包括：三角互证、互补、发展、验证和扩展^[17]。然而在方法的实施过程中，方法之间的融合常常出现在研究的某个过程或阶段，方法之间的混合也并非是从多个方法从整体上存在并行或顺序的关系，会出现不同方法各个步骤之间的混合使用。例如在多

种方法的实施过程中，资料和数据采集阶段都可能会使用到问卷或访谈中的多个步骤。

尽管情报方法创新仍在不断加快，业已成熟的情报方法，包括情报研究特有方法和其他学科引进方法以及不依赖具体学科的一般方法，仍然是情报研究和实践中的主流，并且已经通过大量实践进行了检验。因此，情报方法研究的重心应该逐渐从方法个体研究向方法之间关系的研究进行转变，通过方法之间关系的揭示，综合、规范的运用多种方法解决实际的情报问题，进一步提高方法选用的规范性和方法实施的效率。

3 情报方法实施的一般过程

为使用灵活，并体现方法实施的一般性，本文将方法定义为可以独立解决某一问题的思路、途径、方式和程序。情报方法的应用需经历定义任务、选择方法、实施方法和评估方法四个基本过程（图1）。

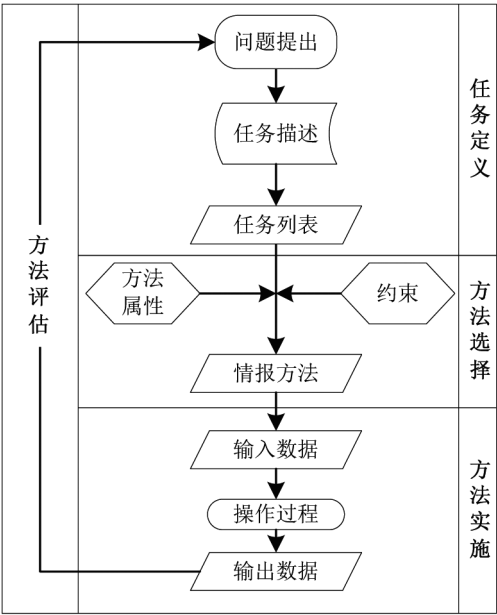


图1 方法实施的基本过程

1 莫衡等. 当代汉语词典 [M]. 上海: 上海辞书出版社. 2001.

3.1 任务定义

情报方法是面向具体业务问题的,任何问题的提出都有可能触发方法的选择和使用,这里的问题从范围来看有可能是面向整个研究的整体问题,也有可能是在研究过程中某个环节面临的阶段性问题;从形式来看有可能是一段描述,也有可能是几项要点。但无论是哪种范围或者形式,都需要将这些问题经过概念化和操作化处理后,转化为一组任务,以明确方法实施的目标。在此阶段,输入内容为问题描述或研究目标,具体处理过程是对问题或研究目标进行概念化和操作化处理,最后形成一组任务列表,作为方法选择和效果评估的依据以及方法实施的目标。

3.2 方法选择

当明确研究任务后,接下来的工作就是如何选择适当的方法去完成这些任务。方法的选择通常受到方法自身属性、任务要求以及外部条件三方面因素的影响。

(1) 方法属性

客观来说,方法本身的属性及其与任务要求相契合的程度是影响方法选择最关键的要素。常见的方法属性包括方法能够解决的主要问题,方法的适用条件,应用方法应满足的假设,方法的局限性,方法对数据的要求等等。

(2) 任务要求

首先是不同的任务类型对方法的要求,从任务本身来看,包括描述型、探索型、解释验证型、预测型、评价型、规划设计型^[17]五种类型;从情报研究的社会研究属性来看,情报方法解决的任务包含解释型和描述型。无论何种划分方法,每种类型都有其自身的特征以及对方法的要求,是方法选择的重要依据之一。

(3) 外部条件

方法选择过程中的外部条件是指除方法的属

性和任务以外,其他影响方法选择的元素。例如评价研究中,如果数据完整、质量较高,同时任务相关概念的界限清晰,能够进行较为精确的数学描述,并且专家资源完备,这时可以考虑德尔非法结合层次分析的方法解决相关问题;如果任务相关指标具有较强的模糊性,专家资源可获得性较差,则应优先考虑模糊综合评判结合客观性指标权重确定的方法。

3.3 方法实施

方法实施是将任务定义、方法选择过程获得的诸多要素,包括任务列表、任务要求、外部条件、方法属性等内容作为输入,经过一系列处理,形成能够满足任务要求的输出的过程。

(1) 方法实施中的要素

方法在实施过程中包含两类要素,分别是数据(DS)和针对数据的操作(OP)(图2)。从总体上来看,方法实施的基本步骤是得到输入数据,对输出数据进行操作,得到操作后的输出数据(图1)。具体来说,操作过程是一系列输入数据经过处理,得到输出数据的过程的总和(图2)。此处提及的数据可能是定性或者定量数据,也有可能是混合类型的数据,是所有作为方法处理对象的资料或者信息的统称。

在方法的实施过程中,任何操作过程都可以被分解为若干处理步骤,在具体问题的解决中,操作是指有明确的输入数据和输出目标的一次性处理过程。这里所指的一次性处理,是指一个明确研究方法、思维方法、计算公式、算法的调用。例如在个案研究中,资料的收集是其中一个必要的步骤,在该步骤中,对个案研究所要解决的问题的描述指标是该步骤的输入,调查研究或其他非介入性方法是该步骤的处理过程,指标对应的数据即输出数据。如此构成一个完整的处理过程。

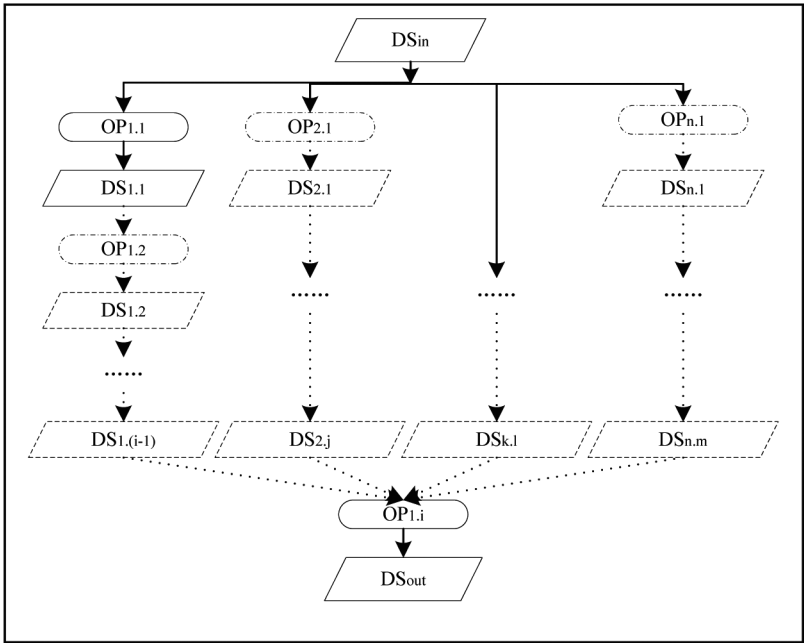


图 2 方法实施的过程

(2) 方法实施中步骤之间的关系

在方法实施过程中，步骤之间存在两种关系，即串联和并联。步骤的串联是方法中前一个步骤的输出数据为后一个步骤的输入数据，或者是后一个步骤的输入数据的一部分；步骤的并联是指一个步骤与其他步骤并行进行，它们有可能共用同一组输入数据，但是各步骤的输出数据互不为输入数据。如图 2 中，OP1.1 与 OPk.1 即为并联关系，而 OP1.1 与 OP1.2 为串联关系。

3.4 方法评估

情报方法的实施通常不是一次性过程，需要对方法实施后得到的结果与最初提出的问题进行对比，以评估方法实施的效果。方法实施的成功与否，一方面取决于通过方法实施得到的结论是否能够有效回答需要方法解决的问题，完成问题

概念化后设立的任务列表，另一方面取决于方法实施的过程，例如方法的选择是否符合方法的基本属性和方法实施的条件约束，方法的实施过程是否规范、合理，效度、信度等必要的检验指标是否达标等。

4 情报方法概念模型

根据情报方法的属性和相互关系，结合方法应用的一般过程，本文构建了情报方法概念模型，其中实体有方法、功能、数据、步骤等。情报方法库的 E-R 模型如图 3 所示。E-R 模型是数据模型设计的最常用方法，由实体、联系、属性三个基本部分组成。实体是现实世界中存在的客观事物，也就是要描述的对象，用方框表示；联系是实体之间的关系，用菱形框表示；实体或联系的性质称为属性，用椭圆形表示^[18]。

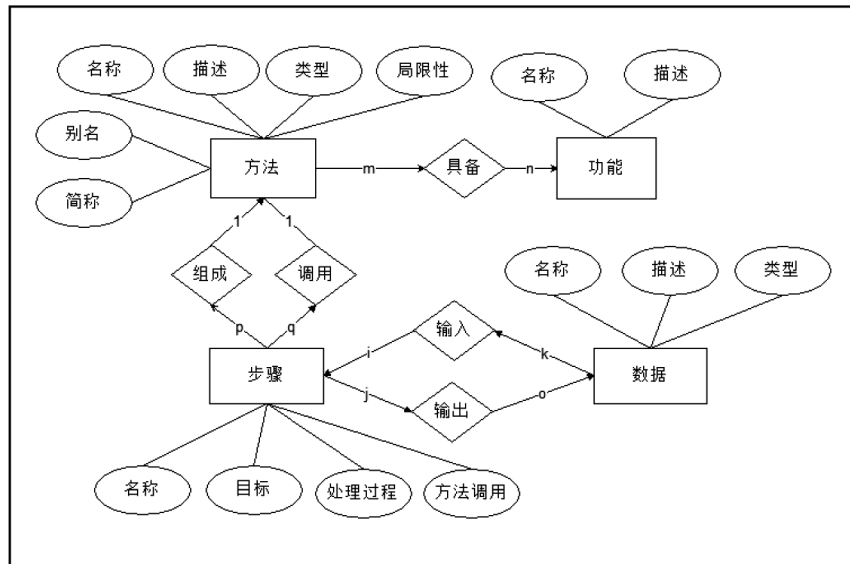


图3 情报方法库 E-R 模型

4.1 方法实体

方法实体是情报方法库中的核心实体，其属性包括对方法的基本描述，包括方法的名称、简称、别名，已经对方法的简述和方法局限性的描述。

4.2 功能实体

方法的功能是方法能够完成的研究目标，在实际研究中，与研究任务直接对应，是查找解决特定任务所需方法的关键检索入口。每项功能都有一个名称，是一个有物理意义的名字，除此之外有一段描述，对功能进行详细的解释。一个方法通常具有一个或多个功能，如数据挖掘中的决策树，既可用于对历史数据的描述，也可用于对未来数据的预测；同时，一个功能也有多个方法可以实现，如完成评估功能既可以使用数据包络分析，又可以使用模糊综合评判，因此，方法和功能之间是多对多的关系。

4.3 步骤实体

步骤是方法的核心构成部分，每个步骤都会有一个物理名称，同时每个步骤都有一个明确的

实施目标，除此之外，每个步骤都对应一个特定的处理过程，这个过程可能是一个简单的信息处理或计算，也有可能是另外一个方法的调用。例如对于评估研究，在指标权重计算所采用的层次分析法中，需要对判断矩阵进行一致性检验，该步骤是一个矩阵计算的过程；同样，在评估指标的确定过程中，可能会用到访谈或问卷等方法，这时相应的步骤应符合访谈或问卷等方法的使用要求。因此，每个步骤都是方法的组成部分，一个方法可能由一个或多个步骤组成，同时，步骤的实施过程中也可能对方法进行调用，并且一个方法可能被多个步骤调用。因此，步骤和方法在组成和调研关系上均是多对一的关系。

4.4 数据实体

情报方法中数据有不同类型，既包含定量数据，也有可能是定性数据，方法实施的一个步骤中可能需要不同类型的多个数据集，每个数据集拥有一个确定的名称。通过方法的实施步骤对输入数据进行加工处理后，会输出一个或多个数据集，可能成为最终结论，也有可能成为其他步骤

的输入。因此步骤和数据在输入和输出关系上均为多对多的关系。

5 情报方法库的结构

根据情报方法库的概念模型，本文设计了情报方法库的结构如图 4 所示，情报方法库共由 7 个主要数据表，其中 MethodDic、Step、DataDic、Fuctions、DataType 分别管理情报方法、方法步骤、方法的输入和输出数据、方法的功能和数据类型，Method2Function 以及 Step2Data 则分别管理方法和功能的关系以及数据和方法步骤的关系。

6 结语

本文在综述情报方法体系研究的基础上，提取了情报方法的属性和相互关系，将多个来源、多种类型的情报方法抽象为一个一般化的情报方法的实施流程，基于此设计了情报方法库数据模型和结构模型。方法库将方法的功能、数据要求、局限性和输出结果等数据进行了规范和统一管理，情报人员根据需解决的问题、数据的类型、特点等已知条件，可以对方法进行有效的选择，避免了方法选择的主观性，同时提高了方法选用的效率，可以成为情报方法管理、情报方法规范化使用以及情报方法创新的基础工具。

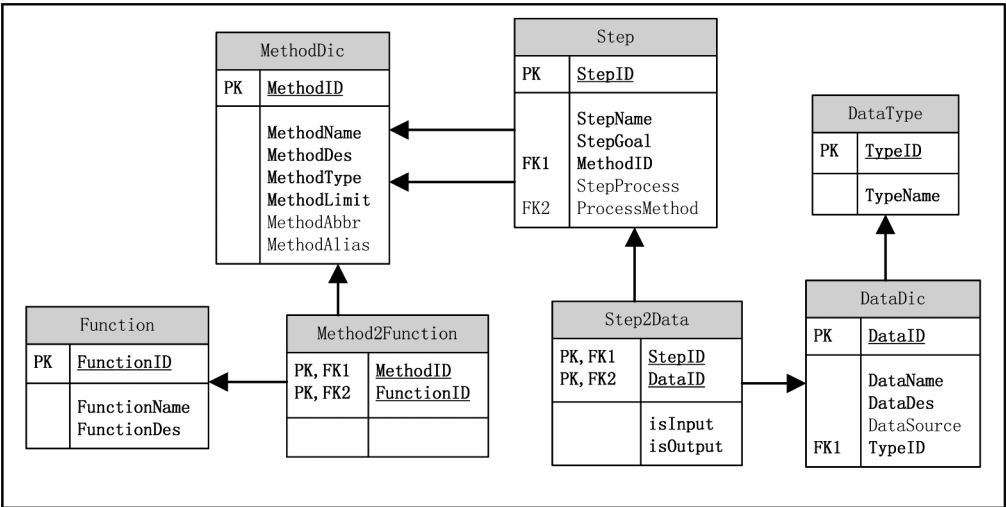


图 4 情报方法库结构

参考文献

[1] 陈晓红. 决策支持系统理论及应用 [M]. 北京: 清华大学出版社, 2000: 72.
[2] 汪作雄. 全生命周期成本估算方法库的设计与实现 [D]. 上海: 东华大学, 2008.
[3] 朱建才. 多目标优化方法库的开发与应用研究 [D]. 西安: 西北工业大学, 2006.
[4] 王华. 社会保险管理学信息系统中方法库的设计与

实现 [D]. 长沙: 湖南大学, 2010.
[5] 王丽珍, 姚绍文. 数据挖掘方法本体研究 [J]. 计算机科学, 2005, 32(3): 197-199.
[6] Vanschoren J, Soldatova L. Exposé: An ontology for data mining experiments [C]. International Workshop on Third Generation Data Mining: Towards Service-oriented Knowledge Discovery, 2010: 31-46.

- [7]Panov P. A modular ontology of data mining [D]. Jo zef Stefan International Postgraduate School, 2012.
- [8]Hilario M, Kalousis A, Nguyen P, Woznica A.A data mining ontology for algorithm selection and meta-mining. International Workshop on Third Generation Data Mining: Towards Service oriented Knowledge Discovery,2009:76-87.
- [9] 彭漪涟, 马钦荣, 冯棉等. 逻辑学大辞典 [M]. 上海: 上海辞书出版社, 2010.
- [10] 王崇德. 情报学研究方法概论 [J]. 情报科学, 1985, 6(6): 1-7.
- [11] 靳娟娟. 情报学方法论研究 [J]. 情报杂志, 1995, 14(3): 27-30.
- [12] 化柏林. 网络海量信息环境下的情报方法体系研究 [J]. 情报理论与实践, 2012, 35(11): 1-5.
- [13] 包昌火. 情报研究方法论 [M]. 北京: 科学技术文献出版社, 1990: 22-27.
- [14] 朱庆华. 信息分析基础、方法及应用 [M]. 北京: 科学出版社, 2004.
- [15] 冷伏海, 冯璐. 情报研究方法发展现状与趋势 [J]. 图书情报工作, 2009, 53(2): 29-33.
- [16] Jarvinen P H. Research Questions Guiding Selection of an Appropriate Research Method [EB/OL]. [2015-01-27] <http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1023&context=ecis2000>.
- [17] 朱庆华, 赵宇翔. 情报学中混合方法研究的理论探索和应用 [J]. 情报学报, 2013, 32(12): 1236-1247.
- [18] 萨师煊, 王珊. 数据库系统概论 [M]. 北京: 高等教育出版社, 2000: 15-21.