

信息分析方法与实践

Information analysis: Method & Practice



杨良斌 / 编著



东北师范大学出版社
NORTHEAST NORMAL UNIVERSITY PRESS

信息分析方法与实践

XINXI FENXI FANGFA YU SHIJIAN



杨良斌 / 编著



东北师范大学出版社
NORTHEAST NORMAL UNIVERSITY PRESS

长春

图书在版编目 (CIP) 数据

信息分析方法与实践 / 杨良斌编著. —长春: 东北师范大学出版社, 2016. 11

ISBN 978 - 7 - 5681 - 2487 - 4

I. ①信… II. ①杨… III. ①信息—分析 IV.

①G202

中国版本图书馆 CIP 数据核字 (2016) 第 273830 号

□策划编辑: 王春彦

□封面设计: 中联学林

□责任编辑: 刘子齐

□内文设计: 中联学林

□责任校对: 赵 滨

□责任印制: 张 允 豪

东北师范大学出版社出版发行

长春市净月开发区金宝街 118 号 (邮政编码: 130117)

销售热线: 0431—84568122

传真: 0431—84568122

网址: <http://www.nenup.com>

电子函件: sdcbcs@mail.jl.cn

北京天正元印务有限公司印装

2017 年 1 月第 1 版 2017 年 1 月第 1 版第 1 次印刷

幅面尺寸: 170mm × 240mm 印张: 20.5 字数: 358 千

定价: 78.00 元

序 言

信息分析是指以社会用户的特定需求为依托,以定性和定量研究方法为手段,通过对社会信息的收集、整理、鉴别、评价、分析、综合等系列化的加工过程,形成新的、增值的信息产品,最终为不同层次的科学决策服务的一项具有科研性质的智能活动。

信息分析是在现代信息分析与咨询活动飞速发展的背景下,于20世纪50年代由情报科学中派生出来的一门新兴学科。近二三十年来,在信息的广泛传播过程中,信息分析得到了迅猛发展,已经有很多优秀的书籍详细地阐述了信息分析的基本原理、步骤、工作程序、划分方法等,如查先进的《信息分析》和《信息分析与预测》,朱庆华的《信息分析基础、方法及应用》,卢小宾、郭亚军的《信息分析理论与实践》,王延飞的《信息分析与决策》等。这些都是优秀的著作,本书很多章节内容都参考了这些著作。但是,本书意在“方法与实践”,主要通过分析方法、进行实证实践来帮助读者理解信息分析,这样就避免了大量篇幅阐述“理论”。

本书把德尔菲法、头脑风暴法、层次分析法、回归分析法、时间序列分析法、科学计量学方法、信息可视化方法、数据挖掘方法、社会网络分析法、竞争情报分析法和大数据分析法等方法都归结于信息分析方法,对每种方法进行了详细论述,以期得到读者的认同。因笔者学术能力和学术成果十分有限,不能为读者贡献足够的科学计量学知识和成果,笔者深感愧疚和遗憾,只能期待在今后的学术科研生涯中,能有更多更好的科研成果回报读者。

特别感谢国际关系学院信息技术创新实践基地(北京市共建项目)提供的出版资金。特别感谢同行者和读者们对本书提出的一系列宝贵意见与建议。特别感谢国际关系学院教务处、科研处和信息科技学院领导及同事对本书出版提供的大量帮助和支持。

此外,感谢选修《社会网络分析》课程的国际关系学院信息科技学院本科生的

帮助和支持。感谢硕士研究生史莹莹对本书的整理。感谢东北师范大学出版社对本书的出版。

限于作者水平和时间有限,疏漏不足之处在所难免,恳请各位专家读者批评指正。

杨良斌

于国际关系学院

2016年6月24日

目 录

CONTENTS

第1章 信息分析方法论 1

1.1 信息分析的定义与特点 1

1.1.1 定义 1

1.1.2 工作流程 2

1.1.3 特点 3

1.1.4 功能 3

1.2 信息分析的方法 4

1.2.1 信息分析方法的研究现状 4

1.2.2 信息分析方法的框架 8

1.3 信息分析的基本方法 12

1.3.1 比较分析法 12

1.3.2 分析与综合法 13

1.3.3 推理法 15

参考文献 22

第2章 德尔菲法 23

2.1 德尔菲法的历史与现状 23

2.1.1 历史 23

2.1.2 现状 25

2.2 德尔菲法的特点与构成要素 27

- 2.2.1 基本特点 27
- 2.2.2 遵守原则 29
- 2.2.3 德尔菲法研究结果的统计要素 29
- 2.2.4 德尔菲法的优缺点 33
- 2.2.5 德尔菲法的现实意义 33
- 2.3 德尔菲法的实施步骤 34
 - 2.3.1 德尔菲法的流程 34
 - 2.3.2 专家知识表示方法 35
 - 2.3.3 专家知识集成方法 37
- 2.4 德尔菲法案例分析 40
 - 2.4.1 新兴产品销售额 40
 - 2.4.2 国美公司部门人员需求 42
 - 2.4.3 手扶拖拉机市场需求变化趋势预测 46
- 参考文献 49

第3章 头脑风暴法 50

- 3.1 头脑风暴法的历史与现状 50
 - 3.1.1 头脑风暴法的提出 50
 - 3.1.2 头脑风暴法的现有研究 51
- 3.2 头脑风暴法的实施流程 55
 - 3.2.1 准备阶段 55
 - 3.2.2 进行阶段 57
 - 3.2.3 整理阶段 60
- 3.3 头脑风暴法案例 61
 - 3.3.1 电线除雪方法 61
 - 3.3.2 新产品命名 62
 - 3.3.3 案例启示 63
 - 3.3.4 头脑风暴注意事项 64
- 3.4 总结与展望 66
 - 参考文献 66

第4章 层次分析法 68**4.1 层次分析法概述 68**

4.1.1 定义 68

4.1.2 应用层次分析法的主要步骤 69

4.1.3 评价指标的选取 71

4.1.4 层次分析法的优点 72

4.1.5 层次分析法的缺点 72

4.2 层次分析法实例与步骤 73

4.2.1 市政工程项目建设决策 73

4.2.2 社会稳定因素排序 80

4.3 扩展与补充 84

4.3.1 模糊层次分析法(FAHP) 85

4.3.2 灰色层次分析法(GAHP) 87

4.3.3 区间层次分析法(IAHP) 91

4.3.4 模糊综合评价法(FCE) 91

参考文献 93**第5章 回归分析法 94****5.1 回归分析法概述 94**

5.1.1 起源 94

5.1.2 简介 95

5.1.3 定义 96

5.1.4 分类 96

5.1.5 回归分析法与相关分析法的联系与区别 96

5.1.6 回归模型 97

5.2 一元线性回归分析 98

5.2.1 数学模型的建立 98

5.2.2 回归方程的显著性检验 99

5.3 多元线性回归分析 101

5.3.1 能够进行多元回归分析的条件 102

5.3.2	使用多元回归分析的一般步骤	102
5.3.3	多元线性回归方程的建立	102
5.3.4	诊断模型	104
5.3.5	假设检验及其评价	104
5.3.6	自变量选择方法	106
5.3.7	多元线性回归分析在现实生活中的应用	107
5.4	非线性回归分析	107
5.4.1	非线性回归模型的直接代换	107
5.4.2	非线性模型的间接代换(对数变换法)	108
5.4.3	不可转换成线性的趋势模型	109
5.4.4	非线性回归应用的几个问题	110
5.4.5	确定非线性模型形式的方法	110
5.4.6	模型的比较	110
5.4.7	实例分析	111
5.5	回归分析法实例	111
5.5.1	一元线性回归分析	111
5.5.2	多元线性回归分析	113
5.5.3	非线性回归分析	113
	参考文献	120

第6章 因子分析法..... 121

6.1	因子分析法概述	121
6.1.1	因子分析法的提出	121
6.1.2	因子分析法的作用	121
6.1.3	因子分析法的目的	123
6.1.4	因子分析法的特点	124
6.1.5	因子分析法与主成分分析法的区别与联系	124
6.2	因子分析法的原理和基本步骤	129
6.2.1	因子分析法的原理	129
6.2.2	因子分析法的基本步骤	133

6.3 因子分析法实例 139

参考文献 141

第7章 时间序列分析法..... 143

7.1 时间序列分析法概述 143

7.2 移动平均法 144

7.2.1 一次移动平均法 144

7.2.2 二次移动平均法 145

7.2.3 加权移动平均法 145

7.2.4 移动平均法相关问题研究 146

7.2.5 移动平均法总结 149

7.2.6 移动平均法应用 149

7.3 指数平滑法 151

7.3.1 一次指数平滑法 151

7.3.2 二次指数平滑法 154

7.3.3 三次指数平滑法 157

7.4 趋势外推法 161

7.4.1 直线趋势延伸法 161

7.4.2 曲线趋势法 162

7.4.3 简单的函数模型 162

7.4.4 趋势外推法的基本假设 164

7.4.5 趋势外推法的主要目标 165

7.4.6 趋势外推法的研究问题 165

7.4.7 趋势外推法的案例分析 165

7.5 时间序列分析法应用实例 169

参考文献 169

第8章 科学计量学..... 171

8.1 科学计量学概述 171

8.1.1 科学计量学的定义 171

- 8.1.2 科学计量学的发展历程 171
- 8.1.3 科学计量学、文献计量学与信息计量学 173
- 8.2 科学计量学的三大定律 174
- 8.3 引文分析法 174
 - 8.3.1 引文分析法的分类 175
 - 8.3.2 引文分析法的内容 175
 - 8.3.3 引文分析法的测度指标 176
 - 8.3.4 引文分析法的局限 177
 - 8.3.5 引文分析法的应用领域 178
- 8.4 共现分析法 179
 - 8.4.1 共现分析法的理论基础 179
 - 8.4.2 共现类型 180
 - 8.4.3 论文共现 180
 - 8.4.4 共词分析法 183
 - 8.4.5 科技论文的主要共现类型 183
- 8.5 科学计量学应用实例 184
 - 8.5.1 应用于科学知识管理中的科学计量学方法和工具 184
 - 8.5.2 利用科学计量学实现科学知识管理的途径:科学知识结构分析 186
 - 8.5.3 作者知识关联网络分析 187
 - 8.5.4 总结 188

第9章 信息可视化方法..... 190

- 9.1 信息可视化概述 190
 - 9.1.1 起源 190
 - 9.1.2 处理阶段及领域 190
 - 9.1.3 研究方向 191
- 9.2 信息可视化的相关技术 191
 - 9.2.1 层次信息可视化 192
 - 9.2.2 文本信息可视化 194
 - 9.2.3 文本内容可视化 194

9.2.4	文本关系可视化	195
9.3	信息可视化工具	196
9.3.1	CiteSpace	196
9.3.2	UCINET	199
9.3.3	VOSviewer	203
9.3.4	Pajek	206
9.4	信息可视化实例	209
	参考文献	211

第10章 数据挖掘方法 212

10.1	数据挖掘简介	212
10.1.1	数据挖掘产生的背景	212
10.1.2	数据挖掘的步骤	212
10.1.3	数据挖掘的基本技术	214
10.1.4	数据挖掘的算法	215
10.1.5	数据挖掘的应用	215
10.1.6	数据挖掘的发展趋势	216
10.2	数据挖掘过程	216
10.2.1	问题定义	217
10.2.2	数据收集与预处理	217
10.2.3	数据挖掘的实施	217
10.2.4	结果解释与评估	217
10.3	聚类分析	218
10.3.1	什么是聚类分析	218
10.3.2	聚类分析方法的特征	218
10.3.3	聚类分析的发展历程	219
10.3.4	系统聚类分析法	219
10.3.5	聚类分析的3种方法	220
10.3.6	系统聚类方法的步骤	221
10.3.7	聚类分析的主要应用	221

10.4	关联分析	222
10.4.1	定义及概述	222
10.4.2	相关算法的解析	222
10.5	文本挖掘	226
10.5.1	文本挖掘的定义	226
10.5.2	文本挖掘的研究现状	226
10.5.3	文本挖掘的主要内容	227
10.5.4	文本挖掘的技术	228
10.5.5	文本挖掘的应用	230
10.6	基于 STCS 的新型信息搜索引擎	231
10.7	Web 挖掘	232
10.7.1	Web 数据挖掘的定义	232
10.7.2	Web 数据挖掘的分类	233
10.7.3	Web 文本挖掘	233
10.7.4	Web 多媒体挖掘	234
10.7.5	Web 使用挖掘	236
10.8	数据挖掘实例	237
	参考文献	239

第 11 章 社会网络分析法 240

11.1	社会网络分析概述	240
11.1.1	社会网络分析的历史	240
11.1.2	社会网络分析的现状	243
11.1.3	对我国社会网络分析研究成果的评价和展望	244
11.2	社会网络分析流程	246
11.3	社会网络分析实例	255
11.3.1	霍桑实验和“扬基城”研究	255
11.3.2	小世界效应	257
11.3.3	Krackhardt 的高科技管理人员	257
11.3.4	Freeman 的电子信息交换系统网络	258

- 11.3.5 国家贸易数据 259
- 11.3.6 Galaskiewicz 的 CEO 和俱乐部网络 263

参考文献 265

第 12 章 竞争情报分析方法 266

- 12.1 竞争情报分析概述 266
 - 12.1.1 竞争情报 266
 - 12.1.2 竞争情报分析 267
- 12.2 SWOT 分析 268
 - 12.2.1 SWOT 简介 268
 - 12.2.2 SWOT 分析方法的特点 269
 - 12.2.3 SWOT 分析方法的步骤 269
 - 12.2.4 SWOT 分析模型 270
 - 12.2.5 分析举例 271
- 12.3 定标比超 272
 - 12.3.1 定标比超简介 272
 - 12.3.2 定标比超的由来 272
 - 12.3.3 定标比超的类型 273
 - 12.3.4 定标比超的方法和步骤 275
 - 12.3.5 定标比超方法的评价 278
- 12.4 价值链分析法 280
 - 12.4.1 价值链分析法的提出 280
 - 12.4.2 价值链分析的基本原理 281
 - 12.4.3 利用价值链分析法开展竞争情报研究的一般步骤 282
- 12.5 波士顿矩阵 285
 - 12.5.1 波士顿矩阵简介 285
 - 12.5.2 波士顿矩阵的特点 287
- 12.6 竞争情报分析实例 288
 - 12.6.1 京东商城简介 288
 - 12.6.2 京东商城竞争情报分析 288

12.6.3 构建京东商城 SWOT 矩阵 290

参考文献 291

第 13 章 大数据分析法 292

13.1 大数据技术简介 292

13.1.1 大数据 292

13.1.2 大数据技术 293

13.1.3 大数据发展趋势 296

13.2 大数据分析流程 297

13.2.1 数据层 298

13.2.2 分析层 299

13.2.3 支撑层 301

13.2.4 结果展示层 301

13.3 自然语言处理 302

13.3.1 定义 302

13.3.2 发展历史 303

13.3.3 研究中遇到的困难 304

13.3.4 深度学习带来的新曙光 305

13.4 大数据分析实例 310

参考文献 312

第 1 章

信息分析方法论

1.1 信息分析的定义与特点

1.1.1 定义

信息分析(Information Analysis)是根据特定问题的需要,对大量相关信息进行深层次的思维加工和分析研究,形成有助于问题解决的新信息的信息劳动过程。其一般以社会用户的特定需求为依托,以定性和定量研究方法为手段,通过对社会信息的收集、整理、鉴别、评价、分析、综合等系列化的加工过程,形成新的、增值的信息产品,是一种深层次或高层次的信息服务,同时也是一项具有研究性质的智能活动。

信息分析最早于二战后从情报科学中派生而出,但却与情报分析有略微不同。通常来说,情报分析是有清晰指向性的,任务明确,情报源相对固定且稳定,主要任务是对已获取的材料进行甄别和判断后,主观判断,客观甄别,辅以其他手段证伪或证实后形成结论,为各种决策提供有力参考;而信息分析和数据分析大同小异,任务性质、研究方法上都日渐趋同,都是在海量且纷繁杂乱、并无明确相关性的材料中,进行已确定,不确定的、无目的或有目的的任务。通过智能、非智能算法工具,对样本进行统计分析,寻找奇异数据,从中发现、归纳、证实某种未知或已知的具有现实意义的征兆、过程、结果。简而言之,情报分析中,主观判断占优势,对分析者的相关经验要求较高,而信息分析则以实证为主,即演绎与归纳的区别。

1.1.2 工作流程

对于信息分析而言,其工作流程是至关重要的。信息分析的流程主要包括规划定向(选题及设计框架)、信息搜集、信息整理与鉴别、信息分析与提炼、产品形成、信息传递几个步骤,接下来将对各个步骤进行具体介绍。

(1)规划定向。规划定向即选择信息课题,明确研究对象、研究目的与研究内容。在明确课题后,对整个分析过程进行预测、规划,完成开题报告、研究框架、工作计划等内容的设计。

(2)信息搜集。根据课题性质选取合适的信息源并从中获取信息。信息源的类型有很多种,通常分为文献信息源与非文献信息源两种。针对文献信息源,通常使用系统检索法、追溯检索法以及浏览检索法等方法;而针对非文献信息源,通常使用问卷调查法、社会调查法、网络调查法等方法。

(3)信息整理与鉴别。信息整理主要是对于已经收集到的信息的形式及内容进行整理,而信息鉴别则是对包括信息的可靠性、新颖性、适用性等信息价值进行鉴别。简言之,整理的过程就是信息的组织过程,目的是使信息从无序变为有序,成为便于利用的形式;而鉴别或者评价的过程是对整理出来的原始信息进行价值评判的过程,目的是筛选出有用的信息,淘汰掉无用或不良的信息。整理与鉴别过程通常交替进行,随着一个过程的深入,另一个过程也进入到更深的层次。其共同作用的结果是使所搜集到的信息成为有序且有用的信息。

(4)信息分析与提炼。信息分析与提炼的结果要与选题的针对性相呼应,应能回答该项研究所要解决的主要问题。一方面要借助思维的分析活动,把研究对象的整体分解成各个能反映整体特征的部分,从中舍弃偶然的、非本质的东西,抽取必然的、本质的东西,并对其分别进行深入细致的考察;另一方面要运用综合的方法,超越时空的限制,将分解出来的无序的、零散的各个部分的本质认识进行重新组合,研究之间的关系,并将蕴含于其中的各种隐含信息和关联关系揭示出来,达到重现整体、推断未知或预测未来的目的。

(5)产品形成。信息分析所形成的产品通常包含消息类产品、数据类产品及研究报告类产品三种,且各具特点。如消息类产品应简洁、新颖、迅速,并具有推荐性质;数据类产品常以手册、要览、年鉴或数据库的形式呈现;研究报告类产品要求做到有述有评,需要同时具有叙述性、综合性、浓缩性和具体性的特点。

(6)信息传递。传递信息具有信源、信宿、信道、信息内容、信息符号及信息保障条件六个要素。传递反馈的方式则通常有传播学模式与情报学模式两种。