

基于 活动情报的 数据融合支持

DATA FUSION
SUPPORT TO
ACTIVITY-BASED
INTELLIGENCE

[美] 理查德·T·安东尼 (Richard T. Antony) 著

朱良 李博 杨文进 等译

朱勇 李传庆 审

 机械工业出版社
CHINA MACHINE PRESS



基于活动情报的 数据融合支持

[美]理查德·T·安东尼 (Richard T. Antony) / 著
朱 良 李 博 杨文进 龙 凯 / 译
朱 勇 李传庆 / 审

机械工业出版社

《基于活动情报的数据融合支持》详细阐述了有关数据融合的理论知识和实践方法,为创建与来源、实体、分类和应用无关的信息融合应用程序,提出了实用、直观且合理的框架,论证了基于文本信息与传统传感器数据的融合,能够提高态势理解的完善性和及时性,适用于条约核查、打击有组织犯罪和保卫国土安全。

本书作为数据融合领域的前沿著作,可为我国从事数据融合、情报制作及相关专业的人士提供有益的参考。

Data Fusion Support to Activity-Based Intelligence, by Richard T. Antony

ISBN: 978-1-60807-845-5

Copyright © 2016 ARTECH HOUSE All rights reserved.

This title is published in China by China Machine Press with license from Artech House, Inc. This edition is authorized for sale in China only, excluding Hong Kong SAR, Macao SAR and Taiwan. Unauthorized export of this edition is a violation of the Copyright Act. Violation of this Law is subject to Civil and Criminal Penalties.

本书由Artech House, Inc授权机械工业出版社在中华人民共和国境内(不包括香港、澳门特别行政区及台湾地区)出版与发行。未经许可之出口,视为违反著作权法,将受法律之制裁。

北京市版权局著作权合同登记 图字: 01-2016-4759号。

图书在版编目(CIP)数据

基于活动情报的数据融合支持/(美)理查德·T·安东尼(Richard T. Antony)著;朱良等译. —北京:机械工业出版社, 2016.11

书名原文: Data Fusion Support to Activity-Based Intelligence

ISBN 978-7-111-55735-7

I. ①基… II. ①理… ②朱… III. ①数据融合—研究
IV. ①TP274

中国版本图书馆CIP数据核字(2016)第311021号

机械工业出版社(北京市百万庄大街22号 邮政编码100037)

策划编辑:孟阳 责任编辑:孟阳

责任校对:黄兴伟 封面设计:马精明

保定市中华美凯印刷有限公司印刷

2017年1月第1版第1次印刷

169mm×239mm·21印张·2插页·336千字

标准书号: ISBN 978-7-111-55735-7

定价: 105.00元

凡购本书,如有缺页、倒页、脱页,由本社发行部调换

电话服务

网络服务

服务咨询热线: 010-88361066 机工官网: www.cmpbook.com

读者购书热线: 010-68326294 机工官博: weibo.com/cmp1952

010-88379203 金书网: www.golden-book.com

封面无防伪标均为盗版

教育服务网: www.cmpedu.com

本书得到“知远战略与防务研究所”的大力支持



“知远战略与防务研究所”是一家独立的战略和防务研究机构，立足于防务动态和学术热点的追踪分析，着眼于长远性、战略性问题的深入研究，并为国家有关机构及决策者提供独立、客观的战略与政策建议。



前言

我于 2005 年首次参与一个社交网络项目时，并不知道能否为跟踪高价值个体并了解其组织形式这一目标做出贡献。作为一个刚刚步入该领域的人，我对社交网络仍一无所知。尽管我已经撰写过一本有关数据融合自动化原理的专著，但事实上，我还不太肯定这本书所涉及的内容在现实中是否可行。这一新项目提出的要求对我而言非常陌生，并且存在无法克服的技术难题。最可能的情况是，相比研发界之前在处理硬目标问题时所采用的方法，对基于文本的信息融合要求采用一种完全不同的分析模式。

我从研究基于文本信息融合的内涵入手。尽管常规技术传感器输出信息与非结构化自然语言之间存在巨大差异，但是否存在一些基本原理，适于在传统多源融合体系结构内利用常规融合方法处理这一“新”的信息来源呢？我在最初几个月里提出了许多疑问，但几乎都没有找到答案。

虽然最初的问题仍有一些未能解决，但已经逐步清晰的是，根据第一性原则的观点，硬传感器和软传感器的融合，确实可以采用类似的方法进行处理。通过早期研究，我将“关联问题”转变为“相似性”搜索，这一概念独立于数据源，因此较为便利。经过一段时间后，我已经确定了 20 个“应用无关”的基本级融合服务，这些服务构成数据源无关融合应用软件的基本积木块。在项目推进的最初三年里，我成功开发出执行上述大多数服务的软件原型，包括一个可根据与高价值个体明确的、推断的和发现的关联，并依据局部情境因素识别候选“关心个体”的工具。

虽然技术界已经投入数十年时间建设“全源”融合系统，但其中很少有能接受基于文本输入信息的系统。在新的世界秩序下，要将这一重要的信息源下放给人工利用，看来是无法接受的。虽然对自由文本的自动化理解仍存在重大困难，但建设可以同时利用基于文本信息源和现有硬传感器输入信息的融合系统却是完全可行的。要将基于文本的信息附加到现有融合软件中，

就必然会提出新的要求，包括有效利用不断提高的自然语言理解能力（例如含义摘要、故事梗概）、语义知识表达式、更高的第 0 级规格化能力以及开发新型支持性融合服务。

本项目以各种方式提出理想化的问题。因为项目经理愿意赋予参与承包的团队较大的自由空间来开发解决方案，所以我相信本项目能够开拓创新并真正推动融合技术的发展进步。如果没有本项目及其管理方法的保证，赖以成书的必要研究和实验就不可能完成。最初，本项目主要研究软传感器输入信息，但此处提出的理论框架也全面包含了传统传感器输入信息。

创作本书的主要目的是，说明基于文本信息与传统传感器数据的融合是一个完全可以实现的目标，它能够提高态势理解的完善性和及时性，尤其适用于反叛乱应用软件、条约核查、打击有组织犯罪和保卫国土安全等现实挑战。

提示：本书呈现的融合软件原型输出信息原本均标出颜色以便用户理解产品，现均以单色表现。对由此可能给读者造成的不便，作者事先致歉。

还需注意的是，颜色是人为添加的，仅用于视觉呈现。在其内部，所有以颜色显示的表达式都储存为矢量。

致 谢

与我第一本有关融合的著作相同，本书也是始于一个零散无序的 Word 文档，我在该文档中随手记录自己的想法和重要问题，并记下一段时间里取得的成果。随着文件所占空间的增长，终于有一天，我发现只需要再做一些工作（实际上还是做了不少工作的）就可以再次推出一本有关数据融合的专著。一开始，这种可能性还很微弱，还是那么遥不可及，但我一直希望退休前能够为自己的专业领域再做一次贡献，正是这一愿望激励着我继续写作。

早年间，我曾与约瑟夫·卡拉科夫斯基（Joseph Karakowski）密切合作。约瑟夫在解决其他问题时采用了我提出的观点。我们两人间频繁而有力的技术探讨，最终帮助我提炼出本书涉及的许多基本概念。在此期间，我和约瑟夫共同撰写了多篇论文，其中有很大一部分都收入本书。

此外，需向为我提供大力支持的雇佣单位 SAIC 表示感谢，并特别感谢两位项目经理。理查德·希勒（Richard Shearer）调任我为首席科学家，保证我能够在资金极为困难的条件下继续开展研发工作。威廉·克劳斯（William Kraus）在内部改组后成为我的上级领导，同样对我的工作抱以信任，并大力鼓励我继续研究和写作。

最后，要向我的妻子表示感谢，她同样给予我重要的支持。出于一些不足为外人道的原因，我设计和构建情境及融合支持系统的灵感多萌发于凌晨两点到四点之间。因此，当我迅速从床上爬起来记下自己最新的“高见”时，就会无数次把妻子吵醒。和大多数作者的感觉一样，即使有所爱的人全力支持，著书并出版仍是一个十分困难的过程。没有家人的坚决支持，这部书稿就不可能完成。

目 录

	前言	IV
	致谢	VI
1	引言和背景	1
	参考文献	13
2	基础理论	15
2.1	第一性原则	15
2.2	行为理解	24
2.3	融合形式	25
2.4	补充性融合服务	32
2.4.1	消息提取服务	32
2.4.2	消息规格化服务	33
2.4.3	位置规格化服务	34
2.4.4	时间规格化服务	35
2.4.5	过滤服务	36
2.4.6	数据库支持	36
2.4.7	人一计算机交互	37
2.4.8	实体支持服务	37
2.4.9	位置支持服务	38
2.4.10	时间支持服务	38

2.4.11	情境支持服务	38
2.5	为指定应用映射融合合成产品	39
2.6	选定问题域融合模型	42
2.7	从合成产品映射功能运行	47
2.7.1	一对多连接	47
2.7.2	多对一连接	47
2.7.3	不连接右侧功能	48
2.7.4	不连接合成产品的右侧功能	49
2.8	映射到不同问题域模型	50
2.8.1	飞机飞行安全域模型	50
2.8.2	机器人车辆控制模型	54
2.9	情境	57
2.9.1	情境作为约束条件	60
2.9.2	情境作为特征	62
2.9.3	约束条件与特征	63
2.10	数据库注意事项	63
2.11	语义信息要求	65
2.11.1	位置和时间	69
2.11.2	副词短语之外的位置描述	73
2.11.3	从非结构化文本进行半自动化事实提取	74
2.11.4	应用示例	77
2.11.5	半自动化句法分析器执行	80
2.12	结合硬、软混合数据	81
	参考文献	82
3	原型执行	83
3.1	位置规格化服务	84
3.1.1	形式方法	88
3.2	空间推理支持服务	92

3.3	时间规格化服务	94
3.4	时间推理支持服务	98
3.5	情境支持服务	102
3.6	第 1 级：灵巧轨迹	105
3.7	第 2 级：实体对实体关系发现	108
3.8	在大型图形数据结构中导航	113
3.9	第 2 代原型	114
3.10	硬—软融合示例	117
4	更高级别推理	121
4.1	可递性	121
4.2	多重假设推理	124
4.3	真值维护	127
4.4	数据挖掘	128
4.5	融合形式的再解读	131
4.5.1	假设生成形式	131
4.5.2	过程控制	133
4.5.3	作为控制功能的融合形式	135
4.6	社交网络空间 / 时间关系可视化	139
	参考文献	141
5	特殊功能	142
5.1	高价值实体发现	142
5.2	关键位置	145
5.3	关心位置发现	147
5.4	报警	151

6	发展方向	153
6.1	初步想法	153
6.2	支持基于活动情报	154
6.3	第 1 级扩展	159
6.4	第 2 级扩展	161
6.5	关系	164
6.6	网络分析	168
6.7	组织发现	170
6.8	更高级别产品抽象	175
6.9	活动利用	176
6.10	角色和行为理解	179
6.11	用户指向性工具开发	181
6.12	事实规格化	182
6.13	事件识别和利用	187
	参考文献	187
7	精炼方面的扩展内容	188
7.1	谱系相关指标	188
7.2	实体解析	190
7.3	精制实体对实体关系产品	193
8	二元融合	196
8.1	概要	196
8.2	相似性量度与实体具体性	198
8.3	深入阐述	200
8.3.1	基于实体融合	200
8.3.2	情境	201

8.4	二元融合类别及其在应用开发中的作用	202
8.4.1	类别 1 融合	205
8.4.2	类别 2 融合	205
8.5	“叙事者”：整合类别 1 和类别 2 融合	209
8.6	位置模式“叙事者”	212
8.7	类别 3 融合	214
8.8	整合融合类别 1~3	217
8.9	小结和探讨	222
	参考文献	224
9	全面集中：执行分析工作流程	225
9.1	情境和融合支持系统工作流程样本	228
9.2	组织理解	239
9.3	行为识别	245
	参考文献	248
10	推理范式	249
10.1	基于主体方法	249
10.2	背景推理范式	250
10.3	基于案例推理	251
10.4	专家系统	252
10.5	神经式网络	252
10.6	贝叶斯网络	253
10.7	一阶逻辑	254
10.8	图形匹配	254
10.9	不确定性管理	255
10.10	多重假设推理与真值维护	255
	参考文献	258

11	空间推理支持	260
11.1	背景	260
11.2	视觉激发空间引擎技术	261
11.3	空间模式匹配	263
11.4	高效率布尔集合运算	266
11.5	空间规格化	268
11.6	人类地理	272
11.6.1	专业技术发现	272
11.6.2	关系发现	273
11.6.3	内容增强	273
11.6.4	人类地理数据的语义描述	274
11.6.5	模糊集合运算	275
11.7	陆地路线计划与推理	278
11.8	语义约束条件构建与情境敏感推理	280
	参考文献	283
12	时间推理支持	284
12.1	时间表达式合成	284
12.2	时间准确性与具体性	286
12.3	相互联系的分布函数与隶属函数	289
12.4	情境敏感性	291
12.5	时间推理支持服务	293
12.6	执行注意事项	295
12.7	情境和融合支持系统执行	296
12.8	相关性指标	298

12.9	时间关系	302
12.10	对时间表述进行句法分析	305
	参考文献	309
13	总结与建议	311
	词汇表	318
	关于作者	320

1

引言和背景

数据融合用途广泛，包括基于活动情报（ABI）（Long, 2013）、场所监控、机器人控制、工厂自动化、装备维修、疾病诊断和环境管理等众多应用方向。这些应用有许多基本要求是相同的，本书重点阐述第一种应用，即通过融合和聚合广泛的信息源以理解实体（包括个体、车辆、组织等）的行为，并最终理解实体的意图。

数十年来，数据融合算法的发展主要集中在跟踪陆地、海下和空中运动对象方面。在理解单个实体及组织的行为方面，必须将对象运动视为特殊的对象活动情况。发送电子邮件、面对面开会、打电话、互联网搜索、社交媒体交流、远程银行操作以及其他大量通用实践与跟踪有形对象运动共同构成对态势理解而言非常重要的活动。

无论应用如何，具有实际意义的决策工作通常要求收集证据以及进行分析和各种形式的评价和评估。考虑到许多应用具有复杂性，从态势感知到天气预报，各方面决策始终需要利用各种相关信息源，应用先前的有关知识并巧妙组合各种证据。因为事实和知识一般都具有一定程度的确定性（通常为未知），所以现实世界的决策工作本质上倾向于非布尔体系。

综合多种信息源以支持态势感知，实际上是所有生物有机体的一种根本生存机制。即使是低等动物也依赖多种感官的融合，才可能找到食物并躲避其他物种的猎捕。高等物种，例如人类，会在通过繁忙路口时融合视觉和听觉信息以避开危险。人类结合嗅觉和味觉就能区别哪些是美味，哪些是毒物。人类利用触觉和视觉小心翼翼地使用微小的物品。

正是因为人类高度依赖多感官融合，开发支持遥感应用的人造融合系统才反映出极其重要的生物过程的自然延伸。虽然大多数技术专家都同意完全有必要为满足既有目标或目标集而提炼浩如烟海、纷繁复杂的信息，但是其中依旧没有解决的一个难题是，我们尚未对融合过程形成一个简明、全面且为业界广泛接受的定义。

我们并不是要提出一个全新的数据融合定义，而是要研究融合过程的根本原则及一般性质。我们的目标是帮助读者更好地体会与开发连续态势感知产品，并支持完善决策有关的根本复杂性事物。正如我们所见，融合极为多样和开放，而融合过程涉及从丰富且各异的数据源中选取、整合、指派、归类、聚合、转换、混合、结合、区分、综合、合并、精制、消除歧义、提炼并抽象化信息，这样一种认识有助于解释为什么难以简明地描述这一过程。

除了要从多样且经常是不成比例的数据中推导出真相外，信息融合还必须保持真实性。数据源的可靠性从确定到不确定，再到有意假信息变化。在具有相当程度的数据源冗余、信息多样化和无体系源偏差的情况下，进行多源融合有可能帮助判断不正确、不准确和不一致，甚至是蓄意欺骗性的信息（例如敌对环境）。另一个问题是，由于真相所处的自然环境总是动态的，真相本身也总是随时间而变化。

总的说来，数据源要么来自于基于物理的远程（硬）传感器，要么来自于基于文本的（软）来源（包括语言、基于文本的电子媒体、书面和口头报告、交流消息）。基于物理的传感器具有多种传感形态，包括雷达、视频、图像、电磁拦截、声学、红外、磁、生物识别和材料探测器材。

上述内容至少存在于两个数据融合算法大分类中，即特征级和实体级。本文并不研究传感器专用特性和传感器数据处理方法，而是重点研究已利用的单个传感器数据的融合。因此，我们将重点严格限定在支持基于对象制作（OBP）的实体级融合上，这也是当前情报制作的一个关键目标。基于对象制作背后的概念是情报应围绕关心对象而组织（Johnston, 2013）。为真正发挥基于对象制作的潜力，就必须开发全面信息共享的方法，打破传统的烟囱体制，从而保证所有可用信息都与单个对象有关。聚焦于关心对象和关心实体，实际上就能够使分析过程专适于具体的问题域和应用。

在运动目标指示器（MTI）数据（雷达、视频）方面，这意味着原始探

测已经转换为实体轨迹。在上空图像方面，我们假设已经（利用人工方法或自动化方法）生成报告，描述对象的性质、对象在场景中的分布及其他相关信息；此类报告可以是结构化形式或自由文本形式。在电子情报方面，已经以一定的准确性确定可能目标的类别、关键识别特征和位置不确定性（uncertainty）（误差椭圆）。在生物识别传感器方面，得到的产品可能是一列具有一定可信度的候选个体。

情境（context）是态势理解中一个被业界广泛认同的关键成分，但它和融合一样难以做出简明定义。情境与暗物质、隐藏变量和隐含约束条件一样，都对理解直接可观察对象至关重要。虽然情境和融合一样很难准确定义，但是其基本概念还是较为明确的。

正如海洋为理解鱼类的行为提供了情境，任何关心域中的实体和活动都同样存在于一个代表其情境的环境（environment）之中。当然，我们现在必须明确“环境”的含义。从我们的目的出发，我们可以说环境倾向于适宜一些行动而不适宜其他一些行动。一般说来，环境可能极其复杂，也可能对数据融合过程产生直白（明显）和隐晦（不明显）的影响。

为说明情境利用的重要性，请参考图 1.1 所示对多个空间接近目标的时间戳雷达探测结果，对困难进行分析。单纯的统计跟踪算法倾向于（根据规



图 1.1 融合多个目标一段时间内的运动轨迹示例