信息资源聚合与数据挖掘

姜灵敏 马于涛 著

XINXI ZIYUAN JUHE YU SHUJU WAJUE

华南理工大学出版社

信息资源聚合与数据挖掘

姜灵敏　马于涛　著

华南理工大学出版社
·广州·

内 容 简 介

本书以网络环境下的资源聚合与优化和数据挖掘为主线，研究在海量数据中提取有效信息、进行高速数据挖掘的方法，为信息管理、查询优化、决策支持、过程控制等提供理论指导和应用工具。本书共分六章，主要从信息融合与数据挖掘、语义 Web 与本体技术、网络信息资源模型、复杂网络环境中信息资源的互操作、网络资源的发现与搜索、Web 数据挖掘六个方面对信息聚集与数据挖掘进行探讨。

本书对从事信息聚合、数据挖掘和信息管理的科技人员具有重要的参考价值，可以用作计算机、信息技术等专业硕士生、本科生的教材。

图书在版编目（CIP）数据

信息资源聚合与数据挖掘/姜灵敏，马于涛著.—广州：华南理工大学出版社，2007.3

ISBN 978-7-5623-2586-4

Ⅰ.信… Ⅱ.①姜…②马… Ⅲ.信息管理：资源管理-数据库系统-研究 Ⅳ.G203-39

中国版本图书馆 CIP 数据核字（2007）第 048124 号

总 发 行：华南理工大学出版社（广州五山华南理工大学 17 号楼，邮编 510640）
营销部电话：020－87113487 87110964 87111048(传真)
E-mail:scutc13@scut.edu.cn **http://www.scutpress.com.cn**
责任编辑：王云昀 黄丹丹
印 刷 者：广州市穗彩彩印厂
开 本：787mm×960mm 1/16 **印张**：12 **字数**：258 千
版 次：2007 年 3 月第 1 版 2007 年 3 月第 1 次印刷
定 价：20.00 元

前　言

随着互联网的高速发展和广泛应用，Internet 上的信息日益丰富，积聚了各种资源，如何发现、捕获和挖掘有效的信息资源一直是信息科学和计算机科学领域的研究热点。由于缺乏整体性和系统化的研究，分布、异构信息的智能聚合问题没有得到有效解决，使得互联网上丰富的信息资源无法有效利用与深度共享，造成了严重的资源浪费，成为制约基于 WWW 应用的不同领域业务发展的瓶颈，进而影响企业的竞争力和发展潜力。以 Web 数据挖掘和信息融合技术为基础，旨在解决异域、异构信息资源的资源共享、语义融合和服务智能化等问题，从而为用户提供准确可信、个性化的信息/知识以及智能应用服务，信息资源聚合研究受到了越来越多的关注。

语义 Web 是下一代互联网的发展方向，其实质就是增强网络资源内容和功能的语义表示，以满足分布式主流计算环境语义互操作的需要，使软件 Agent 对 WWW 上异构、分布的信息进行智能、有效的访问和检索。语义 Web 吸取人工智能、信息论、哲学、逻辑和计算复杂性等学科的研究成果，力图对 Web 上信息的表示和获取方式进行改进，以解决目前使用 Web 时存在的瓶颈。语义 Web 的核心思想是通过增加一些语义信息，使得计算机能参与到自动处理 Web 信息的过程中，并为实现智能化的 Web 应用提供必要的技术基础。最近本体在 Web 上的应用导致了语义 Web 的诞生，是实现语义 Web 的基石和关键所在，在 W3C 的主导下有望解决 Web 信息共享时的语义问题，从而实现世界范围内的知识共享和智能信息集成。

21 世纪，信息资源的管理与服务需要面对的是互联网规模的复杂信息源，这些信息来源于不同的领域，存储在不同的数据源中，针对不同的应用服务要求，用户和依赖的平台及管理系统也各不相同。为了更科学地管理和使用这些信息资源，不但需要能够从多个分布、异构和自治的复杂信息源中聚集信息，而且还要能够保持信息在不同系统中的完整性和一致性，以及不同系统间信息资源的互操作性。单一的信息融合技术已经不能满足新的信息资源利用的需要，特别是对于分布、异构、异域的复杂信息资源的管理与服务，以及不同层次信息资源的聚合，使得信息资源的深度共享与智能化服务、互操作性管理与协同的工作显得十分重要。

如何把数据变成知识，把知识变成决策，把决策变成利润（财富），成为人们持续追求的目标。本书以网络环境下的资源聚合与优化和数据挖掘相结合为主线，研究在海量数据中提取有效信息、进行高速数据挖掘的方法，为信息管理、查询优

化、决策支持、过程控制等提供理论指导和应用工具。本书共分六章，主要从信息融合与数据挖掘、语义 Web 与本体技术、网络信息资源模型、复杂网络环境中信息资源的互操作、网络资源的发现与搜索、Web 数据挖掘六个方面对信息聚集与数据挖掘进行探讨。

武汉大学软件工程国家重点实验室何克清教授，华中科技大学计算机学院李平安教授，武汉科技大学计算机学院陈建勋教授对本书给予了指导和帮助；华中科技大学计算机学院马丽副教授，中国银行业监督管理委员会河南监管局周锋博士，武汉大学软件工程国家重点实验室李兵副教授、彭蓉副教授、梁鹏博士，武汉科技大学计算机学院顾进广副教授、李顺新副教授对本书提出了宝贵意见，在此一并谨致诚挚的感谢。

本书的出版获得了广东外语外贸大学出版基金的资助和华南理工大学出版社的大力支持，在此表示衷心感谢。

由于作者的学识所限，书中可能存在许多不足和疏漏，敬请读者和同行批评指正。

作　者

2007 年 1 月

目　录

第一章　信息融合与数据挖掘

自计算机问世并广泛应用于数据处理以来，人们遇到了两次大的数据危机。第一次危机出现在20世纪60年代，大的物理流伴随着大信息流，传统的文件方式不能适应信息处理的需求，这就是所谓的第一次数据危机。在这次危机中诞生了信息处理大师E. F. Codd，出现了数据库技术。第二次危机出现在20世纪90年代，人类积累的数据量以高于每月15%或每年$(1.15)^{12}=5.3$倍的速度增加，数据海洋不能产生决策意志。于是，出现了这样一个怪圈：为了决策→扩大数据库能力→搜集海量数据→使得决策者难以决策。正如奈斯特所说："We are drowning in information, but starving for knowledge."（人类正被数据淹没，但人类却饥渴于知识）如何把数据变成知识，把知识变成决策，把决策变成利润（财富），成为人们持续追求的目标。

21世纪，信息资源的管理与服务需要面对的是互联网规模的复杂信息源，这些信息来源于不同的领域，存储在不同的数据源中，针对不同的应用服务要求，用户和所依赖的信息平台及管理系统也各不相同。为了更科学地管理和使用这些信息资源，不但需要能够从多个分布、异构和自治的复杂信息源中聚集信息，而且还要能够保持信息在不同系统中的完整性和一致性，以及不同系统间信息资源的互操作性。单一的信息融合技术已经不能满足新的信息资源利用的需要，特别是对于分布、异构、异域的复杂信息资源的管理与服务，以及不同层次信息资源的聚合(aggregation and fusion，聚集与融合)，使得信息资源的深度共享与智能化服务、互操作性管理与协同的工作显得十分重要。

信息聚合作为一个新的概念，目前还没有非常统一的定义。有的学者认为：信息聚合是指从不同的数据源汇集并分析相关信息，解决这些信息在语义方面的异构性，并提供基于数据源之间关系、业务过程的聚合等功能[1]；也有学者认为：信息聚合简单地说，就是异构信息资源的语义集成和协同工作过程[2]。最近的研究表明，在大规模、动态、跨组织的网络环境下，为了实现信息资源的深度共享和有效利用，需要提供一种松散耦合的信息聚合设施，它不仅能够实时地提供一致的信息视图，而且能提供基于工作流过程的信息集成；该设施还要能够适应动态变化的组织关系及业务过程的变化，例如组织的重构、新业务的扩展等[3]。面对上述挑战，迫切需要对信息聚合进行深入、系统化的研究。

本书将信息聚合定义为对互联网上的信息资源进行聚集（挖掘）和融合的过程，旨在解决分布、异构信息资源的资源共享、语义融合和服务智能化等问题，从

而为用户提供准确可信、个性化的信息/知识以及智能应用服务。由此可见，信息聚合的研究主要包括两个方面：一是信息的聚集，即如何收集和挖掘互联网上的信息资源；二是信息的融合，即如何解决异构资源互操作和自主协同问题。如何使信息资源能够互相理解，并根据用户的需求有效、动态、智能地聚集融合各种资源，是信息聚合研究的核心问题。只有信息资源的有序化聚合才能深度共享，才便于可控可管，才利于动态、跨领域的网络信息资源集成，构造个性化的智能应用服务，从而辅助实现知识创新、协同工作、问题解决和决策支持。

1.1 信息融合

1.1.1 信息融合的定义

“信息融合”（Information Fusion）一词起源于1973年美国国防部资助的声呐信号理解课题研究中提出的“数据融合”（Data Fusion）概念，直到1998年这个词才被正式确立为一个学科方向的代名词，但至今尚无统一定义。信息融合首先在军事领域得到了广泛应用，是信息战指挥控制系统的命脉。传统的信息（数据）融合研究主要集中在信息融合涉及的融合算法和技术上，这种狭义的信息融合的前提条件是：传感器数据能够量测同时、目标同域和同维、完全通信（同步、无误码、不破损）。随着互联网的快速发展，除指挥控制系统、精确武器制导、工业过程控制、航空航天、机器人、交通管制等传统领域外，信息融合技术已经广泛应用于远程医疗、遥感监测、生物信息、金融等诸多领域。信息融合概念本身的内涵与外延发生了巨大的改变，由早期对多个传感器数据的综合处理转为向基于互联网和Web的多种信息源数据的集成和分析方向发展。

目前，多传感器信息融合技术获得普遍的关注和广泛的应用，“融合”一词几乎无限制地被众多应用领域所引用。虽然“信息融合”很难给出一个统一的定义，但它是针对一个系统中使用多种传感器（多个和/或多类）这一特定问题而展开的一种信息处理的研究方向，根据国内外研究的成果，“信息融合”比较确切的定义可概括为：充分利用不同时间和空间的多个传感器测得的数据信息，运用现代数学方法和计算机技术，按一定的准则，对这些数据信息进行分析、综合、支配和使用，获得对被测对象的一致解释与描述，进而实现相应的决策与评估的信息处理过程。按照这一定义，多传感器系统是信息融合的硬件基础，多源信息是信息融合的加工对象，协调优化和综合处理是信息融合的核心[4]。

在最早应用的军事领域中，“信息融合”被定义为：一个处理探测、互联、相关、估计以及组织的多源信息和数据的多层次、多方面过程，以便获得准确的状态和身份估计、完整而及时的战场态势和威胁估计。这一定义强调信息融合的核心是

指对来自多个传感器的数据进行多级别、多层次、多方面的处理，从而产生新的有意义的信息（包括较低层次上的状态和身份估计，以及较高层次上的整个战场态势估计），而这种新信息是任何单一传感器所无法获得的[5]。

综合考虑上述两个定义，融合都是将来自多传感器或多源的信息和数据进行综合处理，从而得出更为准确可信的结论。目前，互联网的高速发展给日常生活和工作带来了巨大改变，网络已成为人们获取信息和实时交流的主要工具。在这个信息/知识大爆炸的时代，如何有效获取和利用互联网上的信息资源成为人们关心的热门话题。因此，在本书中“信息融合”被定义为对互联网上的信息资源进行处理的过程，从而获得用户所期望的准确可信的信息/服务。

1.1.2　信息融合技术的产生背景

由于超大规模集成（VLSI）和超高速集成电路（VHSIC）、高精度数控机床、计算机辅助设计和制造以及其他设计和生产的改进，传感器的性能得到了很大的提高，如更高的分辨率、更远距离上更高的探测概率和更快的反应时间等。因此，各种面向复杂应用背景的军事或民用多传感器信息系统也随之大量涌现。例如，海军作战指挥控制系统的信息不只限于舰载传感器收集的实时信息，还包括敌情综合报告、观测事件和战术标图得到的非实时无设备信息，并且信息通常由一群集体参战的战舰、飞机和潜艇共同完成收集和评定[5]。在医疗领域，由于不同的人员使用不同的方法和仪器来诊断，出现了大量的数据冗余和不一致性问题；在工程测量和建筑结构的无损检测领域，对同一检测对象（检测母体）由不同人员测量或采用不同的方法进行测量，都会出现信息表现形式的多样性、信息数量的巨大性、信息关系的复杂性以及要求信息处理的及时性等问题，而这些问题已大大超出了人脑的信息综合能力。因此，越来越迫切地需要新的技术途径对海量的信息进行解释、分析和评估。

数据融合技术，也称为多传感器融合技术，于20世纪70年代应运而生。通过多传感器融合就可以充分利用多传感器的资源，将多个传感器在时间和空间上的互补性或冗余性按照某种算法或准则进行综合，从而增加判断和估计的精确性和可靠性。该技术最早应用于军事领域C3I（指挥Command、控制Control、通信Communication与情报Intelligence）系统中和各种武器平台上以及许多民事领域，这些应用领域主要包括：战场任务和无人驾驶飞机、图像的分析与理解、目标的监测与跟踪、工业过程监控、气象预报、医疗诊断、工程测量及建筑结构的无损检测。20世纪80年代，传感器技术的飞速发展和传感器投资的大量增加，使得军事系统中传感器数量急剧增加；超远程武器的出现和发展，从根本上改变了C3I系统的信息处理方式；军事指挥人员需要更多的信息和数据来判断和决策，更加强调速度和实时性。因此，数据融合的研究工作成为军工生产和高技术开发等多方面关心

的问题。例如，美国国防部早在 1984 年就成立了数据融合专家小组，指导、组织和协调数据融合技术的研究，并在 1988 年将数据融合列入 90 年代重点研究发展的 20 项关键技术之一，且列为最先发展的 A 类。与此同时，数据融合理论逐渐引起世界范围内的广泛关注，成为国际上十分活跃的研究领域之一。

20 世纪 80 年代以来，美国三军总部对应用数据融合的战术和战略监视系统一直给予高度重视，且美国国防部从海湾战争中真实感受到了数据融合技术的巨大潜力。因此，在海湾战争结束后，美国更加重视信息自动综合处理技术的研究，在 C3I 系统中增加了计算机，建立以信息融合为核心的 C4I 系统。20 世纪 90 年代中期，随着互联网技术的快速发展，单纯的数据融合技术已难以满足现代社会的信息处理要求。信息融合技术开始逐步发展成为多方关注的共性关键技术，并出现了许多热门研究方向，不断推动着研究的深入进行。迄今为止，信息融合技术理论上已形成一个全新的方向。发展日趋成熟的信息融合技术以其广阔的时空信息覆盖范围、强大的信息综合和提取能力，成为信息处理领域中强有力的工具。

目前，网络信息空间已经成为信息资源依存和共享服务的主要场所，它产生、积聚了海量、动态、不确定、非规范的各种网络资源（数据、信息、知识）。这些分布、异构、异域的复杂信息资源由于缺乏有效聚合、深度共享与互操作性管理，难以实现语义融合、资源重用和服务共享[6]。同时，随着社会的发展，对网络资源的效益化利用和以用户为中心的人性化、个性化服务提出了更高要求，如电子商务、电子政务、网格计算等的应用需求。但由于缺乏科学有效的系统化理论与方法、处理技术和管理手段，网络信息资源呈急剧膨胀的无序生长态势，难以提供高度协同的共享服务，造成了信息资源的浪费[7]，如当前最好的搜索服务引擎 Google 能搜索到满足用户实际需要的信息不足 20%，网络信息的利用率就更低。

万维网（World Wide Web，简称 WWW 或 Web）是互联网最重要和最广泛的应用之一，利用万维网用户可以浏览互联网上所有的信息资源。但是，万维网存在以下两个明显的不足。

（1）计算机不能理解网页内容的语义。例如，对于网上的一组字符“09.05A”，计算机分不清它代表的是上午时间“九点零五分”还是澳大利亚货币“九点零五元”，因而无法处理。

（2）网上有用信息难以收集和挖掘，即使借助功能强大的搜索引擎，查准率也比较低，且它在帮助用户得到成批相关网页的同时，也夹杂了许多用户不需要的垃圾信息。

存在这些问题的原因在于万维网目前采用的是超文本标记语言（Hyper Text Markup Language，简称 HTML），网页上的内容是设计成专供人类浏览的，并非给计算机理解和处理的，因此无法为用户提供自动处理网上数据或信息的功能。此外，万维网是按“网页的地址”而非“内容的语义”来定位信息资源的，网上所有

信息都是由不同的网站发布的，相同主题的信息分散在全球众多不同的服务器上，同时又缺少有效工具将不同来源的相关信息综合起来，因此形成一个个信息“孤岛”，查找所需的信息就像大海捞针一样困难。这就迫切需要建立相关的信息聚合基础设施（infrastructure），利用互联网把地理位置分布广泛的各种资源（包括计算资源、存储资源、带宽资源、软件资源、数据资源、信息资源、知识资源等）连成一个逻辑整体（如图 1－1 所示），为用户提供一体化信息和应用服务（计算、存储、访问等），最终实现在这个虚拟环境下进行资源共享和协同工作，彻底消除资源“孤岛”，减少信息资源的浪费。

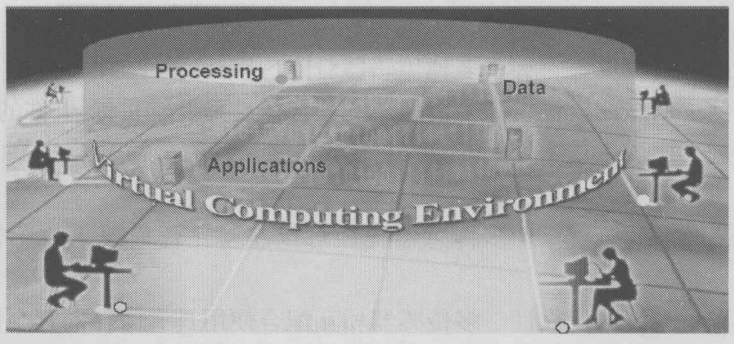

图 1－1　虚拟计算环境

1.1.3　传统信息（数据）融合的基本原理和特点

多传感器信息融合是人类或其他生物系统中普遍存在的一种基本功能。人类通过使用这一功能力把来自人体各个传感器（眼、耳、鼻、四肢）的信息（外物、声音、气味、触觉）组合起来并采用先验知识去统计，从而理解周围环境和正在发生的事件。仿效人脑的这种信息融合能力来处理实际问题，已成为科学和工程学科研究人员的愿望。一般的想法是，在信息处理中从不同的传感器可能获得的那些互为补充的信息经融合以后将得到更精确的观测结果。多传感器信息融合技术的基本原理就像人脑综合处理信息一样，充分利用多个传感器资源，通过对这些传感器及其观测信息的合理支配和使用，把多个传感器在时间和空间上的冗余或互补信息依据某种准则进行组合，以获取被观测对象的一致性解释或描述。数据融合的基本目标是通过数据优化组合导出更多有效信息，它的最终目的是利用多个传感器共同或联合操作的优势，来提高多个传感器系统的有效性[8]。

多传感器数据融合系统与所有单传感器信号处理或低层次的数据处理方式相比，后者是对人脑信息处理的一种低水平模仿，而前者可更大程度地获取被探测目标和环境的信息量。多传感器数据融合与经典信号处理方法之间存在着本质区别：

前者的信息具有更复杂的形式，而且可以在不同的信息层次上出现，包括数据层（像素层）、特征层和决策层（证据层）。因此，多传感器数据融合在解决探测、跟踪和识别问题方面，具有许多优良性能。

（1）增加了系统的生存能力。在若干传感器不能利用或受到干扰，或某个目标/事件不在覆盖范围时，总会有一部分传感器可以提供信息，使系统能够不受干扰连续运行，弱化故障并增加检测概率。

（2）扩展了空间覆盖范围。通过多个交叠覆盖的传感器作用区域，扩大了空间覆盖范围，进而增加了系统的监视能力和检测概率。

（3）扩展了时间覆盖范围。使用多个传感器的协同作用来提高检测概率，某个传感器可以探测其他传感器不能顾及的目标/事件。

（4）提高了可信度。一种或多种传感器对同一目标/事件加以确认。

（5）降低了信息的模糊度。多传感器的联合信息降低了目标/事件的不确定性。

（6）改进了探测性能。对目标/事件的多种测量的有效融合，提高了探测的有效性。

（7）提高了空间分辨率。多传感器孔径可以获得比任何单一传感器更高的分辨率。

（8）改善了系统的可靠性。多传感器相互配合使用具有内在的冗余度。

（9）增加了测量空间维数。系统不易受到敌方行动或自然现象的破坏。

与单传感器系统相比，多传感器系统的复杂性大大增加，由此产生了一些不利因素，如成本提高，设备的尺寸、重量、功耗等物理因素增大。在执行每项具体的任务时，必须将多传感器的优良性能与由此带来的不利因素加以权衡。

1.1.4 数据融合的常用方法

多传感器数据融合的常用方法[9]基本上可概括为随机和人工智能两大类。随机类方法有加权平均法、卡尔曼滤波法、多贝叶斯估计法、Dempster-Shafer（D-S）证据推理、产生式规则等；而人工智能类则有模糊逻辑理论、神经网络、模糊集理论、专家系统等。可以预见，神经网络和人工智能等新概念、新技术在多传感器数据融合中将起到越来越重要的作用。

1．加权平均法

信号级融合最简单、最直观的方法是加权平均法，该方法将一组传感器提供的冗余信息进行加权平均，其结果作为融合值。

2．卡尔曼滤波

卡尔曼滤波主要用于融合低层实时动态的多传感器冗余数据。该方法应用测量模型的统计特性递推地确定融合数据的估计，且在统计意义下是最优的。如果系统可以用一个线性模型描述，且系统与传感器的误差均符合高斯白噪声模型，则卡尔

曼滤波将为融合数据提供唯一的统计意义下的最优估计。滤波器的递推特性使得它特别适合在那些不具备大量数据存储能力的系统中使用。在某些系统中，如果数值不稳定或系统模型为线性的假设不成立，则不能使用传统的卡尔曼滤波器，而采用单位上三角矩阵和对角矩阵协方差量化滤波器或扩展卡尔曼滤波器。使用卡尔曼滤波器对 n 个传感器的测量数据进行融合后，既可以获得系统的当前状态估计，又可以预报系统的未来状态。

3．贝叶斯估计

贝叶斯估计为多传感器融合提供了一种手段，是融合静态环境中多传感器低层信息的常用方法。它使传感器信息依据概率原则进行组合，测量不确定性并以条件概率表示。当传感器组的观测坐标一致时，可以用直接法对传感器测量数据进行融合。但大多数情况下，传感器是从不同的坐标系对同一环境进行描述，这时传感器测量数据要以间接方式采用贝叶斯估计进行数据融合。间接法要解决的问题是，求出与多个传感器读数相一致的旋转矩阵和平移矢量。

在对传感器数据进行融合时，必须确保测量数据代表同一实体，即需要对传感器测量数据进行一致性检验。Mahalanobis 距离在决定数据能否进行融合时非常有用，当两个传感器的测量不一致时，Mahalanobis 距离将变大，通常采用概率距离作为传感器之间的一致性检验。该方法的基本思想是剔除含有误差的传感器信息，而使用“一致的传感器”信息计算融合值。

多贝叶斯估计将每一个传感器作为一个贝叶斯估计，将各个单独物体的关联概率分布合成一个联合的后验的概率分布函数，通过使联合分布函数的似然函数为最小，提供多传感器信息的最终融合值，最后，融合信息与环境的一个先验模型提供关于整个环境的一个特征描述。

4．统计决策理论

与贝叶斯估计不同，统计决策理论中的不确定性为可加噪声，不确定性的适应范围更广。不同传感器观测到的数据必须经过一个鲁棒综合测试以检验它的一致性，经过一致性检验的数据用鲁棒极值决策规则融合。

5．Dempster-Shafer 证据理论

Dempster-Shafer 证据理论是由 Dempster 首先提出、由 Shafer 进一步发展起来的一种不确定推理理论，是贝叶斯估计方法的扩展。贝叶斯估计方法的缺点在于必须给出先验概率，而证据理论则能够处理这种由不明来源引起的不确定性。

Dempster-Shafer 证据推理的基本要点包括：基本概率赋值函数、信任函数和似然函数。该方法的推理结构是自上而下的，分为三级：第一级为目标合成，其作用是把来自独立传感器的观测结果合成为一个总的输出结果；第二级为推断，其作用是获得传感器的观测结果并进行推断，将传感器观测结果扩展成目标报告；第三级为更新，各种传感器一般都存在随机误差，而在时间上充分独立且来自同一传感

器的一组连续报告比任何单一报告可靠，因此，在推理和多传感器合成之前，要先组合（更新）传感器的观测数据。

6．具有置信因子的产生式规则

具有置信因子的产生式规则采用符号表示目标特征和相应的传感器信息之间的联系，与每一个规则相联系的置信因子表示它的不确定性程度。当在同一个逻辑推理过程中的两个或多个规则形成一个联合规则时，就可以产生融合。应用产生式规则进行融合的主要问题是，每个规则的置信因子的定义与系统中其他规则的置信因子相关，如果系统中引入新的传感器，需要加入相应的附加规则。

7．模糊逻辑

模糊逻辑是多值逻辑，通过指定一个 0～1 之间的实数表示真实度，相当于隐含算子的前提，允许将多传感器信息融合过程中的不确定性直接表示在推理过程中。如果采用某种系统化的方法对融合过程中的不确定性进行建模，则可以产生一致性模糊推理。与概率统计方法相比，逻辑推理具有许多优点。由于它对信息的表示和处理更加接近人类的思维方式，在一定程度上克服了概率论所面临的问题，比较适于高层次上的应用（如决策）。但是，这也导致信息的表示和处理缺乏客观性，受主观因素影响大。此外，逻辑推理本身也还不够成熟和系统化。

模糊集合理论对于数据融合的实际价值在于它外延到模糊逻辑，模糊逻辑是一种多值逻辑，隶属度可视为一个数据真值的不精确表示。在多传感器信息融合过程中，存在的不确定性可以直接用模糊逻辑表示，然后使用多值逻辑推理，根据模糊集合理论的各种演算对各种命题进行合并，进而实现数据融合。

8．神经网络

神经网络具有很强的容错性以及自学习、自组织和自适应能力，能够模拟复杂的非线性映射。神经网络的这些特性和强大的非线性处理能力，恰好满足了多传感器数据融合技术处理的要求。在多传感器系统中，各信息源所提供的环境信息都具有一定程度的不确定性，对这些不确定性信息的融合过程实际上是一个不确定性推理过程。神经网络可以根据当前系统所接收到的样本相似性确定分类标准，这种确定方法主要表现在网络的权值分布上，同时可以采用神经网络特定的学习算法来获取知识，得到不确定性推理机制。利用神经网络的信号处理能力和自动推理功能，即实现了多传感器数据融合。神经网络的研究为多传感器集成与融合的建模提供了一种很好的方法。

1.1.5 信息融合的研究方向和应用领域

信息融合不是一门单一的技术，而是一门跨学科的综合理论和方法，并且仍然是一个不很成熟的研究方向，尚处在不断变化和发展的过程中。从目前收集到的国内外资料来看，信息融合的研究方向可归纳为：

(1) 建立数据融合的基础理论。基础理论研究分为两个方面：一是同类信息相融合的数值处理方法，特别是研究各种最优、次优分散式算法；二是不同类型信息相融合的符号处理方法，其理论研究难度较大，目前以专家系统技术、各种人工智能技术为主。但传统的专家系统常常不能满足实际的推理和实时性需要，因此，有必要研究具备学习功能的新人工智能技术（如人工神经网络技术、本体技术等）。

(2) 兼有稳健性和准确性的融合算法和模型的研究。着重研究相关分析、融合处理和系统模拟的算法和模型，强调这些算法和模型的稳健性和准确性，开展对数据融合系统的评估技术和度量标准研究。

(3) 研究数据融合使用的数据库与知识库、高速并行检索和推理机制。

(4) 开发推理系统，尤其是不确定性推理，以进行融合过程中的状态估计和决策分析。

(5) 研究数据融合的分布式数据处理体系结构。

(6) 把处理算法分解成适于在并行机上实现的并行处理。

(7) 将神经网络用于解决探测跟踪、分类和估计等问题。

(8) 数据融合系统的工程化设计方法和系统评估方法。

近 20 多年来，多传感器信息融合技术受到人们的普遍关注，取得了一系列研究成果，并被广泛应用于不同的领域。在大部分融合应用中，基本的系统目标是：探测对象或环境状态，实现对象的识别、分类、跟踪、监控等功能并且确保实时观测。应用领域从广义上可以分为军用和民用两类。

1．军事领域的应用

随着战略、战术、C3I 问题的复杂性增加和工作范围的扩大，人工处理已经不能满足战场决策的需求。这就要求改进数据综合处理的方法，使之能高速、高效处理大量的输入数据，并产生精确的估计。为此，引入了自动化数据融合处理，以便把各种各样的数据合成为一个单一的、连续的战略、战术态势。这样，指挥员才能正确及时地理解我军和敌军情况，才能有效地进行战场管理和决策。

在数据融合中，通常使用多个信息源和传感器收集战术态势信息，这些信息包括事件或目标的检测报告、目标航迹、事实陈述，以及以数据形式表示的不确定性度量。这些数据可用于战术目标和事件的检测、定位和识别。对于不同国籍或型号的目标的识别，可通过任何一个或一组传感变量来实现，这种识别处理能够通过测量目标属性、目标行为或由多个传感器提供的上下文提示来推断目标的性质。

2．电子和信息系统领域的应用

在电子和信息系统学科中，“数据融合”一词中的“数据”具有信息的含义，它包括声、光、电、视像、震动、触摸、力学以及文字类信息在内的各种信息。数据融合系统基于无线网络、有线网络、智能网络、宽带智能综合数字网络，采用光纤数字通信技术、图像传输技术及高吞吐量、并行访问等技术来汇集信息。各种传