

端到端的传感器融合框架调研

四室 -孙霖

目录

1	绪论	2
1.1	数据融合概述	2
1.1.1	数据融合的一般概念与定义	2
1.1.2	数据融合的优势	2
1.1.3	数据融合的应用领域	2
1.2	数据融合系统的模型和结构	3
1.2.1	数据融合的功能模型	3
1.2.2	信息融合的结构模型	3
1.3	常用数据融合技术与方法	4
2	经典数据融合方法	7
2.1	贝叶斯数据融合算法	7
2.1.1	贝叶斯理论基础	7
2.1.2	贝叶斯理论与数据融合	7
2.1.3	传统 Bayes 面临挑战与解决思路	8
2.2	D-S 证据推理方法	8
2.2.1	证据理论基础	8
2.2.2	基于 D-S 证据推理法的融合模型	9
2.2.3	D-S 证据推理法的推广和改进	9
3	现代数据融合方法	10
3.1	基于机器学习的数据融合方法	10
3.1.1	机器学习	10
3.1.2	多模态数据融合	10
3.1.3	讨论	11
3.2	深度多模态学习	11
3.2.1	深度多模态学习与传统多模态学习比较	12
3.2.2	深度多模态学习理论基础	12
3.2.3	总结	14
4	端到端数据融合	15

1 绪论

1.1 数据融合概述

1.1.1 数据融合的一般概念与定义

数据融合又称多源信息融合，是一种多层次、多方面的处理过程，包括对多元数据进行检测、相关、组合和估计，从而提高状态和身份估计的精度，以及对战场态势和威胁的重要程度进行适时完整的评价，是在几个层次上完成对多源信息处理的过程。

1.1.2 数据融合的优势

与单传感器系统相比，多传感器系统主要有以下优势：

- (1) 增强系统的生存能力
- (2) 扩展空间覆盖范围
- (3) 扩展时间覆盖范围
- (4) 提高可信度
- (5) 降低信息的模糊度
- (6) 改进探测性能
- (7) 提高空间分辨率
- (8) 增加了量测空间维度
- (9) 成本低、占空少

同时，制约数据融合向深入发展的因素有

- (1) 信息类型的高度相异性
- (2) 多源信息和多任务引入的固有复杂性
- (3) 目前尚没有数学工具来统一描述和处理此类复杂的问题

1.1.3 数据融合的应用领域

军事应用是多传感器信息融合技术诞生的源泉，主要用于包括军事目标的检测、定位、跟踪和识别，具体应用包括海洋监视、空对空或地对空防御系统等。

近年来，在民用领域中数据融合得到了较快的发展，主要应用领域包括图像融合、工业智能机器人、遥感、刑侦和故障诊断。

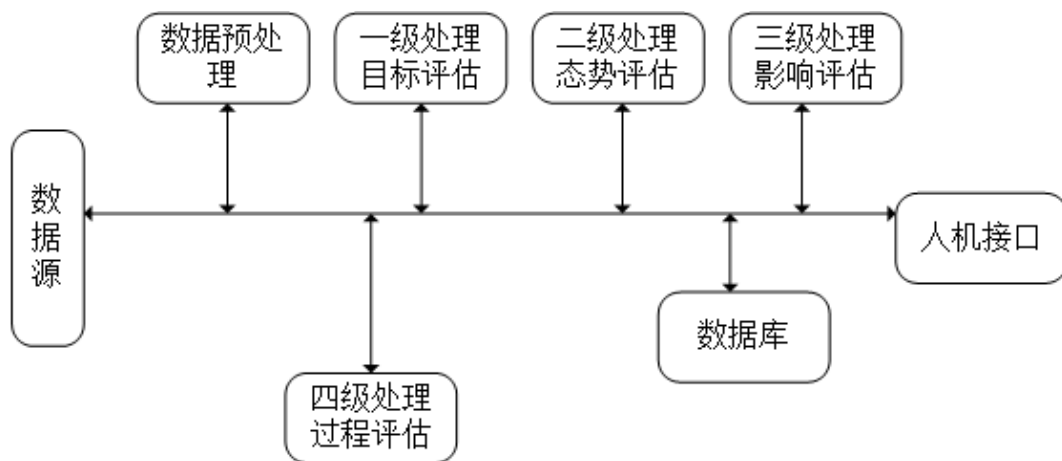


图 1: JDL 数据融合模型

1.2 数据融合系统的模型和结构

1.2.1 数据融合的功能模型

数据融合由美国三军组织——实验室理事综合会（JDL）数据融合组首先提出，模型如图 1 所示，将多传感器数据融合分为五级，分别为数据预处理、目标评估、态势评估、影响评估、过程评估。

第一级处理的是所谓目标评估 (object assessment)，主要功能包括数据配准、数据关联、目标位置和运动学参数估计，以及属性参数估计、身份估计等，其结果为更高级别的融合过程提供辅助决策信息。其中位置估计通常以最优估计技术（线性估计技术、非线性估计技术）为基础，而身份估计一般是以参数匹配技术或者模式识别技术为基础，从比较简单的技术（如多数表决法）到更复杂的统计方法（Bayes 方法、D-S 证据理论）。

第二级处理的是所谓态势评估问题 (situation assessment)，是对整个态势的抽象和评定，其中，态势评定的输入包括事件检测、状态估计以及为态势评定所做的一组假设，态势评定的输出在理论上是所考虑的各种假设的条件概率。

第三级处理的是所谓影响评估问题 (impact assessment)，是一种多层视图处理过程，对参与者设想或预测的行为的影响进行评估。

目前对于二三级的融合处理研究主要集中于基于知识的方法，例如基于规则的黑板模型系统等。

第四级的是所谓过程评估过程 (process assessment)，通过建立优化指标，对整个融合过程进行实时监控与评价，从而实现多传感器自适应信息获取和处理，并最终提高整个实时系统的性能。

1.2.2 信息融合的结构模型

文献 [5] 将数据融合系统的结构模型分为三部分，分别为检测级结构模型、位置融合结构模型以及属性融合结构模型。其中检测级结构模型主要有 4 中，即并行结构、分散式结构、串行结构和树状结构；根据位置融合结构模型，系统结构模型主要有集中式、分布式、混合式和多级式；属性数据融合结构主要有三类：决策层、特征层和数据层融合，下面着重介绍这三种融合结构模型。

按照融合系统中数据抽象的层次，融合可划分为三个级别：数据级融合、特征级融合、决策级融合，各个级别融合处理的结构如图 234 所示。

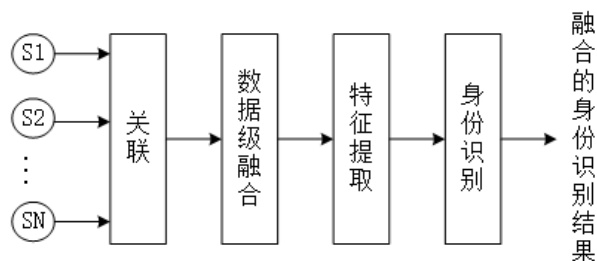


图 2: 数据级融合结构图

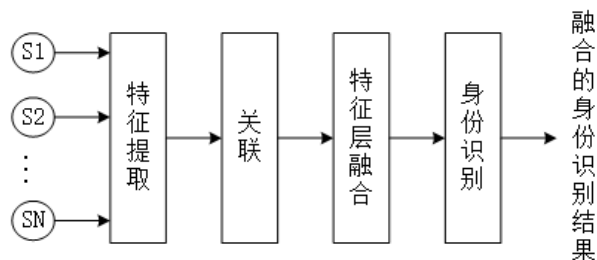


图 3: 特征级融合结构图

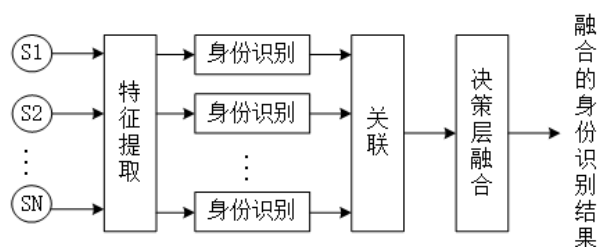


图 4: 决策级融合结构图

表 1对其所属层次、主要特点、方法及应用进行了总结归纳 [1]。

1.3 常用数据融合技术与方法

多传感器数据融合涉及多方面理论和技术，如信号处理、估计理论、不确定性理论、最优化理论、模式识别、神经网络等，图 5对现有常用数据融合方法进行归纳，主要分为经典方法和现代方法两大类。

• 1. 信号处理与估计理论方法

信号处理与估计理论方法包括用于图像增强预处理的小波变换技术、加权平均、最小二乘、Kalman 滤波等线性估计技术，以及扩展 Kalman 滤波 (EFK)、Gauss 滤波 (GSF) 等非线性估计技术等，近年来更多的 UFK 滤波、基于随机采样的粒子滤波和 Markov 链的蒙特卡洛 (Monte Carlo, MCMC) 等非线性估计技术的研究也有了很价值的研究成果。期望最大化 (EM) 算法为求解在具有不完全观测数据情况下的参数估计与融合问题提供了全新的思路；另外，通过建立一定的

表 1: 数据融合结构模型对比

类型	数据级融合	特征级融合	决策级融合
所属层次	最低层次	中间层次	高层次
主要优点	原始信息丰富，并能提供另外两个层次所不能提供的详细信息，精度最高。	实现了对原始数据的压缩，减少了大量干扰数据，易实现实时处理，并具有较高的精度。	所需要的通信量较小，传输带宽要求低，容错能力强，可应用于异质传感器。
主要缺点	需处理的传感器数据量巨大，处理代价高，耗时长，实时性差，原始数据易受噪声污染，仅对产生同类观测的传感器是适用的 [5]。	融合前需对特征进行相关处理，把特征向量分类成有意义的组合。	判决精度低，同时数据处理的代价比较高。
主要方法	HIS 变换、PCA 变换、小波变换和加权平均等。	聚类分析法、贝叶斯估计法、信息熵法、加权平均法、D-S 证据推理法、表决法及神经网络法。	贝叶斯估计法、专家系统、神经网络法、模糊集理论、可靠性理论、逻辑模板法
主要应用	多源图像复合、图像分析和理解	多传感器目标跟踪、融合系统，参数相关和状态向量估计。	为指挥控制和决策提供依据。

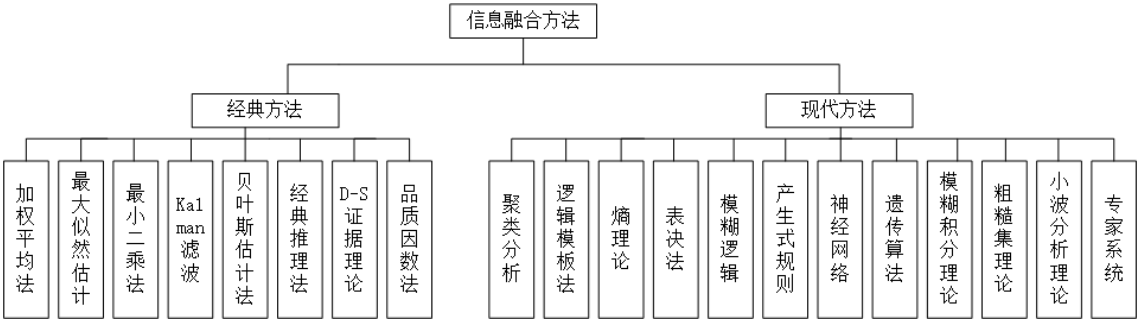


图 5: 常用的数据融合方法

优化指标，可以借助最优化方法来获得参数得最优估计，典型算法有极小化风险法以及极小化能量法等。

- 2. 统计推断方法

统计推断方法包括经典推理、Bayes 推理、证据推理、随机集 (Random set) 理论以及支持向量机理论等。

- 3. 信息论方法

信息论方法运用优化信息度量的手段融合多源数据，典型算法有熵方法、最小描述长度方法 (MDL) 等。

- 4. 决策论方法

决策论方法往往应用于高级别的决策融合。

- 5. 人工智能方法

人工智能方法包括模糊逻辑、神经网络、遗传算法、基于规则的推理以及专家系统、逻辑模板法、品质因数法 (FOM) 等。

- 6. 几何方法

几何方法通过充分探讨环境以及传感器模型的几何属性来达到多传感器融合的目的。

2 经典数据融合方法

信息融合作为对多源信息的综合处理过程, 具有本质的复杂性, 传统的估计理论和识别算法为信息融合技术奠定了不可或缺的理论基础, 同时, 近年来出现的一些新的基于统计推断、人工智能以及信息论的方法, 正成为推动信息融合技术向前发展的重要力量

2.1 贝叶斯数据融合算法

2.1.1 贝叶斯理论基础

经典统计理论不采用先验概率, 将被测参数看作是一个固定值, 没有充分利用其先验信息, 精度和信度不依赖于样本。

贝叶斯网络利用概率计算表示所描述事件的不确定性, 因此也称为贝叶斯条件概率计算, 条件概率叙述为: 在已知事件 B 的情况下, 事件 A 出现的概率为 x , 即 $P(A|B) = x$, 贝叶斯定理为:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \quad (2.1)$$

当融合多源信息的时候, 贝叶斯定理可以进行扩展, 如果用 E 表示多传感器观测矢量, H_i 是假设 H 中的 s 个状态合计, 即 $i = 1, \dots, s$, 则有

$$P(H_i|E) = \frac{P(E|H_i) \cdot P(H_i)}{\sum_{i=1}^s P(E|H_i) \cdot P(H_i)} \quad (2.2)$$

式中, $P(H_i)$ 是先验概率, 即在进行观测之前, 状态 i 的概率, $P(H_i|E)$ 是在已知观测矢量或证据 E 的情况下, 状态 i 的后验概率 [6]。

2.1.2 贝叶斯理论与数据融合

Bayes 点估计的估计误差 $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n)$ 是标量函数 $L(\tilde{x})$ 成为一个损失函数, 设 $\hat{x} = \varphi(z)$, 估计误差 \tilde{x} 的损失函数是 $L(\tilde{x})$, 则风险函数定义为:

$$R(x, \varphi) \triangleq E(L(\tilde{x})|x) = E(L(x - \varphi(z))|x) \quad (2.3)$$

其中 φ 是估计方法, 则 Bayes 风险定义为

$$\mathcal{J}(\varphi) \triangleq E_x[R(x, \varphi)] = E_x\{E_{z|x}[L(x - \varphi(z))|x]\} \quad (2.4)$$

其中 $E_x, E_{z|x}$ 分别表示按分布或条件分布求期望, 而最小 Bayes 风险估计定义为

$$\hat{x}^* = \varphi^*(z), \quad \mathcal{J}(\varphi^*) = \min \mathcal{J}(\varphi) \quad (2.5)$$

假设一个状态空间, 贝叶斯估计器提供了一种计算后验概率分布的方法, k 时刻的概率为 x_k , 已知 k 组测量 $Z_k = z_1, \dots, z_k$ 和先验分布如下:

$$p(x_k|Z^k) = \frac{p(z_k|x_k)p(x_k|Z^{k-1})}{p(Z^k|Z^{k-1})} \quad (2.6)$$

其中, $p(z_k|x_k)$ 为似然函数——基于给定的传感器测量模型, $p(x_k|Z^{k-1})$ 为先验分布函数, $p(Z^k|Z^{k-1})$ 只是作为一种规格化术语, 保证概率密度函数归一化 [3]。

基于 Bayes 统计的目标识别融合一般步骤包括：1) 获得每个传感器单元输出的目标身份说明。2) 计算每个传感器单元对不同目标的身份的不确定性。3) 计算目标身份的融合概率，最后进行 4) 目标识别决策

基于 Bayes 统计的目标识别融合模型如图 6 所示。

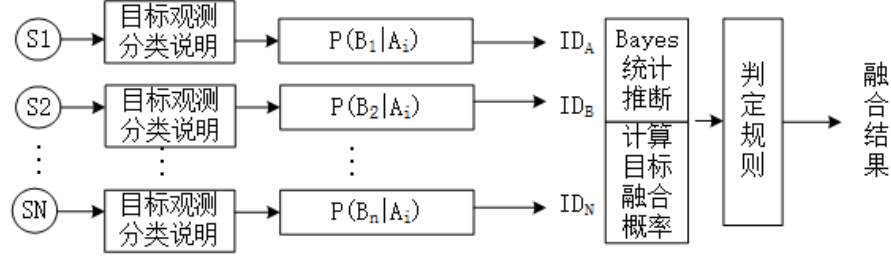


图 6: 基于 Bayes 统计的目标识别融合模型

2.1.3 传统 Bayes 面临挑战与解决思路

在现代信号处理、图像处理、计算机视觉以及自动控制领域存在着大量需要处理的非线性非 Gauss 随即系统，系统的概率分布不再具有 Gauss 特征，即使知道其解析解，但其统计量仍面临着高维积分问题，不便于计算。因此人们提出了各种数值近似算法，如扩展卡尔曼滤波（EFK）、Gauss 滤波（GSF）等次优滤波方法，也有人试图使用多模型混合估计的方法对一类非线性非 Gauss 问题进行求解，但这些算法的稳定性、精度和计算量都难以满足实际需要。总的来说，在非线性非 Gauss 条件下，这些基于模型线性化和 Gauss 假设的方法在系统状态和方差的误差仍然较大，并有可能引起发散。

2.2 D-S 证据推理方法

2.2.1 证据理论基础

证据理论是由 Dempster 和 Shafer 于 20 世纪 60 年代初建立的一套数学理论，是对概率论的进一步扩充，适合于专家系统、人工智能、模式识别和系统决策等领域的实际问题。

证据在本质上就是基于观测对不同假设赋予权值的一种方法，解释如下：

(1) 能够处理任意数量的假设

(2) 能够把证据的权值解释为一种函数，而这个函数把假设的先验概率空间映射到基于观测的假设后验概率空间。

(Dempster-Shafer 合成公式) 设 m_1, m_2 是 \mathcal{H} 上的两个 mass 函数，则

$$m(\emptyset) = 0$$

$$m(A) = \frac{1}{N} \sum_{E \cap F = A} m_1(E) m_2(F), \quad A \neq \emptyset \quad (2.7)$$

是 mass 函数，其中

$$N = \sum_{E \cap F \neq \emptyset} m_1(E) m_2(F) > 0 \quad (2.8)$$

为归一化系数。

设 m 是假设空间 \mathcal{H} 上的 mass 函数, $\mathcal{P}(\mathcal{H})$ 表示 \mathcal{H} 上的所有子集构成的幂集, P 是 \mathcal{H} 上的概率分布, 则有 $v: \mathcal{P}(\mathcal{H}) \rightarrow [0, 1]$, 满足 $v(\emptyset) = 0$, 且

$$v(A) = \frac{m(A) \cdot P(A)}{\sum_{\emptyset \neq B \subseteq \mathcal{H}} m(B) \cdot P(B)} \quad (2.9)$$

以及 $\gamma: \mathcal{P}(\mathcal{H}) \rightarrow [0, 1]$ 满足 $\gamma(\emptyset) = 0$; 且

$$\gamma(A) = \frac{m(A)/P(A)}{\sum_{\emptyset \neq B \subseteq \mathcal{H}} m(B)/P(B)} \quad (2.10)$$

仍是 \mathcal{H} 上的 mass 函数。

2.2.2 基于 D-S 证据推理法的融合模型

在数据融合的应用中, 所有的融合方法都必须面临着处理各种不确定信息的问题, 而 D-S 证据推理方法为不确定信息的表达和合成提供了自然而强有力的理论基础, 这使得它在数据融合等领域获得了广泛的应用。文献 [2] 介绍了 D-S 证据推理方法在数据融合中的应用。D-S 证据推理法以 Dempster-Shafer 证据理论为基础, 其核心是 Dempster 合成规则, 它的主要特点是: 满足比贝叶斯概率理论更弱的条件, 具有直接表达“不确定”“不知道”的能力, 下面对 D-S 证据推理法进行主要论述。

基于 D-S 证据推理法的融合模型如图 7 所示, 首先对来自多个传感器和信息源的数据和信息 (即证据) 进行预处理, 然后计算各个证据的基本概率分配函数、可信度和似然度, 再根据 Dempster 合成规则计算所有证据联合作用下的基本概率分配函数、可信度和似然度, 最后按照一定的判决规则选择可信度和似然度最大的假设作为融合结果。

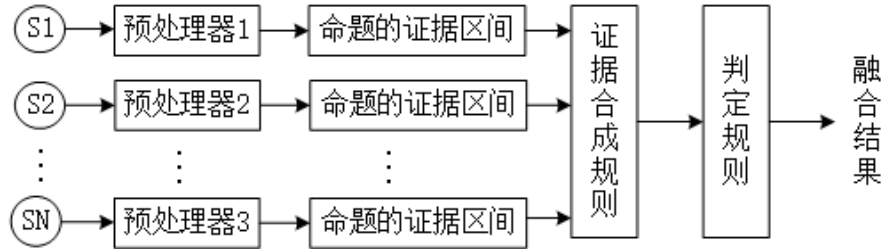


图 7: 基于 D-S 证据推理法的多传感器数据融合模型

2.2.3 D-S 证据推理法的推广和改进

妨碍 D-S 证据推理法在数据融合领域广泛应用的最主要原因是该方法导致的巨大计算量的问题, D-S 的快速计算有两种策略, 第一种策略是针对特殊的证据组织结构, 构造相应的快速算法; 第二种策略是减少焦元个数, 采用近似计算方法。

D-S 方法的扩展主要包括: 如何使用 D-S 证据推理法处理不确定性数据; 如何用证据理论表示规则强度; 如何用推广证据理论等。由于使用 D-S 证据推理法存在以下三个方面的原因, 使得它实际上并不能取得想象中的结果: 要求证据必须是独立的; 证据合成规则没有非常坚实的理论支持; 计算上存在着潜在的组合爆炸问题。从当前的 D-S 证据推理法的发展趋势来看, 将相关理论如模糊集理论、随机集理论、粗糙集理论等与之结合是一条重要途径。

3 现代数据融合方法

3.1 基于机器学习的数据融合方法

由于传感器和微机电技术的发展,传感器采集到的数据量级也越来越大,这也导致了融合这些数据的时候遇到了问题,显然的,传统的数据处理技术难以满足大数据背景下对数据所存在的冗余性、不完整性的要求,文献 [9] 对近期基于机器学习和数据处理层面的技术进行了调查和总结。

文献 [7] 中将多模态机器学习主要可以分为多模态表示学习 (Multimodal Representation), 模态转化 (Translation), 对齐 (Alignment), 多模态数据融合 (Multimodal Fusion) 和协同学习 (Co-learning)。下面主要在多模态数据融合方向进行调研。

多模态数据融合主要有三方面的优势: 第一, 观察同样的对象可以使预测结果更加可靠; 第二, 可以让我们捕捉到相对于单个模态互补的信息; 第三, 当一个模态数据丢失或损坏时, 融合依然可以进行。多模态数据融合有非常广泛的应用, 如视听识别 (AVSR, audio-visual speech recognition)、多模态情绪识别、医学图像分析、多媒体事件检测等。

3.1.1 机器学习

机器学习是目前最火热的课题之一, 属于人工智能和计算级科学的子领域, 它使得计算级系统可以像人类一样看、了解、学习和预测这个世界。

数据在机器学习中扮演了很重要的角色, 数据可以作为采样、训练集、实例和测试集等角色, 基于给定的数据集, 机器学习主要分为监督学习、无监督学习和半监督学习。监督学习主要包括支持向量机 (Support Vector Machine, SVM) 和人工神经网络 (Artificial Neural Network, ANN), 无监督学习主要包括连通性模型、层次聚类、K-means 算法、EM 算法。

3.1.2 多模态数据融合

多模态数据融合是多模态机器学习最原始的目标之一, 之前的调研将多模态数据融合分为早期、后期和混合融合方法, 技术上来说, 多模态数据融合是从多模态的输入中预测输出的方法。在此我们将多模态数据融合方法分为两个主要的策略, 即模型无关方法和模型有关方法, 模型不可知方法是不直接确定一个具体的机器学习方法, 模型有关方法却明确提出了他们的融合结构, 如多核学习 (Multiple kernel learning, MKL)、图形模型 (Graphical models) 和神经网络 (Neural Networks)。

1) 模型无关方法

不可知模型方法可被分为早期、后期和混合融合, 早期融合是在特征提取之后立即进行融合, 这就表示它可以在低等级的特征就进行修正和集成, 同时它只需要训练一个模型, 这使得相比于其他方法容易训练。相反的, 后期融合是在各模态做出决策后在进行融合, 使用单峰决策值并且使用如平均、投票决策、基于权重的通道噪声和信号方差来对决策进行融合, 在不同的模态中可以使用不同的方法, 而且当一个或多个模态信息丢失得情况下后期融合依然可以进行训练和实行。混合融合试图利用上述两种方法的优点应用于同一个结构框架。

模型无关方法的一个优点是, 它们可以使用几乎任何单峰分类器或回归器来实现。

2) 模型有关方法

模型无关方法使用的是单峰机器学习算法, 他们最终使用的技术不是为多模态数据准备的, 下面三个方法是多模态数据准备的技术, 他们的对比如表 2 所示。

表 2: 模型有关方法技术对比

技术	多核学习	图形模型	神经网络
介绍	基于核的支持向量机的扩展，它允许不同的模态的数据使用不同的核来进行学习	可分为两大类：生成-建模联合概率；或判别-建模条件概率	利用神经网络的隐含层来实现信息融合
优点	损失函数是凸函数，允许模型的训练使用标准优化包和全局最优解，可同时被用于分类和回归	容易的利用数据的时间和空间结构，同时支持专家知识和可解释的模型	可以从大容量的数据中进行学习，支持端到端的方式，并且效果显著
缺点	在测试期间依赖于训练数据（支持向量），导致推理速度慢和占用大量内存		缺少可解释性，需要大量的训练数据

多核学习方法是基于核的支持向量机的扩展，它允许不同的模态的数据使用不同的核来进行学习，因为核可以视为数据点和模态之间的功能函数，所以 MKL 对于异构数据有更好的融合效果。MKL 的优点是损失函数是凸函数，这允许模型的训练使用标准优化包和全局最优解。

文献 [8] 中在融合图像信号和 LiDAR 点云信号的过程中，对于图像信号使用预训练的 ResNet-18 网络，而对 LiDAR 信号使用的是定制剩余网络 (customized residual network)，并且分别使用了两种端到端学习的特征级融合架构，第一种是将图像融合到激光雷达中，利用图像特征信息的丰富性来增强 BEV 特征，另一种融合架构是通过将三维检测投影到图像和 BEV 特征图上，得到轴对齐的图像 ROI 和定向的 BEV ROI。

3.1.3 讨论

多模态数据融合已经成为被广泛讨论的课题，也拥有了大量的方法如模型不可知方法、图形模型、多核学习和各种神经网络，这些方法各有优缺点，如一些可以应用于数据量较小的场景，另外的在噪声情况下有不错的效果。

但是目前多模态数据融合依然面对一些挑战，如信号为非时序信号或非一维信号，很难建立互补信息的利用模型，以及不用的模态可能在同一时间的表现或者噪声不同。

文献 [12] 中将数据融合应用于物体追踪算法，并使用端到端训练方法和三级融合架构，经比较发现，将融合放在特征提取器后对各个模态的特征进行融合并喂给训练器进行训练得到的效果最佳。

3.2 深度多模态学习

基于深度学习的数据融合又称深度多模态学习 (Deep Multimodal Learning)，是指利用神经网络技术通过训练、学习来生成和优化最适合的数据融合结构的过程。

文献 [10] 对于近期深度多模态学习的发展及研究做了总结与调查，将深度多模态学习分为三类，即前期融合 (Early Fusion)、后期融合 (Late Fusion) 和中间融合 (Intermediate Fusion)，同时也讨论了将

多模态融合的表达层和深度学习架构相结合的方法，并针对多模态识别和融合架构的生成与优化两个方面进行重点论述。

与神经网络相结合的深度多模态学习具有自学习的特性，数据融合阶段中的每个模态自学习并生成合适的分层表示（hierarchical representation），这是深度多模态学习相对于经典数据融合算法的优势，该技术可以应用于医疗、图像识别、LIDAR(Light detection and ranging) 等众多领域，同时在人体动作识别方面和自动系统（如自动驾驶）等领域作用显著。

3.2.1 深度多模态学习与传统多模态学习比较

传统上来说，数据融合属于前期融合，融合的任务和工作大多为去除模态之间的相关性或在低维子空间中表示融合数据，主要包括主成分分析（PCA）、独立成分分析（independent components analysis）和典型相关分析（canonical correlation analysis）。在二十一世纪后，随着模式识别的兴起，研究者开始将其应用于数据融合并称之为后期融合，这种方式在典型的不同维度的数据源和采样率中效果显著。

深度多模态学习与传统多模态学习的对比如表 3所示。

表 3: 深度多模态学习与传统多模态学习的对比

深度多模态学习	传统数据融合
结构表示和融合表示都由数据学习而得	结构由手动设计且需要对处理问题的先验知识支持
不需要或很少数据的预处理（端到端的训练）	一些技术依靠与数据的预处理
结构可实现隐式降维	特征选择和降维通常是显式的
支持早期、后期和中期融合	通常是早期和后期融合
容易扩展数据种类和大小以及模态的数量	早期融合难以扩展，后期融合的规则需要进行定义
训练阶段可以学习结构模型	结构模型通常被人工确定
深层、复杂的网络需要大量的训练数据	不需要很多数据
大量的超参数调节十分重要	用到的超参数较少

3.2.2 深度多模态学习理论基础

深度多模态学习需要考虑的是以下三个问题：1）何时对不同模态的数据进行融合。2）需要合成是那些模态的数据。3）如何处理丢失或损坏数据和模态。

1) 模型

应用深度多模态学习于一个新问题同时需要包括结构和学习算法两部分，我们称之为模型，根据模型的学习范数将模型分为区分型 (Discriminative models)、生成型 (Generative models) 和混合型 (Hybrid models)。生成型模型支持新数据的加入或者是经过预处理的生成，通常将数据 X 和标签 Y 结合并学习联合概率 $P(X,Y)$ ；相反的，区分型模型则是应用于初级的预测任务，这些模型学习的是 X 和 Y 的条件概率 $P(Y|X)$ 。相比之下，生成型模型在丢失数据的情况下显得更加灵活。三种模型的总结见下表 4。

表 4: 区分型、生成型和混合型模型对比

类型	区分型	生成型	混合型
构建	深层架构直接建模从输入到输出的映射	描述观测或可见数据的高阶相关特性	使用生成型的结果做区分
特点	最小化一般损失函数, 这些模型占到了所提出的多模态学习的大部分	描述可见数据及其相关类的联合统计分布, 最大化各类之间的距离	借助生成型表达输入的多模态数据和区分型完成分类和识别两部分
应用	分类和识别	用于模式分析或合成, 聚类	分类和识别

在文献中将混合型模型根据方法分为联合法 (joint methods)、迭代法 (iterative method) 以及阶段法 (staged method), 我们根据初级学习范数总结了模型的分类, 这样的结构可以有利的解决深度学习的问题, 即整体结构设计的难度。

2) 融合架构

深层结构提供了在早期、中期和晚期融合的灵活性, 在深度学习出现之前数据融合结构通常只是早期融合, 也称之为特征层融合, 随着深度学习的出现, 特征层融合也扩展到了中期融合。

早期融合集成分散的多源数据为一个特征向量输出, 之后将特征向量作为机器学习算法的输入, 被融合的数据为从传感器得到的原始数据, 即传统的数据融合; **后期融合**是指经不同的离散的深度学习模型分类器决策后汇总, 因为分类器的误差往往是不相关的而且该方法独立于特征的, 因此这种融合结构常常会受到青睐。融合规则通常是最大化融合、平均融合或者基于贝叶斯准则的融合, 甚至是使用一个分类器; **在中间层中**, 神经网络通过图层管道映射输入将原始数据转化为更高级别的特征表达作为输出, 之后融合不同的表达层到一个隐含层后学习一个联合多模态表达便是一个可控的操作, 该共享表达层可以处理单个数据或者同时处理来自多个模态的数据。三个融合框架如图 8所示。

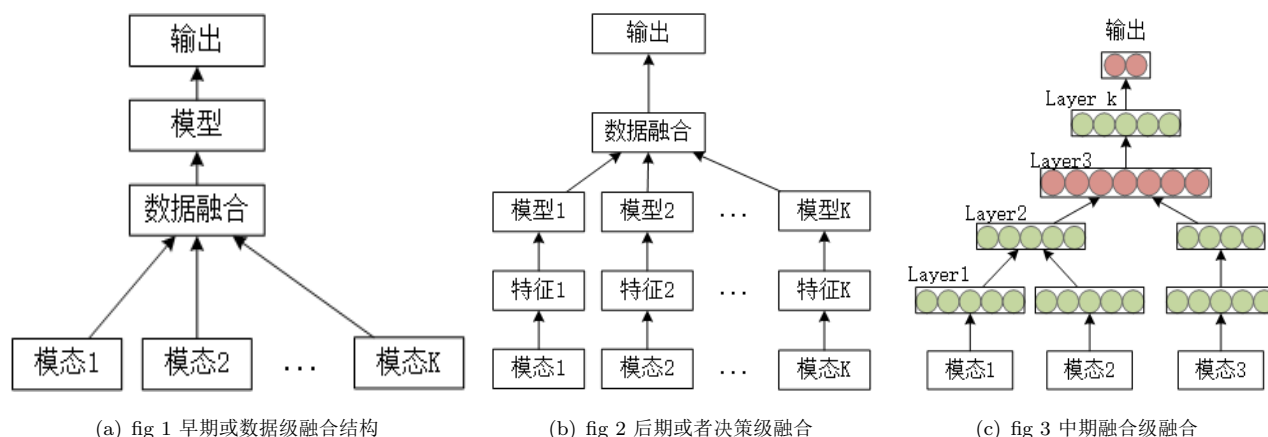


图 8: 三级现代融合架构

文献 [11] 对于自动驾驶方向的雷达和图像两个模态的数据进行了前期、后期、感知层和端到端融合并作出对比, 其中端到端的后期融合的均方误差和准确率最高。

3) 融合架构的学习和优化

大多数多模态深度学习框架是人工制定的, 哪些模态需要被融合、模型的深度为多少通常是基于直

觉，当输入的模态大于等于两个的时候，神经元的最优数量和层数将会具有挑战性，一个自然的过程就是通过将其转化为一个模型搜索或结构学习问题来寻找一个最优的多模态融合体系结构。一个结构过大可能会使得结果很好或者过拟合。

基于遗传算法的结构最优化选择是较早开始进行的神经网络架构生成工作，但是当数据量和网络节点数量大的时候，迭代的时间成本和计算成本就变得很昂贵，同时，贝叶斯体系结构优化是一个离散优化问题，通过使用基于高斯过程的贝叶斯优化搜索所有可能的多模态融合架构，之后使用统计方法来比较搜索空间中的各个结构的质量。

虽然特征设计工程在很大程度上是通过深度表示学习来解决的，但下一个合乎逻辑的步骤将是取消对深层体系结构的精心设计，并寻求能够自动实现这一点的技术。

3.2.3 总结

本小结我们讨论了使用神经网络或者深度学习来对数据融合架构进行学习和生成，这是最能体现神经网络工作改善数据融合的特点，对于一个数据融合方法来说，深度多模态学习的技术可以分为前期融合技术、后期融合技术和中期融合技术，这不仅可以使得数据融合可以更灵活地表示，更可以允许多模态融合的结构深度是可变的。

4 端到端数据融合

数据融合对于整体学习过程可以分为端到端学习和感知层学习，其中，感知层学习 (perception learning) 的任务是生成环境的表示，端到端学习 (end-to-end learning) 目的在于使用一步解释环境并且生成控制动作（输出）。神经网络提供了一个更普遍的框架用于对条件描述进行直接建模。

参考文献

- [1] 黄漫国, 樊尚春, 郑德智, and 邢维巍. 多传感器数据融合技术研究进展. *传感器与微系统*, 29(03):5–8+12, 2010.
- [2] 徐从富, 耿卫东, and 潘云鹤. 面向数据融合的 ds 方法综述. *电子学报*, (03):393–396, 2001.
- [3] 张品, 董为浩, and 高大冬. 一种优化的贝叶斯估计多传感器数据融合方法. *传感技术学报*, 27(05):643–648, 2014.
- [4] 韩崇昭 and 朱洪艳段战胜. 多源信息融合 (第二版). 清华大学出版社, 2010.
- [5] 何友, 彭应宁, 陆大. 多传感器数据融合模型综述. *清华大学学报 (自然科学版)*, (09):14–20, 1996.
- [6] 刘严岩. 多传感器数据融合中几个关键技术的研究. PhD thesis, 中国科学技术大学, 2006.
- [7] T. Baltrušaitis, C. Ahuja, and L. Morency. Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2):423–443, 2019.
- [8] M. Liang, B. Yang, Y. Chen, R. Hu, and R. Urtasun. Multi-task multi-sensor fusion for 3d object detection. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7337–7345, 2019.
- [9] Tong Meng, Xuyang Jing, Zheng Yan, and Witold Pedrycz. A survey on machine learning for data fusion. *Information Fusion*, 57:115 – 129, 2020.
- [10] D. Ramachandram and G. W. Taylor. Deep multimodal learning: A survey on recent advances and trends. *IEEE Signal Processing Magazine*, 34(6):96–108, 2017.
- [11] N. Warakagoda, J. Dirdal, and E. Faxvaag. Fusion of lidar and camera images in end-to-end deep learning for steering an off-road unmanned ground vehicle. In *2019 22th International Conference on Information Fusion (FUSION)*, pages 1–8, 2019.
- [12] L. Zhang, M. Danelljan, A. Gonzalez-Garcia, J. van de Weijer, and F. Shahbaz Khan. Multi-modal fusion for end-to-end rgb-t tracking. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pages 2252–2261, 2019.