

第二篇 感知模态

第4章 多模态感知融合

《智能感知技术导论》



知识点9

多模态数据融合

多模态数据

- 同一事物通常能够被不同模式的数据进行描述，每个模式都可被认为是一种模态，而多种模态的数据的组合称之为多模态数据。

描述雪景的多媒体数据

图像



文本

It snowed in the evening. Flakes of snow were drifting down. If you walked in the snow, you can hear a creaking sound.

音频

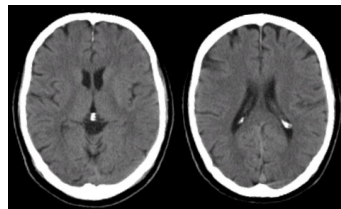


不同来源的医学影像数据

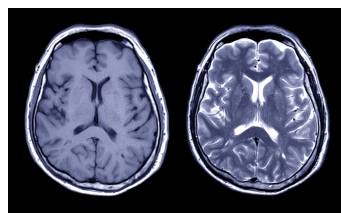
B超



CT

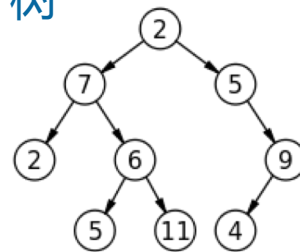


MRI

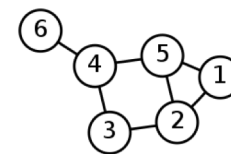


具有不同数据结构的数据组合

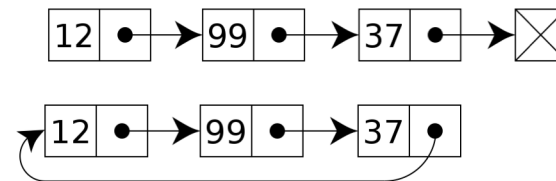
树



图



链表



多模态数据融合

- 多模态数据融合指的是将多个模态的数据进行有效整合，产生比任何单一模态数据所能提供的更一致、准确、有用的信息，从而作出更明智的决策。

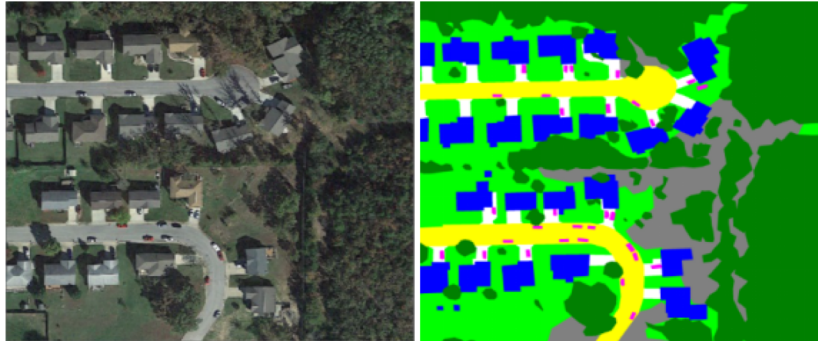
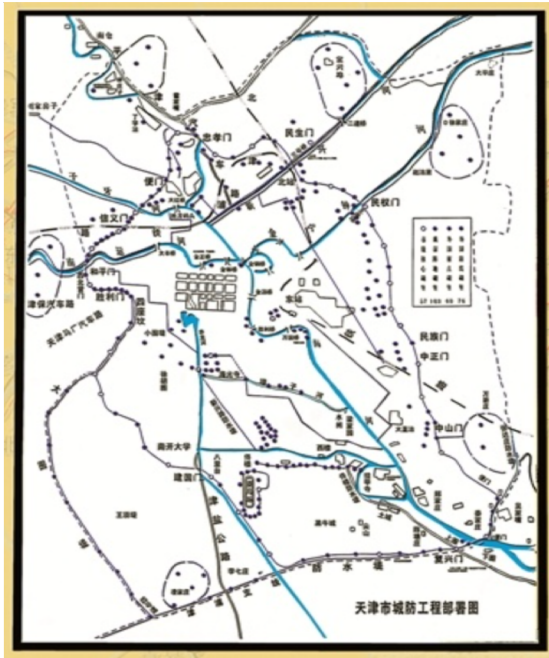


摸到象腿的人说：大象像一根柱子；
摸到身体的人说：大象像一堵墙；
摸到耳朵的人说：大象像一把扇子。

—— 单一模态数据难以对事物提供完整描述，容易“横看成岭侧成峰”。

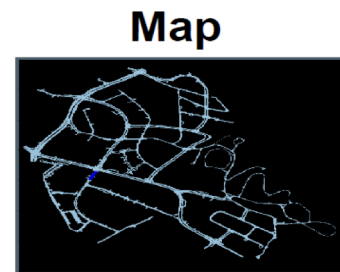
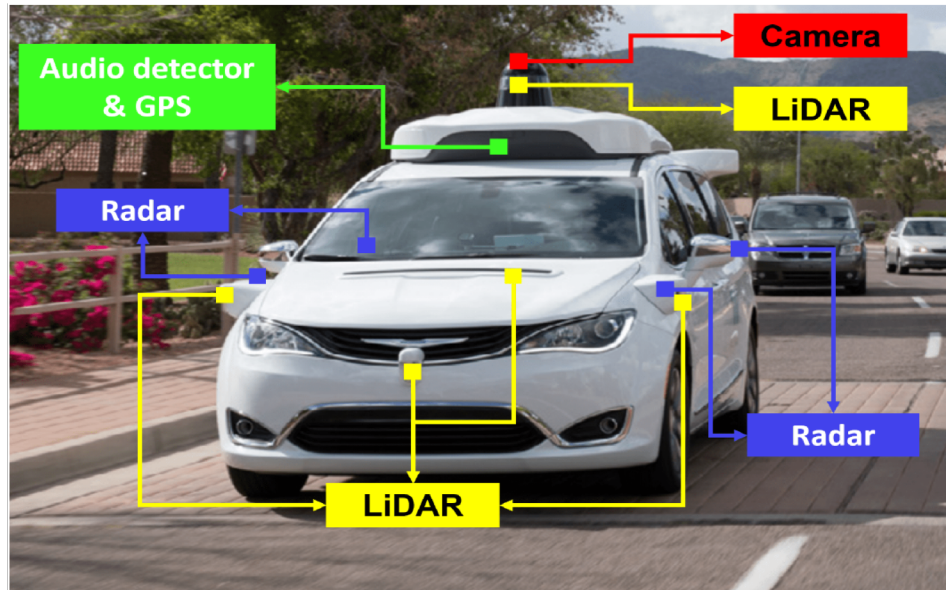
多模态数据融合

- 案例1：军事辅助决策
 - 通过对城防地图、遥感影像、侦察视频、兵力部署、武器装备等情报信息进行全方位分析和融合，帮助判断局势动态，制定作战计划。

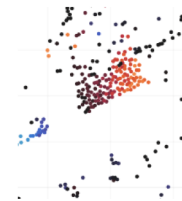


多模态数据融合

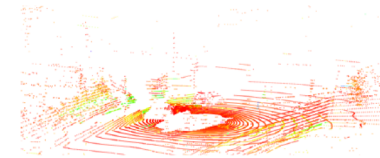
- 案例2：自动驾驶决策^[1]
 - 具备自动驾驶功能的车辆通常配备GPS、摄像头、毫米波雷达、激光雷达等不同模态的传感器，在不同情况下提供互补的感知信息，从而鲁棒、实时地感知目标并作出决策。



Radar Points



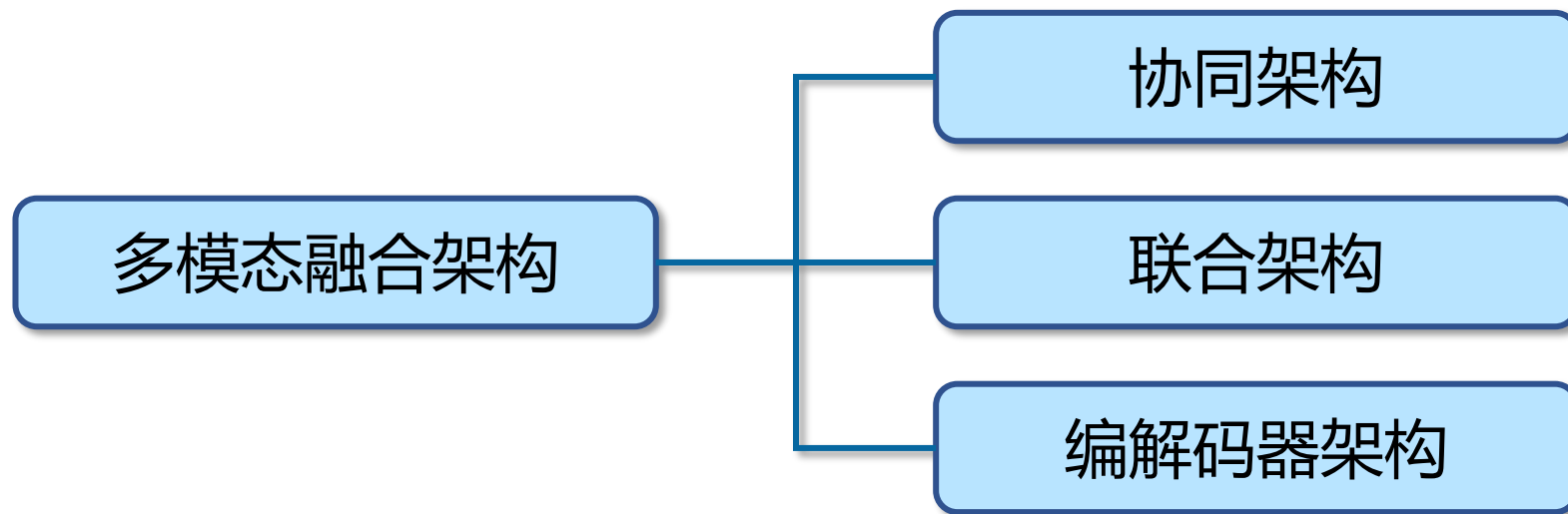
LiDAR Points



[1] <https://zhuanlan.zhihu.com/p/328671301>.

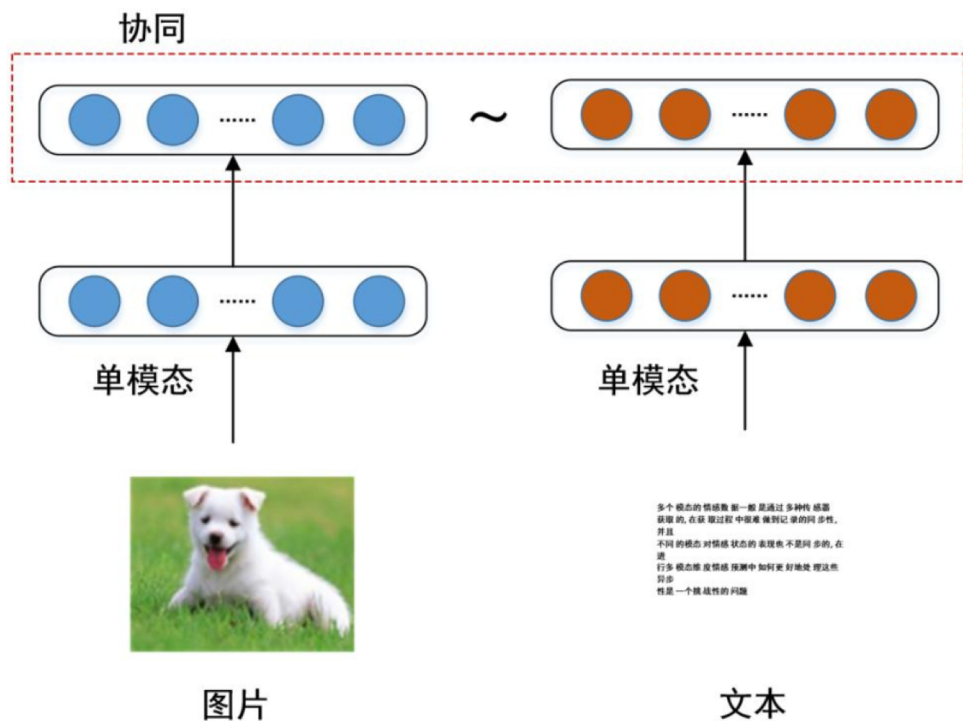
多模态融合架构

- 多模态融合试图找到某种适用于各模态数据的统一表示，其主要目标是缩小模态间的异质性，同时保持各模态特定语义的完整性。
- 多模态融合架构分为三类：协同架构、联合架构、编解码器架构。



多模态融合架构：协同

- **协同架构**：通过查找协同子空间中不同模态之间的关联性，使得在一定约束下各单模态相互协同。



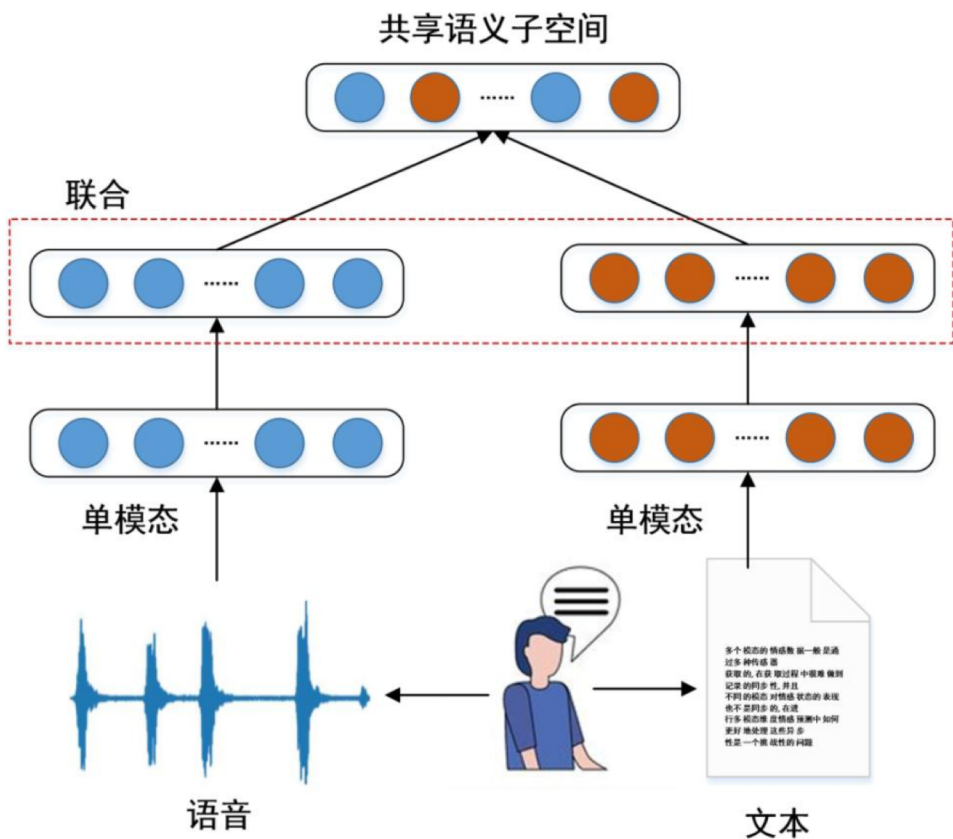
协同方式：跨模态相似模型、典型相关分析。

优点：各模态信息不同，有利于保留单模态独有的特征；各单模态可以独立运行，有利于跨模式迁移学习。

缺点：模态融合难度大，跨模态学习模型不易实现，且难以在两个以上的模态之间实现迁移学习。

多模态融合架构：联合

- 联合架构：将多个单模态特征映射到共享语义子空间，从而融合多个模态的特征表示。



联合方式：

“加”联合 $z = f(w_1^T v_1 + \dots + w_n^T v_n)$

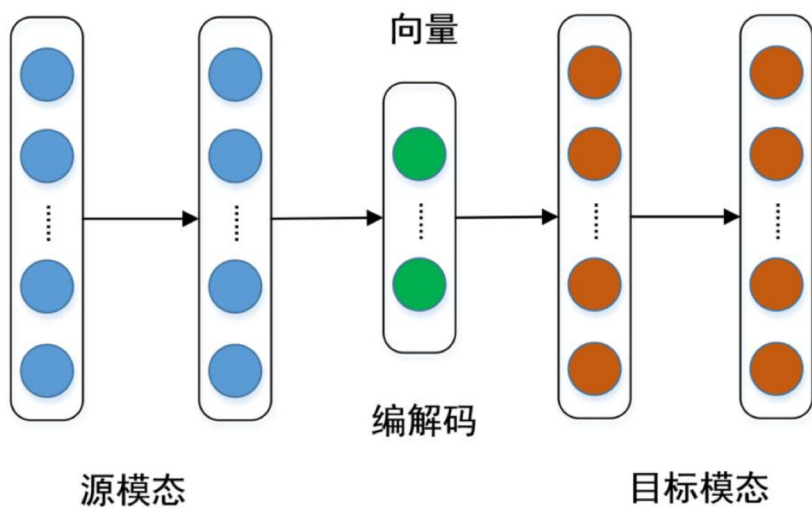
“乘”联合 $z = \begin{bmatrix} v^1 \\ 1 \end{bmatrix} \otimes \dots \otimes \begin{bmatrix} v^n \\ 1 \end{bmatrix}$

优点：融合方式简单，且共享子空间具备语义不变性，有利于模态转化。

缺点：对每个单模态的语义完整性有较高要求，数据不完整或错误的问题在早期较难被发现和处理。

多模态融合架构：编解码器

- **编解码器架构**：通常用于将一种模态映射到另一种模态的多模态转换任务中，主要由编码器和解码器两部分组成。



优点：能够在源模态基础上生成新的目标模态样本。

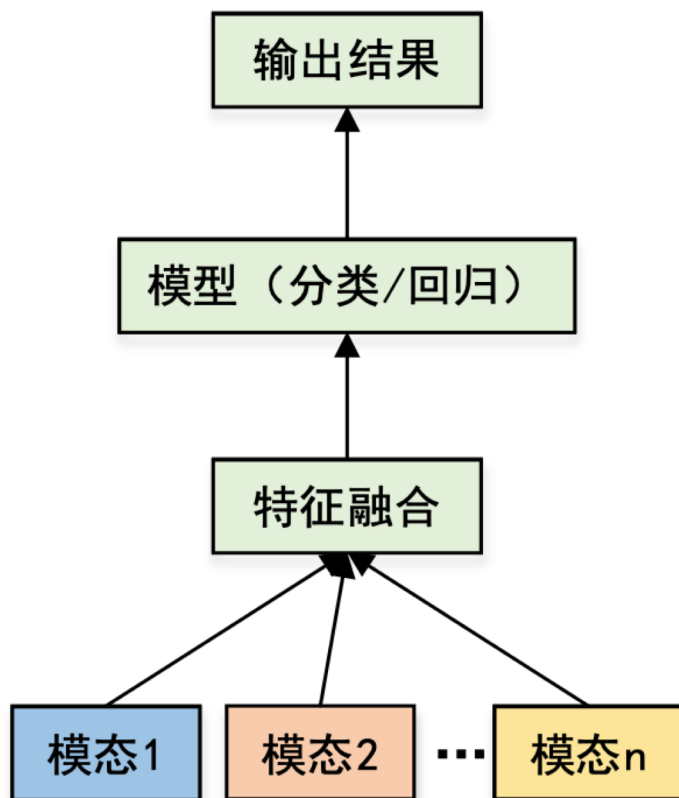
缺点：每个编码器和解码器只能编码一种模态，且决策模块设计复杂。

多模态融合方法

- 模型无关的融合：不直接依赖于特定的机器学习算法。
 - 早期融合（基于特征）：在提取各模态特征后立即融合
 - 晚期融合（基于决策）：在每种模式输出结果后进行融合
 - 混合融合：结合了早期融合方法和单模态预测器的输出
- 基于模型的融合：利用机器学习模型显式地在构造中融合。
 - 多核学习
 - 图像模型
 - 神经网络

多模态融合方法：模型无关

- **早期融合**：也称为**特征级融合**，在提取各模态特征后就进行融合（通常只是简单连接各模态特征）。



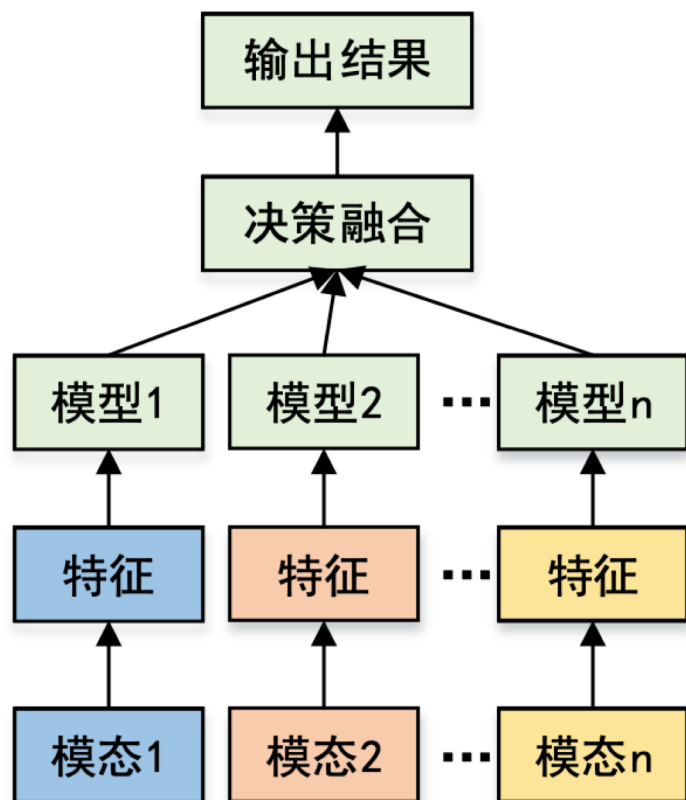
优点：能较好捕捉特征之间的关系。

缺点：

- 容易**过度拟合**训练数据；
- 需要解决不同数据源的**时间同步**问题；
- 可能无法充分展示模态间的互补性，导致**输入冗余**向量。

多模态融合方法：模型无关

- **晚期融合**：也称为**决策级融合**，在各模态作出决策之后再融合，得到最终决策。



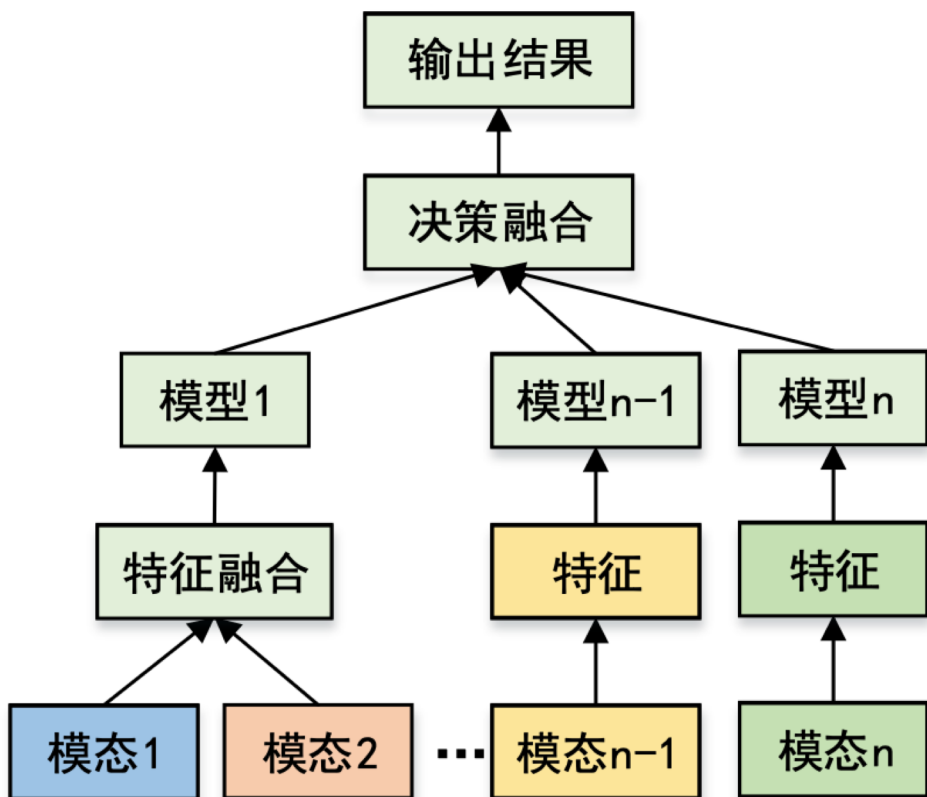
融合方式：基于规则确定不同模型输出结果的组合，即**规则融合**（最大值融合、平均值融合、贝叶斯规则融合、集成学习等）。

优点：能较好处理过拟合问题。

缺点：不允许分类器同时训练所有数据。

多模态融合方法：模型无关

- **混合融合**：结合了早期和晚期融合方法，在综合两者优点的同时，也增加了模型的结构复杂度和训练难度。



优点：方法灵活。

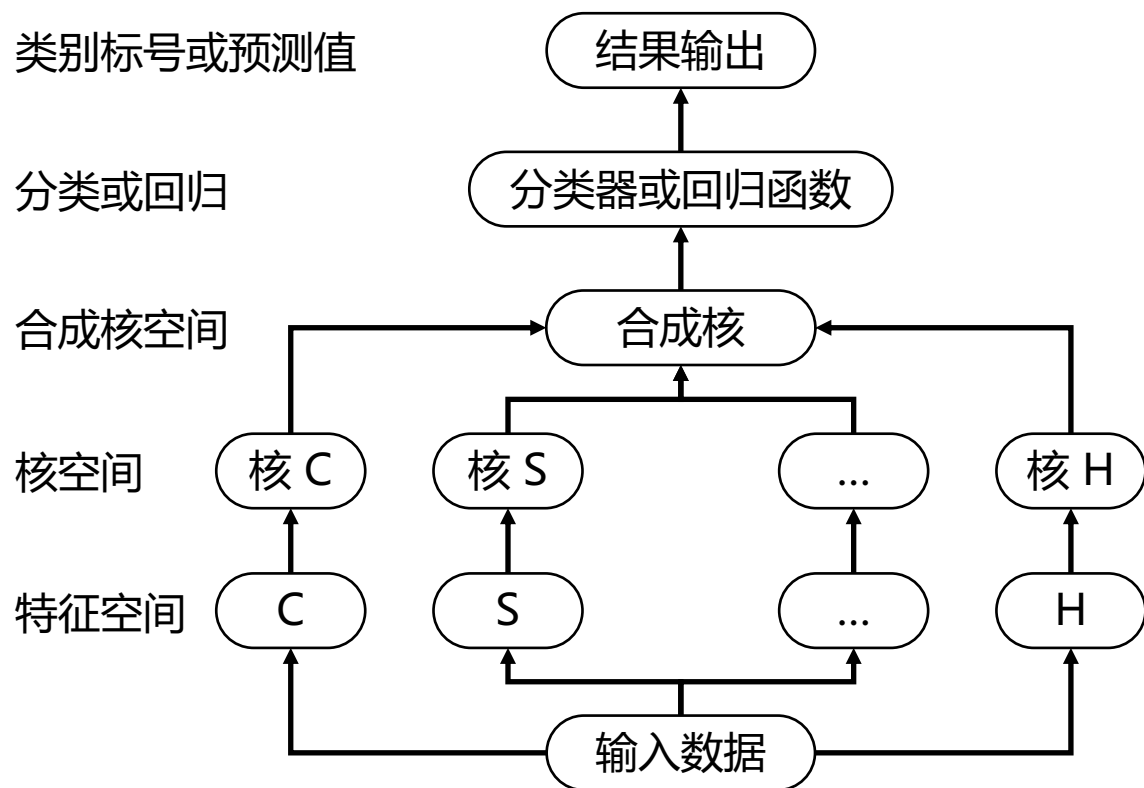
缺点：需要仔细设计组合策略，例如如何、何时和哪些模态可以融合。

多模态融合方法

- 模型无关的融合：不直接依赖于特定的机器学习算法。
 - 早期融合（基于特征）：在提取各模态特征后立即融合；
 - 晚期融合（基于决策）：在每种模式输出结果后进行融合；
 - 混合融合：结合了早期融合方法和单模态预测器的输出。
- 基于模型的融合：利用机器学习模型显式地在构造中融合。
 - 多核学习
 - 图像模型
 - 神经网络

多模态融合方法：基于模型

- **多核学习**是内核支持向量机方法的扩展，允许使用不同的核对应数据的不同视图。



优点：

- 核选择具有灵活性，有利于更好地融合异构数据；
- 损失函数是凸的，允许使用标准优化包和全局最优解进行模型训练，可大大提升深度神经网络模型性能。

缺点：在测试期间需要依赖训练数据，从而导致内存占用较大，性能较低。

多模态融合方法：基于模型

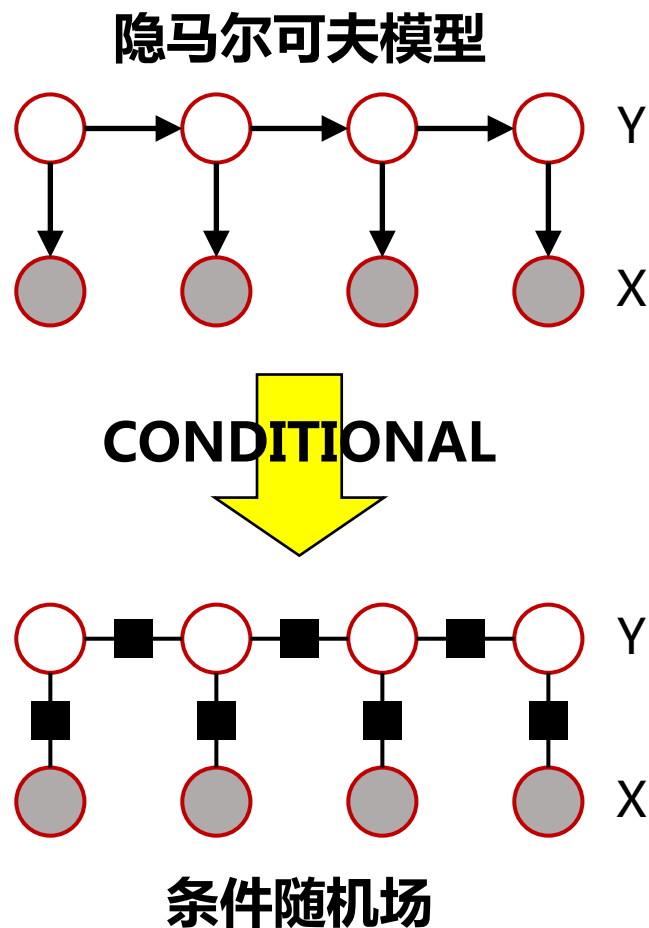
- **图像模型**通过图像分割、拼接和预测对浅层或深度图形进行融合，从而生成模态融合结果。

常见模型：生成模型联合概率（如隐马尔可夫模型）、判别模型条件概率（如条件随机场）。

优点：

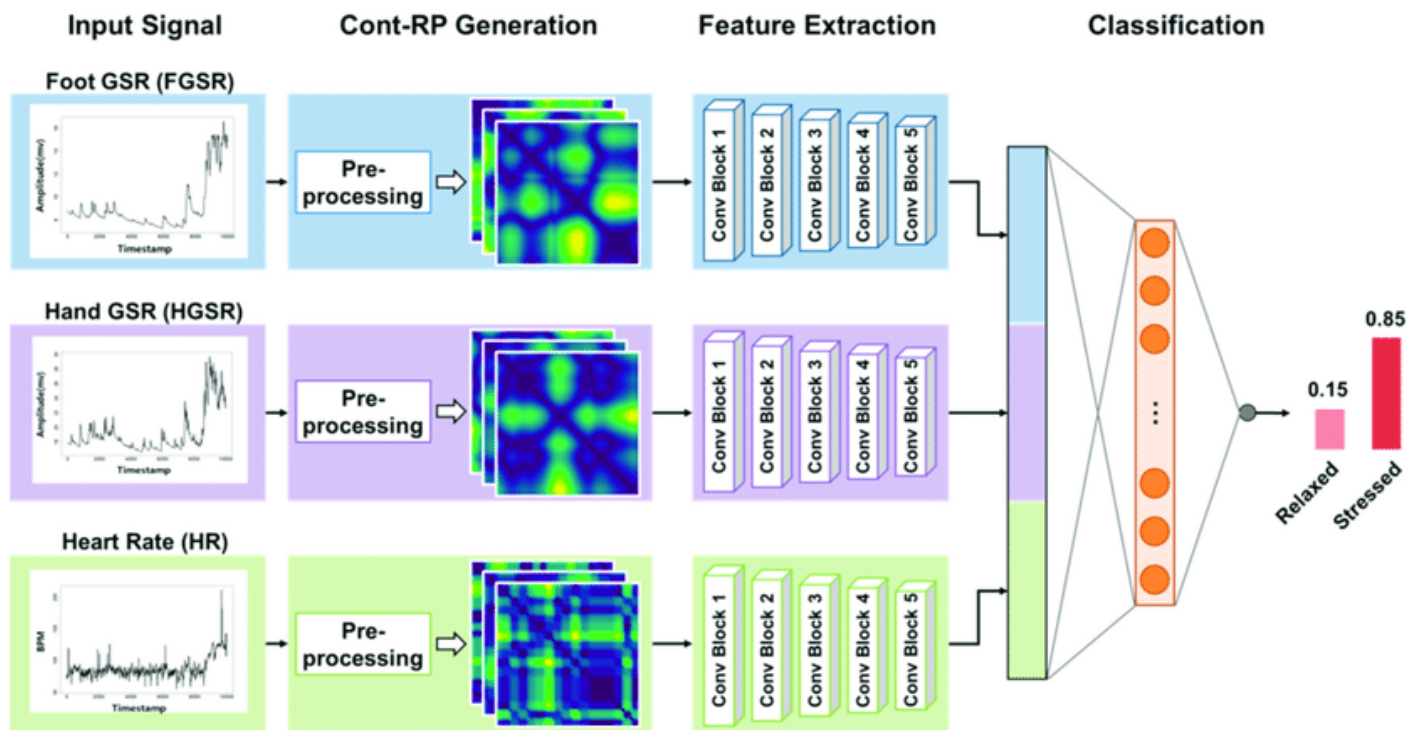
- 能够很容易地利用数据的空间和时间结构，特别适合于与时间相关的建模任务；
- 允许将人类专家知识嵌入到模型中，增强模型的可解释性。

缺点：模型的泛化能力不易提升。



多模态融合方法：基于模型

- **神经网络**是目前多模态融合最广泛的方法之一，已大量用于多模态融合的任务中。



优点：

- 具备从大量数据中学习的能力，且性能优异；
- 神经网络的分层方式有利于不同模态的嵌入，具有较好的可扩展性。

缺点：随着模态增多，模型的可解释性变差。



本章完