

1

自编码器直接使用一层或者多层的神经网络对输入数据进行映射，得到输出向量，作为从输入数据提取出的特征。一般用于降维、去噪和图像生成

通过将输入压缩成一个隐藏空间表示来进行工作，然后通过这种表示重构输出。

编码器：自编码器的前半部分，功能在于把输入变成一个隐藏的空间表示。它可以用一个编码函数 $h = f(x)$ 表示。

解码器：这部分旨在从隐藏空间的表示重构输入。它可以用解码函数 $r = g(h)$ 表示。

2

自编码器由编码器和解码器组成，二者可以被分别定义为变换 ϕ 和 ψ ，使得：

$$\begin{aligned}\phi &: \mathcal{X} \rightarrow \mathcal{F} \\ \psi &: \mathcal{F} \rightarrow \mathcal{X} \\ \phi, \psi &= \arg \min_{\phi, \psi} \|\mathcal{X} - (\psi \circ \phi)\mathcal{X}\|^2\end{aligned}$$

在最简单的情况下，给定一个隐藏层，自编码器的编码阶段接受输入 $\mathbf{x} \in \mathbb{R}^d = \mathcal{X}$ 并将其映射到 $\mathbf{h} \in \mathbb{R}^p = \mathcal{F}$ ：

$$\mathbf{h} = \sigma(\mathbf{W}\mathbf{x} + \mathbf{b})$$

像 \mathbf{h} 通常表示编码、潜变量或潜在表示。 σ 是一个逐元素的激活函数（例如sigmoid函数或线性整流函数）。 \mathbf{W} 是权重矩阵， \mathbf{b} 是偏置向量。权重和偏置通常随机初始化，并在训练期间通过反向传播迭代更新。自编码器的解码阶段映射 \mathbf{h} 到重构 \mathbf{x}' （与 \mathbf{x} 形状一致）：

$$\mathbf{x}' = \sigma'(\mathbf{W}'\mathbf{h} + \mathbf{b}')$$

其中解码器部分的 $\sigma', \mathbf{W}', \mathbf{b}'$ 可能与编码器部分的 $\sigma, \mathbf{W}, \mathbf{b}$ 无关。

自编码器被训练来最小化重建误差（如平方误差），通常被称为“损失”：

$$\mathcal{L}(\mathbf{x}, \mathbf{x}') = \|\mathbf{x} - \mathbf{x}'\|^2 = \|\mathbf{x} - \sigma'(\mathbf{W}'(\sigma(\mathbf{W}\mathbf{x} + \mathbf{b})) + \mathbf{b}')\|^2$$

其中 \mathbf{x} 通常在训练集上平均。

3

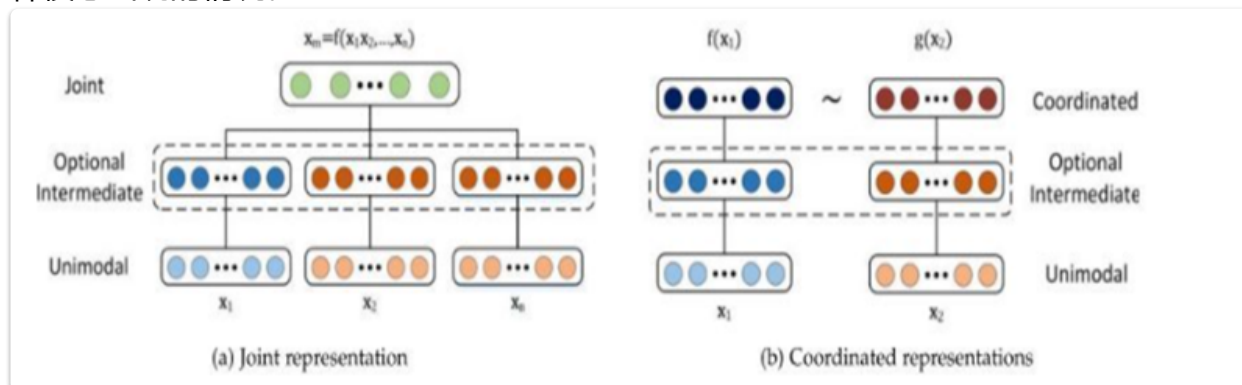
玻尔兹曼机可被视作随机过程的，可生成的相应的Hopfield神经网络。它是最早能够学习内部表达，并能表达和（给定充足的时间）解决复杂的组合优化问题的神经网络。

离散Hopfield神经网络+模拟退火+隐单元=Boltzman机

4

多模态表征方法：联合表示（joint representation），协同表示（coordinated representation）

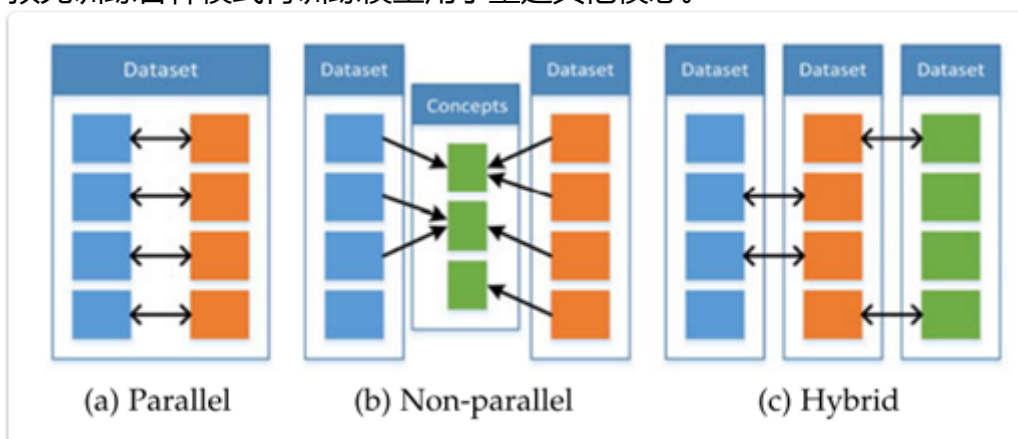
- 联合特征表示方法将多模态数据投影到一个共同的特征表示空间，最适用于推理时所有模态的数据都出现的场景。
- 协同特征表示方法将每个模态投影到分离但相关的空间，这种方法适用于推理时仅有一种模态出现的情况。



5

预先训练单个模态的编码器再训练模型用于重建其他模态。

预先训练各种模式再训练模型用于重建其他模态。



笔记（个人观点）

模态融合的重点主要有：

1. 将两种模态映射到同一潜在空间；
 2. 两种模态处于相同的潜在空间，以保证特征的互补性。
- 现在做大模型再根据下游任务做微调感觉是一种大趋势，比如GPT比如DALL-E2。

reference

Baltrušaitis, T., Ahuja, C., & Morency, L. P. (2018). Multimodal machine learning: A survey and taxonomy. *IEEE transactions on pattern analysis and machine intelligence*, 41(2), 423-443.