

本章使用K均值算法进行聚类。K均值算法随机初始化K个聚类中心，将样本中和聚类中心距离最小的类别作为该样本的聚类类别，并且在每次迭代中使用同一个簇中的所有样本更新聚类中心直至收敛。K均值算法主要受两个因素的影响，K值和距离度量方式，本章使用欧式距离作为度量，K值是需要根据需求进行调节的超参。K均值算法将样本分成K个聚类簇，确保每个聚类中心的样本都有某种相似性，对每一个簇给定一个类别，可以得到对应样本的聚类标签，这时可以使用监督学习的方式进行特征提取。本章使用无标签样本的聚类标签和带标签数据的真实标签进行特征提取，将聚类的相似性和真实的类别信息同时引入深度模型，结合监督信息，使模型学习到适用于融合了监督信息和无监督信息的分类特征。

4.2 结合无监督知识学习的高光谱图像分类方法

4.2.1 原理介绍

聚类算法可获取反映样本相似性结构的伪标记知识，本章通过对无标记数据引入聚类标签构建联合聚类 and 分类的任务学习无监督知识。不同于主动学习模型预测样本伪标签可能会引入标签噪声，同一个簇中的样本一定存在某些相似特性，体现出无标记数据的结构信息。在可用标签十分有限时，标记样本往往不能为模型提供足够的监督信息，使深度模型容易过拟合。本章利用聚类算法得到样本的聚类标签，使用神经网络将对真实标签的分类和对聚类标签的分类嵌入同一个框架，将样本相似性作为先验知识引入模型，使模型能够学习无监督知识，同时为深度模型提供正则化约束，所以本章中的无监督信息本质是通过引入分类损失来实现。

对于给定的高光谱数据集 \mathbf{X} ，对应数据的维度为 $w \times h \times d$ ， w 和 h 表示高光谱图像对应的宽度和高度， d 表示光谱波段数。假设数据集总共包含 n 个样本，整个数据集可以表示为 $X : (x_1, x_2, \dots, x_n)$ ， $x_i \in X$ ，前 l 个标记样本 $X_L : (x_1, x_2, \dots, x_l)$ 对应的标签为 $Y_L : (y_1, y_2, \dots, y_l)$ ， $y_i \in 1, 2, \dots, C$ ，表示数据集中包含 C 个类别，剩余的 $n - l$ 个未标记样本表示为 X_U ，我们的目标是利用 X 和 Y_L 训练一个分类器，能够正确对 X_U 中的样本进行分类。

如图4-1所示，共享的特征提取模块（共享模块）对一个batch中的标记样本和无标记样本进行特征提取，然后将特征输入相应的分类器，利用真实标签和聚类标签的监督信息，进行端到端训练。为了更好地利用无标记样本，本章模型设计了三路分支的网络，通过共享的卷积模块，对标记样本和无标记样本共同进行特征提取通过联合训练将无监督信息有效地传递到监督学习的过程，成功提升高光谱图像分类任务的性

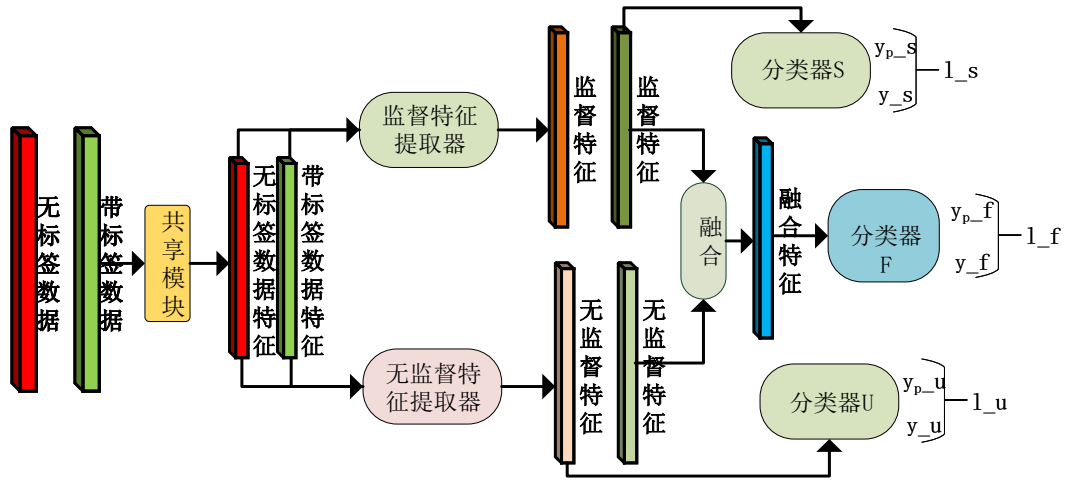


图 4-1: 结合无监督学习的高光谱图像分类方法框架图

能。

模型使用卷积神经网络对数据 X 进行特征提取，对于所有样本 $X = X_L \cup X_U$ ，K均值算法将其分成 K 个聚类簇，为每一个样本赋予一个聚类标签，新得到的训练样本 $X : (x_1, x_2, \dots, x_n)$ ，其对应的标签为 $Y_U : (y_1, y_2, \dots, y_n)$ ， $y_i \in 1, 2, \dots, K$ ，表示聚类标签。模型整体框架如图4-1所示，其中带标签数据和无标签数据共同组成一个batch，经过共享的特征提取模块，得到无标签数据特征和带标签数据特征；而后，它们被送入监督特征提取器，得到带标签数据的监督特征和无标签数据的监督特征；送入无监督特征提取器得到带标签数据的无监督特征和无标签数据的无监督特征。最终，带标签数据的监督特征通过分类器S进行标签预测，和真实标签通过交叉熵得到带标签数据的损失 l_s ；无标签数据的无监督特征通过分类器U预测无标签数据的聚类标签，通过K均值算法得到的聚类标签计算交叉熵得到无标签数据的损失 l_u ；带标签数据的监督特征和无监督特征通过融合模块，得到包含监督信息和无监督信息的特征，通过分类器F，得到带标签数据的预测标签，通过与真实标签计算交叉熵得到带标签数据融合了监督信息和无监督信息的损失 l_f 。

由于K均值算法可以捕获样本的类间相似性和差异性，通过卷积神经网络利用聚类标签进行训练，可以有效将无监督信息通过监督的方式进行强化，同时通过共享的特征提取模块，可以将监督信息和无监督信息高效地融合，使无监督的知识可以在训练的过程中流向监督任务，为网络提供正则化，防止过拟合现象。下面详细介绍本章模型的网络结构和训练过程。

4.2.2 网络结构

如图4-1所示，整个模型主要包含如下几个模块：共享的特征提取模块（共享模块）、监督特征提取模块、无监督特征提取模块、特征融合模块、分类器S、分类器U和分类器F。下面分别介绍相应的模块。

1)共享的特征提取模块使用十层的三维卷积神经网络进行数据的特征提取，具体的网络结构及参数设置如表4-1所示：

层数	卷积核	卷积步长	特征数	padding
Conv1	(3,1,8)	(1,1,3)	64	valid
Conv2	(1,1,3)	(1,1,1)	64	valid
Conv3	(1,1,3)	(1,1,1)	64	same
Conv4	(1,3,3)	(1,1,2)	128	valid
Conv5	(1,1,3)	(1,1,1)	128	valid
Conv6	(1,1,3)	(1,1,1)	128	same
Conv7	(1,1,3)	(1,1,2)	256	valid
Conv8	(1,1,3)	(1,1,1)	256	valid
Conv9	(1,1,3)	(1,1,1)	256	same
Conv10	(1,1, s_s)	(1,1,1)	512	valid

表 4-1: 共享的特征提取模块结构

表4-1中，padding的方式valid表示不进行特征补齐，same表示进行补齐， s_s 表示Conv9层的光谱维特征大小，经过Conv10，光谱特征维度变成一，特征变成一个向量。每一个卷积模块包含一个卷积操作，一个批归一化（BN）操作和一个激活函数（Swish）。

2) 监督特征提取器和无监督特征提取器各自包含一个全连接层，对应的神经元节点个数为256，全连接层后包含一个批归一化（BN）操作和激活函数（Swish）；分类器S包含一个全连接层，将带监督信息的特征转化到真实样本标签的概率分布，其神经元个数为数据集真实的类别个数；分类器U和分类器S类似，不过其输出节点的个数为K均值算法的聚类个数；分类器F的输入包含监督信息和无监督信息的融合特征，输出为数据集真实类别数；特征融合模块包含不同的融合方式，我们使用特征拼接的方

式得到融合特征，对于不同特征融合方式的讨论见4.3.5。

4.2.3 损失函数

整个模型具有三路独立分支，因此损失函数包含三个部分，表示为：

$$\ell_t = \lambda_1 \ell_s + \lambda_2 \ell_u + \lambda_3 \ell_f \quad (4-1)$$

其中， ℓ_t 表示总损失， ℓ_s 表示监督学习一路对带标记数据的分类损失， ℓ_u 表示无监督一路的未标记数据利用聚类标签的分类损失， ℓ_f 表示融合了监督信息和无监督信息的特征对标记数据的分类损失。 λ_i 表示均衡系数。由于三路分支均为分类任务，我们使用交叉熵定义所有的损失函数，具体如下：

$$\begin{aligned} \ell_s &= -\frac{1}{l} \sum_{i=1}^l y_{si} \times \log(\tilde{y}_{si}) \\ \ell_u &= -\frac{1}{n} \sum_{i=1}^n y_{ui} \times \log(\tilde{y}_{ui}) \\ \ell_f &= -\frac{1}{l} \sum_{i=1}^l y_{fi} \times \log(\tilde{y}_{fi}) \end{aligned} \quad (4-2)$$

其中 y_{si}, y_{fi} 均表示带标签数据的真实标签， y_{ui} 表示无标签数据的聚类标签，对应的 $\tilde{y}_{si}, \tilde{y}_{ui}, \tilde{y}_{fi}$ 为模型预测标签。

4.3 实验结果及分析

4.3.1 数据集

本章使用了三个广泛使用的高光谱数据集验证所提方法的有效性，这三个数据是Indian Pines、PaviaU和Salinas Scene，详细的数据集介绍见2.3.1部分。