目 录

[1. 问题描述 2](#_Toc23532033)

[2. 主要思路 2](#_Toc23532034)

[3. 实现方案 3](#_Toc23532035)

[3.1数据集选取 3](#_Toc23532035)

3.2模型构建 3

[3.3距离计算 4](#_Toc23532035)

[3.4 Supervised Learning 5](#_Toc23532035)

[3.5 Zero Shot Learning 5](#_Toc23532035)

[3.6 One Shot Learning 5](#_Toc23532035)

[4. 结果展示 6](#_Toc23532036)

[4.1 Supervised Learning方案 6](#_Toc23532035)

[4.1.1识别准确率 6](https://d.docs.live.net/2a5af699d38e1057/信号识别方案描述v4.docx#_Toc23532035)

[4.1.2向量SNE降维图 6](https://d.docs.live.net/2a5af699d38e1057/信号识别方案描述v4.docx#_Toc23532035)

[4.2 Zero Shot Learning方案 7](#_Toc23532035)

[4.2.1 1类未知样本 7](https://d.docs.live.net/2a5af699d38e1057/信号识别方案描述v4.docx#_Toc23532035)

[4.2.2 2类未知样本 8](https://d.docs.live.net/2a5af699d38e1057/信号识别方案描述v4.docx#_Toc23532035)

[4.3 One Shot Learning方案 8](#_Toc23532035)

[4.3.1识别准确率 8](https://d.docs.live.net/2a5af699d38e1057/信号识别方案描述v4.docx#_Toc23532035)

[5. 潜在问题 8](#_Toc23532037)

[6. 优化方法 9](#_Toc23532038)

[6.1模型优化 9](https://d.docs.live.net/2a5af699d38e1057/信号识别方案描述v4.docx#_Toc23532035)

[6.2距离优化 9](https://d.docs.live.net/2a5af699d38e1057/信号识别方案描述v4.docx#_Toc23532035)

[6.3损失函数优化 9](https://d.docs.live.net/2a5af699d38e1057/信号识别方案描述v4.docx#_Toc23532035)

### 问题描述

有一包含了*n*类的**已知信号集合*K*** *(Known)*，其中每一类信号*Ki**(i=1,2,…,n)*有其相应的样本集*Di**(i=1,2,…,n)*，含有若干个数据对 *{x, y}1:m*，同时设**未知信号集合**为***U*** *(Unknown)*，且集合***U***中的未知类别数亦是未知的。

现需要有一个模型，在仅仅知晓上述已知信号集***K***的条件下，对于不断出现的**新信号样本***I (Instance)*，能够正确判断出该信号样本*I*是属于已知类别***K***还是未知类别***U***，若判断出*I***∈*K***，则应进一步将*IK*正确的分类：即找出*i (i=1,2,…,n)*，使*IK***∈***Ki* ；若判断出*I***∈*U***，则应**记录**下新的*Iu*到**已记录未知信号集*R*** *(Recorded)*，并进一步将*Iu*分类：即添加新未知类别*Rj (j > n)*、或找出*j (j > n)*，使*Iu***∈***Rj*。

### 主要思路

首先，进行简要地分析：这个问题情形，与**开集学习***(Open Set Learning)*问题不同，这里的模型需要在能判别出未知信号类别的前提下，再利用该未知信号进行分类学习并**逐步进化**；也和**零样本学习***(Zero Shot Learning)*问题不同，这里所有的语义向量都需要模型**自行学习**来获得，而不是依靠现成的语义向量及其到类别映射关系；亦和**小样本学习***(Few Shot Learning)*问题不同，这里的模型需要在获得了小样本类别的语义向量的基础上，再利用新来的样本对已获得的语义向量进行**更新和优化**。可以说，这是一个兼具挑战性、创新性和研究价值的领域。

由此，我们考虑，综合上述已有的三种机器学习问题的特点，构建出一个能够持续自进化的（经历从Zero Shot Learning到Open Set Learning再到Few Shot Learning的训练过程）、主动学习新样本和新类别的深度学习模型。我们将其定义为**增量学习**(**Incremental Learning**(候选Agile/Evolving/Accumulate Learning))模型。

首先根据***K***中的已知样本训练出一个成熟的深度学习**模型*M***，将其中的特定隐藏层，作为提取已知类别信号特征的**特征层*F*** *(Features)*，求出每一个已知信号样本在***F***层上的输出即特征向量*V*，按类别分组，求出各个类别的所有样本的特征向量的**均值***Si (i=1,2,…,n)*，并将*Si*作为该类别的**语义向量***(Semantic Vector)*。此后，可通过计算新样本特征向量和已知类别语义向量之间的距离来对新样本进行类别判定。因此，我们预定义了距离**阈值*θ*** *(Threshold)，*以作为区分**已知类**和**未知类**、以及区分**已记录未知类**和**未记录未知类**的分界线。

当有新样本*I*出现时，我们将其喂入模型***M***中，以获得其特征向量*V*，然后对于每个已知类别，计算出向量*V*和*Si*之间距离*di*：

1. 若存在某个*di*，满足*di* < ***θ****K*，则认为该样本属于已知类集合***K***，并进一步找到*i = argmin di*，以划分到特定类别*Ki*；
2. 否则，认为该样本属于未知信号类***U***，并进一步用它来扩充我们的已记录信号集***R***：
   1. 若当前***R***为空，则添加一个新信号类别*Rn+1*到***R***中,并将V作为该类别的语义向量*Sn+1*；
   2. 否则，对于***R***中的每个已记录类，计算向量*V*和*Sj (j>n)*之间的距离*dj*,：
      1. 若存在某个*dj*，满足*dj* < ***θ****R*，则认为该样本属于已记录信号集*R*，进一步找到*j = argmin dj*，将其划分到特定类别*Rj*，并且用*I*更新该类别的语义向量*Sj*；
      2. 否则，添加一个新信号类别*Rk*到***R***中,并将V作为该类别的语义向量*Sk*

### 实现方案

在具体实现过程中，我们针对问题的核心进行了一些分解实验，如Zero Shot Learning方案、Supervised Learning方案和Few Shot Learning方案，以期对模型获得一个更为全面的认识。

#### 3.1 数据集选取

我们使用公开数据集RadioML 2016a *(*[*https://www.deepsig.io/datasets*](https://www.deepsig.io/datasets)*)*中的数据进行模拟。数据信息如图3.1.1



图3.1.1

考虑到较大的信噪比的样本数据拥有更小的噪音，这里我们选择其中SNRs>=16的数据作为样本，将其中的11类划分为已知类和未知类。

Train set为已知类的85%的数据，Validation set为已知类的5%的数据，而Test set则为已知类和未知类各10%的数据。

此外，我们对信号数据使用了MinMaxScaler进行预处理：

#### 3.2 模型构建

我们搭建了如图3.2.1所示的SRCNN网络作为模型进行训练

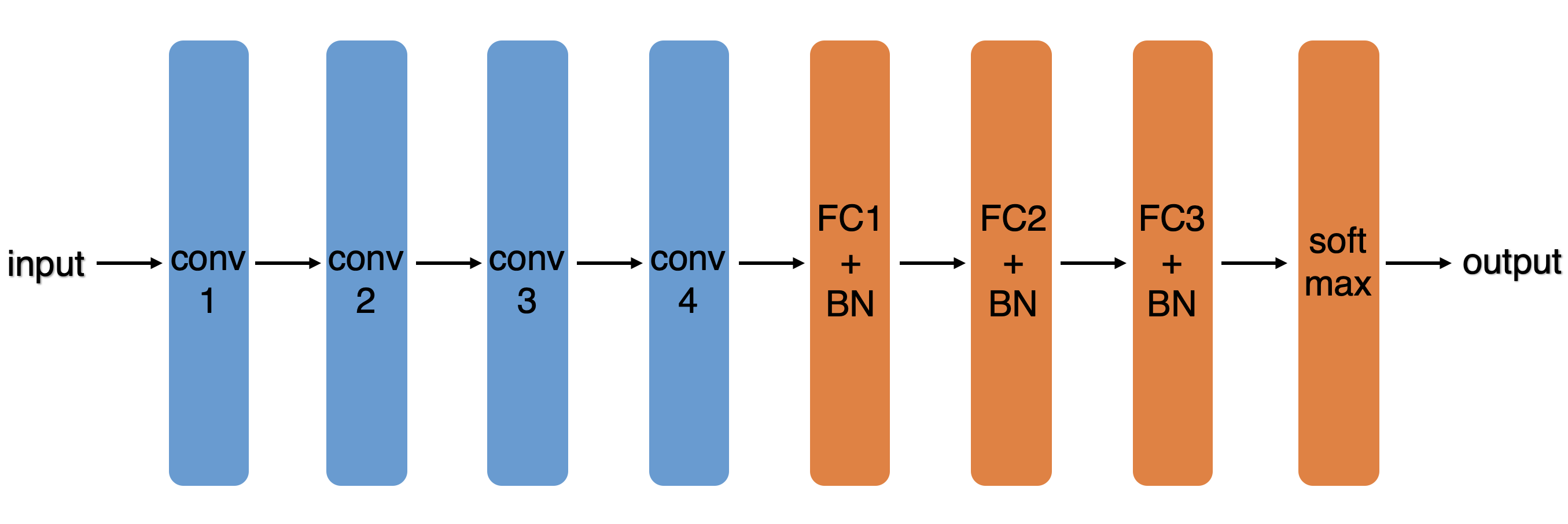


图3.2.1

模型的总损失函数定义如下：

为Cross Entropy Loss，为Autoencoder Loss，为Center Loss，代表其相应的权重。

，其中为x的解码函数。

，其中为第类所有样本的特征向量的聚类中心，为样本的特征向量。

我们将SRCNN模型的**FC3+BN**层的作为特征层。将样本输入到网络后，即可由此特征层的输出，获得该样本的特征向量。

#### 3.3 距离计算

实验中，我们共使用了共五类距离，其中，前四类距离可以统一采用如下的广义距离公式定义：

，其中M为变换矩阵，*v*为新样本的特征向量，*u*为某类别的语义向量。

1. M=Σ：Σ为某一特定类别的所有样本在属性维度上的协方差矩阵，此时距离*D(v)*成为**马氏距离***(Mahalanobis Distance)*
2. M=Λ：Λ为取Σ的对角线元素形成的对角线矩阵
3. M=σ2I：其中，*tr()*表示矩阵的迹，*l*为属性的维度数，I为单位矩阵
4. M=I：此时距离*D(v)*即成为**欧氏距离***(Euclidean Distance)*
5. ：此时距离*D(v)*即**余弦距离***(Cosine Distance)*的**倒数**

#### 3.4 Supervised Learning

进行此项实验，有助于获取Few Shot Learning方案的识别准确率的上界*(Upper Bound)*。

此时Train set将从全部11个类上产生，将其喂入模型以进行监督学习式的训练，并且在特征层上通过聚类的方法来获取各个已知类的语义向量。在测试时，计算测试样本的特征向量与所有已知类的语义向量之间的距离（采用效果目前最好的欧氏距离），取最小距离所属类为当前样本的预测类型。

#### 3.5 Zero Shot Learning

此项实验是为了模拟模型的在初见到未知类样本时的情形。

将Train set喂入模型中，在特征层上通过聚类的方法获得各个已知类的语义向量。在测试时，分别计算测试样本的特征向量与所有已知类的语义向量之间的距离，如果最小距离大于阈值，则认为它属于未知类，反之则为最小距离所属类。

未知类只有一类时，问题将退化为Open Set Learning。实验中我们会考虑当未知类大于等于两类的情况，不过这里我们只对未知类等于两类的情况进行说明：当两个未知类的语义向量都不存在时，定义第一个被测试的未知类为未知类1，剩下的那一类为未知类2。当只存在未知类1时，计算样本与未知类1的距离，如果距离小于最小距离，则判断属于未知类1并更新语义向量，反之则属于未知类2。当两个未知类的语义向量都存在时，可采用后面的Few Shot Learning方案来进行预测，并逐步更新语义向量。这里，我们只对未知类的语义向量进行更新。

#### 3.6 Few Shot Learning

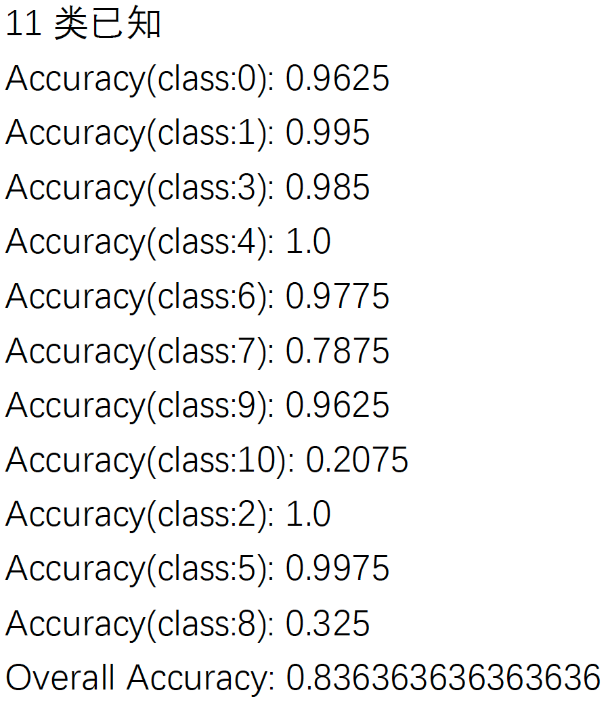
在类Zero Shot Learning方案中，随着某未知类的新样本增加，即可逐步构建起该类的语义向量，并过渡到Few Shot Learning阶段。

将Train set喂入模型中，在特征层上通过聚类的方法获得各个已知类的语义向量。同时每个未知类随机取2-10个样本，在特征层上进行聚类从而获得未知类的语义向量（可动态更新）。测试时，计算样本特征向量与所有类的语义向量的距离，取最小距离所属类为当前样本的预测类型。

### 结果展示

#### 4.1 Supervised Learning方案

##### 4.1.1识别准确率



可见，除了第10类和第8类的准确率稍逊之外，其余类别相对表现良好。

##### 4.1.2向量SNE降维图

Train Set (图中红星标识该类别的语义向量，即该类所有样本的特征向量的聚类中心)

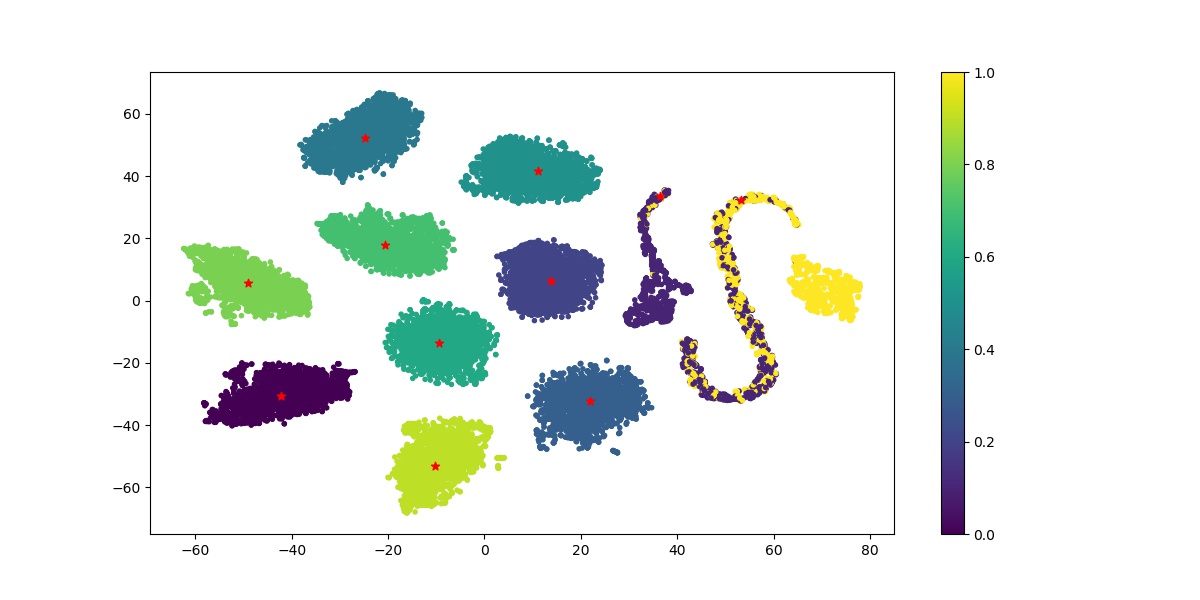


图4.1.2.1

Test Set

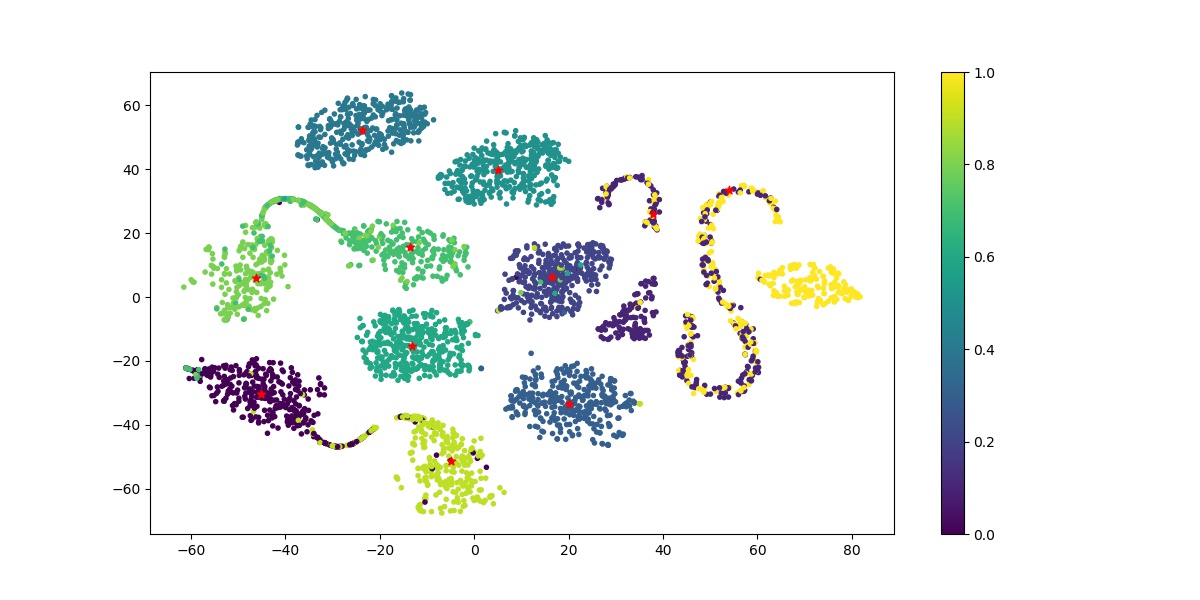


图4.1.2.2

从降维后的图中可以看出，第10类和第0类距离很近，而第8类和第7类距离较近 (在测试集上有所体现)，由此，模型很容易在预测这两类数据时出错，这便解释了第10类和第8类的准确率相对较低的情况。

#### 4.2 Zero Shot Learning方案

##### 4.2.1 1类未知样本

选取第2类为未知样本，其余10类为已知样本

判断为未知样本的准确率为

0.9951690821256038

判断已知样本并正确识别分类的准确率：

0.8384345208228801

##### 4.2.2 2类未知样本

选取第2类和第5类为未知样本，其余9类为已知样本

判断为未知样本并正确识别分类的准确率为

Class (2):0.99019608

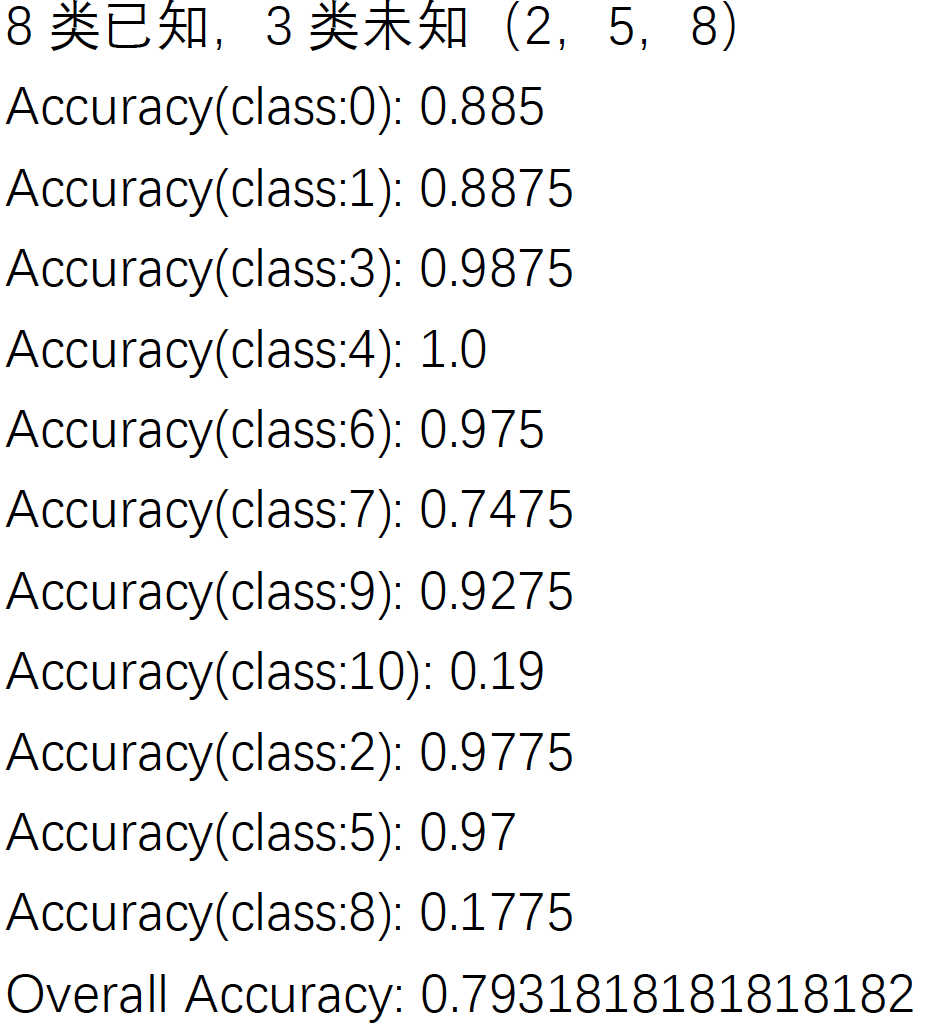
Class (5):0.98536585

判断已知样本并正确识别分类的准确率：

0.8207705192629816

#### 4.3 Few Shot Learning方案

##### 4.3.1识别准确率



此时，仍然在第10类和第8类的准确率上稍微不太令人满意，不过相对于Supervised Learning方案的上界准确率，并未发生根本性的暴跌。

### 潜在问题

经过大量实验，Zero Shot Learning的表现结果相对有些不稳定，其准确率和输入测试样本的顺序有关，在80%的情况下可以达到当前的准确率，而有20%的情况下，准确率可能有较大幅度的退化。

### 6. 优化方法

#### 6.1 模型优化

应用至实际中时，会根据信号数据的规模对模型进行调整，如果数据量较大，将考虑更换Backbone为ResNet 18。

#### 6.2 距离优化

在未知类别只有单个样本时，可能需要重新定义出更合理的距离计算公式

#### 6.3 损失函数优化

在Loss Function中添加新项及其权重，如添加Focal Loss等以优化那些表现相对欠佳的类别的准确率。（结果还不是没有用上Focal Loss）