



# 基于循环神经网络的语音 识别实验

指导老师：凌震华

实验助教：杜荟鹏



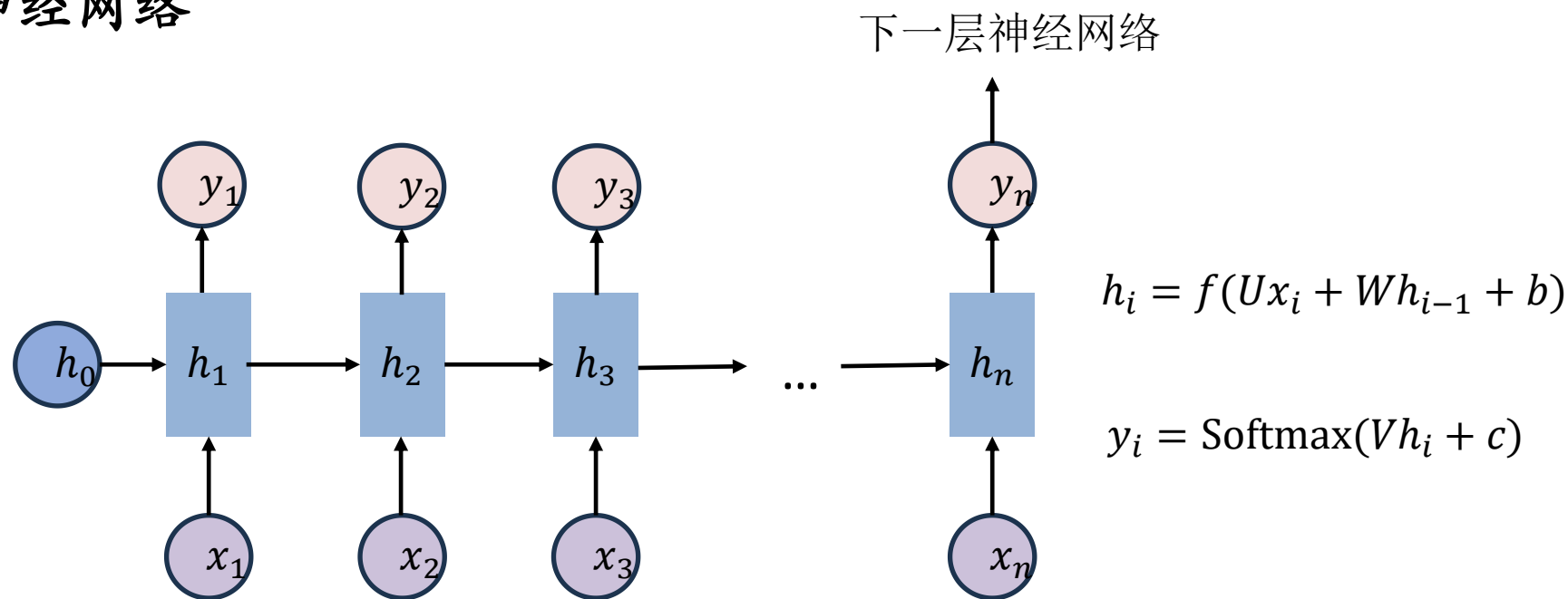
# 实验目的

- 理解循环神经网络 (Recurrent Neural Network, RNN) 的运算原理和过程
- 掌握自动语音识别 (Automatic Speech Recognition, ASR) 的概念和原理



# 实验原理

## 循环神经网络

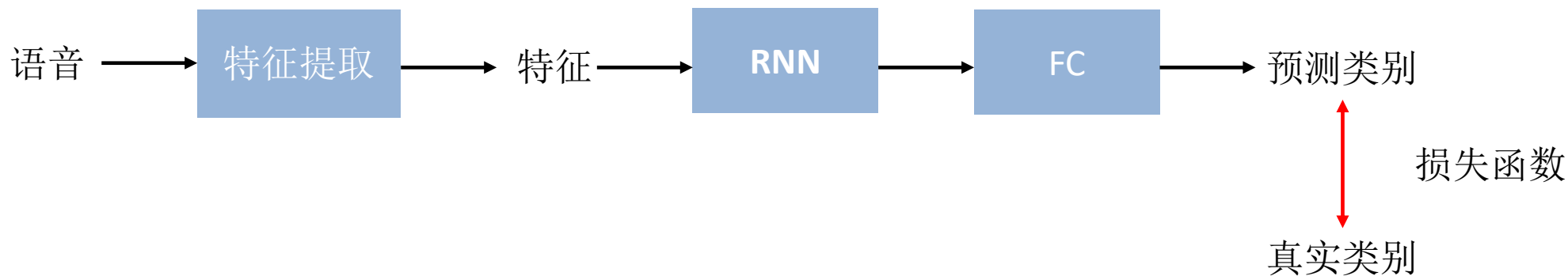


- 对于输入序列  $\mathbf{x} = [x_1, x_2, \dots, x_n]$ ，经过RNN的处理得到输出过程如图所示
- 此次实验样例代码中给出的循环神经网络，是使用最后一个时间步的输出作为下一层网络的输入



# 实验原理

## 命令行识别

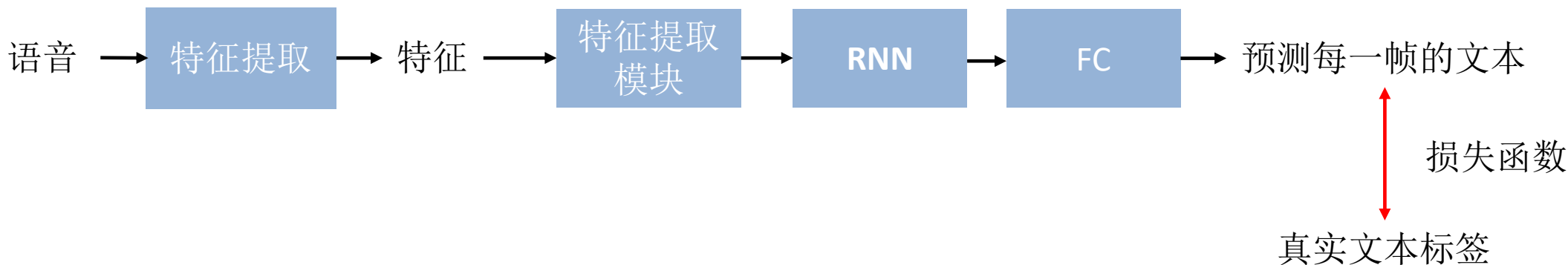


- 语音先经过数字信号处理中的方法提取特征，形成矩阵形式输入神经网络
- 特征经过RNN处理再经过全连接（Fully Connected，FC）层映射到类别维度
- 预测类别和真实类别使用交叉熵损失函数进行训练



# 实验原理

## 端到端语音识别



- 通样地，语音先经过数字信号处理中的方法提取特征
- 特征经过特征提取模块处理后，再送入RNN处理再经过全连接层
- 和命令词识别不同的是，此任务对于特征的每一个时间步（帧）都会预测一个类别
- 由于对齐问题，预测文本和真实文本使用Connectionist Temporal Classification损失函数进行训练



# 实验内容（第一周）



## 命令行识别

- 对于speechcommands数据集，进行一个语音分类任务。即输入语音在有限类别的词汇中找出其对应的词。
- 以正确率为指标，要求准确率达到20%以上。
- 补全模型代码，提交集群生成的.out 文件，并根据此.out 文件里的实验数据分析结果。



# 实验内容（第二周）



## 命令行识别进阶

- 在第一周实验的基础上做出改进。
- 以正确率为指标，要求准确率达到80%以上。
- 补全模型代码，提交集群生成的.out 文件，并根据此.out 文件里的实验数据分析结果。



# 实验内容（第三、四周）



## 端到端语音识别

- 实现端到端语音识别，即输入一段语音，模型能够生成其对应的文本。
- 补全指标测量代码，损失函数代码，完成模型的训练。
- 词错误率指标达到40%以下，提交集群生成的.out 文件，并根据此.out 文件里的实验数据分析结果。





# 评分细则

- 代码完整性 20%
- 实验结果与分析 50%
- 思考题 30%
- \*扩展实验 10%（工作量较大，不鼓励所有人都做此部分）



# 注意事项

- 阅读实验手册和观看视频会对实验起到一定的帮助
- 代码出现bug可以询问助教





谢谢!

