

## 一. 单选题（共10题，37分）

1. LSTM(长短期记忆网络)相比普通RNN的核心改进是：

- A. 使用更大的批量训练数据
- **B. 引入了门控机制和记忆细胞**
- C. 增加了更多的隐藏层
- D. 删除了反向传播过程

2. 以下哪种任务不适合用RNN变体(如LSTM)处理？

- **A. 图像分类(如ResNet)**
- B. 语音识别
- C. 时间序列预测
- D. 机器翻译

3. 门控机制(如GRU/LSTM)主要解决了普通RNN的什么问题？

- **A. 梯度消失/爆炸**
- B. 无法处理高维输入
- C. 需要大量标注数据
- D. 计算速度慢

4. 在序列生成任务中，束搜索(beam search)的作用是：

- A. 减少模型参数量
- B. 加速模型训练
- C. 动态调整学习率
- **D. 生成多个候选序列并保留最优路径**

5. 双向RNN的最终隐藏状态是：

- A. 随机选择其中一个方向
- B. 只保留前向隐藏状态
- C. 两者的平均值
- **D. 前向和后向隐藏状态的拼接**

6. 深层神经网络通过堆叠多个RNN层是为了：

- A. 仅用于图像分类任务
- B. 减少训练时间
- C. 避免使用梯度下降
- **D. 提取更高层次的时序特征**

7. GRU的更新门(update gate)的数学形式通常为：

- **A. Sigmoid函数**
- B. ReLU函数
- C. Tanh函数
- D. Softmax函数

8. LSTM中，遗忘门(forget gate)的作用是：

- **A. 决定从记忆细胞中丢弃哪些信息**

- B. 计算损失函数
  - C. 直接输出预测结果
  - D. 生成候选记忆细胞
9. GRU(门控循环单元)中，重置门(reset gate)的主要作用是：
- A. 计算候选隐藏状态
  - B. 直接输出最终预测结果
  - **C. 决定丢弃多少过去的信息**
  - D. 控制隐藏状态的更新幅度
10. 双向循环神经网络(BiRNN)适用于以下哪种场景？
- A. 无法处理时序数据
  - **B. 需要同时利用过去和未来信息的任务(如机器翻译)**
  - C. 只能处理固定长度的序列
  - D. 仅支持单层网络结构

## 二. 多选题（共10题，37分）

11. 双向RNN的局限性包括：
- A. 无法用于语言模型
  - **B. 不能实时处理流式数据(如实时语音输入)**
  - **C. 必须已知完整序列后才能计算**
  - D. 计算复杂度低于单向RNN
12. 机器翻译中，编码器-解码器架构通常包含：
- A. 无需注意力机制
  - **B. 解码器逐步生成目标序列**
  - C. 必须使用双向RNN
  - **D. 编码器将输入序列压缩为上下文向量**
13. GRU的候选隐藏状态计算依赖于：
- **A. 前一时刻隐藏状态**
  - B. 学习率
  - **C. 当前输入**
  - **D. 重置门**
14. 序列到序列学习中，束搜索(beam search)与贪心搜索的区别是：
- A. 两者计算复杂度相同
  - B. 束搜索一定比贪心搜索效果好
  - **C. 束搜索保留多个候选路径**
  - **D. 贪心搜索每一步选择概率最高的词**
15. 以下哪些技术可用于改善RNN训练？
- **A. 梯度裁剪**
  - **B. 权重初始化(如Xavier)**

- C. 增加批量大小至整个数据集
- **D. 使用Adam优化器**

16. 以下哪些是RNN变体的常见应用场景？

- **A. 文本生成**
- B. 图像风格迁移
- **C. 视频动作识别**
- **D. 股票价格预测**

17. 深层RNN可能遇到的问题包括：

- **A. 训练速度变慢**
- B. 无法处理序列数据
- **C. 梯度消失**
- **D. 过拟合**

18. 关于门控机制的描述正确的是：

- **A. 可以动态控制信息流动**
- **B. 通过Sigmoid函数输出0~1之间的值**
- C. 与残差连接(ResNet)功能相同
- D. 仅在LSTM中使用

19. 关于GRU和LSTM的正确描述是：

- **A. 两者均能缓解梯度消失问题**
- **B. GRU比LSTM参数更少**
- **C. GRU没有记忆细胞(cell state)**
- **D. LSTM有输入门、输出门和遗忘门**

20. LSTM的记忆细胞(cell state)的特点是：

- **A. 可以长期保留信息**
- B. 只能存储二进制值
- C. 直接作为最终输出
- **D. 通过门控机制选择性更新**

21. (简答题) 为什么双向RNN不适合实时任务(如实时语音识别)？

因为双向RNN需要访问整个序列的未来信息，无法边输入边输出，导致无法满足实时任务的低延迟需求。

22. (简答题) 序列到序列学习(seq2seq)模型训练过程和预测过程有什么不一样？

序列到序列(seq2seq)模型在训练过程中采用“教师强制”(teacher forcing)，即每一步解码器的输入是真实的上一个目标词；而在预测过程中，解码器只能使用自己上一步生成的词作为输入，不能看到真实答案，因此预测过程更容易出现误差传播。

23. (简答题) 编码器-解码器架构在机器翻译任务中是如何工作的?(结合RNN变体说明)

在机器翻译中，编码器-解码器架构通过RNN变体（如LSTM或GRU）将源语言序列编码成一个上下文向量，捕捉整体语义信息；然后解码器使用另一个RNN变体，从该上下文向量出发，逐步生成目标语言的词语序列，实现从源语言到目标语言的映射。

24. (简答题) 解释LSTM中遗忘门和输入门如何协同工作更新记忆细胞。

在LSTM中，遗忘门决定当前时刻要从上一个记忆细胞中“遗忘”多少旧信息，而输入门控制要将多少新的候选信息写入细胞；两者协同作用，通过加权融合旧记忆和新信息来更新记忆单元，使模型既能保留长期依赖，又能灵活引入新内容。