

实验题目	循环神经网络			实验日期	2022.5.24
班级	1903104	学号	1190300321	姓名	郑晟赫

CS32262 模式识别与深度学习实验

实验报告

一、实验目的（介绍实验目的）

1. 利用 Pytorch 自己实现 RNN、GRU、LSTM 和 Bi-LSTM
2. 利用上述四种结构进行文本多分类
3. 任选上述一种结构进行温度预测

二、实验环境（介绍实验使用的硬件设备、软件系统、开发工具等）

1. 硬件设备：CPU：i7-9750H；GPU：NVIDIA GeForce GTX 1650；16G RAM；512GHD Disk
2. 软件系统：Windows 10；Python 3.7.10；PyTorch 1.9.0
3. 开发工具：PyCharm

三、实验内容（介绍实验过程、设计方案、实现方法、实验结果等）

这一部分主要介绍实验实现过程以及结果分析。

1. 数据集分析

实验涉及两个数据集，分别是文本分类数据集和天气数据集。

文本分类数据集涉及 10 个类别（书籍、平板、手机、水果、洗发水、热水器、蒙牛、衣服、计算机、酒店）的网购评论数据。在分类过程中根据用户评论信息推测是该条信息属于哪一个类别。数据集划分中共 62774 条数据，按给定顺序编号 $i=1, 2, \dots, 62774$ ， $i\%5=4$ 做验证集， $i\%5=0$ 做测试集，其余作为训练集。

天气数据集每 10 分钟记录一次气压、气温、风速等天气数据，共 420551 条记录，历时八年（2009~2016）；取前六年做训练集，后两年做测试集；对测试集分段，每七天作为一段，对于每一段而言，给定前五天的数据，预测后两天的数据。

2. 代码编写

2.1 数据读取

自定义 dataset 类，使用 PyTorch 中的 Dataloader 读取。

2.2 网络搭建

本实验搭建的是四种网络，分别是 RNN，LSTM，GRU，Bi-LSTM。四种网络本质上都是循环神经网络，在一个 cell 输出的数据均需要输入下一个 cell，且每个 cell 之间共享参数。

RNN 图解可见图 1，实现过程中主要分为 cell 内部的一个线性层和一个非线性层。在处理输出的时候，将最后一个 cell 的输出视为最后的输出结果，由于本实验第一个任务要处理的是 10 个类别的分类，因此在最后一个输出之后接一个输出维度为 10 的全连接层。因此 RNN 中需要训练的参数就是一个全连接层中的矩阵和偏置，以及最后输出的全连接层，相较于后续实现的 LSTM 等网络来说，相对简单，因此训练速度较快，对于长距离依赖问题也不能很好的解决。因此 RNN 在较长句的处理中性能较差。

实验题目	循环神经网络			实验日期	2022.5.24
班级	1903104	学号	1190300321	姓名	郑晟赫

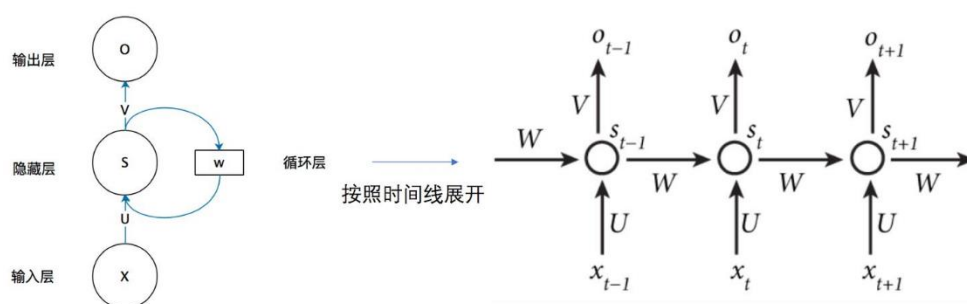


图 1. RNN 结构图

LSTM 相较于 RNN 的不同在于 LSTM 引入了遗忘门、输入门和输出门。其中遗忘门的引入需要处理的就是为了解决 RNN 处理长序列容易出现的梯度消失或者梯度爆炸。遗忘门可以根据输入和细胞状态决定是否丢弃一些信息，从而保证了长距离依赖。LSTM 的结构可见图 2。

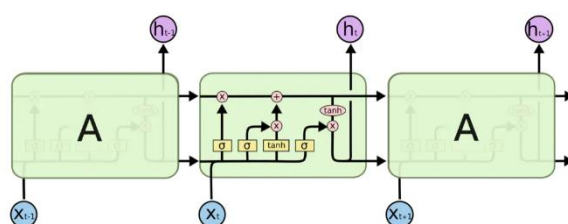


图 2. LSTM 结构图

双向 LSTM 事实上就是两个 LSTM，一个是从前到后，一个是从后向前，将两个 LSTM 的输出连接之后作为输出结构。双向 LSTM 结构可见图 3。

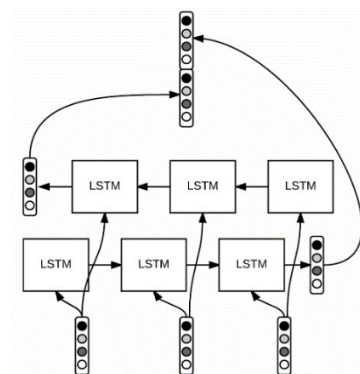


图 3. 双向 LSTM 结构图

GRU 是 LSTM 的一个变体，在实验结果与 LSTM 相近的同时简化计算过程。从而使得 GRU 训练时间相较于 LSTM 大大缩短。GRU 的结构和计算公式可见图 4。

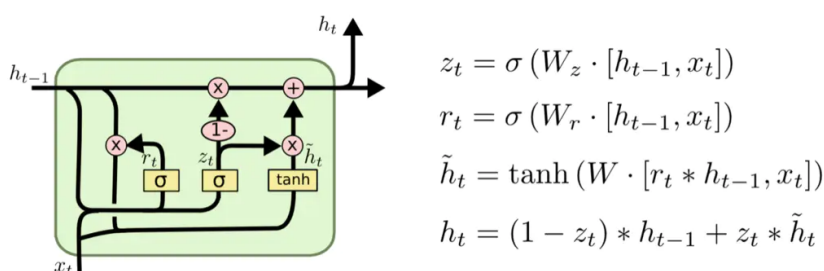


图 4. 双向 LSTM 结构图

实验题目	循环神经网络			实验日期	2022.5.24
班级	1903104	学号	1190300321	姓名	郑晟赫

2.3 优化器

两个任务均采用 Adam 优化器。

2.4 损失函数

由于任务 1 是多分类问题，因此任务 1 采用交叉熵作为损失函数，直接调用 PyTorch 中的函数。而由于任务 2 是一个回归问题，因此任务 2 采用 L1 loss 作为损失函数，计算预测值与真实值的绝对误差之和。

3. 实验验证

3.1 实验设置

Batch_size: 64

Epoch: 2

学习率: 0.001

3.2 实验结果

任务 1 各算法性能对比如表 1 所示。

表 1 各算法性能比较

算法	准确率 acc	F1 score	训练时间/s
RNN	0.7068	0.5703	88.0450
LSTM	0.8120	0.7629	298.100
BiLSTM	0.8388	0.8026	488.852
GRU	0.8277	0.7439	130.807 s

观察表中的测试结果可以发现，在四种算法中，双向 LSTM 的性能最好，这是因为双向 LSTM 充分利用了正反两个方向的信息，但是与之相对的，训练的速度也是这几种方法中最慢的；而 GRU 性能与 LSTM 相近，但是训练速度大大上升，这是因为 GRU 简化了每个 cell 内部的计算方式，因此训练速度加快。

任务 2 采用双向 LSTM，训练结果预测效果与真实结果比较选取其中一段展示如图 4 所示。其中红线为预测效果，而蓝线为真实气温。可以发现在一定程度上预测结果和真实气温接近。说明双向 LSTM 的能力相对来说还是比较好的。

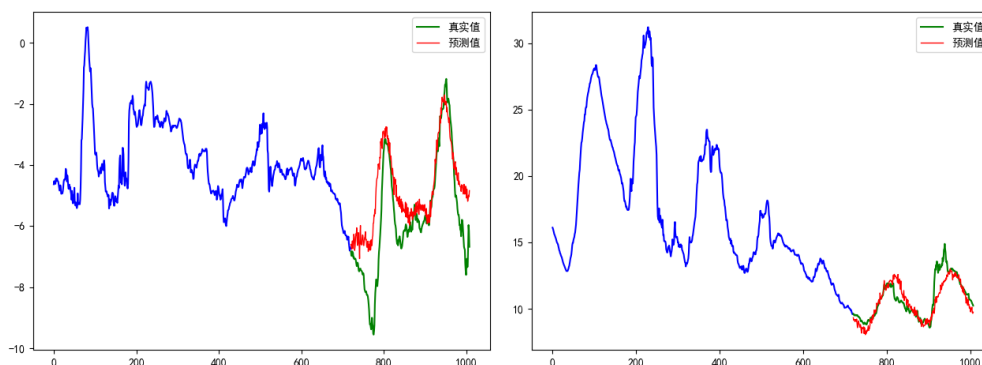


图 4 任务 2 效果展示

任务 2 的各项训练结果如表 2 所示。

表 2 任务 2 训练性能

	平均误差	中位误差
图 4 左图	0.823	0.546
图 4 右图	0.612	0.361

实验题目	循环神经网络			实验日期	2022.5.24
班级	1903104	学号	1190300321	姓名	郑晟赫

可以发现任务2的性能还有较大的提升空间，这可能是因为没有充分利用到数据集中提供的多种测量指标。

四、实验结论（总结实验发现及结论）

本次实验实现了四种比较常用的循环神经网络，经过实验发现 LSTM 的性能相对于 RNN 有比较明显的提升，但是训练速度也下降的比较明显。而 GRU 则进行折中，在性能变化不明显的前提下，简化了计算过程，加速了训练过程。经过本次实验，对于循环神经网络有了更深的认识。