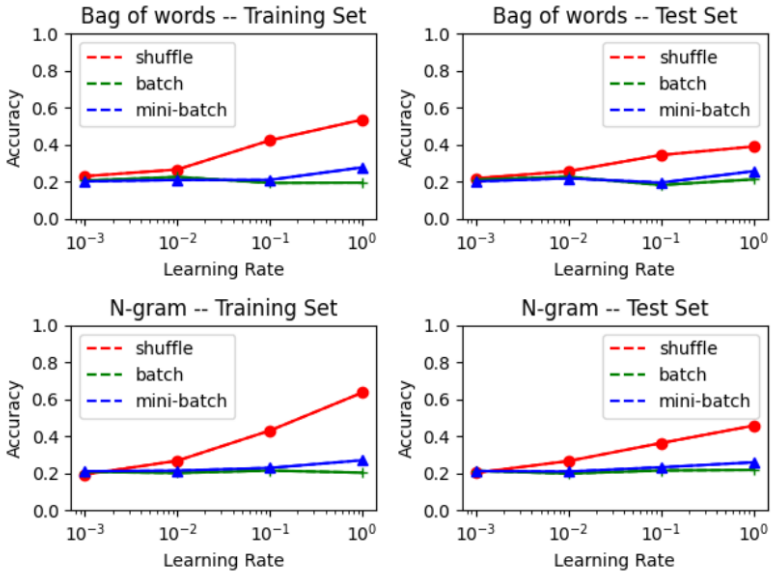
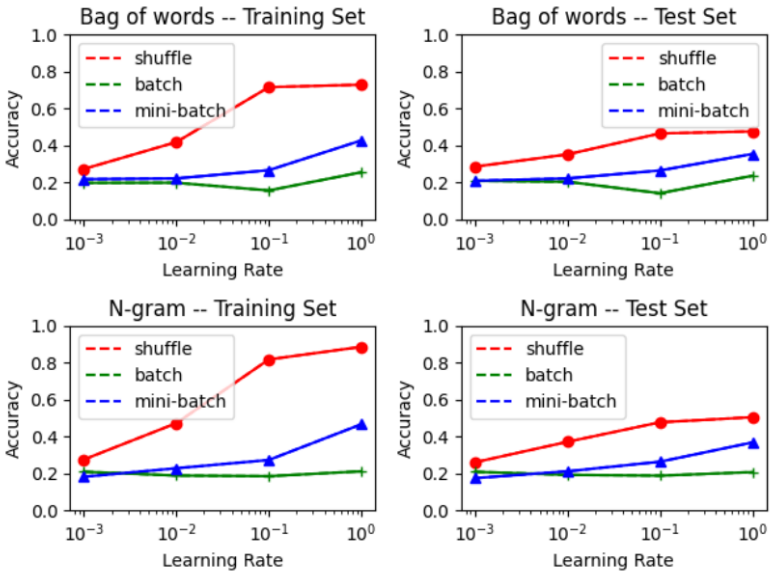
新生代码任务

- NLP任务
  - 任务一：基于机器学习的文本分类
    - 实验设置：
      - 样本个数：10000
      - 训练集占比：80%
      - 学习率：[0.001, 0.01, 0.1, 1]
      - Mini-Batch大小：样本的1%即100
      - 回归模型：Softmax回归（Logistic回归通常应用于二分类）
    - 实验结果：
      - 一、计算梯度次数总数为10000的实验



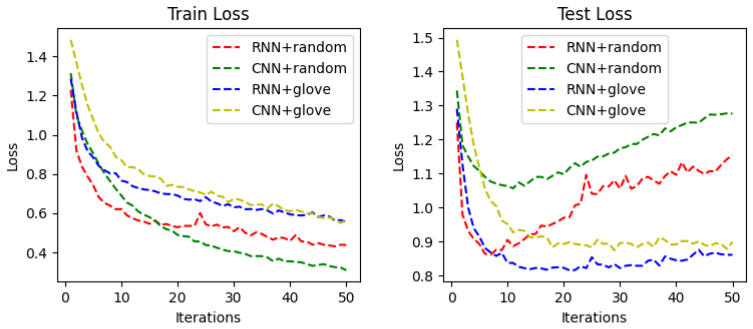
        - N元模型明显要优于词袋模型，这是因为N元模型相较于词袋模型充分考虑了句子中的词序。
        - 较小的学习率无法使模型收敛，相比之下较大的学习率表现较好。
        - mini-batch和batch的表现几乎相同，而shuffle的表现最好（可能是因为N元模型和词袋模型所构建的特征通常是高维稀疏的，batch和mini-batch在更新时会遍历较多样本，从而使梯度方向容易被高频词所主导，最终导致了较差的预测精度。）
      - 二、计算梯度次数总数为100000的实验

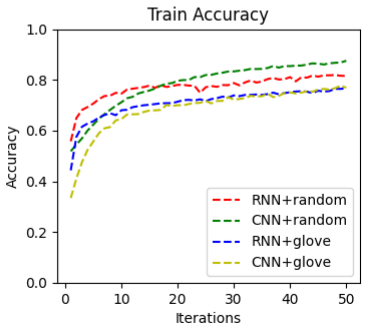


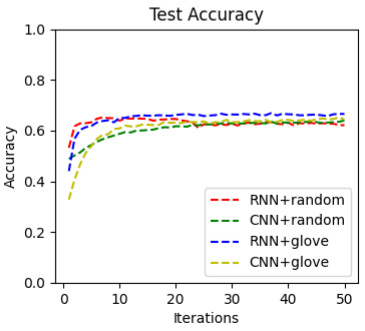
- 实验结果和实验一的结论基本相符。
- 训练集的准确率已超过80%，但这意味着随着训练次数的增加，模型在训练集上的拟合效果也变得越来越好。
- 无论是实验一还是实验二，测试集的最高准确率都在50%附近。
- 这意味着实验二所增加的梯度计算次数最终仅导致了模型在训练集上的过拟合，并未对模型的泛用性产生提升效果。

任务二：基于深度学习的文本分类

- 实验设置：
  - 样本个数：约150000个
  - 训练集占比：70%
  - 特征抽取模型：CNN，RNN
  - 嵌入初始化方法：随机初始化，GloVe预训练模型初始化
  - 学习率：0.001
  - 嵌入维度：50
  - Batch Size：500
- 实验结果：
  - 总体效果：

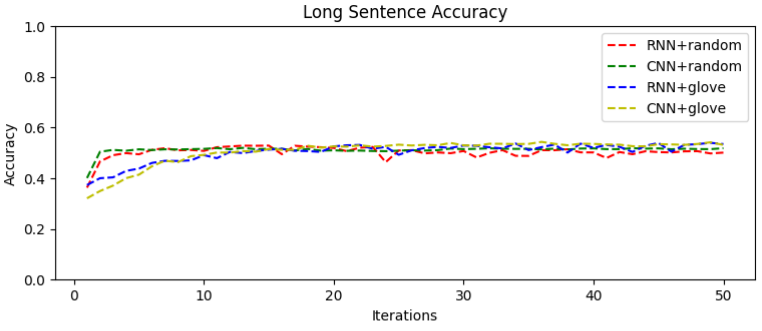


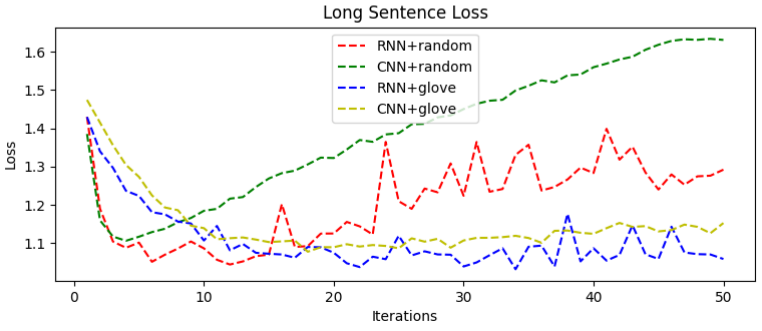




- 可以看到RNN在测试集的准确率（最大值）比CNN都要高，且测试集的损失值（最小值）也要比CNN的要低。
- 在同种模型下，GloVe初始化也要比随机初始化的效果好，即在测试集准确率大、测试集损失值小。

- 长句子分类效果：

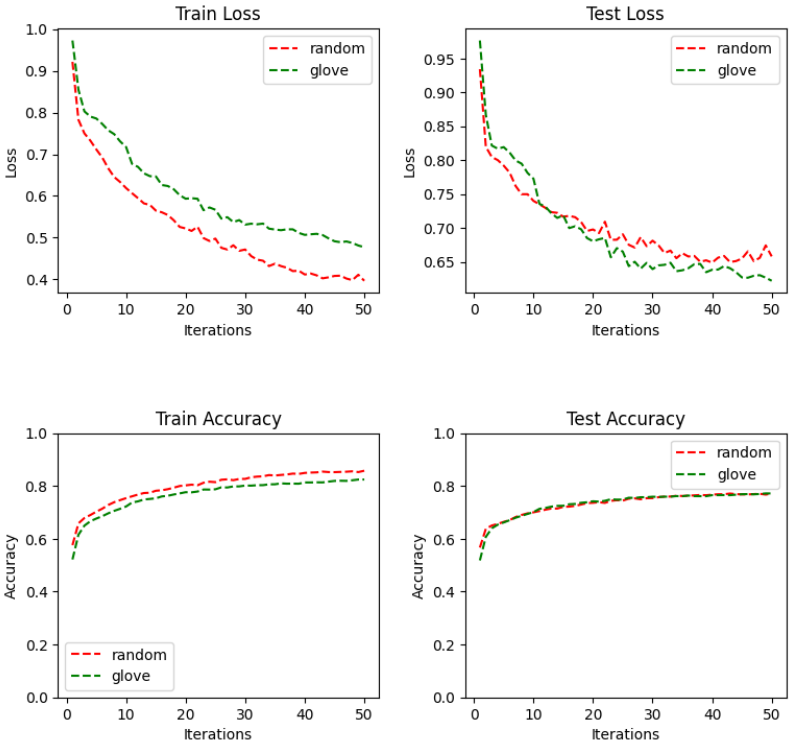




- 特别关注测试集中单词数大于20的句子的损失值和正确率

        - RNN的效果并不比CNN好，而且无论是CNN还是RNN，长句子的情感分类准确率也只有55 % 左右，比总体的平均正确率低了约 10%。

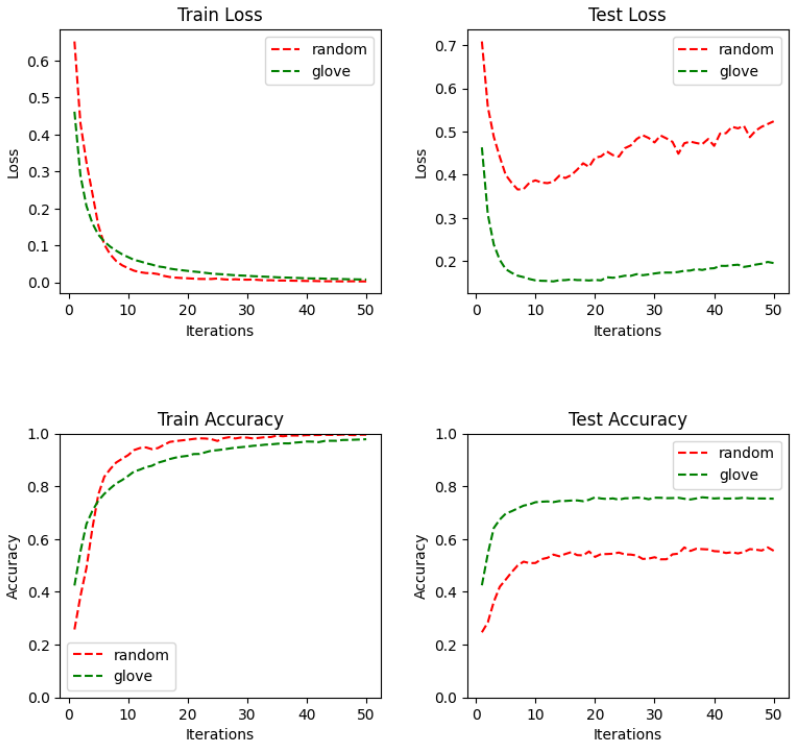
- **任务三：基于注意力机制的任务匹配**
    - 实验设置：
        - 样本个数：约550000
        - 训练集占比：70%
        - 基线模型：ESIM
        - 嵌入初始化方法：随机初始化，GloVe预训练模型初始化
        - 学习率：0.001
        - 嵌入维度：50
        - Batch Size：1000
    - 实验结果：



    - 实验效果整体较好，预测准确率接近0.8，反映出LSTM对于处理长距离依赖和提取上下文特征具有重要作用，在句子匹配中能够更好地捕捉语义信息。

- **任务四：基于LSTM+CRF的序列标注**
    - 实验设置：
        - 训练集与测试集：train.txt，test.txt
        - 嵌入初始化方法：Random / GloVe
        - 学习率：0.001
        - 嵌入维度：50
        - Batch_size：32

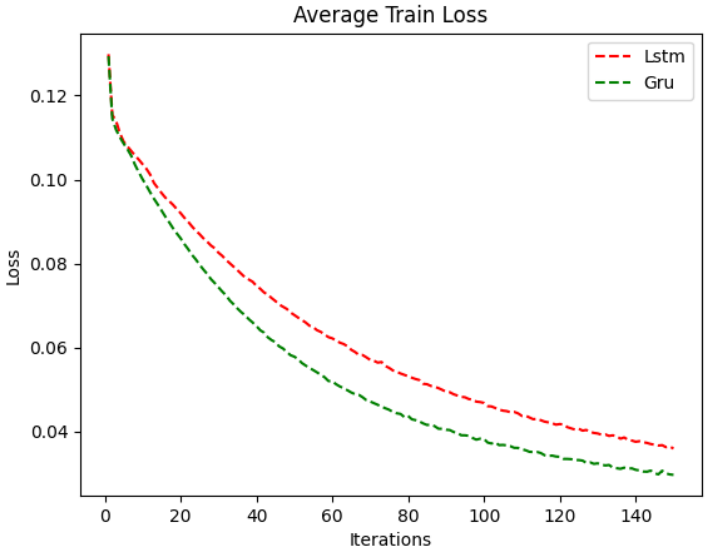
    - 训练轮次：50
- 实验结果：



    - 可以看到GloVe初始化比随机初始化明显更好，随机初始化标注正确率随机最多只有55%，而GloVe最高能到76.23%。
    - CRF作为无向图模型，通过建模标签之间的依赖关系来提高预测的准确性，显著提升了序列标注任务的性能

任务五：基于神经网络的语言模型

- 实验设置：
    - 数据集：poetryFromTang.txt（少部分诗句夹杂有英文字母，已采用人工方式清洗，最终得到163首古诗）
    - 嵌入初始化方法：Random（数据集中语言为中文，目前暂无较合适的预训练模型，因此采用随机初始化）
    - 学习率：0.001
    - 嵌入维度：50
    - Batch_size：1（数据集较小，仅包含136条数据）
    - 训练轮次：150
- 实验结果：



- 可以看出GRU的效果明显比LSTM效果更好

- 诗句展示
  - 生成固定诗句（0向量初始化）
    - 水亭凉气多，池养右军开。
    - ，池养右军鹅。
    - ，自怜越成周。
    - 。
  - 生成随机诗句（随机向量初始化）
    - 排僚愁自多，移宵始半堂平
    - 至竟息亡开，离下问前叨。
    - 山藏伯禹穴，城压伍胥涛。
    - 今来青鸟东，无怜越能心。
    - 水中作黍，栖栖深名名。
    - ，枯涧为君君不。
  - 生成固定藏头诗（0向量初始化）
    - 春江山路江天，天子青台深。
    - 夏与君不见，唐华各驰骤。
    - 秋月山来溪天，天子青溪寒。
    - 冬与君不见，行乐截为极。
  - 生成随机藏头诗
    - 春儋共开中未休心。
    - 夏消牝嫌峒小麦熟，背勇伏囚拘。
    - 秋添甫院学皂穴，荣期以穷自。
    - 冬升其冷。

- 可以看出模型在生成藏头诗时的效果比生成随机诗句更好