

2025 年春季《自然语言处理》期末综合设计

一、基于深度模型的情感分析和逻辑推理任务

1. 任务一：情感分类

- (1) 任务介绍：基于 IMDB-10 数据集构建深度模型实现十分类情感分类。
- (2) 数据集：IMDB-10 是电影评论的十分类情感（评分 1-10）标注的数据集。
- (3) 评价指标：F1 值（macro-F1，主要）、准确度（Accuracy）、均方根误差（Root Mean Squared Error, RMSE）
- (4) 数据样本

Text	Label
i excepted a lot from this movie , and it did deliver . <sssss> there is some great buddhist wisdom in this movie . <sssss> the real dalai lama is a very interesting person , and i think there is a lot of wisdom in buddhism . <sssss> the music , of course , sounds like because it is by philip glass . <sssss> this adds to the beauty of the movie . <sssss> whereas other biographies of famous people tend to get very poor this movie always stays focused and gives a good and honest portrayal of the dalai lama . <sssss> all things being equal , it is a great movie , and i really enjoyed it . <sssss> it is not like taxi driver of course but as a biography of a famous person it is really a great film indeed .	10
i agree , i also do not like paris , i was there and it was dirty , criminal , hasty and expensive . <sssss> the film tries its best to show paris in its best colors . <sssss> out of 16 episodes , the half is terrible weak , dull , boring and largely uninteresting . <sssss> often , they lack cohesion and they seem to go simply nowhere . <sssss> the worst ones are the zany vampire ditty with elijah wood and several more . <sssss> the one with nick nolte begs for continuation but fails too . <sssss> the other half is a winner , the parts with natalie portman , steve buscemi and then bob hoskins are brilliant and great ... and then there is the best one about the poor killed african american guitar man who was a victim of the white gang . <sssss> this short episode stand out and is the gem alone .	5
crouching tiger , hidden dragon is such a bad movie . <sssss> the love stories don ' t make sense at all . <sssss> the cute princess is so annoying . <sssss> the violence does not make any sense either . <sssss> all in all a perfectly pointless movie . <sssss> why should we care about these stupid characters who act in such meaningless ways ?	1

【提示】

- 该数据集中，大部分文本长度超过 512 个 Token，在进行数据处理时，可以只取其中 512 个 Token 作为模型输入

- 【<sssss>】字符串表示文本中的句子分割符，可在数据处理时将其过滤掉
- 提供的数据集中包括用户 ID（第 1 列）和产品 ID（第 2 列）信息，可在数据处理时将其忽略

2. 任务二：蕴含识别

（1）任务介绍：基于 SNLI 数据集构建深度模型实现自然语言推理任务。自然语言推理是通过已知的前提文本来确定能否推断目的文本。自然语言推理也可以理解为识别文本的蕴含的任务。

（2）数据集：Stanford Natural Language Inference（SNLI）是一项自然语言推理任务数据集。数据样本为一对文本，分别为前提和假设文本，标签分为 3 类，分别是：标签 0（蕴含，*Entailment*）-可以通过前提推断出假设；标签 2（矛盾，*Contradiction*）-可以推断出与假设相反；标签 1（中立，*Neutral*）-所有其他情况。

（3）评价指标：F1 值（macro-F1，主要）、准确度（Accuracy）

（4）数据样本

Text 1 (premise)	Text 2 (hypothesis)	Label
A person on a horse jumps over a broken down airplane.	A person is training his horse for a competition.	1
A person on a horse jumps over a broken down airplane.	A person is at a diner, ordering an omelette.	2
A person on a horse jumps over a broken down airplane.	A person is outdoors, on a horse.	0

【提示】

- 该数据集中 Text 1 和 Text 2 存在先后顺序约束
- 提供的数据集为 parquet 格式，请自行查阅资料实现数据读取，或从 huggingface 官网（<https://huggingface.co/datasets/stanfordnlp/snli>）进行读取

二、总体要求

1. 本次实验需要利用 PyTorch 深度学习框架及第三方自然语言处理工具库完成情感分类及蕴含推理两个任务，包括：实现数据的下载及加载、预训练词向量的下载及加载、模型的搭建、模型的训练及推理。主要考查学生综合利用现有知识库进行任务数据的获取与处理、模型的搭建与训练，结果测试与评估；

2. 分别实现（1）基于预训练 GloVe（网络获取）作为词向量特征的 BiLSTM 模型；（2）基于预训练 BERT 作为词向量特征的 BiLSTM 模型；（3）基于预训练 BERT 的下游任务精

调模型。

3. 要求画出模型图，实验结果截图，源代码 (*.py) 与电子实验报告 (姓名_学号_实验报告.pdf) 进行打包 (姓名_学号_期末考核.zip 或 rar) 提交至雨课堂。纸质实验报告于 2025 年 6 月 30 日之前统一交给周洲有同学，并送于信息学院 1518 室。

4. 基于训练集优化网络、验证集用于超参选取、测试集用于结果预测。

5. 对实验过程中的问题和收获进行总结。

三、题目评分依据

基于 python 实验平台进行实验，提交的实验报告以个人为单位，并实现以下功能（100 分）：

- （1）所设计的系统正确性（10%）
- （2）所设计的系统功能完整，可以实现需要的相关功能（40%）
- （3）所编写的实验报告是否整洁，缩进和大小写是否正确(10%)
- （4）所实现的结果是否正确（30%）
- （5）程序报告的版面和书写质量（10%）