
深度学习与自然语言处理第五次作业

李文雯 ZY2203106

作业要求

任意选取 3~4 个目前前沿语言大模型（不限中文或者英文），通过提示工程的方法来检验和对比不同模型下游任务上的性能，可选择 3~5 个不同的自然语言下游任务来进行测试。

摘要

随着深度学习技术的不断发展，自然语言处理领域涌现出了众多强大的预训练模型。这些模型通常采用 Transformer 架构，经过大量无标签文本数据的预训练，进而在各种下游任务中进行微调。本文将重点比较 ChatGPT、Alpaca-LoRA 和文心一言三个模型在机器翻译、情感分析、代码生成和数学推理四种任务上的性能。

实验内容

1. 机器翻译

例子：

自然语言处理是计算机科学领域与人工智能领域中的一个重要方向。它研究能实现人与计算机之间用自然语言进行有效通信的各种理论和方法。NLP 是一门融语言学、计算机科学、数学于一体的科学，涉及自然语言，所以它与语言学的研究有着密切的联系，但又有重要的区别。但 NLP 并不是一般地研究自然语言，而在于研制能有效地实现自然语言通信的计算机系统，特别是其中的软件系统，因此又和计算机科学密不可分。

ChatGPT 的回复：

Natural Language Processing (NLP) is an important field in computer science and artificial intelligence. It focuses on studying various theories and methods that can achieve effective communication between humans and computers using natural language. NLP is a scientific discipline that integrates linguistics, computer science, and mathematics, involving natural language. Therefore, it has close ties with linguistic research, but also has significant differences. NLP is not just the study of natural language in general, but rather the development of computer systems, especially software systems, that can effectively implement natural language communication. Hence, it is closely related to computer science as well.

Alpaca-LoRA 的回复：

Output

Natural Language Processing (NLP) is an important direction in both the Computer Science and Artificial Intelligence fields. It studies how to effectively communicate between humans and computers using various theories and methods. NLP is closely related to linguistics, but it also has important differences. However, NLP is not just researching natural language, but also developing computer systems that can effectively implement natural language communication, especially software systems. Therefore, it is also closely related to computer science and cannot be easily distinguished.

文心一言的回复：

Natural language processing (NLP) is an important direction in the fields of computer science and artificial intelligence. It studies various theories and methods that enable effective communication between humans and computers in natural language. As a discipline that integrates linguistics, computer science, and mathematics, NLP involves the study of natural language, making it closely related to linguistics, but with important distinctions. However, NLP does not generally study natural language per se, but rather focuses on developing computer systems, especially software systems, that can effectively implement natural language communication. Therefore, it is closely related to computer science.

2. 情感分析

例子：

- (1) 这件衣服大小、款式以及价格都令我很满意；
- (2) 这部电影虽然时长很短、但剧情跌宕起伏、迷雾重重，值得一看。
- (3) 最近由于加班严重，导致身体健康出现了问题，以后要养成规律作息。

ChatGPT 的回复：

- (1) 这句话是积极、正面的意见，表达了用户对衣服的满意。
- (2) 这句话也是积极、正面的意见，虽然描述了电影时长不长，但是以其迷人的情节引起了观众的兴趣。
- (3) 这句话强调了用户注意身体健康与规律作息的重要性，是一种自我管理和体贴自己的态度。

Alpaca-LoRA 的回复：

Output

Sentence 1: Positive
Sentence 2: Negative
Sentence 3: Positive

文心一言的回复：

(1) 这件衣服大小、款式以及价格都令我很满意；

这句话表达了作者对这件衣服的满意程度，可以理解为积极的情感。

(2) 这部电影虽然时长很短、但剧情跌宕起伏、迷雾重重，值得一看。

这句话表达了作者对这部电影的评价，认为虽然时长较短，但剧情紧凑、情节复杂，值得一看。可以理解为积极的情感。

(3) 最近由于加班严重，导致身体健康出现了问题，以后要养成规律作息。

这句话表达了作者对加班导致身体健康问题的担忧，并提出了养成规律作息的建议。可以理解为消极的情感。

3. 代码生成

例子：用 python 写一段随机森林分类的代码

ChatGPT 的回复：

以下是使用Python中sklearn库实现随机森林分类的示例代码，其中主要包括数据处理、模型训练与测试三部分。

```
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
import pandas as pd

# 加载数据集
data = pd.read_csv("your_dataset_file_path.csv", header=None)

# 提取特征和类别标签
X = data.iloc[:, :-1]
y = data.iloc[:, -1]

# 将数据集分割为训练集和测试集
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

# 定义随机森林分类器
rfc = RandomForestClassifier(n_estimators=100, max_depth=5, random_state=42)
```

python 复制代码

```
# 在训练集上拟合随机森林分类器
rfc.fit(X_train, y_train)

# 在测试集上进行预测并计算准确率
y_pred = rfc.predict(X_test)
acc = accuracy_score(y_test, y_pred)
print("测试集准确率: {:.2f}%".format(acc*100))
```

在上述代码中，我们首先利用pandas读取数据集文件，并通过 `.iloc` 方法提取输入特征 `x` 和类别标签 `y`。然后，我们使用 `train_test_split()` 函数将数据集按比例分割成训练集和测试集。接着，我们定义了一个随机森林分类器，并通过 `fit()` 方法在训练集上进行拟合。最后，我们使用该分类器在测试集上进行预测，并计算出准确率。

需要注意的是，由于数据集不同，随机森林中的超参数可能需要调整以获得更好的效果。

Alpaca-LoRA 的回复：

Output

```
import numpy as np
from sklearn.ensemble import RandomForestClassifier

X = np.array([[1, 2, 3], [4, 5, 6], [7, 8, 9]])
y = np.array([1, 2, 3])

clf = RandomForestClassifier()
clf.fit(X, y)

print(clf.predict(X))
```

文心一言的回复：

python

[复制代码](#)

```
1 from sklearn.ensemble import RandomForestClassifier
2 from sklearn.datasets import make_classification
3 from sklearn.model_selection import train_test_split
4 from sklearn.metrics import accuracy_score
5
6 # 创建一个随机森林分类器
7 rf = RandomForestClassifier(n_estimators=100, random_state=42)
8
9 # 创建一个虚拟数据集
10 X, y = make_classification(n_samples=1000, n_features=10, n_informative=5, n_
11
12 # 将数据集分为训练集和测试集
13 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, rand
```

```
15 # 训练模型
16 rf.fit(X_train, y_train)
17
18 # 在测试集上进行预测
19 y_pred = rf.predict(X_test)
20
21 # 计算模型的准确率
22 accuracy = accuracy_score(y_test, y_pred)
23 print("Accuracy:", accuracy)
```

这段代码使用 `sklearn` 库中的 `RandomForestClassifier` 类来创建一个随机森林分类器，并使用 `make_classification` 函数生成一个虚拟数据集。然后，将数据集分为训练集和测试集，使用训练集来训练模型，并在测试集上进行预测。最后，计算模型的准确率并输出结果。

4. 数学推理

例子：

一把锁的密码，由三个数字组成。请根据下面五个条件推出开锁密码。

- 1.三位数 246，有 1 个号码正确，且位置正确。
- 2.三位数 258，有 1 个号码正确，但位置不正确。
- 3.三位数 692，有 2 个号码正确，但位置都不正确。
- 4.三位数 174，没有一个号码正确。
- 5.三位数 419，有 1 个号码正确，但位置不正确

ChatGPT 的回复：

根据以上五个条件，我们可以推断出开锁密码为 2 4 6。

解释如下：

1. 246 这个数字有 1 个号码正确，且位置正确。由此我们可以确定 2 是正确的数字，而且在第一个位置上。
2. 258 这个数字有 1 个号码正确，但位置不正确。结合第一条信息，我们知道正确的数字是 2，并且它不能是第二个数字，因此必须是第三个数字。
3. 692 这个数字有 2 个号码正确，但位置都不正确。这就意味着 6 和 9 都是正确的数字，但它们不能分别出现在第二个和第三个位置上，因为这样会与条件 2 相矛盾。因此，6 必须是第一个数字，而 9 需要出现在第二个数字位置或第三个数字位置。
4. 174 这个数字没有一个号码正确，因此 1、7 和 4 都不可能是正确的数字。
5. 419 这个数字有 1 个号码正确，但位置不正确。根据前面得出的 2、6 和 9 都是正确的数字，从剩下的数中选择一个满足条件就只剩下了 4 和 1 两个数字了。根据条件 1，4 和 6 不可能同时出现在正确的位置上，所以我们得到的是 1。于是 4 必须在第二个数字位置上。

Alpaca-LoRA 的回复：

Output

The unlock code for the lock is 246.

文心一言的回复：

根据条件 1，这个三位数的密码是 246。因此，我们可以排除 246 这个数字。

根据条件 2，我们可以推断出正确的数字是 5，但它的位置不正确。因此，我们可以排除数字 5。

根据条件 3，我们可以推断出正确的数字是 9 和 2，但它们的位置都不正确。因此，我们可以排除数字 9 和 2。

根据条件 4，我们可以推断出这个三位数的密码不可能是 174。因此，我们可以排除数字 1、7 和 4。

根据条件 5，我们可以推断出正确的数字是 4，但它的位置不正确。因此，我们可以排除数字 4。

综上所述，这个三位数的密码不可能是 1、2、4、5、7 或 9。因此，可能的数字组合是：26、28、62、68、82、86。但是，由于题目中没有提到数字 6 的情况，我们无法确定哪个数字组合是正确的。

实验结果

根据三大模型在不同下游任务上生成的结果，我们可以得出：

- (1) 在机器翻译任务中，ChatGPT、Alpaca-LoRA 和文心一言都能给出较为准确的汉译英结果，但 Alpaca-LoRA 模型的翻译速度相对于其他两个模型更慢一些，且流畅度不如其他两个模型。
- (2) 在情感分析任务中，ChatGPT 和文心一言都能很好地分辨出情感类型并给出合理的解释，Alpaca-LoRA 在分析消极情感时出现一些偏差。
- (3) 在代码生成任务中，ChatGPT 和文心一言生成的代码较为全面，包括 python 包的引入、数据集加载、训练、测试以及代码注释，而 Alpaca-LoRA 生成的代码更为简单，需要进行修改。
- (4) 在数学推理任务中，ChatGPT 和 Alpaca-LoRA 都没有给出正确的推理结果，文心一言的推理思路是正确的，但没有给出唯一的答案。

结论

1. 通过使用相同 prompt 对三个大模型的各方面性能进行测试，可以发现对于简单的机器翻译、代码生成任务中三个模型水平相当，Alpaca-LoRA 由于参数量相对较少而在情感分析任务中表现较差，三者数学推理题上的表现都不好，说明大模型目前还没有具备完善的数理逻辑。

2. 总体而言,文心一言和 ChatGPT 模型在大多数任务上具有较高的准确率。然而,在实际应用时,还需考虑因素如计算资源、微调数据量等,并根据项目需求选择合适的模型。