1. System-Original

完整版提示词引导

* 1. Content

你是一位大学教师教授，需要对学生提交的毕业设计论文进行评估。

请评估以下<报告文本/总结报告文本>在描述<结构完整性>，<逻辑清晰度>， <语言连贯性>， <内容独特性和创新性>， <参考文献规范性>，<课程知识掌握度>方面的表现，

并根据各指标<占比比例>进行打分与点评，打分范围0-10分。并最终按照<打分模版>给出学生报告打分结果与评价”。打分模板如下：

最终打分：<> (范围0-10分)

1. 结构完整性得分：<>, 占比20%，原因如下：<>

2. 逻辑清晰度得分：<>, 占比20%，原因如下：<>

3. 语言连贯性得分：<>, 占比20%，原因如下：<>

4. 内容独特性和创新性得分：<>, 占比20%，原因如下：<>

5. 参考文献规范性得分：<>, 占比10%，原因如下：<>

6. 课程知识掌握度得分：<>, 占比10%，原因如下：<>

修改意见：<>

请严格按照以下格式返回结果，最终打分一行、6个维度各自一行、修改意见一行，不要擅自添加换行：

最终打分：<> (范围0-10分)

1. 结构完整性得分：<>, 占比20%，原因如下：<>

2. 逻辑清晰度得分：<>, 占比20%，原因如下：<>

3. 语言连贯性得分：<>, 占比20%，原因如下：<>

4. 内容独特性和创新性得分：<>, 占比20%，原因如下：<>

5. 参考文献规范性得分：<>, 占比10%，原因如下：<>

6. 课程知识掌握度得分：<>, 占比10%，原因如下：<>

'修改意见：<>

* 1. Result

完整提示词引导下的少样本微调结果

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 模型 | 1-shot | | | 2-shot | | |
| MAE | MSE | PCC | MAE | MSE | PCC |
| 通义千问-Plus | 0.42 | 0.27 | 0.27 | 0.36 | 0.21 | 0.46 |
| 通义千问2.5-14B-1M | 0.32 | 0.18 | 0.27 | 0.32 | 0.18 | 0.24 |
| 通义千问-Turbo | 0.50 | 0.36 | 0.27 | 0.47 | 0.37 | 0.07 |
| DeepSeek-V3 | 0.32 | 0.18 | 0.30 | 0.32 | 0.18 | 0.30 |
| DeepSeek-R1 | 0.42 | 0.32 | 0.32 | 0.44 | 0.31 | 0.40 |

1. system-holistic

大模型考虑六个维度，然后直接进行总分评分 不需要先分别对六个维度评分

* 1. Content

你是一位大学教师教授，需要对学生提交的毕业设计论文进行评估。

请评估以下<报告文本/总结报告文本>在描述<结构完整性>，<逻辑清晰度>， <语言连贯性>， <内容独特性和创新性>， <参考文献规范性>，<课程知识掌握度>方面的表现，

参考上述各个指标，综合六个维度给出最终分数。并最终按照<打分模版>给出学生报告打分结果与评价”。打分模板如下：

最终打分：<> (范围0-10分)

修改意见：<>

请严格按照以下格式返回结果，最终打分一行、6个维度各自一行、修改意见一行，不要擅自添加换行：

最终打分：<> (范围0-10分)

'修改意见：<>

* 1. Result

直接进行总分评分情况下的模型结果

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 模型 | 0-shot | | | 1-shot | | | 2-shot | | | 3-shot | | |
| MAE | MSE | PCC | MAE | MSE | PCC | MAE | MSE | PCC | MAE | MSE | PCC |
| 通义千问  Plus | 0.65 | 0.68 | 0.16 | 0.46 | 0.33 | 0.30 | 0.42 | 0.26 | 0.32 | 0.40 | 0.26 | 0.17 |
| 通义千问  2.5-14B-1M | 0.48 | 0.37 | 0.30 | 0.35 | 0.20 | 0.12 | 0.33 | 0.17 | 0.31 | 0.35 | 0.19 | 0.43 |
| 通义千问  Turbo | 0.79 | 1.01 | 0.04 | 0.61 | 0.59 | 0.23 | / | / | / | / | / | / |
| DeepSeek  V3 | 0.37 | 0.24 | 0.37 | 0.35 | 0.19 | 0.40 | 0.34 | 0.19 | 0.25 | 0.35 | 0.21 | 0.50 |
| DeepSeek  R1 | 0.56 | 0.48 | 0.29 | 0.30 | 0.14 | 0.52 | 0.31 | 0.17 | 0.34 | 0.39 | 0.24 | 0.56 |

通义千问-Turbo在2-shot、3-shot的情况下会出现大量模型回复结果混乱的情况，反复尝试后无明显改进，以3-shot为例，结果如下：

Report 1：

最终打分，可以评估增分数据集对模型性能的影响。4.3 BiGAN 模型构建1）生成器：使用一个随机生成器，能够生成器生成的样本可以有效用于垃圾分类器生成器生成的数据，可以提升模型的泛化能力更强的特征。220见过的数据上表现得更加优秀。在本实验中，生成器的训练设置本实验中使用的模型为：图像尺寸：64×64×64×3，以以以以以以以以以以以以以以以以以以以以以以以以以以以以以以以以减少计算成本。优化器：Adam 优化器进行比较。实验设置为2000个epoch，

Report 2：

最终打分：进行零填充，9.0 (范围00对拼接操作将)将两者合并，)

修改意见： 1. \*\*结构完整性架构设计完整性较好，逻辑清晰，结构完整性和逻辑清晰度良好。 逻辑清晰晰度较好，逻辑结构清晰，能够到到到到到到到数据处理、模型参数调调调调调调调调调调调调调调调调调调调调调调参和稳定性。模型的参数设置合理，模型训练流程描述较为详细，语言表达较为连贯。 语言连贯性性性性性性性性性性性性性性性性性性性性性性性性性性性性较强，但创新性不足足足足足足足足足足足足足足足足足足性，且通过过过过过过过过过过过过过过过过过过过过过样例，模型能够准确识别并处理样本数据，但创新性稍弱，模型在测试阶段表现稳定。 参考文献使用规范，引用格式符合学术标准。 课程知识掌握良好，能够较好地结合生成对抗网络与变分自编码器的混合模型，能够生成高质量的样本数据。 最终打分：

Report 3：

最终的潜在空间概念，

1. system\_w-o-role

不给身份信息

* 1. Content

请评估以下<报告文本/总结报告文本>在描述<结构完整性>，<逻辑清晰度>， <语言连贯性>， <内容独特性和创新性>， <参考文献规范性>，<课程知识掌握度>方面的表现，

并根据各指标<占比比例>进行打分与点评，打分范围0-10分。并最终按照<打分模版>给出学生报告打分结果与评价”。打分模板如下：

最终打分：<> (范围0-10分)

1. 结构完整性得分：<>, 占比20%，原因如下：<>

2. 逻辑清晰度得分：<>, 占比20%，原因如下：<>

3. 语言连贯性得分：<>, 占比20%，原因如下：<>

4. 内容独特性和创新性得分：<>, 占比20%，原因如下：<>

5. 参考文献规范性得分：<>, 占比10%，原因如下：<>

6. 课程知识掌握度得分：<>, 占比10%，原因如下：<>

修改意见：<>

请严格按照以下格式返回结果，最终打分一行、6个维度各自一行、修改意见一行，不要擅自添加换行：

最终打分：<> (范围0-10分)

1. 结构完整性得分：<>, 占比20%，原因如下：<>

2. 逻辑清晰度得分：<>, 占比20%，原因如下：<>

3. 语言连贯性得分：<>, 占比20%，原因如下：<>

4. 内容独特性和创新性得分：<>, 占比20%，原因如下：<>

5. 参考文献规范性得分：<>, 占比10%，原因如下：<>

6. 课程知识掌握度得分：<>, 占比10%，原因如下：<>

'修改意见：<>

* 1. Result

无身份信息提示情况下的模型结果

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 模型 | 0-shot | | | 1-shot | | | 2-shot | | | 3-shot | | |
| MAE | MSE | PCC | MAE | MSE | PCC | MAE | MSE | PCC | MAE | MSE | PCC |
| 通义千问  Plus | 0.57 | 0.44 | 0.34 | 0.41 | 0.26 | 0.26 | 0.39 | 0.24 | 0.36 | 0.37 | 0.22 | 0.18 |
| 通义千问  2.5-14B-1M | 0.68 | 0.76 | 0.22 | 0.33 | 0.18 | 0.20 | 0.32 | 0.17 | 0.31 | 0.43 | 0.27 | 0.40 |
| 通义千问  Turbo | 0.76 | 0.86 | 0.20 | 0.40 | 0.26 | 0.44 | / | / | / | / | / | / |
| DeepSeek  V3 | 0.35 | 0.20 | 0.46 | 0.29 | 0.15 | 0.49 | 0.36 | 0.21 | 0.28 | 0.28 | 0.14 | 0.61 |
| DeepSeek  R1 | 0.47 | 0.34 | 0.32 | 0.32 | 0.16 | 0.48 | 0.35 | 0.18 | 0.43 | 0.46 | 0.33 | 0.31 |

通义千问-Turbo在2-shot、3-shot的情况下会出现大量模型回复结果混乱的情况，反复尝试后无明显改进，以3-shot为例，结果如下：

Report 1：

最终生成的优化问题尤其重要。在每次更新的002，Beta1 和 Beta2 的值分别为 0.9和0.999，确保了优化过程的稳定。最终打分：8.5 (范围0-10分)

1. 结构完整性得分：8, 占比20%, 原因如下：报告的结构较为完整，包括了模型生成样本的能力。4.4对比实验4.4.。

2. 逻辑清晰度得分：8.5, 占比20%, 原因如下：报告逻辑较为清晰，各部分内容按照数据处理、分析到可视化展示的顺序逐步展开，...

Report 2：

最终输出图像的范围一致。3.最终的模型训练与完完编码器的输入，没有没有输出潜在向量量作为输出，具体结构，使用LeNet-5层卷积网络进行特征提取，然后使用LeakyReLU 和 和BatchNormalization 进行行向量的维度一致。2662.3.3.1.2.2.2.2.2.20,潜在向量为输入，使用用以检测生成的样本生成：首先通过预训练的模型：加载预训练的模型。2600000次迭代：2.1）加载真实数据和对应的潜在空间，向量z，通过生成器生成器器生成新样本。2.2）判别器：通过编码器对对真实样本，通过编码器生成潜在潜潜在空量。2.

Report 3：

最终衡量模型的性能，具体公式为分类准确率 = (TP+TN)/(TP+FP+FN+TN这里 TP 表示模型性能的手段，TP（True Positive，正样本的分类任务。通过增加样本的准确率 = TN/(TP+TN)。TP 表试样本数的百分比。公式为：分类准确率 = TP/(TP+FN+FP+TN \* 100000%。204.3分类准确率的提高分类准确地识别和分类各类垃圾。4.3.3模型性能的关键指标。4.4实验中，分别对原始数据集和增广后的测试集，来判断增广的效果。4.4.1 数据集选取在实验中，模型训练后的测试集准确率提升，说明增广数据集的合理性。4.4.220的数据集，可以提升模型的泛化能力更强的特征。21