

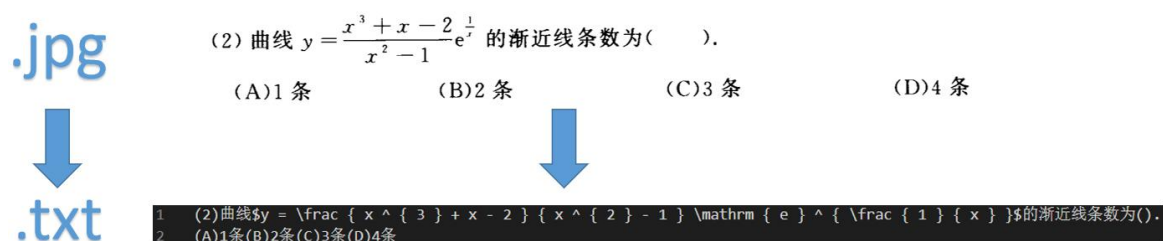
课程设计

这学期开始取消了考试，改为一个接近完整科研或工程项目的课程设计实战，占课程总评成绩的50%，希望同学们能从中学习到人工智能落地的项目经历。欢迎同学们对于课程设计的建议和意见，帮助教学团队优化教学质量。

项目：基于Encoder-Decoder架构的神经网络公式识别

问题定义：

- 给定一张数学试题图片，我们希望输出为图片内容对应的文本形式，其中数学公式使用LaTeX标记语言来描述公式的内容和空间布局。



数据集：高考数学、物理、化学、生物的公式

评价指标：

EditDistance	即Levenshtein距离，以取反的百分数呈现，越大越好。例：80%的EditDistance代表需要改动20%的内容才能达到groundtruth 参考： https://www.jianshu.com/p/a617d20162cf
ExactMatch	当预测结果和gt一样时才算100%准确，否则为0%

分组：4-5人，实验代码等数据以组为单位上交，实验报告单独撰写，单独上交，注明分工内容。

工作内容：

- 熟悉并使用labelImg软件提取公式图片（群里会提供教程）。为了减少数据获取难度，本次实验会收集同学们提取的公式图片，通过mathpix识别后，取mathpix的识别结果作为ground truth，再发回给大家作为数据集来训练。
- Encoder用CNN，Decoder用RNN：这个模型的结构与实验3的结构相似，模型详情参考上次实验文档。值得注意的是，近年来通过在Encoder和Decoder之间添加注意力机制能大大提升模型的效果[1][4]，若添加有注意力与无注意力的对比分析可酌情加分。
- Encoder用ResNet，Decoder用Transformer[2][3]：近年来Transformer从NLP领域向其他深度学习领域发展，并在各个领域都取得不错的成果，被证明具有强大的表达能力。实现建议使用高级接口，可参考[5]。

时间跨度：3周

评分标准：

标注：40%，完成对应标注任务即可得到该部分的85分（100分制），额外标注一份即可得到100分（100分制）。

答辩时评估(需要准备ppt)：

① CNN+RNN：15%，评分根据参考代码实现是否正确、实验结果评测指标、代码可复现性和注释等方面来考虑。

② ResNet+Transformer：15%，评分根据参考代码实现是否正确、实验结果评测指标、代码可复现性和注释等方面来考虑。

课程论文：10%，实验完成后分析以下问题，模型是否存在过拟合？从输出结果来看还存在哪些不足？

完成上面4项基本内容可得80%，下面加分项20%：

1. 实验拓展：优化模型、优化数据预处理、讨论任务瓶颈等。

如果同学们围绕这几个问题进行额外分析与实验，可以视情况得到该部分的优秀及以上分数实验结果好的可以考虑加分更多。

2. “基线”指标：ExactMatch=85%，EditDistance=85%，如果你模型的结果超过了这个“基线”，我们会进行加分。另外，欢迎你记录实验过程中的理解和思考，任何有价值的实验记录都会加分鼓励。

3. 有UI的根据UI的美观、实用性等方面酌情加分

参考资料：

[1] <https://arxiv.org/abs/1609.04938v1>

[2] <https://arxiv.org/abs/1706.03762>

[3] <https://arxiv.org/abs/2103.06450#:~:text=Full%20Page%20Handwriting%20Recognition%20via%20Image%20to%20Sequence,of%20handwritten%20or%20printed%20text%20without%20image%20segmentation.>

[4] https://github.com/LinXueyuanStdio/LaTeX_OCR_PRO

[5] <https://github.com/kingyiusuen/image-to-latex>