

人工智能程序设计 课程大作业

选题一 中文分词算法

一、题目概述

根据现有的中文分词算法，实现一个中文分词系统，不允许调用现有分词工具。

二、题目要求

1. 实现一个**传统的匹配算法**（前向/后向最大匹配等）；
2. 实现一个**不含神经网络的机器学习算法**（HMM/CRF 等）；
3. 二选一：
 - 1) 实现一个**基于深度学习神经网络的算法**，不使用 BERT 等预训练模型，即不允许通过 huggingface 等开源网站直接引用权重（发现需重新完成）。可以自己动手搭建 RNN、Transformer 等结构从头开始训练，可结合 1-2；
 - 2) 使用华为 MindSpore 框架实现**基于预训练模型 BERT 或 BERT+CRF 的分词算法**；
4. 1-3 为基本要求，有能力的同学可以额外实现 2-3 个算法进行比较；
5. 实验报告中需要给出所使用的算法介绍，并且对比其在测试集上的结果。另外，要求在测试集上对比一个现有分词工具的实验结果；
6. 设计一个**UI 界面**，要求可以在 UI 中输入一个句子，输出分词结果，并在实验报告中附图展示；
7. 算法和 UI 均使用 python 进行编程。

三、参考资料

数据集：<http://sighan.cs.uchicago.edu/bakeoff2005/>

数据集参考：<https://zhuanlan.zhihu.com/p/45243642>

现有分词工具包含 jieba / thulac / ltp 等

四、评分细则

1. 完成所有基础要求（除 4 以外）可获得大作业总分的 85%，各项分值占比如下：

传统匹配算法	10%
机器学习算法	20%
基于深度学习/预训练模型的算法	30%
UI 界面	10%
实验报告	15%

2. 加分项 1：额外实现 2-3 个算法，视实现的效果给分，最多 +15%；
3. 加分项 2：实验报告分析有亮点，最多 +5%；
4. 加分项 3：复现任意一篇 2022-2023 年发表的与中文分词相关的论文（能复现论文结果即可，不要求从头开始编写。需要附论文原文），视复现程度给分，最多 +15%；
5. 加分项 4：调研中文分词领域的最新研究成果（2020 年之后发表），并撰写调研报告，其中至少包括 5 篇论文，每篇论文要具体分析其优点和缺点，最多 +10%；
6. 以上加分项的分数分别独立累积计入大作业总分中，最终大作业得分不超过满分的 120%。

选题二 图像语义分割

一、题目概述

使用多种方法，实现图像的语义分割系统，不允许直接整体调用训练完成的语义分割模型。

二、题目要求

1. 实现一个**不含神经网络的语义分割算法**（图匹配/聚类/SVM 等）；
2. 实现两个**基于神经网络的语义分割算法**，要求两种算法的特征提取 Backbone 和上层模块均为不同的网络结构；
3. 有能力的同学可以额外实现 2-3 个不同算法进行比较；
4. 使用 **SBD 数据集** 进行本次实验，实验报告中需要给出所使用的算法介绍，并对比算法在验证集上的分割效果，评价指标至少需要使用平均 IoU；
5. 设计一个 UI 界面，要求可以传入本地的图像文件，输出可视化的语义分割结果，即与输入大小相同的图像，并在实验报告中附图展示；
6. 算法和 UI 均使用 python 进行编程。

三、参考资料

SBD 数据集下载链接为：

<http://home.bharathh.info/pubs/codes/SBD/download.html>

四、评分细则

1. 完成所有基础要求（除 3 以外）可获得大作业总分的 85%，各项分值占比如下：

不含神经网络的语义分割算法	25%
基于神经网络的语义分割算法	35%
UI 界面	10%
实验报告	15%

2. **加分项 1:** 额外实现 2-3 个算法，视实现的效果给分，最多 +15%;
3. **加分项 2:** 实验报告分析有亮点，最多 +5%;
4. **加分项 3:** 复现任意一篇 2022-2023 年发表的与图像语义分割相关的论文（能复现论文结果即可，不要求从头开始编写。需要附论文原文），视复现程度给分，最多 +15%;
5. **加分项 4:** 调研图像语音分割领域的最新研究成果(2020 年之后发表)，并撰写调研报告，其中至少包括 5 篇论文，每篇论文要具体分析其优点和缺点，最多 +10%;
6. 以上加分项的分数分别独立累积计入大作业总分中，最终大作业得分不超过满分的 120%。

注:

1. 请从上述两个选题中选择其一完成;
2. 不允许抄袭;
3. 给出源代码及实验报告，实验报告中至少需要包含算法原理介绍、算法实现说明、实验结果与分析以及 UI 界面截图;
4. 请将源代码和实验报告打包提交，命名格式为:

姓名_学号_大作业.zip

例如:

李好雨_2022214293_大作业.zip