

小作业一 语言模型PPL计算

截止时间 2021 10月 17, 23:59 之前 **得分** 10 **提交** 一份上传文件 **文件类型** pdf
可用 2021 9月 27 22:00 至 2021 10月 17 23:59 20 天

此作业锁定于 2021 10月 17 23:59。

本次作业需要同学们用训练完的语言模型计算PPL。

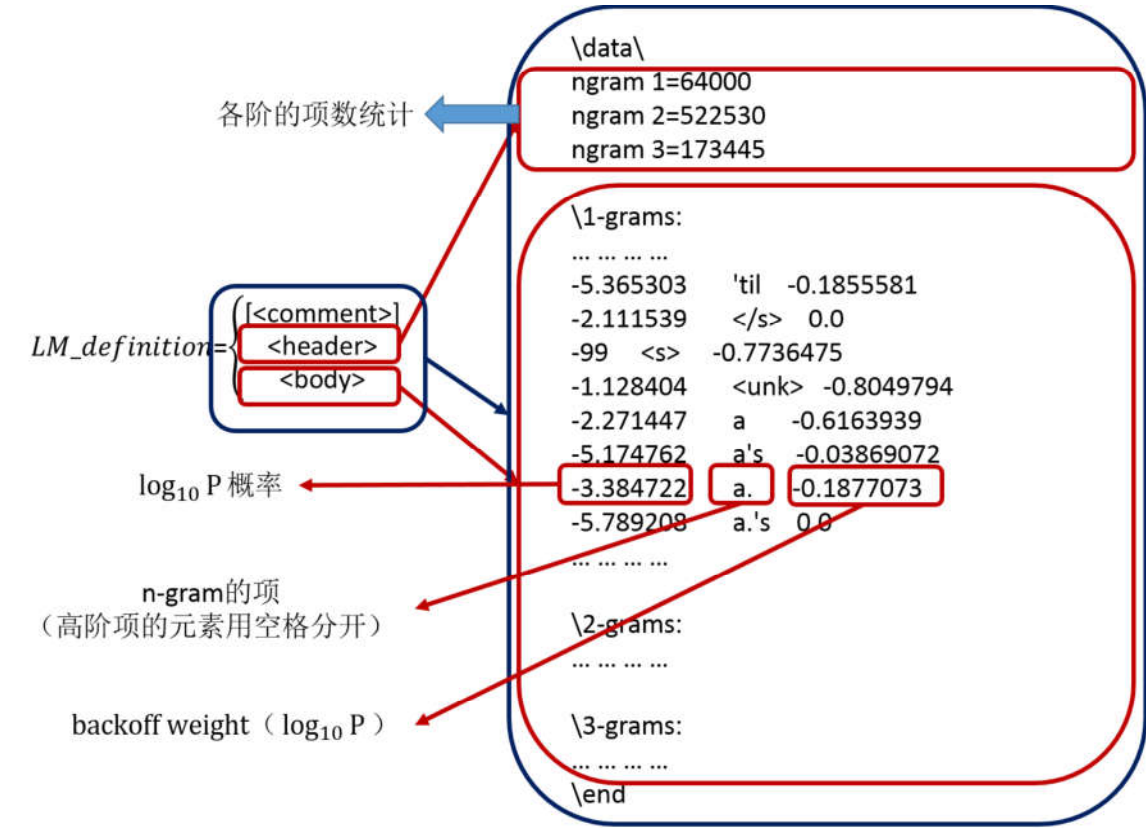
我们以同学们的学号为语料库，训练了一个tri-gram语言模型（模型已经上传，可通过[此链接](#)获取）。

请大家用上述的tri-gram语言模型计算四位助教学号的PPL：

- 021033210023
- 019033910051
- 120033910006
- 120033910013

注：

1. tri-gram以单个数字为token
2. tri-gram的起始符和终止符分别为<s>和</s>（所以计算时也需要加上）
3. tri-gram文件中的权重表示以10为底的概率的对数
4. **若某项back-off weight未在ARPA中出现，默认该项的back-off weight为0**
5. 语言模型的格式为ARPA，格式如下图所示



提交形式

上交一份PDF格式的报告(不超过一页)，报告中包含每位助教学号对应的PPL以及计算过程。