

## 第十三周周记

周一	
完成内容	看法律文本的数据，并 debug 程序，查看法律文本的数据处理过程。
内容描述	1、发现法律文本中的好多数据并没有实际的用处，如每一条文本的第一句“公诉机关襄城县人民检察院。被告人张某西，男，生于 1973 年 2 月 20 日。”这样的句子可以想办法去掉，以减少模型的负担。2、应从每条文本的“指控”开始，是这条文本比较重要的数据。
未解决问题	无

周二	
完成内容	看 RF Breiman L. Random forests. Machine Learning, 2001; 45(1):5-32. <a href="https://doi.org/10.1023/A:1010933404324">https://doi.org/10.1023/A:1010933404324</a>
内容描述	
未解决问题	无

周三	
完成内容	看论文，学习用 RF 实现的一个法律判决预测模型： Katz DM, Bommarito MJ, II, Blackman J(2017) A general approach for predicting the behavior of the Supreme Court of the United States. PLoS ONE 12(4): e0174698.
内容描述	该判决预测模型用了美国最高法院的大量数据，最后的正确率可以达到 70.2%，用随机森林进行预测以及其他的功能。
未解决问题	文章的数据集的处理还不是很明白

周四	
完成内容	看论文，学习用 RF 实现的一个法律判决预测模型： Katz DM, Bommarito MJ, II, Blackman J(2017) A general approach for predicting the behavior of the Supreme Court of the United States. PLoS ONE 12(4): e0174698.
内容描述	1、下载了该论文的实现程序，查看其中的数据保存和提取方式
未解决问题	无

周五	
完成内容	抽取四万条训练数据，并测试
内容描述	1、修改训练数据，抽取出四万条训练数据中的刑期文本，并将其加到该条法律文本的最后，并以“\t”分隔。

	2、测试，将四万条数据放入程序中进行测试训练并预测，训练结果的正确率达 0.661。但最后的预测结果并不理想。
未解决问题	预测了 50 条数据，发现这五十条数据的预测结果都是一致的，该程序某些地方有一些问题需要修改。

周末	
完成内容	
内容描述	
未解决问题	

工程汇总	
任务描述	1、抽取训练数据。2、测试程序。
代码量	
未解决问题	无

论文汇总	
论文列表	[1] Katz DM, Bommarito MJ, II, Blackman J(2017) A general approach for predicting the behavior of the Supreme Court of the United States. PLoS ONE 12(4): e0174698. [2] Breiman L. Random forests. Machine Learning, 2001; 45(1):5–32. <a href="https://doi.org/10.1023/A:1010933404324">https://doi.org/10.1023/A:1010933404324</a>
论文摘要	
未解决问题	无

下周任务	
工作	1、将法律文本进行再处理，修改文本长度，使得模型能加快处理速度。 2、学习 RF，尝试能否修改模型，使得训练的结果有所提高。 3、修改程序，是的预测结果能正常。 4、改进程序，加入犯罪类型的预测。 5、实现模型。
论文	[1] Katz DM, Bommarito MJ, II, Blackman J(2017) A general approach for predicting the behavior of the Supreme Court of the United States. PLoS ONE 12(4): e0174698.
其他	
汇总	

日期:18/3/25 - 18/3/31