无论研究什么，都首先要对研究的对象有所了解，文本分类也是如此。在参考了相关论文后，我觉得lstm结构处理文本问题比较好，所以决定先使用基于开源的tensorflow环境下的Word2vec和lstm，来处理。先将论文处理为词向量，然后再将词向量作为输入导入lstm神经网络结构中，之后通过训练生成模型，对真假论文进行判断。

假的论文相较于真论文的差别基本上就在于文章有没有逻辑和实际意义。假的论文仅仅只是符合语法，不可能和真论文一样存在逻辑，所以通过训练生成的模型可以判断是否具有具体意义，从而将真假论文分类。