汉语词性标注任务

1. 方法：采用一阶隐马尔可夫模型计算参数，采用维特比算法解码
2. 训练语料来源：北大新闻分词词性标注语料（人民日报1998年1月）
3. 测试：先计算每一句话的正确率（标注正确的词数除以总词数），再对所有句子的正确率求平均，得到平均正确率
4. 缺陷：
   1. 没有处理未登录词。如果输入的词在训练语料中没有出现过，就无法给出整个句子的词性
   2. 断句不合理。程序中只是按照语料中的一行作为一个句子。实际上一行应该是一篇文档，包含多个句子。
   3. 未调用分词。程序所需的输入是字符串数组，该数组按顺序包含了一句话中的每个词。