1. **作业2 - 概率图模型练习**

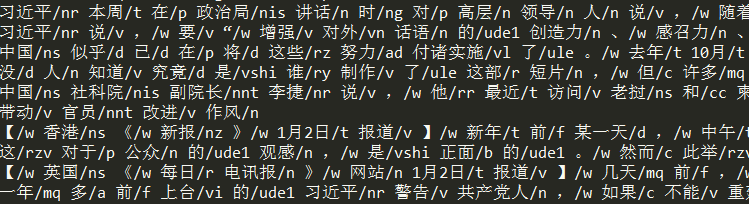
**任务：词性标注**

**词性标注**是为分词后的每个单词标注一个正确的词性。

**要求：**用隐马尔科夫模型实现词性标注。

**数据资源：**语料来自1998年1月份人民日报语料。

**语料格式：**



**语料使用方法：**

（1）采用合理的方式对数据进行预处理；

（2）按照8:2比例，将数据划分为训练数据和测试数据；训练数据部分用来训练模型参数，测试部分用来验证模型效果。

**模型验证：**

采用交叉验证方式，**指标：**准确率。

**模型说明：**

推荐使用隐马尔可夫模型（HMM），

**使用方式可选择：**

（1）自己编码实现（鼓励有能力、感兴趣的同学采用这种方式）；

（2）利用开源工具包：

**隐马尔科夫模型：**[**http://htk.eng.cam.ac.uk/**](http://htk.eng.cam.ac.uk/)

**作业截止日期：2018-4-17**