

2021年自然语言处理_期末真题_回忆版

【author】Maxpicca-Li

【写在前面】

1. 考试时间：2021年12月24日14:00~16:00。
2. 试卷整体简单，开卷，复习的时候还是把书上的一些重点看一下，要不然都不能秒翻到内容。不能秒翻到内容的话，就可能面临着抄书抄不完的情况。考试感慨：NLP开卷考试=抄书+手撕CNN和HMM。
3. 本回忆版真题于2021年12月24日19点20分写成，因为情绪不好，不想和朋友说话。
4. 计院专业课的试卷似乎都不准老师发出来，希望有学弟学妹们能将我“回忆试卷”的习惯传承下去！
——from VayneDuan
5. 其余专业课的回忆版试卷如下：记得 star & follow, 会持续更新的!

Source: [VayneDuan/cqu-cs-learning-materials](https://github.com/VayneDuan/cqu-cs-learning-materials)

Fork: [Maxpicca-Li/cqu-cs-learning-materials](https://github.com/Maxpicca-Li/cqu-cs-learning-materials)

——【写在前面】基于原作者VayneDuan修改

1. 填空题（10道*2分=20分）

1. 命名实体识别的定义
2. LSTM中，**输出门**用于表示到下一个细胞状态的信息
3. $p_1 = 1/2, p_2 = 1/4, p_3 = 1/8, p_4 = 1/8$, 求熵: $7/4$
4. 无向图中，不能使用条件概率密度作为参数化，应该使用 **联合概率密度** 的参数化
5. WordNet中，“猫”和“哺乳动物”属于 **上下位关系**
6. “他将来学计算机”，属于**组合型歧义**
7. **余弦**函数用于评估两个向量之间的相似度。
8. 文本分类性能评价中，**召回率**表示分类正确的个数在所有标准答案中的比例。
9. . . .
10. . . .

2. 简答题（4道*5分=20分）

1. 自然语言处理的概念、主要研究内容和方法
2. 最大似然估计的基本原理
3. 交叉熵的概念、物理意义和应用
4. 请阐述语言知识库与典型语言知识库

3. 计算题（2道*15分=30分）

1. CNN相关（考的很简单）
 1. 步长为1无pad的卷积
 2. 步长为2有外围0 pad的卷积
 3. 平均池化
 4. relu激活函数
2. HMM相关
 1. viterbi算法，前向递推两个时刻

2. 回溯，计算最优隐状态序列

4. 综合题 (3道*10分=30分)

1. 列举文本分类中特征选择的几种方法，并分析其特点、存在的问题及解决方法
2. 对两种向量表示方法——向量空间模型vsm和word2vec进行综合分析并对比
3. 命名实体识别中的基于多特征的识别方法，请用该方法对人名进行实体识别，并写出具体的步骤
(模型以及模型参数)