

愿中国青年都摆脱冷气，只是向上走，不必听自暴自弃者说的话。
能做事的做事，能发声的发声。
有一分热，发一分光。
就令萤火一般，也可以在黑暗里发一点光。
——不必等候炬火
此后如竟没有炬火，我便是唯一的光

大 连 理 工 大 学

姓名：_____

学号：_____

部 (院): _____

_____级_____班

课程名称：自然语言处理 试卷：A 考试形式：闭卷

授课部 (院)：电信学部 考试日期：2023 年 4 月 28 日 试卷共 5 页

	一	二	三	四	五						总分
标准分	20	20	18	18	24						100
得 分											

得	
分	

一、填空题（每题 2 分，共 20 分）

- _____是研究模仿人类语言认知过程的自然语言处理方法和实现技术的一门学科。
- _____通常通过一些特殊的语句或者语言现象的研究啦得到对人的语言认知能力，而这些语句和语言现象在实际应用中并不常见。
- _____常被用来衡量两个随机变量分布的差距。
- _____目标是优化噪声信道中信号传输的吞吐量和准确率。
- 如何区分不同上下文中词汇的确切语义，就是_____问题。
- 语言模型设计任务是寻找_____最小的模型。
- _____是精确描述语言及其结构的手段。
- 图灵机与有限自动机（FA）的区别：_____可以通过其读/写头改变输入带的字符。
- 语料库的划分有：异质的、同质的、_____的、专用的。
- 由字构词的分词方法将分词过程看做字的_____问题。

得	
分	

二、判断题（每题 2 分，共 20 分）

- （ ）NLP比计算语言学更侧重基础理论和方法的研究。
- （ ）自然语言处理的内容包括对词法、句法、语义、语用等信息的识别、分类、提取、转换、生成等技术
- （ ）越是不确定的随机变量越要更多的信息确定
- （ ）设计语言模型时候一般用交叉熵衡量好坏
- （ ）线性带限自动机与I型文法等价，其读写头不超越原来输入带上字符串的初始终结位置。

6. () FA完成状态转移时候产生输出，而FST系统只实现了状态转移，不产生输出。
7. () 如果文法G无二义的，它的任何句子的最左推导和最右推导生成语法书可能相同
8. () 依次图和依存树的约束包括：可变父节点、连通、无环、可投射
9. () 生成式比判别式模型丰富，研究单类问题比判别式灵活性强，模型通过增量学习，且可用于数据不完整的情况。
10. () PARSEVAL评测方法区分能力不是很强，而且在特殊情况下计算出的正确率和召回率存在较大偏差。

得分	
----	--

三、多项选择题（每题 3 分，共 18 分）

- 语音学的组成 ()
 - 发言语音学
 - 声学语音学
 - 描述语音学
 - 听觉语音学
- 语系的划分包括 ()
 - 分析语
 - 联合语
 - 屈折语
 - 黏着语
- NLP面临的挑战 ()
 - 普遍存在不确定性
 - 未知语音现象的不可预测性
 - 面临数据的充分性
 - 语音知识表达的复杂性
- 分词中基于词典的切分方案有 ()
 - 最大匹配算法
 - 最短路径算法
 - 由字构词
 - 语音模型
- 语料库的构建原则有 ()
 - 真实性
 - 可靠性
 - 分布性
 - 流通性

6. 短语结构分析方法中，对比评估指标有（ ）

- A. 树的相似性 B. 结构一致 C. 熵的大小 D. 语法评估兴趣小组的测评方案

得分

四、简答题（每题 3 分，共 18 分）

1. 最大匹配法的优缺点？
2. 中文姓名识别的一般步骤？
3. 依存句法分析器的构建步骤？

得分

五、计算题（共 24 分）

1. 文法 $G = (\{E, T, F\}, \{a, +, *, (,)\}, P, E)$

$P: E \rightarrow E+T \mid T$

$T \rightarrow T * F \mid F$

$F \rightarrow (E) \mid a$

输出句子 $a * a + a$ ，请写出两种推导以及相应的分析树

2. 有基于PCFG的句法分析 $G(S)$ ：

$S \rightarrow NP, VP$ 1.00 $NP \rightarrow children$ 0.20 $NP \rightarrow hands$ 0.15

$NP \rightarrow play$ 0.05 $NP \rightarrow NP, PP$ 0.4 $NP \rightarrow balls$ 0.15

$NP \rightarrow PT, NP$ 0.05 $DT \rightarrow The$ 1.00 $PP \rightarrow P, NP$ 1.00

$P \rightarrow with$ 1.00 $V \rightarrow play$ 1.00 $VP \rightarrow V, NP$ 0.7

$VP \rightarrow VP PP$ 0.3

计算句子 The children play balls with hands 的分析树并给出最佳分析树

参考答案:

一、填空

1. 自然语言理解
2. 理性主义
3. 相对熵
4. 噪声信道模型
5. 语义消歧
6. 困惑度
7. 形式语言
8. 图灵机
9. 系统的
10. 分类

二、选择

1. 错
2. 对
3. 对
4. 错
5. 对
6. 错
7. 对
8. 错
9. 对
10. 对

三、多选

1. ABD
2. ACD
3. ABD
4. ABC
5. ABCD
6. ACD

四、简答

1. 优点:

程序简单易行, 开发周期短;

仅需要很少的语言资源(词表), 不需要任何词法、句法、语义资源;

缺点:

歧义消解的能力差;

切分正确率不高, 一般在95%左右。

2. 看PPT


3. 看PPT

参考答案:

五、计算

1. (相信大家都会)


2. 以下答案来自ChatGPT

 GPT-4 >

2. 有基于PCFG的句法分析G(S):

$S \rightarrow NP, VP$ 1.00 $NP \rightarrow children$ 0.20 $NP \rightarrow hands$ 0.15
 $NP \rightarrow play$ 0.05 $NP \rightarrow NP, PP$ 0.4 $NP \rightarrow balls$ 0.15
 $NP \rightarrow PT, NP$ 0.05 $DT \rightarrow The$ 1.00 $PP \rightarrow P, NP$ 1.00
 $P \rightarrow with$ 1.00 $V \rightarrow play$ 1.00 $VP \rightarrow V, NP$ 0.7
 $VP \rightarrow VP, PP$ 0.3

计算句子 The children play balls with hands 的分析树并给出最佳分析树

 This bot may make incorrect statements. It does not have knowledge of events after 2021. Tap to learn more.

首先, 我们需要将句子"The children play balls with hands"分解成词汇单元: The, children, play, balls, with, hands。然后, 我们将使用给定的PCFG (概率上下文无关文法) 规则来生成可能的分析树。根据给定的规则, 我们可以得到以下概率:

1. $S \rightarrow NP VP$ (1.00)
2. $NP \rightarrow children$ (0.20)
3. $NP \rightarrow hands$ (0.15)
4. $NP \rightarrow play$ (0.05)
5. $NP \rightarrow NP PP$ (0.4)
6. $NP \rightarrow balls$ (0.15)
7. $NP \rightarrow DT NP$ (0.05)
8. $DT \rightarrow The$ (1.00)
9. $PP \rightarrow P NP$ (1.00)
10. $P \rightarrow with$ (1.00)
11. $V \rightarrow play$ (1.00)
12. $VP \rightarrow V NP$ (0.7)
13. $VP \rightarrow VP PP$ (0.3)

现在, 我们可以生成句子的可能分析树并计算它们的概率。以下是一个可能的分析树及其概率:

1. S
 - NP
 - DT (The) (1.00)
 - NP (children) (0.20)
 - VP
 - V (play) (1.00)
 - NP (balls) (0.15)
 - PP
 - P (with) (1.00)
 - NP (hands) (0.15)

计算概率:
 $(1.00) * (0.20) * (1.00) * (0.15) * (1.00) * (0.15) = 0.0045$

根据给定的规则, 这是"The children play balls with hands"的最佳分析树, 其概率为0.0045。

仅供参考，若发现不足之处请联系我，若你也想分享经验也可以联系我
邮箱 994182204@qq.com

😭回来吧，我的 nlp😭

💡我最骄傲的信仰💡

⚡历历在目的早八⚡

😭眼泪莫名在流淌😭

💡依稀记得熵计算💡

👍还有复杂的平滑👍

⚡幻想今天必过关⚡

✨通宵熬夜都不累✨