

填空：

1

乔姆斯基形式语法定义为四元组：

$G = (V_n, V_t, S, P)$

V_n :非终结符(non-terminals)的有限集合

V_t :终结符(terminals)的有限集合

S : V_n 中的初始符号

P :重写规则(rewriting rules)，又称生成规则(production rules)

2

Chomsky根据重写规则的形式，把形式语法分为4级：

无约束文法

上下文有关文法

上下文无关文法

正则文法

3

语言模型是用来计算一个句子概率的概率模型

4

语言模型计算概率的方法：最大似然估计

5

语言模型数据平滑的方法

加1法(Additive smoothing)

减值法/折扣法 (Discounting)

删除减值法:低阶代替高阶

6

条件随机场的特征函数：边特征函数 t_i ，节点特征函数 S_k

7

CRF模型建模关键任务:特征选择

8

中文的词法分析任务(孤立语)：分词，未登录词识别，词性标注

9

切分歧义问题分类：交集型歧义，组合型歧义

10

未登录词处理将识别问题转化成标注问题

11

未登录词处理对于输入句子中的每个汉字，定义四个标记：

不属于未登录词O，未登录词首字B，未登录词尾字E，未登录词中间字I

12

词性标注是在给定句子中判定每个词的语法范畴，确定其词性并加以标注

13

词法分析的评价指标

正确率 (precision, P)

召回率 (Recall ratio, R)

F-度量值

14

句法分析任务是确定句子的句法结构或句子中词汇之间的依存关系

15

Coke-Younger-Kasami (CYK) 算法通过构造识别矩阵进行分析

16

概率上下文无关文法(PCFG) 是CFG的概率拓广, 可以直接统计语言学中词与词、词与词组以及词组与词组的规约信息，并且可以由语法规则生成给定句子的概率。

17

局部句法分析主要任务是第二层名词短语、动词短语等的识别

18

依存结构分析: 分析出词与词之间的依存关系,揭示句子的结构。

19

句子中心词：

动词是句子的中心，它支配着别的成分，而不受其他成分支配。直接受动词支配的有名词词组和副词词组，名词词组形成“行动元”(不超过三个: 主语、宾语1、宾语2);副词词组形成“状态元”(可以无限个); 动词支配行动元的个数称为 该动词的“价”

20

统计依存句法分析方法：生成式的分析方法，判别式的分析方法，决策式的(确定性的)分析方法

21

篇章:由一个以上的句子(sentence)或语段(utterance)构成的有组织、有意义的自然语言文本整体。

22

RST理论中最关键的两个成份: 关系和结构段

23

NLP中常用中心理论做指代消解理论

24

中心理论:每个语篇单位有三个中心: 前看中心, 优先中心, 回视中心

25

中心理论根据 回指中心的变化状态 将毗连着的语句关系分为四种, 来界定语篇结构的衔接性

简答:

一

解释什么是Chomsky范式:

任何的由上下文无关文法生成的语言, 均可由重写规则为 $A \rightarrow BC$ 或者 $A \rightarrow x$ 的文法生成, 其中 $A, B, C \in V_n$, $x \in V_t$ 具有这样的重写规则的上下文无关文法。它的推导树均可简化为二元形式, 这样就可以用二分法来分析自然语言, 采用二叉树来表示自然语言的句子结构。上述重写规则称为Chomsky范式。

二

解释隐马尔可夫模型五元组:

1. 隐藏状态 s : 一个系统的(真实)状态, 可以由一个马尔科夫过程进行描述(如, 天气)。
2. 观察状态 o : 在这个过程中‘可视’的状态(例如, 海藻的湿度)。
3. 状态转移概率矩阵 $A = a_{ij}$: 包含了一个隐藏状态到另一个隐藏状态的概率。
4. 观察概率矩阵 $B = b_j(k)$: 从隐藏状态 S_j 观察到某一特定符号 v_k 的概率分布矩阵。
5. 初始状态的概率分布为: $p = p_i$

三

未登录词识别的困难:

很多未登录词都是由普通词汇构成的, 长度不定, 也没有明显的边界标志词。专有名词的首词和尾词可能与上下文中的其他词汇存在交集型歧义切分

四

解释什么是词之间的依存关系:

参加组成一个结构的成分(词)之间是不平等的，一些成分从属于另一些成分，每个成分只能从属于至多一个成分。从属与被从属的关系就决定了结构的性质。

五

引入激活函数的目的是什么？

为了增强网络的表达能力，我需要引入连续的非线性激活函数，因为连续非线性激活函数可以可导的，所以可以用最优化的方法来求解。

六

梯度消失问题：

在神经网络中误差反向传播的迭代公式为

$$\delta^l = \sigma'(z^l) \cdot (W^{l+1})^T \delta^{l+1}$$

其中需要用到激活函数的导数误差从输出层反向传播时每层都要乘激活函数导数。

当用 sigmoid 或 tanh 函数时：

$$\sigma'(x) = \sigma(x)(1 - \sigma(x)) \in [0, 0.25]$$

$$\tanh'(x) = 1 - (\tanh(x))^2 \in [0, 1]$$

这样误差经过每一层传递都会不断衰减，当网络很深时甚至消失

七

递归神经网络基本思想：

将处理问题在结构上分解为一系列相同的“单元”，单元的神经网络可以在结构上展开，且能沿展开方向传递信息。

八

递归神经网络有什么优势？

循环神经网络对于结构展开的非线性句法关系表示能力差，而自然语言处理中语句内部的句法关系通常是非线性的，递归神经网络有助于分析句法信息。

九

词向量的优势：

词向量使用低维的分布式向量表示词，相似词的词向量距离相近，这就让基于词向量设计的一些模型自带平滑功能。

十

CBOW和Skip-gram

CROW模型采用上下文的词向量预测当前词 $w(t)$ 的方法。

Skip-gram采用当前词 $w(t)$ 预测上下文的词向量的方法。