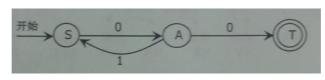
国科大-宗成庆-自然语言处理

- 一. 完成下列各题(本题满分20分)
- (1) 请给出一个汉语句子,说明连词引起歧义结构。
- (2) 请给出字符串 acomnad 与单词 command 之间的编辑距离。
- (3) 下图为有限状态自动机 M:



G={N, Σ ,P,S} N={S,A} Σ ={0,1} P = {S->0A, A->1S|0}

请构造正则文法 G 使 L(G)=T(M)。1、Vn=Q, Vt=Σ, S=q0 2、S->0A 3、B->a

(4) 已知 A、B 两个复杂特征集:

$$A = \begin{bmatrix} POS = NN \\ Lex = boy \\ Sem = Agent \end{bmatrix} \quad B = \begin{bmatrix} Lex = boy \\ Num = Single \\ Case = Subject \end{bmatrix}$$

请给出 $A \cup B = ?$

- (5) 请写出句子"I study English every day." 所有的 3 元文法。(Tri-grams)
- 二. 阐述题(本题满分10分):

请设计一种新的自动评价机器翻译系统译文质量的方法。注意:该方法必须是可计算的。请详细描述你的方法,并阐述其合理性。

- 三. 完成以下两道题(本题满分15分):
- 1. 有如下一句话:

2013年轻工业产品质量大幅度提升。

正确的分词和词性标注结果(标准答案)

2013 年/NT 轻工业/NN 产品/NN 质量/NN加升幅度 句式是否 而某分词和词性标注系统给出的结果是: 词现象)

引入句法分析来对译文质量进行评估。

首先我们可以将系统译文转变成句法树评估句法是否 合理,是否满足句法规则,并按照某种打分标准对句 法的正确程度进行打分。

其次分析其语义信息是否正确。可以将其语义特征信息与参考译文的语义特征信息进行比较,计算其相似不性,最后再按照一定的权重将语义和句法分别加权相 加打命是bleu只者處了n-gram的吻合程度,而无法对

加打縮_度以βμ只_捷處了wvgramβl吻合程度,而无法对 句式是否正确、语义上是否相似(有些词语存在同义 词现象)等标准作出评判。形式化:

Score = w1 * f(x) + w2 * f(y),其中语义相似性可以使 2013/QQ 年轻/AA 工业/NN 产品/NN 质量/NN 大倫介螺旋体制力量提升分V。/PU

请计算该分词系统的准确率、召回率和 F1 值。 $\frac{1}{2}$

2. 下面的图 1 是句子"这将成为解决伊朗核问题的动力"正确的依存关系树(答案):

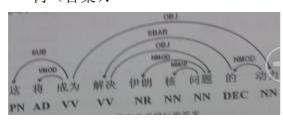
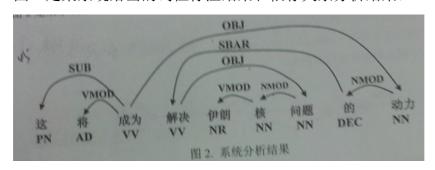


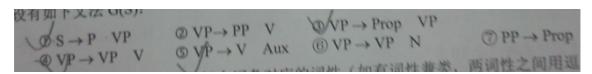
图 1. 依存关系树标准答案

图 2 是某系统给出的词性标注结果和依存关系分析结果:



请计算如下性能指标:

- (1)无标记依存正确率(unlabeled attachment score, UA); 8/9
- (2) 带标记依存正确率(labeled attachment score, LA): 7/9
- (3)依存正确率(dependency accuracy, DA). 7/8
- 四. 请完成下面的题目(本题满分15分): 设有如下文法G(S):



有一部词典包含如下词条和每条词条对应的词性(如有词性兼类,两词性之间用逗号隔开):

# 从: Prop	# 了解: V	# 题: N	# 学会: V, N
# 从小: Prop	# 解: V, N	# 小: A	•••••
# 会: V, N	# 解题: V	# 小学: N	
# 了: Aux	# 他: P	# 学: V	

给定句子: 他从小学会了解题。(假设不考虑句号)。请完成下面的题目:

- (1) 请分别给出利用正向最大分词算法和逆向最大分词算法对给定句子的分词结果: 他 从小 学会 了解 题; 他 从小 学会 了 解题
- (2) 请用逆向最大分词算法给出的分词结果和线图句法分析算法(Chart Parser)分析给定句子,请写出句法分析过程,即分别给出每一步线图 (Chart)、代理表(Agenda)和活动边集(ActiveArc)的内容,给出一种句法分析树。如果还有其他可能的句法分析树结构,请直接给出来(不必写具体的生成过程)。