山东大学 计算机科学与技术 学院

自然语言处理 课程实验报告

学号: 201600181073 | 姓名: 唐超 | 班级: 智能 16

实验题目: 句法分析与依存句法分析

实验学时: 实验日期: 2019.4.1

实验目的:

对于给定的 chinese_sen. txt (中文 5 句)与 english_sen. txt (英文 3 句)中的句子,分别进行句法分析和依存句法分析。

软件环境:

Jupyter Notebook&python3.6

实验步骤与内容:

使用 StanfordCoreNLP 尝试对中英文两个语料进行句法分析和依存句法分析,使 用 LTP 仅对中文语料进行依存句法分析,并对比一下 StanfordCoreNLP 与 LTP 在中文语料上 进行依存句法分析结果的不同。

结论分析与体会:

实验结果:

用 stanfordCoreNLP 在英文文档上做句法分析与依存句法分析,结果分别如下:

```
Constituency Parsing: (ROOT
(S
(NP (NN Nothing))
(VP (VBZ is)
(ADJP (JJ impossible)
(PP (TO to)
(NP (DT a) (JJ willing) (NN heart)))))
(. .)))

Dependency Parsing: [('ROOT', 0, 3), ('nsubj', 3, 1), ('cop', 3, 2), ('case', 7, 4), ('det', 7, 5), ('amod', 7, 6), ('nmod', 3, 7), ('punct', 3, 8), ('ROOT', 0, 11), ('case', 5, 1), ('case', 5, 2), ('compound', 5, 3), ('compound', 5, 4), ('advcl', 11, 5), ('punct', 11, 6), ('csubj', 11, 7), ('dobj', 7, 8), ('cop', 11, 9), ('advmod', 11, 1 0), ('punct', 11, 12), ('ROOT', 0, 3), ('amod', 2, 1), ('nsubj', 3, 2), ('case', 5, 4), ('nmod', 9, 5), ('punct', 9, 6), ('amod', 8, 7), ('nsubj', 9, 8), ('ccomp', 3, 9), ('case', 11, 10), ('nmod', 9, 11), ('punct', 3, 12)]
```

用 stanfordCoreNLP 在中文文档上做句法分析与依存句法分析,结果分别如下:

```
Constituency Parsing: (ROOT
  (IP
    (IP
      (LCP
        (IP
          (NP (NN 学历))
          (VP (VV 造假)
            (NP
              (NP (NN 风波))
              (QP (CD —))
              (NP (NN 月)))))
        (LC 后))
      (PU , )
      (VP (VV 翟天)
        (IP
          (VP (VV 临)
            (IP
              (VP
                (VP
                  (PP (P 与)
                    (NP (NR 辛芷蕾)))
                  (VP (VV 牵)
                    (NP (NN 手))))
                (VP (VV 回家)))))))
      (PU 。)
      (IP
        (NP
          (ADJP (JJ 武磊))
          (NP (NN 替补)))
        (VP (VV 登场))))
```

Dependency Parsing: [('ROOT', 0, 2), ('nsubj', 2, 1), ('nsubj', 18, 3), ('dep', 5, 4), ('advmod:loc', 18, 5), ('case', 5, 6), ('punct', 18, 7), ('dep', 18, 8), ('ccom p', 8, 9), ('case', 11, 10), ('nmod:prep', 12, 11), ('conj', 9, 12), ('dobj', 12, 1 3), ('conj', 12, 14), ('punct', 18, 15), ('amod', 17, 16), ('nsubj', 18, 17), ('ccom p', 2, 18), ('punct', 2, 19), ('name', 21, 20), ('nsubj', 24, 21), ('dep', 24, 22), ('mark:clf', 22, 23), ('conj', 2, 24), ('dobj', 24, 25), ('nsubj', 27, 26), ('conj', 24, 27), ('dobj', 27, 28), ('punct', 2, 29), ('compound:nn', 31, 30), ('compound:nn', 33, 31), ('compound:nn', 33, 32), ('nsubj', 34, 33), ('conj', 2, 34), ('punct', 37, 3 5), ('compound:nn', 37, 36), ('nsubj', 40, 37), ('punct', 37, 38), ('advmod', 40, 3 9), ('ccompound:nn', 34, 40), ('compound:nn', 43, 41), ('compound:nn', 43, 42), ('dep', 46, 47) 9), ('ccomp', 34, 40), ('compound:nn', 43, 41), ('compound:nn', 43, 42), ('dep', 46, 43), ('case', 45, 44), ('nmod:prep', 46, 45), ('ccomp', 40, 46), ('punct', 2, 47), ('compound:nn', 50, 48), ('compound:nn', 50, 49), ('nmod:tmod', 60, 50), ('punct', 6 0, 51), ('nsubj', 54, 52), ('advmod', 54, 53), ('amod', 58, 54), ('mark', 54, 55), ('amod', 57, 56), ('compound:nn', 58, 57), ('nsubjpass', 60, 58), ('auxpass', 60, 5 9), ('conj', 2, 60), ('ccomp', 60, 61), ('ccomp', 61, 62), ('compound:nn', 65, 63), ('compound:nn', 65, 64), ('dobj', 62, 65), ('punct', 2, 66), ('det', 68, 67), ('nsub j', 70, 68), ('advmod', 70, 69), ('conj', 2, 70), ('nummod', 73, 71), ('mark:clf', 71, 72), ('nsubj', 74, 73), ('ccomp', 70, 74), ('nummod', 78, 75), ('mark:clf', 75, 75) 6), ('compound:nn', 78, 77), ('dobj', 74, 78), ('punct', 2, 79), ('nsubj', 81, 80), ('conj', 2, 81), ('punct', 81, 82), ('amod', 84, 83), ('nmod:assmod', 88, 84), ('cas e', 84, 85), ('nummod', 88, 86), ('mark:clf', 86, 87), ('dep', 90, 88), ('name', 90, 89), ('dep', 97, 90), ('punct', 97, 91), ('name', 93, 92), ('dep', 97, 93), ('punct', 97, 94), ('cc', 97, 95), ('name', 97, 96), ('nsubj', 98, 97), ('ccomp', 81, 98), ('au x:asp', 98, 99), ('nmod', 102, 100), ('case', 100, 101), ('dobj', 98, 102), ('punct', 2, 103)]

另外,用 pyltp 在中文文档上做依存句法分析的结果如下:

図家 210 成胜 西班牙人 で で で で で で で で で	学历	造假	风波	一月	后	,	翟	天临	与	辛芷蕾	牵手
2:0 战胜 威影少 西班牙人 官方 微博 肉地 宣布 名件 《复仇者 联盟 4 》 正式 定档 名用 3月 29日 35 所 ACM 2019年 0 3月 29日 36 所 高校 37 2019年 0 本种 37 37 38 37 38 37 38 37 38 37 38 37 38 38 38 38 38 38 38 38 38 38 38 38 38	回家		74.17		16	V.L			_	-	50 db
一次 一次 一次 一次 一次 一次 一次 一次					梅西	独	中	内	兀	助	巴萨
定档 4月 24日 在 内地 上映 小 呼声 极 高 的 人工智能 专业 被 列入 新增 审批 本科 专业 名单 , 名单 , 高校 获 首 批 建设 资格。 新增 审批 本科 专业 名单 , 全国 共有 35 所 高校 获 首 批 建设 资格。 不 位 创造者 Yoshu Bengio , Yann LeCun , 以及 Geoffrey Hinton 获得 了 2019年 的 图灵奖。 图表表表表表表表表表表表表表表表表表表表表表表表表表表表表表表表表表表表表							-	70/ -E		**	h
专业 初月 29日 下午 , 呼声 极 高 的 人工智能 35 所 高校 获 首 批 建设 资格 。 ACM 宣布 , 深度 学习 的 三 位 创造者 Yoshu a Bengio , 2019年 N LeCun , Ecun , U及 Geoffrey Hinton 获得 7 2019年 的 图灵奖 N N W Hinton 获得 8 Bengio , 2019年 N LeCun , UA UA W Hinton 获得 7 2019年 的 D N V N UA N <td>漫</td> <td></td> <td></td> <td></td> <td></td> <td>**</td> <td>复仇者</td> <td>联盟</td> <td>4</td> <td>>></td> <td>止式</td>	漫					**	复仇者	联盟	4	>>	止式
专业 被 列入 新増 审批 本科 专业 名单 、全国 共有 35 所 高校 获 首 批 建设 资格。 。 ACM 宣布 , 深度 学习 的 三 位 创造者 Yoshu T 2019年 的 图灵奖 。 Bengio , Yann LeCun , 以及 Geoffrey Hinton 获得 T 2019年 的 图灵奖 。 Bengio , Yann LeCun , 以及 Geoffrey Hinton 获得 T 2019年 的 n	定档				内地			_			
ACM 宣布 京校 京校 学习 的 三 位 创造者 Yoshu 本 Yoshu 京校 学习 的 三 位 创造者 Yoshu 京校 で で で で で で で で で						- /			的		
ACM 宣布 、 Yann LeCun 学习 的 三 位 创造者 Yoshu 获得 了 2019年 的 国灵奖 。 以及 Geoffrey Hinton 获得 n v n nt nd wp nh v p nh v v wp nh v wp nh v p nh v v mp nh v wp nh d v m q v nz m v n wp v n nz p nl v wp n n m wp a v nt nt p nl n	•		7 4.		审批				,	全国	共有
Bengio	35			获							
了 2019年的 的 图灵奖 。 n v n nt nd wp nh v p nh v v wp nh v wp nh d v m q v nz m v n wp v n n nz v wp n n m wp a v nt nt p nl v wp v nt nt nt p nl		ACM	宣布	,	深度	学习		Ξ	位	创造者	
n v n nt nd wp nh v p nh v v wp nh d v m q v nz m v n wp v n n nz v wp n n m wp a v nt nt p n1 v wp n d a u n np n1 n n p v v v n n n wp n w m q v n n m wp n wp n wp n wp n wp n wp n </td <td></td> <td>Bengio</td> <td></td> <td></td> <td>LeCun</td> <td>,</td> <td>以及</td> <td>Geoffre</td> <td>У</td> <td>Hinton</td> <td>获得</td>		Bengio			LeCun	,	以及	Geoffre	У	Hinton	获得
v wp nh v v wp nh d v m q v nz m v n wp v n n nz v wp n n m wp a v nt nt p n1 v wp v nt nt nt nt p n1 v mp v v v n n n wp n v mp q v n n mp m	了	2019年	的	图灵奖	0						
v nz m v n wp v n n nz v wp n n n wp a v nt nt p n1 v wp v nt nt nt p n1 n np n1 v mp v v v m q v n wp n wp n v mp q v m q v n m wp n ws ws wp c ws ws v u nt u n ws ws wp c ws ws v u nt u n ws ws wp c ws ws <t< td=""><td>n</td><td>v</td><td>n</td><td>nt</td><td>nd</td><td>wp</td><td>nh</td><td>V</td><td>р</td><td>nh</td><td>V</td></t<>	n	v	n	nt	nd	wp	nh	V	р	nh	V
wp n n m wp a v nt nt p n1 v wp v nt nt nt wp n d a u n n p v v v n n n wp n v m q v n q v n wp ws wp v wp n v u m q n ws ws wp ws ws wp c ws ws v u nt u n ws ws wp c ws ws v u nt u n ws ws wp c ws ws v u nt u n 11:C00 8:WP 15:SBV 8:C00 15:WP 20:SBV 20:ADV <th< td=""><td>V</td><td>wp</td><td>nh</td><td>V</td><td>V</td><td>wp</td><td>nh</td><td>d</td><td>V</td><td>m</td><td>q</td></th<>	V	wp	nh	V	V	wp	nh	d	V	m	q
v wp v nt nt nt nt wp n d a u n n p v v v n n n wp n v m q v n q v n wp ws wp ws wp v wp n v u m q n ws ws wp ws ws wp c ws ws v u nt u n n ws ws wp ws ws wp c ws ws v u nt u n <	V	nz	m	V	n	wp	V	n	n	nz	V
n n p v v v v n n n wp n v m q v n q v n wp ws ws v wp n v u m q n ws ws wp ws ws ws ws v u nt u n n wp ws ws ws ws ws v u nt u n ws ws ws ws ws ws ws v u nt u n n n n n ws ws ws ws ws ws ws v u nt u n n n n n n n n n n n n n n n n n	wp	n	n	m	wp	a	V	nt	nt	р	n1
v m q j v m q v n wp ws v wp n v u m q n ws ws wp ws ws ws ws v u nt u n 2:ATT 3:ATT 5:ATT 8:ADV 5:WP 8:SBV 0:HED 11:ADV 9:POB 8:COO 11:COO 8:WP 15:SBV 20:SBV 20:ADV 15:COO 22:ATT 23:SB V 8:COO	V	wp	v	nt	nt	nt	wp	n	d	a	u
v wp n v u m q n ws ws wp ws ws ws ws v u nt u n ws u nt u nt u n ws u nt u nt u n ws u nt u nt u nt u ws u nt st u nt	n	n	p		V	V	n	n	n	wp	n
ws ws wp c ws ws v u nt u n wp 2:ATT 3:ATT 5:ATT 5:ATT 8:ADV 5:WP 8:SBV 0:HED 11:ADV 9:POB 8:COO 11:COO 8:WP 15:SBV 8:COO 15:WP 20:SBV 20:ADV 15:COO 22:ATT 23:SB V 20:VOB 23:VOB 26:ADV 23:VOB 26:VOB 8:WP 30:ATT 31:ATT 32:ATT 33:SB V 8:COO 37:WP 36:ATT 37:ATT 45:SBV 37:WP 45:ADV 45:SBV 42:ATT 45:AD V 45:ADV 43:POB 33:VOB 8:WP 8:COO 49:ATT 50:ATT 47:VOB 8:WP 54:SB V 54:ADV 57:ATT 54:RAD 57:ATT 59:FOB 59:ADV 8:COO 59:VOB 62:ATT 63:AT T 64:ATT 60:VOB 8:WP 67:SBV 8:COO 69:ATT <td>V</td> <td>m</td> <td>q</td> <td>j</td> <td>V</td> <td>m</td> <td>q</td> <td>v</td> <td>n</td> <td>wp</td> <td>WS</td>	V	m	q	j	V	m	q	v	n	wp	WS
wp 2:ATT 3:ATT 5:ATT 5:ATT 8:ADV 5:WP 8:SBV 0:HED 11:ADV 9:POB 8:COO 11:COO 8:WP 15:SBV 8:COO 15:COO 15:WP 20:SBV 20:ADV 15:COO 22:ATT 23:SB V 20:VOB 23:VOB 26:ADV 23:VOB 26:VOB 8:WP 30:ATT 31:ATT 32:ATT 33:SB V 8:COO 37:WP 36:ATT 37:ATT 45:SBV 37:WP 45:ADV 45:SBV 42:ATT 45:AD V 45:ADV 43:POB 33:VOB 8:WP 8:COO 49:ATT 50:ATT 47:VOB 8:WP 54:SB V 54:ADV 57:ATT 54:RAD 57:ATT 59:FOB 59:ADV 8:COO 59:VOB 62:ATT 63:AT T 64:ATT 60:VOB 8:WP 67:SBV 8:COO 69:ATT 70:ATT 67:DBL 67:VOB 73:AT T 75:ATT 75:ATT 71:WP	V	wp	n	v	u	m	q	n	WS	WS	wp
2:ATT 3:ATT 5:ATT 5:ATT 8:ADV 5:WP 8:SBV 0:HED 11:ADV 9:POB 8:COO 11:COO 8:WP 15:SBV 8:COO 15:COO 15:WP 20:SBV 20:ADV 15:COO 22:ATT 23:SB V 20:VOB 23:VOB 26:ADV 23:VOB 26:VOB 8:WP 30:ATT 31:ATT 32:ATT 33:SB V 8:COO 37:WP 36:ATT 37:ATT 45:SBV 37:WP 45:ADV 45:SBV 42:ATT 45:AD V 45:ADV 43:POB 33:VOB 8:WP 8:COO 49:ATT 50:ATT 47:VOB 8:WP 54:SB V 54:ADV 57:ATT 54:RAD 57:ATT 59:FOB 59:ADV 8:COO 59:VOB 62:ATT 63:AT T 64:ATT 60:VOB 8:WP 67:SBV 8:COO 69:ATT 70:ATT 67:DBL 67:VOB 73:AT T 75:ATT 75:ATT 71:WP 78:SBV	WS	WS	wp	c	WS	WS	V	u	nt	u	n
11:COO 8:WP 15:SBV 8:COO 15:COO 15:WP 20:SBV 20:ADV 15:COO 22:ATT 23:SB V 20:VOB 23:VOB 26:ADV 23:VOB 26:VOB 8:WP 30:ATT 31:ATT 32:ATT 33:SB V 8:COO 37:WP 36:ATT 37:ATT 45:SBV 37:WP 45:ADV 45:SBV 42:ATT 45:AD V 45:ADV 43:POB 33:VOB 8:WP 8:COO 49:ATT 50:ATT 47:VOB 8:WP 54:SB V 54:ADV 57:ATT 54:RAD 57:ATT 59:FOB 59:ADV 8:COO 59:VOB 62:ATT 63:AT T 64:ATT 60:VOB 8:WP 67:SBV 8:COO 69:ATT 70:ATT 67:VOB 73:AT T 75:ATT 75:ATT 71:WP 78:SBV 71:COO 78:WP 81:ADV 85:ATT 81:RA D 84:ATT 85:ATT 87:ATT 87:ATT 95:SBV 87:WP 90:ATT 95:SBV 90:WP 94:LA	wp										
V 20:VOB 23:VOB 26:ADV 23:VOB 26:VOB 8:WP 30:ATT 31:ATT 32:ATT 33:SB V 8:COO 37:WP 36:ATT 37:ATT 45:SBV 37:WP 45:ADV 45:SBV 42:ATT 45:AD V 45:ADV 43:POB 33:VOB 8:WP 8:COO 49:ATT 50:ATT 47:VOB 8:WP 54:SB V 54:ADV 57:ATT 54:RAD 57:ATT 59:FOB 59:ADV 8:COO 59:VOB 62:ATT 63:AT T 64:ATT 60:VOB 8:WP 67:SBV 8:COO 69:ATT 70:ATT 67:VOB 73:AT T 75:ATT 75:ATT 71:WP 78:SBV 71:COO 78:WP 81:ADV 85:ATT 81:RA D 84:ATT 85:ATT 87:ATT 87:ATT 95:SBV 87:WP 90:ATT 95:SBV 90:WP 94:LA	2:ATT	3:ATT	5:ATT	5:ATT	8:ADV	5:WP	8:SBV	0:HED	11:ADV	9:POB	8:C00
V 8:COO 37:WP 36:ATT 37:ATT 45:SBV 37:WP 45:ADV 45:SBV 42:ATT 45:AD V 45:ADV 43:POB 33:VOB 8:WP 8:COO 49:ATT 50:ATT 47:VOB 8:WP 54:SB V 54:ADV 57:ATT 54:RAD 57:ATT 59:FOB 59:ADV 8:COO 59:VOB 62:ATT 63:AT T 64:ATT 60:VOB 8:WP 67:SBV 8:COO 69:ATT 70:ATT 67:DBL 67:VOB 73:AT T 75:ATT 75:ATT 71:VOB 71:WP 78:SBV 71:COO 78:WP 81:ADV 85:ATT 81:RA D 84:ATT 85:ATT 87:ATT 87:ATT 95:SBV 87:WP 90:ATT 95:SBV 90:WP 94:LA	11:COO	8:WP	15:SBV	8:C00	15:COO	15:WP	20:SBV	20:ADV	15:COO	22:ATT	23:SB
V 45:ADV 43:POB 33:VOB 8:WP 8:COO 49:ATT 50:ATT 47:VOB 8:WP 54:SB V 54:ADV 57:ATT 54:RAD 57:ATT 59:FOB 59:ADV 8:COO 59:VOB 62:ATT 63:AT T 64:ATT 60:VOB 8:WP 67:SBV 8:COO 69:ATT 70:ATT 67:DBL 67:VOB 73:AT T 75:ATT 75:ATT 71:VOB 71:WP 78:SBV 71:COO 78:WP 81:ADV 85:ATT 81:RA D 84:ATT 85:ATT 87:ATT 87:ATT 95:SBV 87:WP 90:ATT 95:SBV 90:WP 94:LA	V	20:VOB	23:VOB	26:ADV	23:VOB	26:VOB	8:WP	30:ATT	31:ATT	32:ATT	33:SB
V 54:ADV 57:ATT 54:RAD 57:ATT 59:FOB 59:ADV 8:COO 59:VOB 62:ATT 63:AT	V	8:C00	37:WP	36:ATT	37:ATT	45:SBV	37:WP	45:ADV	45:SBV	42:ATT	45:AD
T 64:ATT 60:VOB 8:WP 67:SBV 8:COO 69:ATT 70:ATT 67:DBL 67:VOB 73:ATT 75:ATT 75:ATT 71:VOB 71:WP 78:SBV 71:COO 78:WP 81:ADV 85:ATT 81:RA D 84:ATT 85:ATT 87:ATT 87:ATT 95:SBV 87:WP 90:ATT 95:SBV 90:WP 94:LA	V	45:ADV	43:POB	33:VOB	8:WP	8:C00	49:ATT	50:ATT	47:VOB	8:WP	54:SB
T 75:ATT 75:ATT 71:VOB 71:WP 78:SBV 71:COO 78:WP 81:ADV 85:ATT 81:RA D 84:ATT 85:ATT 87:ATT 87:ATT 95:SBV 87:WP 90:ATT 95:SBV 90:WP 94:LA	V	54:ADV	57:ATT	54:RAD	57:ATT	59:FOB	59:ADV	8:C00	59:VOB	62:ATT	63:AT
T 75:ATT 75:ATT 71:VOB 71:WP 78:SBV 71:COO 78:WP 81:ADV 85:ATT 81:RA D 84:ATT 85:ATT 87:ATT 87:ATT 95:SBV 87:WP 90:ATT 95:SBV 90:WP 94:LA	T	64:ATT	60:VOB	8:WP	67:SBV	8:C00	69:ATT	70:ATT	67:DBL	67:VOB	73:AT
D 84:ATT 85:ATT 87:ATT 87:ATT 95:SBV 87:WP 90:ATT 95:SBV 90:WP 94:LA		75:ATT	75:ATT	71:VOB	71:WP	78:SBV	71:C00	78:WP	81:ADV	85:ATT	81:RA
D 94:ATT 95:SBV 78:VOB 95:RAD 99:ATT 97:RAD 95:VOB 8:WP		84:ATT	85:ATT	87:ATT	87:ATT	95:SBV	87:WP	90:ATT	95:SBV	90:WP	94:LA
	D	94:ATT	95:SBV	78:VOB	95:RAD	99:ATT	97:RAD	95:VOB	8:WP		

结论分析与体会:

StanfordCoreNLP 的句法分析与依存句法分析在中英文文档上都可以使用,且效果都比较好。