

论文写作的易读性原则

案例分析：基于Seq2Seq的对话数据增广

报告人：刘一佳

合作者：侯宇泰、车万翔、刘挺

<http://yjliu.net/cv/res/2018-08-19-nlpcc-sws.compressed.pdf>

学术报告中的一些设计技巧

报告人：刘一佳

导师：秦兵、车万翔

错误地利用 报告与论文结构的相似性

Challenges and Contribution

- The first challenge is deriving an optimal alignment in ambiguous situations.
- The second challenge is how to automatically extract word alignment from the alignment graph.
- The third challenge is how to use the rule-based and unsupervised alignment to derive the alignment with downstream tasks learning.
- We proposed an automatic aligner based on random forest model.

简介

Overview



模型

Our aligner algorithm

- Extracting aligner with random forest
- Producing alignment

模型

Our oracle parser

模型

Experiments

- We conduct experiments on LDC2004T22
- We evaluate the alignment F-score and Smatch of resulted parser

实验

Conclusion

- We propose a new HMM aligner which is used by a novel transition-based HMM parser. Our aligner is able to extract word alignment from ambiguous situations.
- Both the new aligner and parser show the effectiveness in extracting word alignment from ambiguous situations.
- We also develop transition-based HMM parser based on our aligner and random forest, and it achieves a performance of 88.3 Smatch F1 score on automatic word alignment with only words and POS tags as input.

结论

思考题

- 为什么做学术报告
 - 为了更好地交流
- 做怎样的学术报告
 - ☐ “向听众展示我对问题的深入理解”
 - ☐ “让听众明白我的论文中的技术”
 - ☐ “引起听众的兴趣”

思考题

- 为什么做学术报告
 - 为了更好地交流
- 做怎样的学术报告
 - ☐ “向听众展示我对问题的深入理解”
 - ☐ “让听众明白我的论文中的技术”
 - ☒ “引起听众的兴趣”

听众模型

理想中的听众

- 领域专家
- 已经读过你的论文
- 对于你的工作非常感兴趣

现实中的听众

- 来自其他领域
- 刚刚了解到你的工作
- 这个时段没什么可听的，恰巧发现这屋子网络比较好

类比审稿人模型

审稿

你以为审稿人应该这样审稿的：

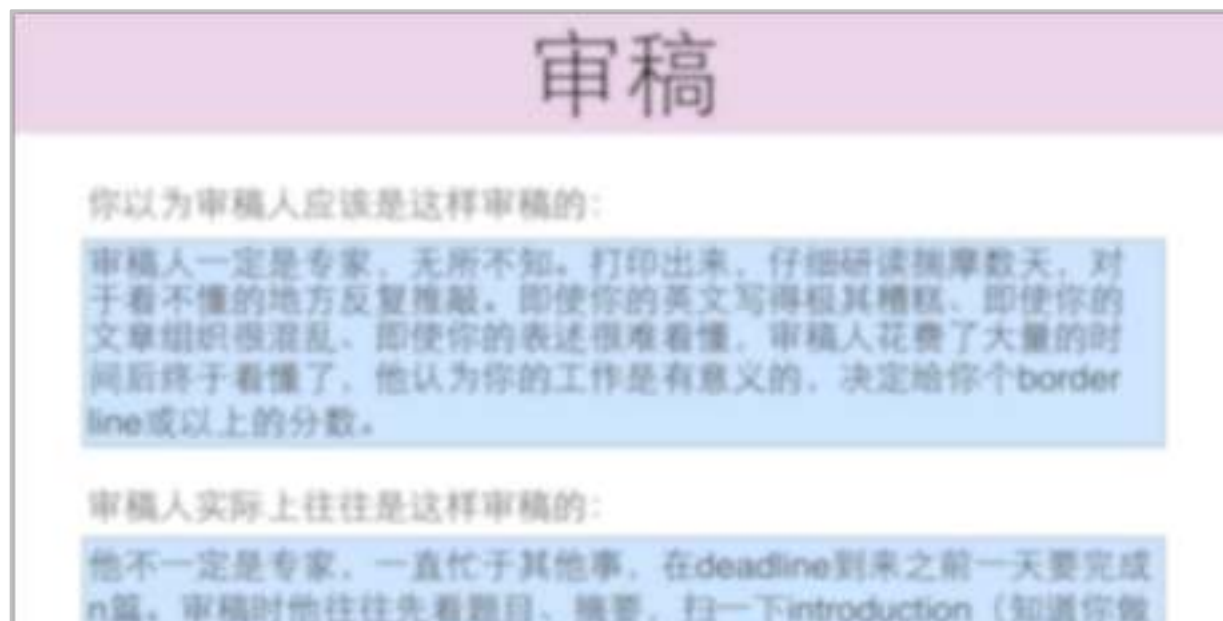
审稿人一定是专家，无所不知。打印出来，仔细研读揣摩数天，对于看不懂的地方反复推敲。即使你的英文写得极其糟糕、即使你的文章组织很混乱、即使你的表述很难看懂，审稿人花费了大量的时间后终于看懂了，他认为你的工作是有意义的，决定给你个border line或以上的分数。

审稿人实际上往往是这样审稿的：

他不一定是专家，一直忙于其他事，在deadline到来之前一天要完成n篇。审稿时他往往先看题目、摘要，扫一下introduction（知道你做什么），然后直接翻到最后找核心实验结果（做得好不好），然后基本确定录还是不录（也许只用5分钟！）。如果决定录，剩下就是写些赞美的话，指出些次要的小毛病。如果决定拒，下面的过程就是细看中间部分找理由拒了。

第一印象定录拒，5分钟内打动审稿人


类比审稿人模型



“You have **two minutes** to engage your audience before they start to doze.” -- Simon Peyton Jones in *How to give a great research talk*

简介部分：展示最好的部分

(Zhang and Nivre 2011, Martins et al. 2013)



Our Work

- A neural network based dependency parser!

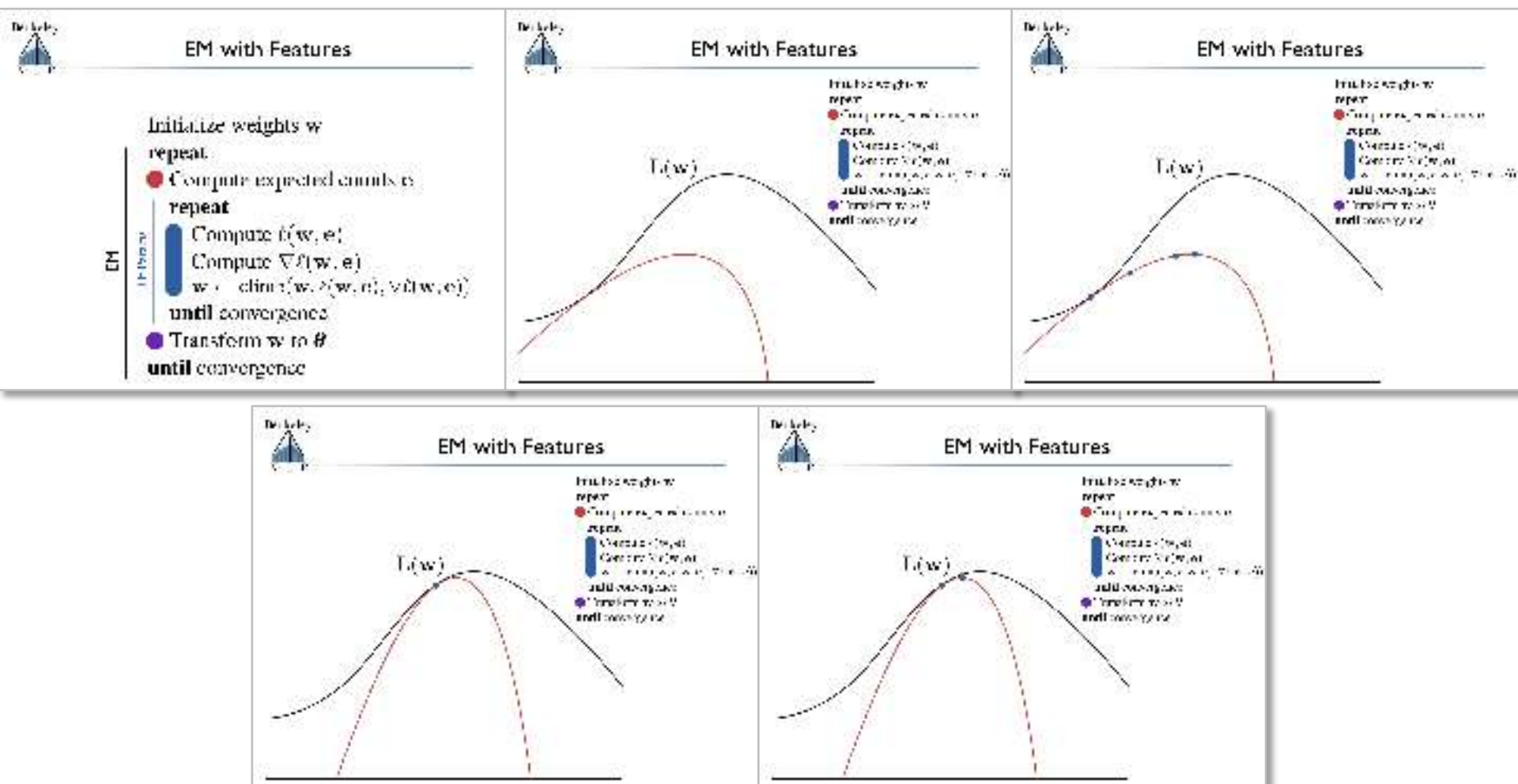
Parsing on English Penn Treebank (§23):

| | | Unlabeled attachment score (UAS) | sent / s |
|----------------------|----------------------------|----------------------------------|-------------|
| Transition -based | MaltParser (greedy) | 89.9 | 560 |
| | Our Parser (greedy) | 92.0 | 1013 |
| | Zpar: beam = 64 | 92.9* | 29* |
| Graph -based | MSTParser | 92.0 | 12 |
| | TurboParser | 93.1* | 31* |

A Fast and Accurate Dependency Parser using Neural Networks

4

模型部分：多用例子



Taylor Berg-Kirkpatrick, Alexandre Bouchard-Côté, John DeNero, and Dan Klein.
2010. Painless Unsupervised Learning with Features, 第28到54页

模型部分：反例

| Transition | Current State | Resulting State | Description |
|------------|--|---|--|
| DROP | $[\sigma s_0, \delta, b_0 \beta, A]$ | $[\sigma s_0, \delta, \beta, A]$ | pops out the word that doesn't convey any semantics (e.g., function words and punctuations). |
| MERGE | $[\sigma s_0, \delta, b_0 b_1 \beta, A]$ | $[\sigma s_0, \delta, b_0 \cdot b_1 \beta, A]$ | concatenates a sequence of words into a span, which can be derived as a named entity (name) or date-entity. |
| CONFIRM(c) | $[\sigma s_0, \delta, b_0 \beta, A]$ | $[\sigma s_0, \delta, c \beta, A]$ | derives the first element of the buffer (a word or span) into a concept c. |
| ENTITY(c) | $[\sigma s_0, \delta, b_0 \beta, A]$ | $[\sigma s_0, \delta, c \beta, A \cup \text{relations}(c)]$ | a special form of CONFIRM that derives the first element into an entity and builds the internal entity AMR fragment. |
| NEW(c) | $[\sigma s_0, \delta, b_0 \beta, A]$ | $[\sigma s_0, \delta, c b_0 \beta, A]$ | generates a new concept c and pushes it to the front of the buffer. |
| LEFT(r) | $[\sigma s_0, \delta, b_0 \beta, A]$ | $[\sigma s_0, \delta, b_0 \beta, A \cup \{s_0 \xleftarrow{r} b_0\}]$ | links a relation r between the top concepts on the stack and the buffer. |
| RIGHT(r) | $[\sigma s_0, \delta, b_0 \beta, A]$ | $[\sigma s_0, \delta, b_0 \beta, A \cup \{s_0 \xrightarrow{r} b_0\}]$ | |
| CACHE | $[\sigma s_0, \delta, b_0 \beta, A]$ | $[\sigma, s_0 \delta, b_0 \beta, A]$ | passes the top concept of the stack onto the deque. |
| SHIFT | $[\sigma s_0, \delta, b_0 \beta, A]$ | $[\sigma s_0 \delta b_0, [], \beta, A]$ | shifts the first concept of the buffer onto the stack along with those on the deque. |
| REDUCE | $[\sigma s_0, \delta, b_0 \beta, A]$ | $[\sigma, \delta, b_0 \beta, A]$ | pops the top concept of the stack. |

实验部分：图比表格好

LDC2014T12 Experiments

- alignment F-score

| Aligner | Alignment F1 (on hand-align) | Oracle's Search (on dev. dataset) |
|---------|---------------------------------|--------------------------------------|
| JAMR | 90.6 | 91.3 |
| Our | 95.2 | 94.7 |

- parser improvements

| model | news14 | all |
|----------------------------------|--------|------|
| JAMR parser: Word, POS, NER, DEP | | |
| + JAMR aligner | 71.3 | 65.9 |
| + Our aligner | 73.1 | 67.6 |
| CAMR parser: Word, POS, NER, DEP | | |
| + JAMR aligner | 68.4 | 64.6 |
| + Our aligner | 68.8 | 65.1 |

Aligner Experiments: Two Open-sourced AMR Parsers



实验部分：图比表格好

信息元素的易理解度



图

| Query | Query | Top-10 F1 score | Top-10 Recall |
|-------|-------|-----------------|---------------|
| Q1 | Q1 | 0.71 | 0.71 |
| Q2 | Q2 | 0.71 | 0.71 |
| Q3 | Q3 | 0.71 | 0.71 |
| Q4 | Q4 | 0.71 | 0.71 |
| Q5 | Q5 | 0.71 | 0.71 |
| Q6 | Q6 | 0.71 | 0.71 |
| Q7 | Q7 | 0.71 | 0.71 |
| Q8 | Q8 | 0.71 | 0.71 |
| Q9 | Q9 | 0.71 | 0.71 |
| Q10 | Q10 | 0.71 | 0.71 |

表格

Shift-reduce parsing is efficient but suffers from parsing errors caused by syntactic ambiguity. Figure 3 shows two Spanish derivations for a dependency tree. Consider the item on the top, this item can either apply a shift action to move a new item or apply a reduce action to obtain a bigger structure. This is often referred to as conflict in the shift-reduce dependency parsing literature (Ulring et al., 2009). In this work, the shift-reduce parser does four types of conflict:

正文

$$\begin{aligned} & \frac{\partial L(\theta)}{\partial \theta_k} \\ &= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(x|y, \theta) \frac{\partial L(x, y, \theta)}{\partial \theta_k} \\ &= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(x|y, \theta) \frac{\partial L(x, y, \theta)}{\partial \theta_k} \end{aligned}$$

公式

Algorithm 1: A simple algorithm for the shift-reduce parser. The algorithm takes as input a sentence s and a dependency tree T . It returns a list of conflicts. The algorithm works as follows: 1. Initialize a list of conflicts. 2. For each node n in T , do: 3. If n is a root node, then add it to the list of conflicts. 4. If n is a child node, then check if its parent is in the list of conflicts. If yes, then add n to the list of conflicts. 5. If n is a leaf node, then check if its parent is in the list of conflicts. If yes, then add n to the list of conflicts. 6. Return the list of conflicts.

算法

Proof of Theorem 1: Let n^k be the weight before the k -th update. To prove that $n^k \geq 0$, suppose the k -th update is made at the node i . Let $n^k = (n^k_1, n^k_2, \dots, n^k_N)$. It follows from the definition of n^k that $n^k_1 = 0$, $n^k_2 = 1$, $n^k_3 = 0$, $n^k_4 = 0$. We take the product of both sides with the vector $\mathbf{1}$. Let $\mathbf{1} = (1, 1, \dots, 1)$. Then $\mathbf{1}^T n^k = 1$, where the inequality follows because of the property of $\mathbf{1}$ assumed in Eq. 3. Since $n^k \geq 0$, we therefore have $n^k_1 = 0$. It follows from the definition of n^k that $n^k_2 = 1$, $n^k_3 = 0$, $n^k_4 = 0$. Suppose $\mathbf{U} = (n^k_1, n^k_2, n^k_3, n^k_4)$. It follows that $\|\mathbf{U}\|_2 \geq \sqrt{1}$.

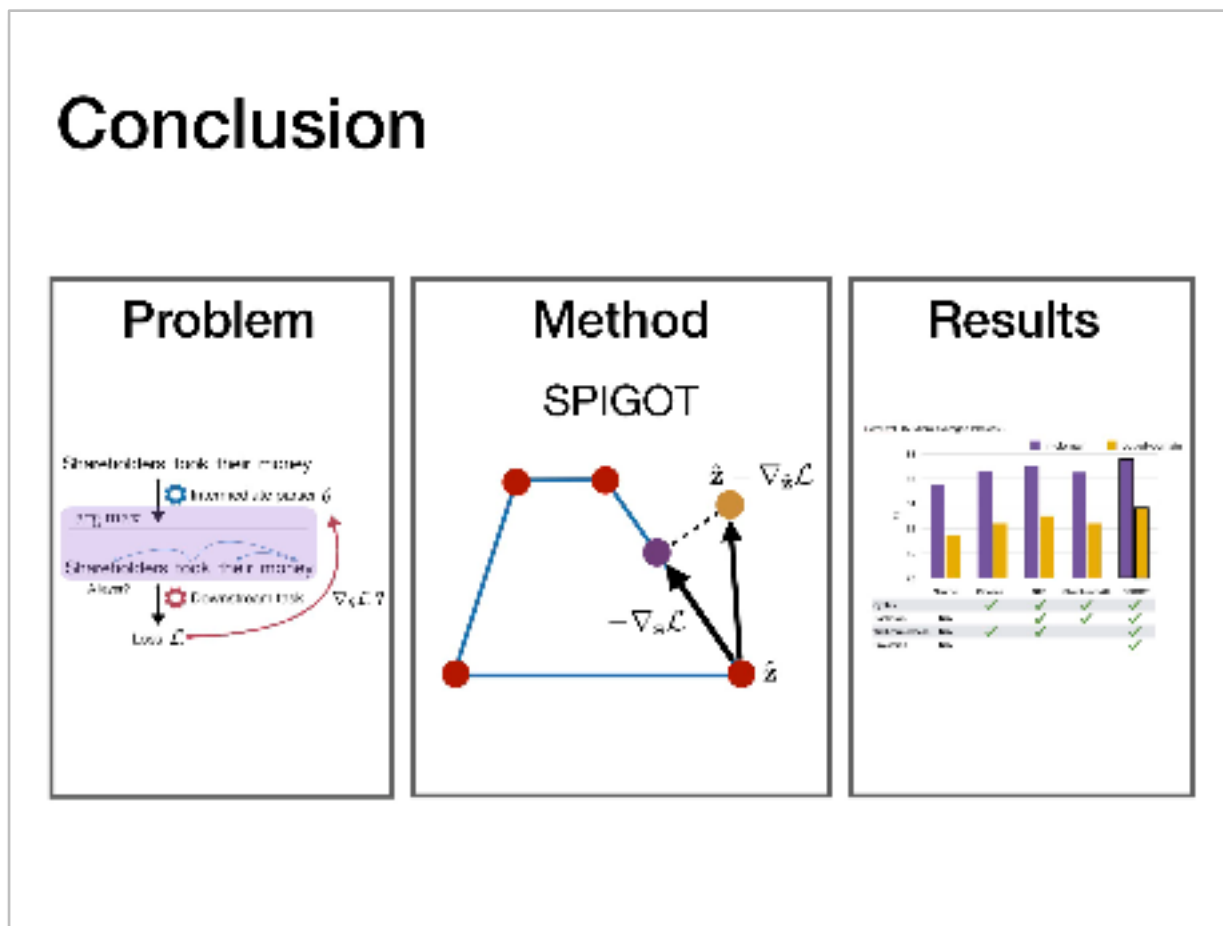
证明

实验部分：图比表格好



用图与例子来描述方法和实验

结论部分：新的展现形式

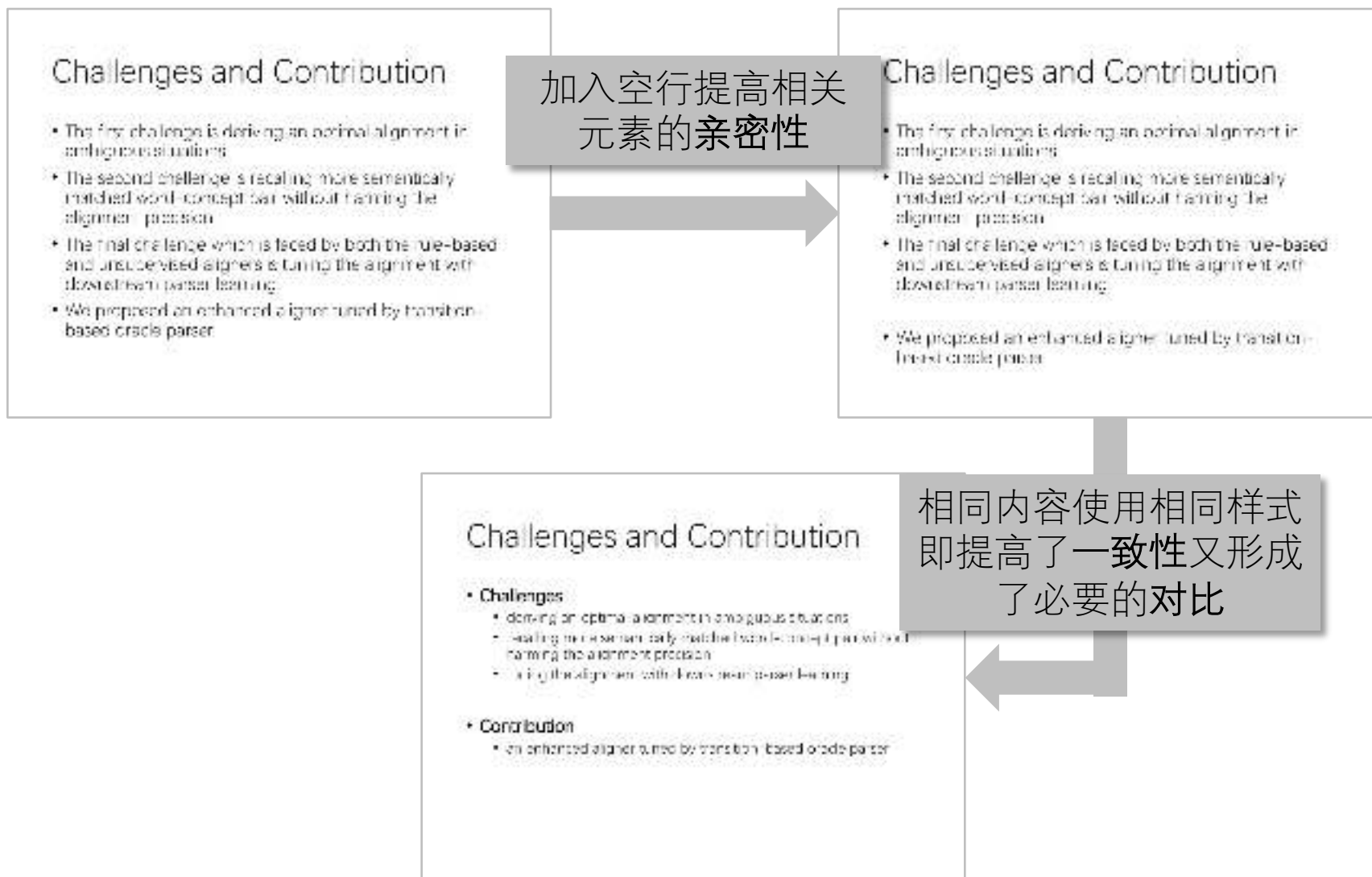


设计原则

- 亲密性：相关的元素应该组织到一起
- 重复：相同的内容达到形式的统一
- 对比：如果两项不完全相同，就应使之截然不同
- 对齐：使元素之间产生关联，有关联的都应对齐



根据设计原则做幻灯片



避免不对齐

Our aligner algorithm

- Enhancing aligner with rich semantic resources
- Producing multiple alignments

```
Input:  $A_0$ : AMR graph with a set of graph fragments  $C$ ;  
a sentence  $W$ ; a set of matching rules  $P_M$ ; and  
a set of updating rules  $P_U$ .  
Output: a set of alignments  $A_0$ .  
1 for  $c \in C$  do  
2    $A_c \leftarrow \emptyset$ .  
3 for  $\rho_M \in P_M$  do  
4   for  $\langle s_1, s_2 \rangle \leftarrow \text{query}(W)$  do  
5     for  $e \in C$  do  
6       if  $\rho_M(e, \langle s_1, s_2 \rangle)$  then  
7          $A_c \leftarrow A_c \cup \{ \langle s_1, s_2, e \rangle \}$ .  
8 updated  $\leftarrow$  true;  
9 while updated or  $C \neq \emptyset$  do  
10  updated  $\leftarrow$  false;  
11  for  $\rho_U \in P_U$  do  
12    for  $\langle s_1, s_2 \rangle \in C \times C$  do  
13      for  $\langle s_1, s_2, e \rangle \in A_c$  do  
14        if  $\rho_U(\langle s_1, s_2, e \rangle) \cap \langle s_1, s_2, e' \rangle \in A_c$  then  
15           $A_c \leftarrow A_c \cup \langle s_1, s_2, e' \rangle$ .  
16          updated  $\leftarrow$  true;  
17  $A \leftarrow \emptyset$ .  
18 for  $\langle a_1, \dots, a_n \rangle \in \text{CartesianProduct}(A_1, \dots, A_n)$  do  
19   legal  $\leftarrow$  true;  
20   for  $a \in \{a_1, \dots, a_n\}$  do  
21      $\langle s_1, s_2, e \rangle \leftarrow a$ .  
22      $\langle s_1', s_2', e' \rangle \leftarrow a_{op}$ .  
23     if  $s_1 \neq s_1' \vee s_2 \neq s_2'$  then  
24       legal  $\leftarrow$  false;  
25   if legal then  
26      $A \leftarrow A \cup \{ \langle a_1, \dots, a_n \rangle \}$ .
```

“乱” 的原因：视线跳动过多

Experiments

- We conduct experiments on LDC2014T12
- We evaluate the alignment F-score and Smatch of resulted parsers

| Aligner | Alignment F1 (on hand-align) | Oracle's Smatch (on dev. dataset) |
|---------|---------------------------------|--------------------------------------|
| JAMR | 90.6 | 91.7 |
| Our | 95.2 | 94.7 |

| model | newswire | all |
|----------------------------------|----------|------|
| JAMR parser: Word, POS, NER, DEP | | |
| + JAMR aligner | 71.3 | 65.9 |
| + Our aligner | 73.1 | 67.6 |
| CAMR parser: Word, POS, NER, DEP | | |
| + JAMR aligner | 68.4 | 64.6 |
| + Our aligner | 68.8 | 65.1 |

| model | newswire | all |
|---------------------------------------|----------|-------------|
| Our single parser: Word only | | |
| + JAMR aligner | 68.6 | 63.9 |
| + Our aligner | 69.3 | 64.7 |
| Our single parser: Word, POS | | |
| + JAMR aligner | 68.8 | 64.6 |
| + Our aligner | 69.8 | 65.2 |
| Our ensemble: Word only + Our aligner | | |
| x3 | 71.9 | 67.4 |
| x10 | 72.5 | 68.1 |
| Our ensemble: Word, POS + Our aligner | | |
| x3 | 72.5 | 67.7 |
| x10 | 73.3 | 68.4 |

“乱” 的原因：视线跳动过多

Experiments

- We conduct experiments on LDC2014T12
- We evaluate the alignment F-score and Smatch of resulted parsers

| Aligner | Alignment F1 (on hand-align) | Oracle's Smatch (on dev. dataset) |
|---------|---------------------------------|--------------------------------------|
| JAMR | 90.6 | 91.7 |
| Our | 95.2 | 94.7 |

| model | newswire | all |
|----------------------------------|----------|------|
| JAMR parser: Word, POS, NER, DEP | | |
| + JAMR aligner | 71.3 | 65.9 |
| + Our aligner | 73.1 | 67.6 |
| CAMR parser: Word, POS, NER, DEP | | |
| + JAMR aligner | 68.4 | 64.6 |
| + Our aligner | 68.8 | 65.1 |

| model | newswire | all |
|---------------------------------------|----------|------|
| Our single parser: Word only | | |
| + JAMR aligner | 68.6 | 63.9 |
| + Our aligner | 69.3 | 64.7 |
| Our single parser: Word, POS | | |
| + JAMR aligner | 68.8 | 64.6 |
| + Our aligner | 69.8 | 65.2 |
| Our ensemble: Word only + Our aligner | | |
| x3 | 71.9 | 67.4 |
| x10 | 72.5 | 68.1 |
| Our ensemble: Word, POS + Our aligner | | |
| x3 | 72.5 | 67.7 |
| x10 | 73.3 | 68.4 |

“乱” 的解法：重新组织内容

Experiments

- We conduct experiments on LDC2014T12
- We evaluate the alignment F-score and Smatch of resulted parsers

| Aligner | Alignment F1 (on hand-align) | Oracle's Smatch (on dev. dataset) |
|---------|---------------------------------|--------------------------------------|
| JAMR | 90.6 | 91.7 |
| Our | 95.2 | 94.7 |

| model | newswire | all |
|----------------------------------|----------|------|
| JAMR parser: Word, POS, NER, DEP | | |
| + JAMR aligner | 71.3 | 65.9 |
| + Our aligner | 73.1 | 67.6 |
| CAMR parser: Word, POS, NER, DEP | | |
| + JAMR aligner | 68.4 | 64.6 |
| + Our aligner | 68.8 | 65.1 |

| model | newswire | all |
|---------------------------------------|----------|------|
| Our single parser: Word only | | |
| + JAMR aligner | 68.6 | 63.9 |
| + Our aligner | 69.3 | 64.7 |
| Our single parser: Word, POS | | |
| + JAMR aligner | 68.8 | 64.6 |
| + Our aligner | 69.8 | 65.2 |
| Our ensemble: Word only + Our aligner | | |
| x3 | 71.0 | 67.4 |
| x10 | 72.5 | 68.1 |
| Our ensemble: Word, POS + Our aligner | | |
| x3 | 72.5 | 67.7 |
| x10 | 73.3 | 68.4 |

LDC2014T12 Experiments

- alignment F-score
- parser improvements

| Aligner | Alignment F1 (on hand-align) | Oracle's Smatch (on dev. dataset) |
|---------|---------------------------------|--------------------------------------|
| JAMR | 90.6 | 91.7 |
| Our | 95.2 | 94.7 |

| model | newswire | all |
|----------------------------------|----------|------|
| JAMR parser: Word, POS, NER, DEP | | |
| + JAMR aligner | 71.3 | 65.9 |
| + Our aligner | 73.1 | 67.6 |
| CAMR parser: Word, POS, NER, DEP | | |
| + JAMR aligner | 68.4 | 64.6 |
| + Our aligner | 68.8 | 65.1 |

视线跳动在论文写作中的作用

信息流的变化



参考文献

- Simon Peyton Jones: How to give a great talk
- 写给大家看的设计书
- 机器翻译学术论文写作方法与技巧
- 知乎专栏：跟我学个P

总结

(Zhong and Herve 2011, Martins et al 2013)

Our Work

- A neural network based dependency parser!

Parsing on English Penn Treebank (§23):

Unlabeled attachment score (UAS) sent / a

| Transition-based | MultiParser (greedy) | 88.9 | 560 |
|------------------|----------------------------|-------------|-------------|
| | Our Parser (greedy) | 92.0 | 1013 |
| | Zpar beam = 64 | 92.9 | 29 |
| Graph-based | MSTParser | 92.0 | 12 |
| | TurboParser | 93.1 | 31 |

$\times 1.8$

A Fast and Accurate Dependency Parser using Neural Networks

为了抓住听众，把最好的部分前置

EM with Features

模型部分有取舍，用好图和例子

Conclusion

Problem

Method

SPIGOT

Results

“结论”也有新思路

Challenges and Contributions

- Challenges**
 - non-trivial task: finding the right word
 - finding more semantically matched words to be used without losing the original meaning
 - finding the right word for downstream tasks
- Contributions**
 - an unsupervised algorithm for finding the right word

亲密性

重复

对比

对齐

四项设计的基本原则

祝大家产出优秀的学术工作