基于图的一致性翻译方法

刘树杰 李志灏 李沐 周明
微软亚洲研究院

翻译一致性问题 （1）

- 翻译一致性:
  - 如果一个翻译候选被其它的翻译候选支持，则该翻译候选称为一致的翻译候选
- 前人工作:
  - 最小贝叶斯风险重排序: 使用同一个N-Best列表收集一致性信息
  - 最小贝叶斯风险系统融合:使用来自不同解码器的同一个源语言句子的N-Best翻译候选收集一致性信息
- 最小贝叶斯风险重排序和系统融合均从同一个源语言句子的翻译候选中收集一致性信息来改善翻译结果

翻译一致性问题 （2）

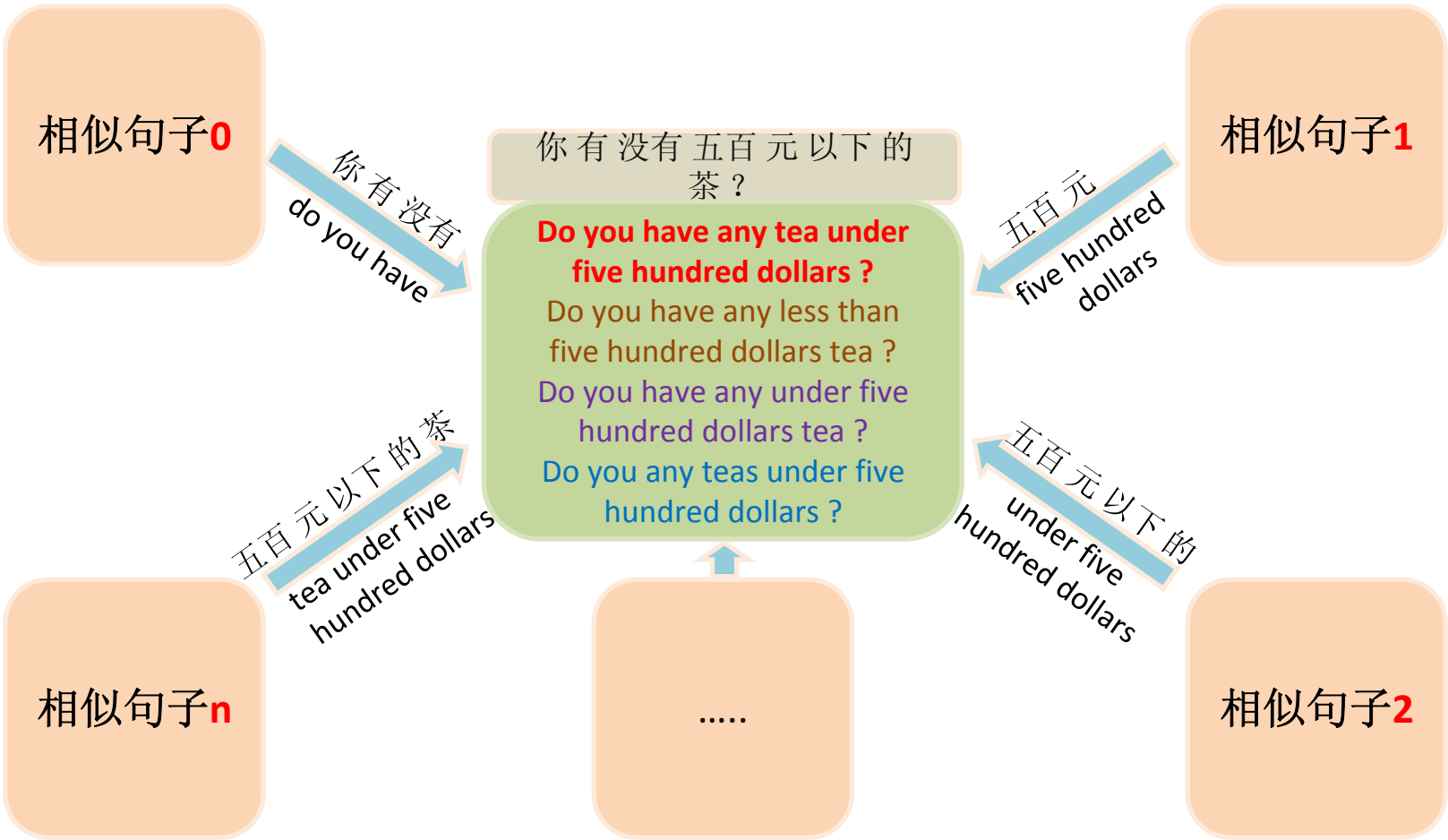
- 为什么不从相似的句子收集一致性信息?
  - 如果两个源语言句子比较相似，那么他们的目标语言句子也应当比较相似

IWSLT Chinese to English Translation Task	
Src	*你 有没有 五百 元 以下 的 茶 ?*
Ref	*Do you have any tea under five hundred dollars ?*
Best1	*Do you have any **less than five hundred dollars tea** ?*
Src	*我 想要 五百 元 以下 的 茶 .*
Ref	*I world like some tea under five hundred dollars .*
Best1	*I would like **tea under five hundred dollars** .*

第一个句子的正确翻译结果出现在N-Best翻译候选中，但是并没有被选为最好的翻译候选

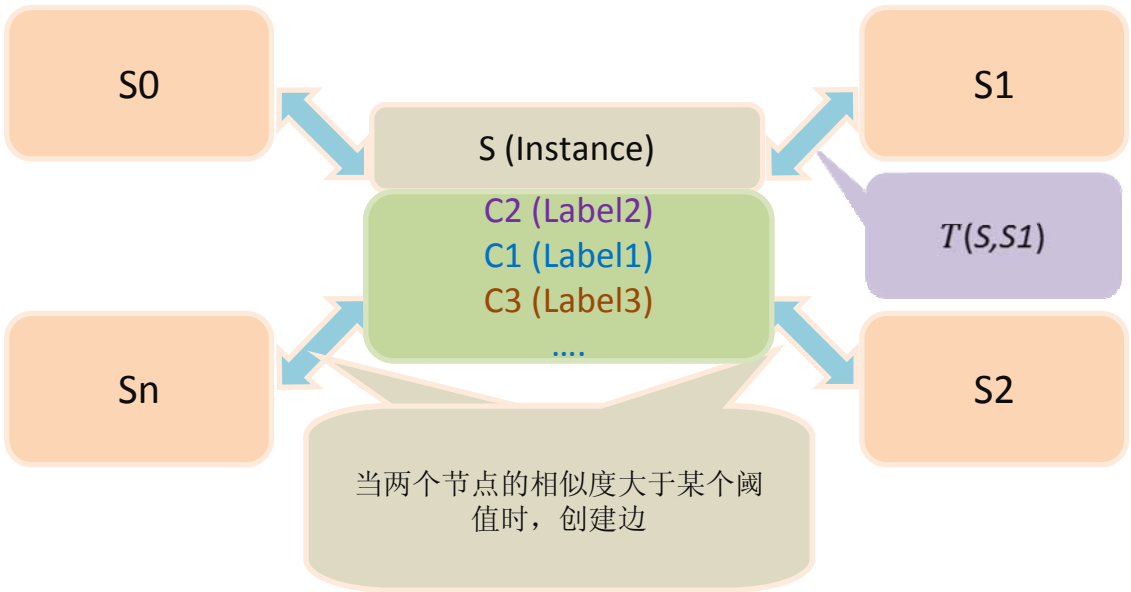
第二个句子对该相同片段的翻译结果可以被用来作为第一个句子翻译候选排序的参考信息

基本想法


相似句子0
你有没有
do you have
你有没有五百元以下的茶？
Do you have any tea under five hundred dollars ?
Do you have any less than five hundred dollars tea ?
Do you have any under five hundred dollars tea ?
Do you any teas under five hundred dollars ?
五百元
five hundred dollars
相似句子1
五百元以下的茶
tea under five hundred dollars
五百元以下的
under five hundred dollars
相似句子n
.....
相似句子2

基于图的学习方法（1）

- 基于图的学习方法很适合用来解决这样的问题
  - 基于图的学习方法可以通过考虑相似样本的标记来对当前样本进行预测
  - 原则：**如果两个样本比较相似，那么他们的标记倾向于相同**
- 标记传播算法
  - 对节点*i*的某个标记*l*的概率$P_{i,l}$可以通过考虑节点*i*的邻居节点集合*N(i)* 中所有的邻居节点对应的概率来更新



$$p_{f,e}^{t+1} = \sum_{f' \in N(f)} T(f,f') p_{f',e}^{t}$$

$$p(C2|S) = p(C2|S0) * T(S,S0) + p(C2|S1) * T(S,S1) + p(C2|S2) * T(S,S2) + \dots$$

基于图的学习方法（2）

- 标记传播算法

$$p_{i,l}^{t+1} = \sum_{j \in N(i)} T(i,j) p_{j,l}^{t} \qquad T(i,j) = \frac{w_{i,j}}{\sum_{j' \in N(i)} w_{i,j'}}$$

  - 通过相似的已标记样本来为未标记样本估计一个比较正确的标记
- 标记传播算法应用于机器翻译
  - **问题**：对于结构学习（机器翻译）问题，不同的样本（源语言句子）不太可能具有相同的标记 l（翻译结果），无法进行概率更新。

$$p(C2|S) = p(C2|S0) * T(S,S0) + p(C2|S1) * T(S,S1) + \cdots$$

$p(C2|S)$ =0，$p(C2|S0)$ =0，$p(C2|S1)$ =0

  - **解决方案**：一个能够使用不同的标记进行概率估算的更新规则

结构化的标记传播算法

- 结构化的标记传播更新规则

标记的传播概率

$$p_{f,e}^{t+1} = \sum_{f' \in N(f)} T_s(f,f') \sum_{e' \in H(f')} T_l(e,e') p_{f',e'}^{t}$$

$$p(C2|S) = (p(C1'|S0) * T_l(C1',C2) + p(C2'|S0) * T_l(C2',C2) + \cdots) * T_s(S,S0) + (p(C1''|S1) * T_l(C1'',C2) + p(C2''|S1) * T_l(C2'',C2) + \cdots) * T_s(S,S1) + \cdots$$

$$T_l(e,e') = \frac{sim(e,e')}{\sum_{e'' \in H(f')} sim(e,e'')}$$

标记的相似度

原始更新规则是新规则在如下定义下的特殊形式：

$$sim(e,e') = \begin{cases} 1 & if\ e = e'; \\ 0 & otherwise; \end{cases}$$

原始规则

$$p_{f,e}^{t+1} = \sum_{f' \in N(f)} T(f,f') p_{f',e}^{t}$$

基于图的一致性翻译模型

- 基于图的一致性翻译模型：在常用对数线性模型的基础上引入了基于图的一致性特征和局部一致性特征

$$p(e|f) = \frac{\exp(\sum_i(\lambda_i \psi_i(e,f)))}{\sum_{e' \in H(f)}(\exp(\sum_i(\lambda_i \psi_i(e',f))))}$$

- 基于图的一致性特征:通过相似的源语言句子的翻译候选获得的一致性特征
- 局部一致性特征:通过相同源语言句子的N-Best翻译候选获得的一致性特征
- 其它常用特征：正向/反向翻译概率，正向/反向词汇特征，调序模型概率，长度惩罚，语言模型概率

基于图的一致性特征

- 基于图的一致性特征:使用结构化标记传播得到的基于图的一致性置信度

$$GC(e,f) = \log\left(\sum_{f' \in N(f)} T_s(f,f') \sum_{e' \in H(f')} T_l(e,e') p_{f',e'}\right)$$

- 标记(翻译候选) 相似度:N-Gram的Dice系数

$$sim(e,e') = Dice(NGr_n(e), NGr_n(e'))$$

- 样本(源语言句子) 相似度: 对称化的句子级BLEU

$$w_{f,f'} = \frac{1}{2}(BLEU_{sent}(f,f') + BLEU_{sent}(f',f))$$

局部一致性特征

- **局部一致性特征**：定义在N-Best翻译候选上的一致性特征

$$LC(e,f) = \log\left(\sum_{e' \in H(f)} p(e'|f) T_l(e,e')\right)$$

MBR得分

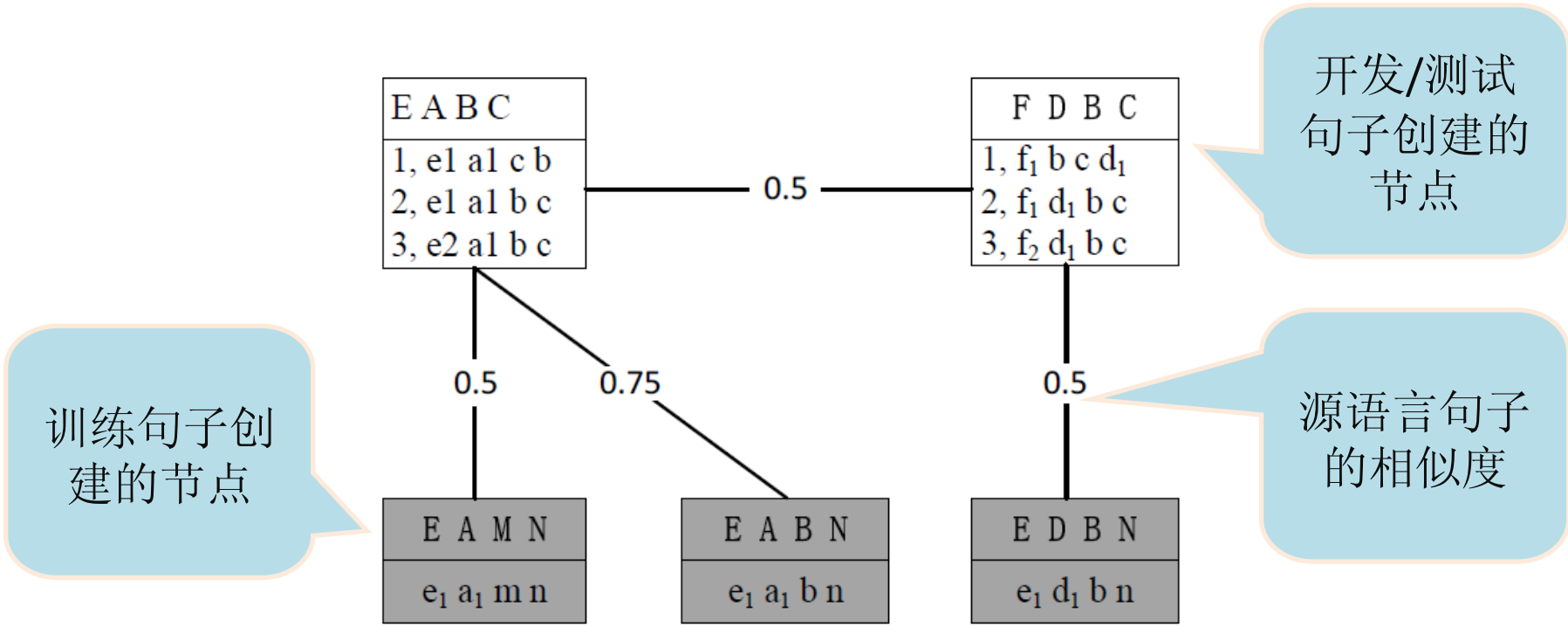
- 局部一致性特征收集同一N-Best列表内部的支持信息

机器翻译重排序
一致性图的创建（1）

- 训练集、开发集和测试集中的每个句子对应一个节点
- 训练集句子对应节点的标记是固定的，视为标注数据
- 训练集句子对应节点之间没有边
- 开发集和测试集句子对应节点的可能标记由解码器生成的N-Best翻译候选给定
- 一个测试集/开发集句子对应节点可以同其它任意节点存在边，只要其相似度大于规定阈值

机器翻译重排序
一致性图的创建（2）

- 用于机器翻译重排序的一致性图的一个示例

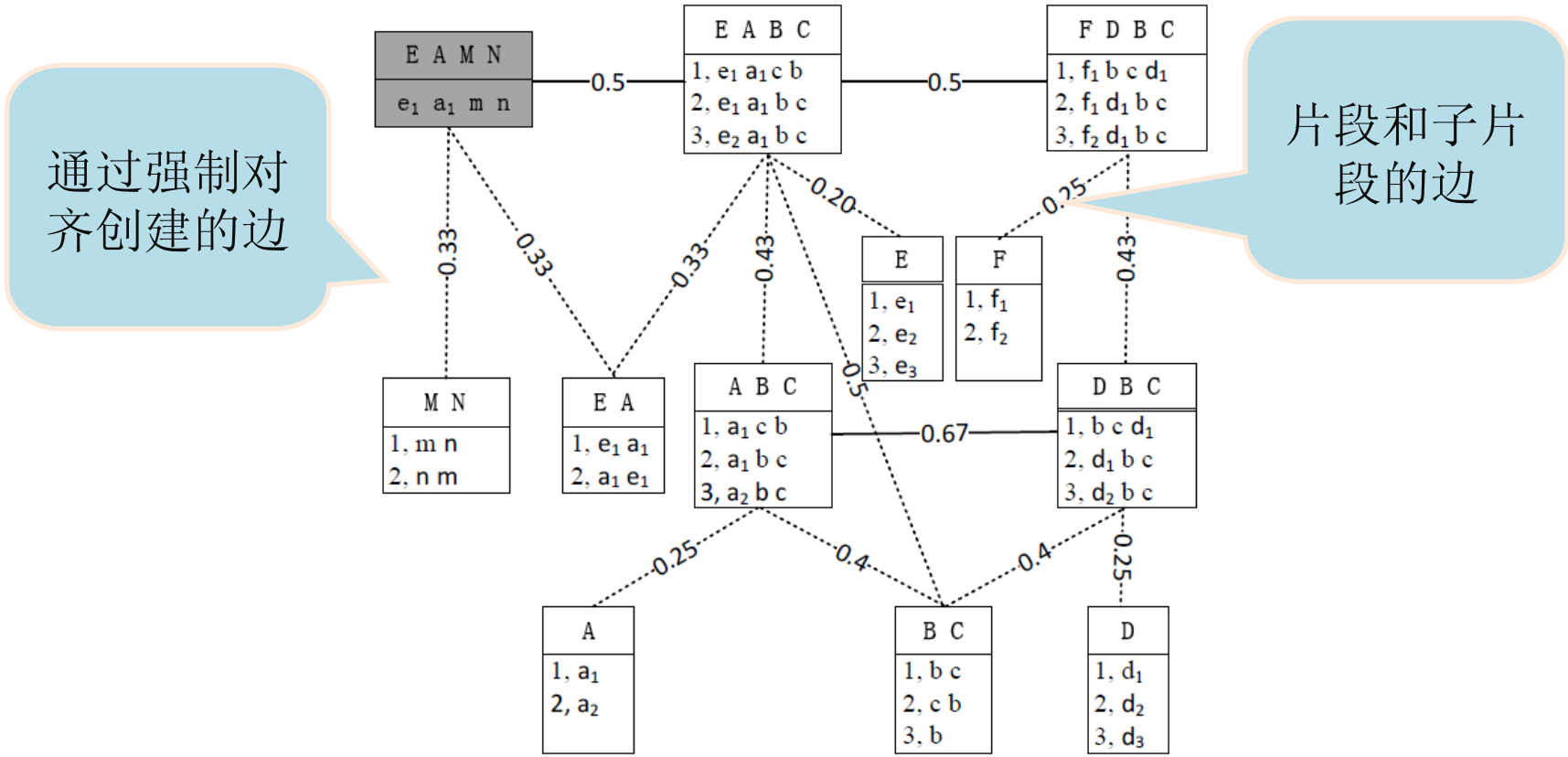


机器翻译解码
一致性图的创建（1）

- 通过对解码过程中片段的翻译候选列表使用基于图的一致性信息进行重新排序，基于图的一致性方法同样可以用于解码过程
- 训练集句子片段的翻译候选通过强制对齐获得
- 开发集/测试集句子片段的翻译候选在CKY解码过程中动态生成
- 如果一个片段是另一个片段的子片段，则强制在对应节点间创建相似边

机器翻译解码
一致性图的创建（2）

- 用于机器翻译解码的一致性图的示例



半监督训练方法

- 一致性图和解码器之间的相互依存关系
  - 机器翻译解码器依赖于基于图的标记传播得到的一致性特征。
  - 一致性图需要解码器提供标记候选（翻译候选），以及作为初始的标记概率的翻译后验概率。

$GC^0 = 0;$

$\lambda^t = MERT(S^{dev}, T^{dev}, GC^0);$

while not converged **do**

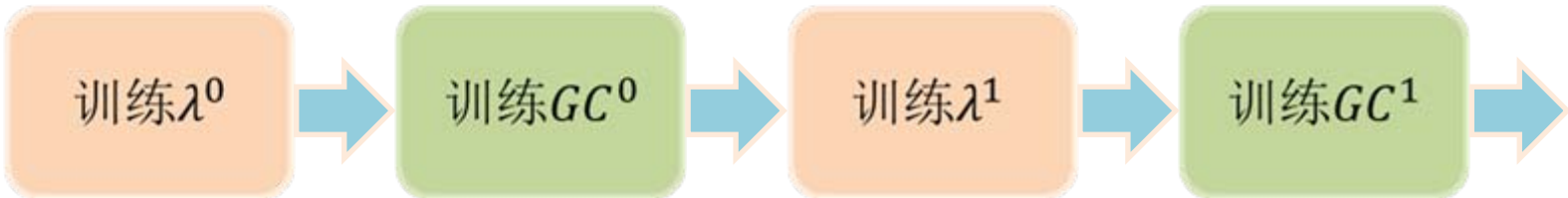
$G^t = CreatG(S^{train}, T^{train}, S^{dev}, S^{test}, \lambda^t);$

$GC^{t+1} = StructLP(G^t);$

$\lambda^{t+1} = MERT(S^{dev}, T^{dev}, GC^{t+1});$

end while

return last (GC^t, λ^t);



实验结果

- 数据集: IWSLT
  - 训练数据: 81K 句对, 655K 中文词，806K 英文词
  - 开发数据: devset8+dialog
  - 测试数据: devset9

	devset8+dialog	devset9
Baseline	48.79	44.73
Struct-LP	49.86	45.54
Rerank-GC	50.66	46.52
Decode-GC	**51.20**	**47.31**

  - Struct-LP : 单纯使用基于图的一致性信息（GCC）做翻译重排序
  - Rerank-GC:使用基于图的一致性信息和其它特征做翻译重排序
  - Decode-GC: 使用基于图的一致性信息和其它特征做解码

结论

- 针对翻译结果的一致性问题，我们使用了基于图的半监督方法，将SMT看做结构学习问题，并针对该问题提出了结构化的标记传播算法。
- 进一步地将结构化标记传播算法获得的一致置信度作为特征，应用于常用的对数线性模型中，将该模型应用于SMT的重排序和解码，提高了SMT的性能。

Thanks