

基于短语和句法的翻译模型

肖桐 朱靖波

xiaotong@mail.neu.edu.cn
zhujingbo@mail.neu.edu.cn

东北大学 自然语言处理实验室
<http://www.nlplab.com>



基于单词有哪些问题

- 基于单词的翻译模型有哪些优点？
 - ▶ 比较符合人的思维、简单直接、易于实现

单词翻译表	P
我 → I	0.6
喜欢 → like	0.3
绿 → green	0.9
茶 → tea	0.8

$s =$ 我 喜欢 绿 茶

基于单词有哪些问题

- 基于单词的翻译模型有哪些优点？
 - ▶ 比较符合人的思维、简单直接、易于实现

单词翻译表	P
我 → I	0.6
喜欢 → like	0.3
绿 → green	0.9
茶 → tea	0.8

$s =$ 我 喜欢 绿 茶
 | / /
 $t =$ I like green tea

基于单词有哪些问题

- 基于单词的翻译模型有哪些**优点**？
 - 比较符合人的思维、简单直接、易于实现

单词翻译表	P					
我 → I	0.6	$s =$	我	喜欢	绿	茶
喜欢 → like	0.3			↗	↘	
绿 → green	0.9	$t =$	I	like	green	tea
茶 → tea	0.8					

- 基于单词的翻译模型有哪些**缺点**？
 - 很多 - 这符合中国人的思维，缺点总是很多：)
 - 独立性假设：单词之间相对独立，没有考虑搭配
 - 调序：较弱的调序建模
 - ...

单词翻译表	P	$s =$ 我 喜欢 红 茶				
我 \rightarrow I	0.6					
喜欢 \rightarrow like	0.3					
红 \rightarrow red	0.8					
红 \rightarrow black	0.1					
茶 \rightarrow tea	0.8					

基于单词有哪些问题

- 基于单词的翻译模型有哪些**优点**？
 - 比较符合人的思维、简单直接、易于实现

单词翻译表	P	
我 → I	0.6	$s =$ 我 喜欢 绿 茶 $t =$ I like green tea
喜欢 → like	0.3	
绿 → green	0.9	
茶 → tea	0.8	

- 基于单词的翻译模型有哪些**缺点**？
 - 很多 - 这符合中国人的思维，缺点总是很多：)
 - 独立性假设：单词之间相对独立，没有考虑搭配
 - 调序：较弱的调序建模
 - ...

单词翻译表	P	
我 → I	0.6	$s =$ 我 喜欢 红 茶 $t =$ I like red tea
喜欢 → like	0.3	
红 → red	0.8	
红 → black	0.1	
茶 → tea	0.8	

基于单词有哪些问题

- 基于单词的翻译模型有哪些**优点**？
 - 比较符合人的思维、简单直接、易于实现

单词翻译表	P					
我 → I	0.6	$s =$	我	喜欢	绿	茶
喜欢 → like	0.3			└─┘	└─┘	
绿 → green	0.9	$t =$	I	like	green	tea
茶 → tea	0.8					

- 基于单词的翻译模型有哪些**缺点**？
 - 很多 - 这符合中国人的思维，缺点总是很多：)
 - 独立性假设：单词之间相对独立，没有考虑搭配
 - 调序：较弱的调序建模
 - ...

单词翻译表	P					
我 → I	0.6	$s =$	我	喜欢	红	茶
喜欢 → like	0.3			└─┘	└─┘	└─┘
红 → red	0.8	$t =$	I	like	red	tea
红 → black	0.1					
茶 → tea	0.8					

“红 茶”为一种搭配，
翻译为“black tea”。

"红 茶"为一种搭配，应该翻译为"black tea"

引入更大的翻译单元

- 简单的单词翻译似乎不行，咋办？

单词翻译表	P					
我 → I	0.6	$s =$	我	喜欢	红	茶
喜欢 → like	0.3					
红 → red	0.8	$t =$	I	like	red	tea
红 → black	0.1					
茶 → tea	0.8					



引入更大的翻译单元

- 简单的单词翻译似乎不行，咋办？ - 引入更大的翻译单元

单词词串翻译表	P	
我 → I	0.6	$s =$ 我 喜欢 红 茶
喜欢 → like	0.3	
红 → red	0.8	
红 → black	0.1	$t =$ I like red tea
茶 → tea	0.8	
我喜欢 → I like	0.3	
我喜欢 → I liked	0.2	
绿茶 → green tea	0.5	
绿茶 → the green tea	0.1	
红茶 → black tea	0.6	
...		



引入更大的翻译单元

- 简单的单词翻译似乎不行，咋办？ - 引入更大的翻译单元

单词词串翻译表	P					
我 → I	0.6	$s =$	我	喜欢	红	茶
喜欢 → like	0.3					
红 → red	0.8					
红 → black	0.1	$t =$	I	like	red	tea
茶 → tea	0.8					
我喜欢 → I like	0.3					
我喜欢 → I liked	0.2	$s =$	我	喜欢	红	茶
绿茶 → green tea	0.5					
绿茶 → the green tea	0.1					
红茶 → black tea	0.6					
...		$t =$	I	like	black	tea

No

Yes

引入更大的翻译单元

- 简单的单词翻译似乎不行，咋办？ - 引入更大的翻译单元

单词词串翻译表	P					
我 → I	0.6	s =	我	喜欢	红	茶
喜欢 → like	0.3					
红 → red	0.8					
红 → black	0.1	t =	I	like	red	tea
茶 → tea	0.8					
我喜欢 → I like	0.3					
我喜欢 → I liked	0.2					
绿茶 → green tea	0.5	s =	我	喜欢	红	茶
绿茶 → the green tea	0.1					
红茶 → black tea	0.6					
...		t =	I	like	black	tea

No

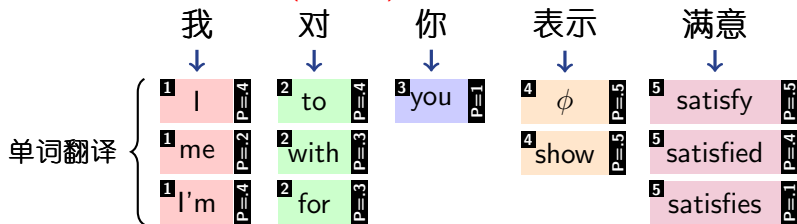
Yes

- 优点很多
 - 翻译时候可以考虑更大范围的上下文信息
比如：“红茶”中的“红”如果和“茶”搭配 ...
 - 更好的局部调序，比如：短语中有“的”字 → ... of ... 结构
 - 更大范围的目标语连续词串的使用，有利于 n -gram 语言模型选择译文

回顾 - 基于词的机器翻译

- 对每个单词的翻译进行（任意）组合

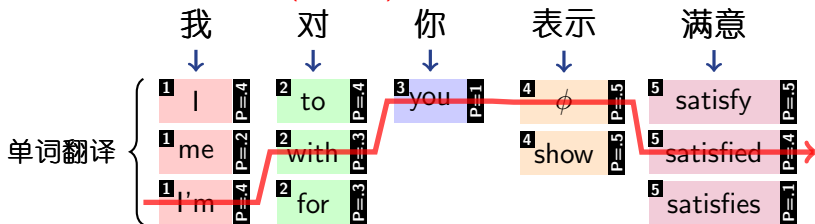
待翻译句子(已经分词):



回顾 - 基于词的机器翻译

- 对每个单词的翻译进行（任意）组合

待翻译句子(已经分词):

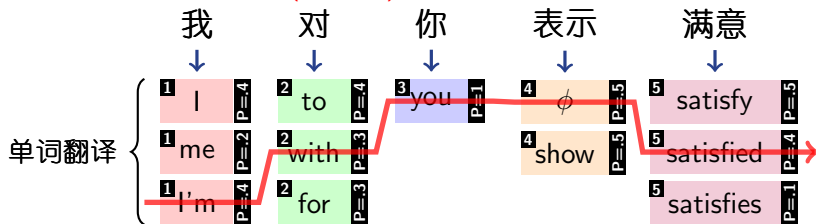


翻译路径 (仅含有单词): \rightarrow

基于“短语”的机器翻译

- 对每个单词及连续词串的翻译进行（任意）组合

待翻译句子(已经分词):

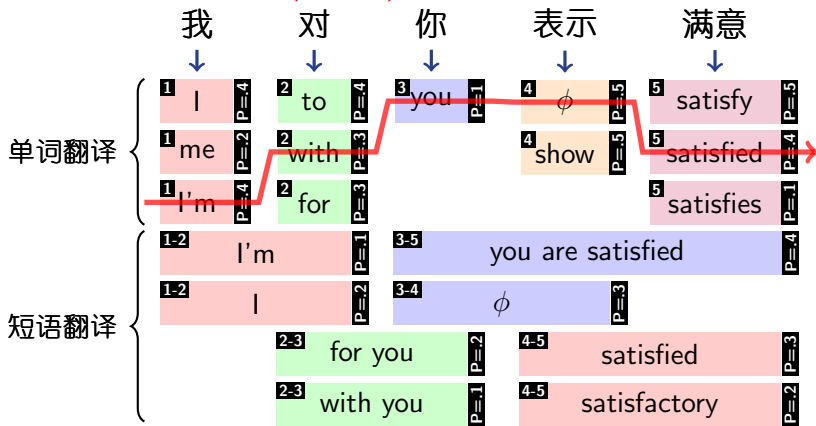


翻译路径（仅含有单词）：→

基于“短语”的机器翻译

- 对每个单词及连续词串的翻译进行（任意）组合

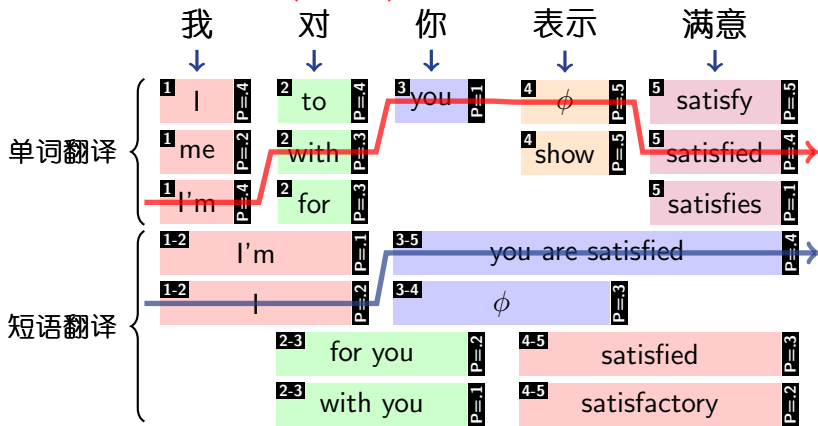
待翻译句子(已经分词):



基于“短语”的机器翻译

- 对每个单词及连续词串的翻译进行（任意）组合

待翻译句子(已经分词):



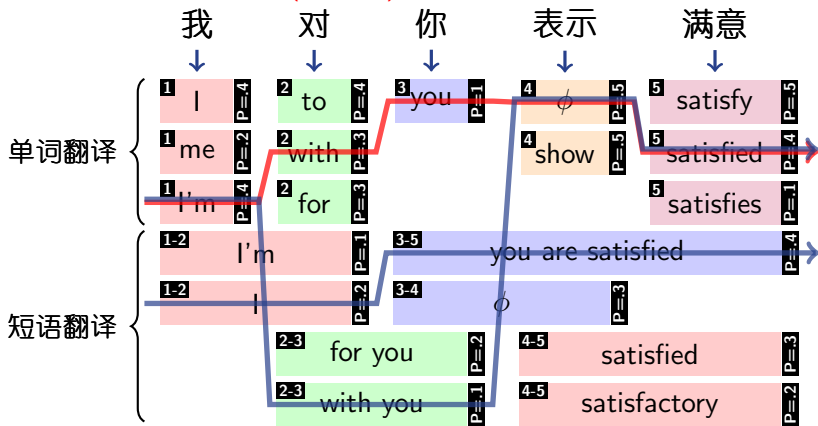
翻译路径（仅含有单词）： →

翻译路径（含有短语）： →

基于“短语”的机器翻译

- 对每个单词及连续词串的翻译进行（任意）组合

待翻译句子(已经分词):



翻译路径（仅含有单词）：→

翻译路径（含有短语）：→