

东北大学自然语言处理实验室

词汇翻译概率



东北大学自然语言处理实验室
<http://www.nlplab.com>

NiuTrans 团队
niutrans@mail.neu.edu.cn

词汇翻译概率

- 问题描述

给定经过分词的源语言（中文）文本、目标语言文本（英文）、以及词对齐（源语言->目标语言）文本，计算词汇翻译概率

- 示例

源语言： A B C D B A A

目标语言： a b c d

词对齐： 0-0 1-1 2-2 3-3

则源语言词汇翻译为目标语言词汇的词汇翻译概率为：

$$\omega(e_i | f_j) = \frac{\text{count}(f_j, e_i)}{\sum_{e_i} \text{count}(f_j, e_i)}$$

以(A,a)为例，即：

$$\omega(a | A) = \frac{\text{count}(A, a)}{\text{count}(A)} = \frac{1}{3}$$

$$\omega(NULL | A) = \frac{\text{count}(A, NULL)}{\text{count}(A)} = \frac{2}{3}$$

- 注意

词汇翻译概率在整篇文档中进行统计

- 数据

当前文件夹 sample-data\目录下

祝好运！