东北大学自然语言处理实验室
对空概率统计

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

对空概率统计

- **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，进行源语言词汇对空概率统计

- **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 A 的对空概率为:

$$\Pr(A_{unaligned}) = \frac{count(A_{unaligned})}{count(A)}$$

即:

$$\Pr(A_{unaligned}) = \frac{2}{3}$$

- **注意**

词汇对空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！