

Instituto Tecnológico y de Estudios Superiores de Monterrey



**Tecnológico
de Monterrey**

Construyendo cuerpo y mente (Gpo 120)

Tarea 4-Investigación II de N-gramas

Facundo Colasurdo

A01198015

Campus Monterrey

Fecha: 12/10/2024

Tarea 3

Investigar las técnicas de (1) absolute discounting, y (2) Smoothing kneser-Ney para n-gramas. ¿Que consisten, que resuelven, algún ejemplo y expresión matemática

• absolute discounting este consiste en restar un descuento fijo D , a cada recuento distinto a Cero, y redistribuir esta masa de probabilidad a los N-gramas con recuento Cero

$$P(w_i | w_{i-1}) = \frac{\text{Count}(w_{i-1}, w_i) - d}{\text{Count}(w_{i-1})} + \lambda(w_{i-1})P(w)$$

• Busca resolver, cuando se le asigna 0 a cualquier N-grama que no se haya visto

ej: ("I eat cake")

- el trigramma cont es 8

- el bigramma cont es 23

- $D = 0.75$

$$P(\text{cake} | \text{eat}) = \frac{(\text{I eat cake}) - D}{(\text{I eat})} + \lambda(\text{eat})P(\text{eat})$$

$$\frac{8 - 0.75}{23} = 0.3152 + \lambda P(\text{cake} | \text{eat})$$

knieser-Ney Smoothing, cambia notaciones de descuentos con un modelo back off

$$P_{KN}(w_i | w_{i-1}) = \frac{\text{Count}(w_{i-1}, w_i) - d}{\text{Count}(w_{i-1})} + \lambda(w_{i-1})P_{\text{backoff}}(w)$$

Busca resolver:

• Toma en cuenta información que usualmente no es tomada en cuenta, en la que aparece en distintos textos

$$\text{ej: } P(\text{San} | \text{if you going to}) = \frac{C(\text{if you're going to San})}{C(\text{if you're going to})}$$

- trigramma 5

- Bigramma 20

- Pseudocount 10

- Unicas 50

- $D = .75$

$$\frac{(5 - 0.75)}{20} + \lambda \frac{10}{60} = \frac{4.25}{20} + \lambda \times 0.16$$