TAREA 4 - INVESTIGACIÓN II DE N-GRAMAS

① Absolute discounting

a. ¿En qué consiste? ¿Qué problemática resuelve?

Esta técnica ajusta las cuentas (frecuencias) de los N-gramas restando un valor fijo a todas las cuentas mayores que cero. Este valor fijo es el Absolute discounting.

Y el problema que resuelve esta técnica es la escasez de datos. Absolute discounting ajusta las probabilidades para que las secuencias no vistas tengan una probabilidad mayor que cero, sin distorsionar significativamente las secuencias vistas.

b. ¿Cuál es su expresión matemática?

$$P(w_i|w_{i-n+1},\dots,w_{i-1}) = \begin{cases} \dfrac{C(w_{i-n+1},\dots,w_i)-D}{C(w_{i-n+1},\dots,w_i)}, & \text{si } C(w_{i-n+1},\dots)>0 \\ \lambda(w_{i-n+1},\dots,w_{i-1})P(w_i|w_{i-n+2},\dots,w_{i-1}), & \\ \quad \text{si } C(w_{i-n+1},\dots,w_i)=0 & \end{cases}$$

$D \longrightarrow$ Descuento absoluto, típicamente entre 0 y 1.

$C(w_{i-n+1},\dots,w_i) \longrightarrow$ Cuenta de la secuencia completa en el conjunto de entrenamiento.

$\lambda(w_{i-n+1},\dots,w_{i-1}) \longrightarrow$ Factor de normalización que asegura que las probabilidades sumen 1.

$P(w_i|w_{i-n+2},\dots,w_{i-1}) \longrightarrow$ Probabilidad en un modelo de n-1-gramas.

c. Ejemplo.

Supongamos que tenemos un modelo de trigrama con la secuencia "Soy un artista". Si el trigrama "un artista famoso" ha aparecido 2 veces, y aplicamos un descuento absoluto de D = 0.75, la probabilidad se calcula como:

$$P(\text{"famoso"} \mid \text{"un artista"}) = \frac{2 - 0.75}{C(\text{"un artista"})}$$

Si el trigrama "un artista mundial" nunca ha aparecido, se usa una interpolación con el módelo de dosgramas para estimar su probabilidad.

② Smoothing Kneser-Ney

a. ¿En qué consiste? ¿Qué problema resuelve?

Es un método que se utiliza para calcular la distribución de probabilidad de n-gramas en un documento en función de sus historias.
Se considera como uno de los métodos más eficaces de suavizado al usar Absolute Discounting para omitir n-gramas con frecuencias más bajas.
Al igual que otros métodos de smoothing, busca resolver el problema de oscasos de datos. Evitando probabilidades cero a secuencias nuevas.

b. ¿Cuál es su expresión matemática?

$$P_{KN}(w_i | w_{i-n+1}, \ldots, w_{i-1}) = \frac{\max(C(w_{i-n+1}, \ldots, w_i) - D, 0)}{C(w_{i-n+1}, \ldots, w_{i-1})}$$

$$+ \lambda(w_{i-n+1}, \ldots, w_{i-1}) P_{KN}(w_i | w_{i-n+2}, \ldots, w_{i-1})$$

Donde:

$C(w_{i-n+1}, \ldots, w_i) \rightarrow$ Es la cuenta del n-grama

$D \rightarrow$ Descuento absoluto

$\lambda(w_{i-n+1}, \ldots, w_{i-1}) \rightarrow$ Factor de normalización

$P_{KN}(w_i | w_{i-n+2}, \ldots, w_{i-1}) \rightarrow$ Probabilidad en un modelo de $(n-1)$-gramas.

La diferencia del Kneser-Ney Smoothing es que las probabilidades de $(n-1)$-gramas son ajustadas en función de cuántos contextos únicos han precedido a una palabra, en lugar de cuántas veces ha aparecido la palabra en total.

c. Ejemplo

Retomando el ejemplo del modelo "Soy un artista". Para calcular la probabilidad suavizada de Kneser-Ney, primero se aplica el Absolute Discounting a las cuentas vistas:

$$P_{KN}(\text{"famoso"} | \text{"un artista"}) = \frac{2 - D}{C(\text{"un artista"})} + \lambda P_{KN}(\text{"famoso"} | \text{"artista"})$$

Si la secuencia "un artista mundial" no ha aparecido, Kneser-Ney asigna probabilidad usando cómo la palabra "mundial" aparece en otros contextos, capturando mejor la naturaleza de las palabras en combinaciones nuevas.