



Procesamiento de Lenguaje Natural

Modelos de Lenguaje

Mauricio Toledo-Acosta
mauricio.toledo@unison.mx

Departamento de Matemáticas
Universidad de Sonora



Section 1

Introducción



Referencias

- **Chapter 3.** Jurafsky, D., Martin, J. H. (2019). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.
- **Chapter II.6.** Eisenstein, J. (2018). Natural language processing. Jacob Eisenstein.



Objetivo

En el parque, los niños juegan con ...



El agua hierve a 100 grados ...



- *café el en mesa la sobre libro un dejé olvidé y*



- *café el en mesa la sobre libro un dejé olvidé y*
- *dejé un libro sobre la mesa en el café y lo olvidé*



6/35





- 1 **Modelos de Lenguaje Estadísticos.** Predicen la siguiente palabra dadas las palabras que le preceden, esto lo hacen usando conteos de co-ocurrencias. El modelo de n -gramas es el más sencillo.
- 2 **Modelos de Lenguaje Neuronales.** Predicen la siguiente palabra usando embeddings que capturan diversos fenómenos lingüísticos a partir del uso de redes neuronales.
 - Modelos neuronales *clásicos*: Word2Vec, FastText, GloVe, ConceptNet, ...
 - Modelos basados en mecanismos de atención: Bert, LLaMa, GPT, Claude, ModernBert, DeepSeek, ...



¿Cómo se calculan las probabilidades?

Procesamiento
de Lenguaje
Natural

Introducción

Modelo de
 n -gramas

Generalizaciones

Evaluando
Modelos de
Lenguaje

Aspectos
adicionales

Algunas
aplicaciones

Queremos calcular la probabilidad de una secuencia de palabras $W = w_1, w_2, \dots, w_n$. Esto lo hacemos con la probabilidad conjunta:

$$P(W) = P(w_1, w_2, \dots, w_n).$$

Una tarea relacionada es calcular la probabilidad condicional

$$P(w_n \mid w_1, w_2, \dots, w_{n-1})$$

Ambas están relacionadas por la regla de la cadena

$$P(W) = P(w_1)P(w_2|w_1)P(w_3|w_1, w_2) \cdots P(w_n \mid w_1, w_2, \dots, w_{n-1})$$



10/35



La distribución de probabilidad del valor futuro de una variable aleatoria depende únicamente de su valor presente, siendo independiente de la historia de dicha variable.

Podemos hacer las siguientes aproximaciones:

$$P(w_n | w_1 w_2 \dots w_{n-1}) \approx P(w_n)$$

$$P(w_n|w_1w_2...w_{n-1}) \approx P(w_n|w_{n-1})$$

$$P(w_n | w_1 w_2 \dots w_{n-1}) \approx P(w_n | w_{n-2} w_{n-1})$$

• • •



Section 2

Modelo de n -gramas



Al decir n -gramas nos referimos a las secuencias o al modelo predictivo que asigna probabilidades.

◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ↺ 🔍 ↻ 13/35



Ejemplo

Procesamiento de Lenguaje Natural

Introducción

Modelo de n -gramas

Generalizaciones

Evaluando Modelos de Lenguaje

Aspectos adicionales

Algunas aplicaciones

	me	gusta	comer	tacos	tamales	mucho	hoy
me	5	50	0	2	1	0	3
gusta	2	0	40	1	4	6	0
comer	0	2	0	30	25	0	0
tacos	1	0	3	0	5	2	1
tamales	0	1	2	5	0	1	0
mucho	4	6	0	0	1	0	8
hoy	2	0	1	1	0	7	0



Ejemplo: probabilidades

Procesamiento de Lenguaje Natural

Introducción

Modelo de *n*-gramas

Generalizaciones

Evaluando Modelos de Lenguaje

Aspectos adicionales

Algunas aplicaciones

	me	gusta	comer	tacos	tamales	mucho	hoy
me	0.07	0.72	0	0.03	0.01	0	0.17
gusta	0.04	0	0.80	0.02	0.08	0.12	0
comer	0	0.04	0	0.55	0.45	0	0
tacos	0.10	0	0.30	0	0.50	0.10	0
tamales	0	0.14	0.29	0.71	0	0.14	0
mucho	0.22	0.33	0	0	0.06	0	0.39
hoy	0.18	0	0.09	0.09	0	0.64	0



Algunas ventajas

Procesamiento de Lenguaje Natural

Introducción

Modelo de n -gramas

Generalizaciones

Evaluando Modelos de Lenguaje

Aspectos adicionales

Algunas aplicaciones

- Los n -gramas ayudan a capturar la información contextual y la semántica dentro de una secuencia de palabras, proporcionando una comprensión más matizada del lenguaje.
- En tareas de recuperación de información (information retrieval), los n -gramas ayudan a emparejar y clasificar documentos según la relevancia de los patrones de n -gramas.
- Los n -gramas sirven como características poderosas en la clasificación de texto y el análisis de sentimientos, capturando patrones significativos que contribuyen a la caracterización de diferentes clases o sentimientos.





the in	rain Spain	in falls
Spain mainly	falls on	mainly the
on plain		



n -gramas sintácticos

Procesamiento
de Lenguaje
Natural

Introducción

Modelo de
 n -gramas

Generalizaciones

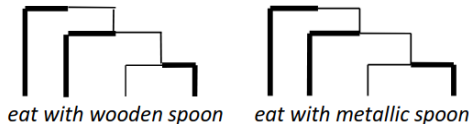
Evaluando
Modelos de
Lenguaje

Aspectos
adicionales

Algunas
aplicaciones

A diferencia de los n -gramas donde las subsecuencias se toman en el orden en el que aparecen en el texto, en los **n -gramas sintácticos** los vecinos se toman siguiendo las relaciones sintácticas de los árboles de dependencia sintáctica.

eat with wooden spoon eat with metallic spoon



¿Cuántos bigramas y bigramas sintácticos en común tienen?



22/35



Perplejidad

Procesamiento de Lenguaje Natural

Introducción

Modelo de n-gramas

Generalizaciones

Evaluando Modelos de Lenguaje

Aspectos adicionales

Algunas aplicaciones

La **perplejidad** es una métrica para evaluar el rendimiento de un modelo de lenguaje. Mide la incertidumbre del modelo para predecir la próxima palabra en una secuencia. Cuanto menor sea la perplejidad, mayor será la capacidad del modelo para predecir la palabra siguiente.

$$\begin{aligned} \text{Pp}(W) &= \sqrt[N]{\frac{1}{P(w_1 w_2 \cdots w_N)}} \\ &= \sqrt[N]{\prod_i \frac{1}{P(w_i | w_{i-1})}} \end{aligned}$$

W es la secuencia entera de palabras de un conjunto de prueba.



Considerar textos escritos en AAVE (African American Vernacular English):



- Al calcular la perplejidad, el modelo de n -gramas debe construirse sin ningún conocimiento del conjunto de prueba. De otra forma la perplejidad puede ser artificialmente baja.
- La perplejidad de dos modelos lingüísticos sólo es comparable si utilizan vocabularios idénticos.
- Una mejora en la perplejidad (m. intrínseca) no garantiza una mejora del rendimiento de una tarea de PLN como el reconocimiento del habla o la traducción automática (m. extrínseca). Sin embargo, la perplejidad suele estar correlacionada con dichas mejoras.



Suavizado: El papel del corpus de entrenamiento

Procesamiento de Lenguaje Natural

Introducción

Modelo de n -gramas

Generalizaciones

Evaluando Modelos de Lenguaje

Aspectos adicionales

Algunas aplicaciones

1 gram	-To him swallowed confess hear both. Which. Of save on trail for are ay device and rote life have -Hill he late speaks; or! a more to leg less first you enter
2 gram	-Why dost stand forth thy canopy, forsooth; he is this palpable hit the King Henry. Live king. Follow. -What means, sir. I confess she? then all sorts, he is trim, captain.
3 gram	-Fly, and will rid me these news of price. Therefore the sadness of parting, as they say, 'tis done. -This shall forbid it should be branded, if renown made it empty.
4 gram	-King Henry. What! I will go seek the traitor Gloucester. Exeunt some of the watch. A great banquet serv'd in; -It cannot be but so.

1 gram	Months the my and issue of year foreign new exchange's september were recession exchange new endorsed a acquire to six executives
2 gram	Last December through the way to preserve the Hudson corporation N. B. E. C. Taylor would seem to complete the major central planners one point five percent of U. S. E. has already old M. X. corporation of living on information such as more frequently fishing to keep her
3 gram	They also point to ninety nine point six billion dollars from two hundred four oh six three percent of the rates of interest stores as Mexico and Brazil on market conditions



Si un n -grama aparece un número suficiente de veces, podemos tener una buena estimación de su probabilidad. Sin embargo, es probable que falten algunas secuencias de palabras perfectamente aceptables. Tendremos casos de n -gramas de probabilidad 0 que deberían tener probabilidad distinta de 0.

denied the allegations:	5
denied the speculation:	2
denied the rumors:	1
denied the report:	1

denied the offer: 0
denied the loan: 0



Suavizado: Palabras desconocidas

Procesamiento de Lenguaje Natural

Introducción

Modelo de *n*-gramas

Generalizaciones

Evaluando Modelos de Lenguaje

Aspectos adicionales

Algunas aplicaciones

¿Qué pasa con palabras que nunca ha visto en el entrenamiento?

En un sistema de **vocabulario cerrado** el conjunto de prueba no contiene palabras desconocidas. En un sistema de **vocabulario abierto**, tenemos que lidiar con palabras que no hemos visto antes, a las que llamaremos **palabras fuera de vocabulario (OOV)**. Podemos lidiar con estas palabras desconocidas añadiendo una pseudopalabra llamada **<UNK>**.



- Convertir un sistema abierto en uno cerrado:
 - Escoger un vocabulario fijo.
 - Convertir cualquier palabra OOV en <UNK>.
 - Estimar las probabilidades para <UNK> de la forma usual, como si fuera una palabra *normal*.
- Crear un vocabulario fijo implícito. Reemplazamos palabras por <UNK> en el entrenamiento basándonos en su frecuencia.



- Add- k smoothing
- Kneser-NeySmoothing
- ...

	i	want	to	eat	chinese	food	lunch	spend
i	6	828	1	10	1	1	1	3
want	3	1	609	2	7	7	6	2
to	3	1	5	687	3	1	7	212
eat	1	1	3	1	17	3	43	1
chinese	2	1	1	1	1	83	2	1
food	16	1	16	1	2	5	1	1
lunch	3	1	1	1	1	2	1	1
spend	2	1	2	1	1	1	1	1





◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ↺ 🔍 ↻ 34/35



Una aplicación: Generación y predicción de texto

Procesamiento de Lenguaje Natural

Introducción

Modelo de
 n -gramas

Generalizaciones

Evaluando
Modelos de
Lenguaje

Aspectos
adicionales

Algunas
aplicaciones