

Procesamiento del Lenguaje Natural

Problemas de morfología y etiquetado sintáctico

1. Dadas las siguientes tablas de estadísticas idealizadas a partir del Brown Corpus:

Primera etiqueta	Segunda etiqueta					
	AT	BEZ	IN	NN	VB	PER
AT	0	0	0	48636	0	19
BEZ	1973	0	426	187	0	38
IN	43322	0	1325	17314	0	185
NN	1067	3720	42470	11773	614	21392
VB	6072	42	4758	1476	129	1522
PER	8016	75	4656	1329	954	0

Palabra	AT	BEZ	IN	NN	VB	PER
bear	0	0	0	10	43	0
is	0	10065	0	0	0	0
move	0	0	0	36	133	0
on	0	0	5484	0	0	0
president	0	0	0	382	0	0
progress	0	0	0	108	4	0
the	69016	0	0	0	0	0
.	0	0	0	0	0	48809

Se pide calcular las probabilidades:

$$P(\text{AT NN BEZ IN AT NN} \mid \text{the bear is on the move})$$

$$P(\text{AT NN BEZ IN AT VB} \mid \text{the bear is on the move})$$

2. Dadas las tablas anteriores, se pide aplicar el algoritmo de Viterbi para predecir la secuencia de etiquetas para la frase: “the bear is on the move”

3. Se la tabla de datos para palabras desconocidas siguiente:

Atributo	Valor	NNP	NN	NNS	VBG	VBZ
desconocida	sí	0,05	0,02	0,02	0,005	0,005
	no	0,95	0,98	0,98	0,995	0,995
mayúsculas	sí	0,95	0,10	0,10	0,005	0,005
	no	0,05	0,90	0,90	0,995	0,995
sufijo	-s	0,05	0,01	0,98	0,00	0,99
	-ing	0,01	0,01	0,00	1,00	0,00
	-tion	0,05	0,10	0,00	0,00	0,00
	otro	0,89	0,88	0,02	0,00	0,01

Se pide calcular las siguientes probabilidades:

$$P(\text{fenestration} \mid t^k)$$

$$P(\text{fenestrates} \mid t^k)$$

$$P(\text{palladiot} \mid t^k)$$

$$P(\text{Palladios} \mid t^k)$$

$$P(\text{guestimating} \mid t^k)$$

