

# Examen de Traducción Automática

MIARFID, Universitat Politècnica de València, 11 de febrero de 2019

Apellidos:

Nombre:

## Cuestiones

Marca cada recuadro con una única opción de entre las dadas. Cada 3 errores restan una respuesta correcta. Las respuestas en blanco ni suman ni restan. Tiempo de duración del examen: 1 hora.

- 1 ☐ A En un modelo de traducción basado en alineamientos estadísticos, indicar qué afirmación es falsa:
- A) Los alineamientos son una relaciones entre palabras de la frase fuente y palabras de la frase destino.
  - B) El número de alineamientos, tanto si se definen como una relación como si se definen como una función de las posiciones de salida en las de la entrada, es exponencial con la longitud de la frase destino.
  - C) El número de alineamientos si se definen como una relación es exponencial con la longitud de la frase destino y la longitud de la frase fuente.
  - D) El número de alineamientos si se definen como una función de las posiciones de salida en las de la entrada, es exponencial con la longitud de la frase destino y polinómico con la longitud de la frase fuente.
- 2 ☐ D En la traducción estadística basada en palabras con el modelo 1, qué afirmación es correcta sobre los parámetros principales del modelo:
- A) Están formados por el diccionario estadístico y los alineamientos estadísticos, ambos se aprenden a partir de un conjunto de datos bilingües
  - B) Están formados por los alineamientos estadísticos que se aprenden a partir de un conjunto de datos bilingües
  - C) Están formados por un diccionario estadístico que se fija a priori sin utilizar un conjunto de entrenamiento
  - D) Definen el diccionario estadístico que se aprenden a partir de un conjunto de frases bilingües por un conteo ponderado
- 3 ☐ D En la traducción estadística basada en palabras, indicar qué afirmación es incorrecta.
- A) El modelo M2 y HMM tienen en común el diccionario estadístico
  - B) El modelo de alineamiento estadístico en el modelo 2 no depende del contexto
  - C) El modelo HMM, el modelo de alineamiento estadístico tiene cierta dependencia del contexto
  - D) El coste computacional de calcular  $\Pr(y | x)$  es el mismo tanto si se utiliza el modelo M2 como si se utiliza HMM
- 4 ☐ A En un sistema de traducción estadístico basado en “phrases” y en un modelo log-lineal, indicar qué afirmación es falsa:
- A) En el proceso del aprendizaje de las “phrases” solo se utiliza el modelo basado en palabras M1
  - B) El número de “phrases” en la entrada y en la salida deben ser iguales
  - C) En el proceso del aprendizaje de las “phrases” se utilizan alineamientos en los dos sentidos
  - D) En los modelos lexicalizados del modelo log-lineal se utilizan diccionarios estadísticos
- 5 ☐ D En el proceso de traducción con modelos basados en “phrases” utilizando el algoritmo “stack decoding” indicar qué afirmación es correcta:
- A) Hay solo una “pila”
  - B) En general, se van expandiendo prefijos de tamaño creciente de la entrada
  - C) El proceso de traducción finaliza cuando se genera un símbolo final de oración
  - D) En la práctica hay una “pila” por cada tamaño del conjunto de cobertura
- 6 ☐ A Indicar qué afirmación es falsa, en general, con respecto a la generación de “word-embeddings”:
- A) El conjunto de entrenamiento debe ser bilingüe
  - B) El conjunto de entrenamiento debe puede ser monolingüe
  - C) Existen métodos basados en predecir el contexto de una palabra
  - D) Existen métodos basados en predecir la palabra siguiente de una secuencia de palabras
- 7 ☐ C Indicar qué afirmación es falsa con respecto a la técnica “byte-pair encoding” en traducción neuronal
- A) Permite reducir el número de palabras desconocidas
  - B) Es una técnica de compresión de datos
  - C) Permite reducir el tiempo de entrenamiento y el de traducción
  - D) Da lugar a secuencias más largas que si no se utiliza esa técnica

8 ☒ A El modelo de atención en un traductor neuronal “enconder-decoder” es importante por una de las afirmaciones siguientes, indicar cuál:

- A) El modelo de atención permite ponderar la contribución de cada palabra fuente en la generación de una palabra destino
- B) El modelo de atención permite alinear las palabras de la frase fuente entre si.
- C) El modelo de atención permite alinear las palabras de la frase destino entre si.
- D) El modelo de atención elimina las palabras de la frase destino en función de su importancia

9 ☒ C Los modelos de traducción neuronal basados en “Transformer” están considerados como el estado del arte en este área, indicar qué afirmación es correcta:

- A) Es un modelo recurrente en la entrada y en la salida
- B) Es un modelo recurrente en la entrada pero no en la salida
- C) Es un modelo no recurrente basado en modelos de auto-atención en la entrada y en la salida y en un modelo de atención entre la entrada y la salida
- D) Es un modelo no recurrente basado en modelos atención entre la entrada y la salida

10 ☒ B En un sistema de traducción interactiva indicar qué afirmación es correcta:

- A) La interactividad permite que el sistema genere traducciones más rápidas
- B) La interactividad permite que el sistema genere una traducción alternativa cada vez que el humano corrige un error
- C) La interactividad permite reducir el TER hasta un cierto punto
- D) La interactividad permite obtener un BLEU por debajo de un cierto umbral

11 ☒ B La post-edición y la traducción interactiva son dos aproximaciones que permiten obtener traducciones de alta calidad y en menos tiempo que la traducción realizada totalmente por humanos. Indicar qué afirmación de las siguientes es cierta:

- A) La traducción interactiva permite modificar los modelos conforme el usuario valida las traducciones, pero la post-edición no
- B) La traducción interactiva y la post-edición pueden permitir modificar los modelos conforme el usuario valida las traducciones
- C) La post-edición permite modificar los modelos conforme el usuario valida las traducciones, pero la traducción interactiva no
- D) Ni la traducción interactiva ni la post-edición permiten modificar los modelos conforme el usuario valida las traducciones

12 ☒ C Dada la SITG  $G = \{(p/2, S \rightarrow [SS]), (1 - p, S \rightarrow \langle SS \rangle), (p/2, S \rightarrow a/x)\}$  indica la probabilidad de par  $(aa, xx)$ :

- A) No es posible calcular dicha probabilidad.
- B)  $p^3/8$
- C)  $(1 - p/2) p^2/4$
- D) La probabilidades de dicha SITG no están correctamente definidas.

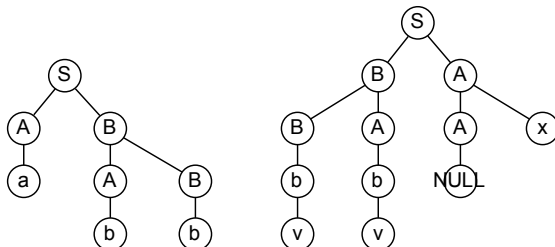
13 ☒ B Dada la SITG  $G = \{(p/4, S \rightarrow [SS]), (p/2, S \rightarrow \langle SS \rangle), (p/8, S \rightarrow a/x), (p/16, S \rightarrow b/y), (p/16, S \rightarrow b/\epsilon)\}$ , donde  $\epsilon$  representa la cadena vacía, indica el par de cadenas cuya derivación es de máxima probabilidad:

- A)  $(aba, yxx)$
- B)  $(aba, xyx)$
- C)  $(aba, xxy)$
- D)  $(aba, yyx)$

14 ☐ ANULADA Dadas las siguientes tablas de reordenamiento, inserción y traducción asociadas a un traductor “tree-to-string”:

r-table			n-table				t-table			
original order	reorder	P(reorder)	parent	S	A	B	w	p(ins-w)	a	b
A B	A B	0.7	node	A	A	A	x	0.6	z	0.6
	B A	0.3	P(None)	0.6	0.5	0.4	y	0.4	v	0.3
B A	A B	0.6	P(Left)	0.3	0.3	0.4			NULL	0.4
	B A	0.4	P(right)	0.1	0.2	0.2				NULL

indica la probabilidad de convertir el árbol de la izquierda en el árbol de la derecha:



- A) 0.0052
- B) 0.000324
- C) 0.0004
- D) 0.0032

15 ☒ A Dadas las tablas de reordenamiento, inserción y traducción asociadas al traductor “tree-to-string” de la pregunta anterior y el árbol de la izquierda, ¿es posible obtener la cadena “zzz”?:

- A) No es posible.
- B) Habría que cambiar las probabilidades.
- C) Si es posible, y se puede obtener con varios árboles.
- D) Si, y se puede obtener con un solo árbol.