

Examen de Traducción Automática

MIARFID, Universitat Politècnica de València, 10 de febrero de 2023

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 Supongamos que tenemos un vocabulario fuente $\{a, b\}$ y otro destino $\{1, 2\}$ y se quiere traducir la cadena $x_1^2 = "a a"$ en la cadena $y_1^2 = "1 1"$ con el denominado Modelo 1 de la traducción estadística definido en las tablas adjuntas. Indicar la opción correcta.

- A) La traducción de "a a" en "1 1" tiene una probabilidad de 0.0139. $p(I|x_1^2) = \begin{cases} 0.0 & \text{para } I < 2 \text{ \& } I > 3 \\ 0.5 & \text{para } I = 2 \text{ o } I = 3 \end{cases} \quad \forall x_1^J$
- B) La traducción de "a a" en "1 1" tiene una probabilidad de 0.25.
- C) La traducción de "a a" en "1 1" tiene una probabilidad de 0.05.
- D) La traducción de "a a" en "1 1" tiene una probabilidad de 0.0277. $x \in \{a, b\} \text{ \& } y \in \{1, 2\}$

$l(y x)$	x_0	a	b
1	0.3	0.1	0.6
2	0.7	0.9	0.4

- 2 ☒ D Supongamos que tenemos un vocabulario fuente $\{a, b\}$ y otro destino $\{1, 1\}$ y se quiere obtener el alineamiento de Viterbi A de la traducción $x_1^2 = "a b"$ a $y_1^2 = "1 2"$ mediante el denominado Modelo 1 de la traducción estadística definido en las tablas adjuntas. Indicar la opción correcta.

- A) El alineamiento óptimo es $A[1] = 1$ y $A[2] = 2$
- B) El alineamiento óptimo es $A[1] = 2$ y $A[2] = 2$
- C) El alineamiento óptimo es $A[1] = 0$ y $A[2] = 0$
- D) El alineamiento óptimo es $A[1] = 2$ y $A[2] = 1$

$$p(I|x_1^J) = \begin{cases} 0.0 & \text{para } I < 2 \text{ \& } I > 3 \\ 0.5 & \text{para } I = 2 \text{ o } I = 3 \end{cases} \quad \forall x_1^J$$

$$x \in \{a, b\} \text{ \& } y \in \{1, 2\}$$

$l(y x)$	x_0	a	b
1	0.3	0.1	0.6
2	0.7	0.9	0.4

- 3 ☒ B En la traducción estadística basados en segmentos ("phrases") monótonos y no monótonos, indicar la afirmación correcta:

- A) La complejidad computacional que presentan los modelos monótonos es inferior a los no monótonos aunque la calidad de las traducciones que producen los modelos monótonos es siempre superior.
- B) La complejidad computacional que presentan los modelos monótonos es inferior a los no monótonos aunque la calidad de las traducciones que producen los modelos monótonos pueden ser inferior.
- C) La complejidad computacional que presentan los modelos no monótonos es siempre inferior a los monótonos aunque la calidad de las traducciones que producen los modelos no monótonos puede ser inferior.
- D) La complejidad computacional que presentan los modelos no monótonos es inferior siempre a los monótonos aunque la calidad de las traducciones que producen los modelos no monótonos puede ser superior.

- 4 ☒ B En el aprendizaje de los modelos de traducción estadísticos basados en segmentos ("phrases") indicar qué afirmación es correcta:

- A) Primero hay que seleccionar el conjunto de "phrases" bilingües, las correspondientes distribuciones de probabilidad y el modelo de lenguaje de salida utilizando un conjunto de entrenamiento paralelo y luego los coeficientes del modelo log-lineal con el mismo conjunto de datos de entrenamiento.
- B) Primero hay que seleccionar el conjunto de "phrases" bilingües, las correspondientes distribuciones de probabilidad y el modelo de lenguaje de salida, entre otros, utilizando un conjunto de entrenamiento paralelo y luego hay que calcular los coeficientes del modelo log-lineal con un conjunto de datos distintos del de entrenamiento.
- C) El modelo de lenguaje de salida se entrena necesariamente con un conjunto de datos distinto de los que se utilizan para la selección de "phrases".
- D) El modelo de lenguaje de entrada se entrena necesariamente con el mismo conjunto de datos utilizados para la selección de "phrases".

- 5 ☒ A En base al Modelo 1 de la cuestión 2, indicar la respuesta correcta con respecto al modelo lexicalizado $p_{lex}("1 2"|"a b", A)$ donde A es alineamiento de Viterbi obtenido de la cuestión 2.

- A) $p_{lex}("1 2"|"a b", A) = 0.54$
- B) $p_{lex}("1 2"|"a b", A) = 0.95$
- C) $p_{lex}("1 2"|"a b", A) = 0.10$
- D) $p_{lex}("1 2"|"a b", A) = 1.00$

- 6 **D** En un sistema de traducción codificador-decodificador basado en redes recurrentes bidireccionales indicar cuál es la respuesta errónea:
- A) El codificador está compuesto por dos LSTMs que analizan la oración fuente de izquierda a derecha y de derecha a izquierda.
 - B) El mecanismo de atención se aplica sobre la secuencia de estados resultantes de concatenar las salidas del LSTM de izquierda a derecha y las salidas del LSTM de derecha a izquierda.
 - C) El LSTM de salida genera sucesivos estados que se proyectan en un espacio vectorial de dimensión talla del vocabulario de salida seguido de la aplicación de una función softmax.
 - D) El decodificador está compuesto por dos LSTMs que generan la traducción de izquierda a derecha y de derecha a izquierda.
- 7 **A** En un sistema de traducción basado en LSTMs y un modelo de atención basado solo en el productor escalar de dos vectores estado (sin matrices), el codificador ha generado una secuencia de vectores $\mathbf{h}_1^e = (-1, 0, -1)^t$, $\mathbf{h}_2^e = (0, -1, 0)^t$, y $\mathbf{h}_3^e = (0, 1, 0)^t$ y el decodificador ha generado $\mathbf{h}_1^d = (0, 1, 0)^t$ y $\mathbf{h}_2^d = (0, 0, 1)^t$, indicar la opción correcta:
- A) El modelo de atención genera para la posición $i = 3$ el vector $\mathbf{u}_3 = (-0.155, 0.000, -0.155)^t$.
 - B) El modelo de atención genera para la posición $i = 3$ el vector $\mathbf{u}_3 = (-0.155, 0.422, -0.155, 0.422)^t$.
 - C) El modelo de atención genera para la posición $i = 3$ el vector $\mathbf{u}_3 = (0.000, -0.155, 0.000)^t$.
 - D) El modelo de atención genera para la posición $i = 3$ el vector $\mathbf{u}_3 = (-0.155)^t$.
- 8 **B** En un sistema de traducción basado en la arquitectura Transformer indicar qué opción es la errónea:
- A) La codificación de la posición (“position embedding”) en la oración fuente tiene como objetivo introducir información sobre el orden de los tokens de entrada en la representación de los mismos.
 - B) La codificación de la posición (“position embedding”) en la oración fuente tiene como misión distinguir las posiciones pares de las impares.
 - C) La codificación de la posición (“position embedding”) en la oración traducida tiene como objetivo introducir información sobre el orden de los tokens traducidos en la representación de los mismos.
 - D) La codificación de la posición (“position embedding”) debe tener la misma dimensión que la codificación de los tokens (“word embedding”).
- 9 **B** En un sistema de traducción basado en la arquitectura Transformer indicar qué opción es la errónea:
- A) La cabezas múltiples (“multi-head”) en los mecanismos de atención en el codificador y en el decodificador tiene como uno de los objetivos que el mecanismo de atención se realice en un espacio de menor dimensión.
 - B) La cabezas múltiples (“multi-head”) en los mecanismos de atención en el codificador y en el decodificador tiene como objetivo aumentar la dimensión de la representación de los tokens fuente y destino.
 - C) La cabezas múltiples (“multi-head”) en los mecanismos de atención en el codificador y en el decodificador tiene como uno de los objetivos la captura de características distintas de los tokens fuente y destino.
 - D) La cabezas múltiples (“multi-head”) se aplican a los mecanismo de autoatención del codificador y del decodificador y al mecanismo de atención cruzada.
- 10 **C** En la inferencia de la traducción basada en LSTMs o en Transformer indicar cuál es la opción correcta:
- A) La búsqueda voraz (“greedy search”) es más rápida y produce traducciones de calidad semejante a la búsqueda en haz (“beam search”) en general.
 - B) La búsqueda voraz (“greedy search”) es igual de rápida que la búsqueda en haz (“beam search”) pero produce mejores traducciones en general.
 - C) La búsqueda voraz (“greedy search”) es más rápida pero produce traducciones de menor calidad que la búsqueda en haz (“beam search”) en general.
 - D) La búsqueda voraz (“greedy search”) es más rápida y produce traducciones de mejor calidad que la búsqueda en haz (“beam search”) en general.
- 11 **A** En el entrenamiento de los modelos de traducción basados en Transformers, indicar cuál es la opción errónea:
- A) El factor de aprendizaje va variando durante el entrenamiento en función del tamaño del conjunto de muestras que forman parte del “minibatch” utilizado.
 - B) El factor de aprendizaje puede ir variando durante el entrenamiento en función de la evolución del gradiente en el entrenamiento.
 - C) El factor de aprendizaje puede ir variando durante el entrenamiento en función de la evolución de las variaciones de los pesos.
 - D) El factor de aprendizaje varía de distinta forma según el algoritmo de optimización escogido (“Adam”, “Adadelta”, ...)

- 12 **B** En el proceso de corrección de los errores de traducción realizado por un humano indicar qué afirmación es correcta:
- A) La traducción interactiva es otra forma de denominar a la post-edición.
 - B) En la traducción interactiva el sistema de traducción interviene cada vez que el humano realiza una operación de edición mientras que en la post-edición el sistema de traducción solo interviene al principio.
 - C) La adaptación online de los modelos de traducción solo se puede realizar en el marco de la traducción interactiva.
 - D) La adaptación online de los modelos de traducción solo se puede realizar en el marco de la post-edición.
- 13 **D** Sobre la traducción interactiva indicar cuál es la opción errónea:
- A) La interacción basada en prefijos se reduce a un problema de buqueda sobre el espacio de sufijos.
 - B) La regla utilizada en la interacción basada en prefijos $\operatorname{argmax}_{y \in \Sigma} \sum_{y_s} Pr(y y_s | x, y_p)$ garantiza que el número de interacciones es mínimo (Σ es el vocaulario de salida).
 - C) La interacción basada en segmentos se reduce a un problema de buqueda sobre el espacio de segmentos de traducción.
 - D) La regla utilizada en la interacción basada en prefijos $\operatorname{argmax}_{y_s} Pr(y_s | x, y_p)$ garantiza que el número de interacciones es mínimo.
- 14 **D** Indicar cuál es la afirmación errónea sobre el proceso de evaluación de un sistema de traducción interactiva:
- A) Las métricas automáticas usuales no miden el esfuerzo cognitivo del humano que realiza las correcciones,
 - B) La métrica agnóstica como el BLEU no mide el esfuerzo humano de corrección/supervisión humana.
 - C) Las métricas automáticas usuales solo miden el esfuerzo físico de corrección/supervisión humana.
 - D) Un sistema de traducción interactiva es mejor que otro si el BLEU/TER que se obtiene del primero es superior/inferior al del segundo.
- 15 **B** En los modelos de lenguaje pre-entrenados basados en Transformer indicar qué afirmación es errónea:
- A) El modelo BERT está basado en el codificador del Transformer y su entrenamiento se basa en el enmascaramiento de tokens de entrada.
 - B) Los modelos auto-regresivos como GPT tienen todos los componentes del decodificador de la arquitectura codificador-decodificador del Transformer.
 - C) El modelo BART está basados en un Tranformer completo y su entrenamiento se basa en el enmascaramiento de tokens de entrada.
 - D) Los modelos de lenguaje pre-entrenados multilingües como mBART se pueden utilizar como traductores para pares de lenguas con pocos recursos.

Algunas fórmulas

$$1 \quad P_{M1}(y | x) = \frac{p(I | x)}{(J+1)^I} \prod_{i=1}^I \sum_{j=0}^J l(y_i | x_j) \quad \text{T2:21.}$$

$$2 \quad \text{Para } 1 \leq i \leq I, \quad A[i] := \operatorname{argmax}_{0 \leq j \leq J} l(y_i | x_j) \quad \text{T2:27.}$$

$$5 \quad p_{lex}(\tilde{y} | \tilde{x}, \tilde{a}) = \prod_{i=1}^{|\tilde{y}|} \frac{1}{|\{j | (i, j) \in \tilde{a}\}|} \sum_{(i, j) \in \tilde{a}} l(y_i | x_j) \quad \text{con } A = \tilde{a} = \{(1, 2), (2, 1)\}, \tilde{y} = "1 \ 2" \text{ y } \tilde{x} = "a \ b" \quad \text{T2:57.}$$

$$7 \quad u_i = \sum_{j=1}^J f_j(h_1^e, \dots, h_J^e, h_{i-1}^d) h_j^e \quad \text{con } f_j(h_1^e, \dots, h_J^e, h_{i-1}^d) = \frac{\exp(h_{i-1}^d \cdot h_j^e)}{\sum_{j'=1}^J \exp(h_{i-1}^d \cdot h_{j'}^e)} \quad 1 \leq j \leq J \ \& \ 1 \leq i \leq I \quad \text{T3:34}$$