

Redes neuronales recurrentes para el desarrollo de las habilidades conversacionales de un asistente de aprendizaje

Erik Carbajal-Degante¹, Omar Terrazas Razo¹, Jackeline Bucio García¹,
Jimena Olveres², Boris Escalante-Ramírez², Guadalupe Vadillo¹

¹ Universidad Nacional Autónoma de México,
Coord. de Universidad Abierta, Innovación Educativa y Educación a Distancia,
México

² Centro de Estudios en Computación Avanzada,
Universidad Nacional Autónoma de México,
México

{erikycd, guadalupe.vadillo}@gmail.com

Resumen. Este trabajo presenta los resultados de un procedimiento para desarrollar las habilidades conversacionales de un asistente de aprendizaje a través del diseño e implementación con algunas bases de datos. Se utilizan técnicas tradicionales de procesamiento de lenguaje natural así como una arquitectura conocida como *seq2seq* de redes neuronales recurrentes para la generación del lenguaje. Los resultados alcanzados proporcionan un indicio de buen rendimiento y son complementados con experimentos subjetivos. Esta propuesta constituye un elemento importante de interacción humano-máquina de los sistemas de tutoría inteligente que enriquece la posibilidad de personalizar el entorno del estudiante además de apoyarle en su formación.

Palabras clave: Asistente de aprendizaje, chatbot, redes neuronales recurrentes, sistemas de tutoría inteligente.

Recurrent Neural Networks for the Development of Conversational Skills of a Learning Assistant

Abstract. In this work, we highlight the importance of the written conversational skills of learning assistants and address the design and implementation with some databases. Natural language processing techniques in combination with a deep neural network architecture known as *seq2seq* is used to generate language. Results provide a clue of good performance that are complemented with subjective assessments. This proposal represents an important element of the human-machine interaction of intelligent tutoring systems that fosters the learners' environment customization and supports the learning process.

Keywords: Learning assistant, chatbot, recurrent neural networks, intelligent tutoring systems.

1. Introducción

La inclusión de la inteligencia artificial (IA) en la educación es cada vez mayor. En una revisión de la literatura de 2009 a 2021, [3] señalan que la investigación publicada sobre sistemas de tutoría inteligente (STI) ocupa uno de los primeros lugares y que la frecuencia y cantidad de publicaciones relativas a ella manifiestan una tendencia de crecimiento en el periodo analizado.

Los STI se conciben, de acuerdo con Ubani y Nielsen [20], retomando los lineamientos de Graesser y su equipo, como entornos de aprendizaje dentro de sistemas de cómputo que aportan a los aprendices modelos personalizados de retroalimentación o instrucciones a partir de modelos computacionales.

Permiten dar seguimiento detallado a diversos estados psicológicos (como emociones, habilidades o nivel de conocimiento) y adaptarse a los resultados para potenciar los aprendizajes de cada estudiante. Requieren un componente de interacción con el usuario y resulta deseable que se comuniquen de la forma más humana posible, por lo que se busca que desarrollen habilidades conversacionales. En la siguiente sección se detallan los tipos de componentes disponibles.

1.1. Chatbots, agentes y asistentes

Se conoce como chatbot a una herramienta de software que interactúa con los usuarios sobre un tema determinado o en un dominio específico de forma natural y conversacional utilizando texto o voz [4]. Para muchos propósitos diferentes, los chatbots se han utilizado en una amplia gama de dominios los cuales incluyen marketing, servicio a cliente, soporte técnico así como educación y capacitación. Se estima que el 70 % de las empresas de tecnología poseerán la capacidad de diseñar su propio sistema de conversación digital a mediados del 2022 [5].

En este sentido, la mayoría de los chatbots serán construidos con el objetivo de brindar una mejor experiencia al usuario, proveer servicios especializados, facilitar y automatizar ciertos procesos, y reducir en gran medida el costo de la interacción humana así como proporcionar disponibilidad y simplicidad de uso, por lo que se puede hablar de una transición generacional de estas entidades con funciones avanzadas donde la inteligencia artificial juega un papel muy importante.

Chatbots. En su forma más simple, los chatbots suelen seguir un conjunto de reglas o flujos establecidos para contestar a las preguntas realizadas por el usuario. Estas reglas o flujos les permiten responder de manera efectiva a las solicitudes dentro de un dominio específico de diálogo, pero no son eficientes para responder preguntas cuyo patrón no coincide con las reglas para el que está capacitado dado que el entrenamiento utilizado por estos sistemas no es tan elevado.

Agentes virtuales. Por otro lado, los agentes virtuales son entidades más avanzadas que los chatbots pero que igualmente utilizan el procesamiento de lenguaje natural (PLN) como forma de interacción, así mismo hacen uso de técnicas relacionadas con el entendimiento de lenguaje natural (ELN) y generación de lenguaje natural (GLN). Para estas entidades, los datos son cruciales puesto que sus habilidades son resultado de utilizar modelos entrenables del aprendizaje automático y del aprendizaje profundo, por lo que se dice que los agentes virtuales poseen cierto nivel de inteligencia (ver Figura 1 para mayor detalle comparativo).

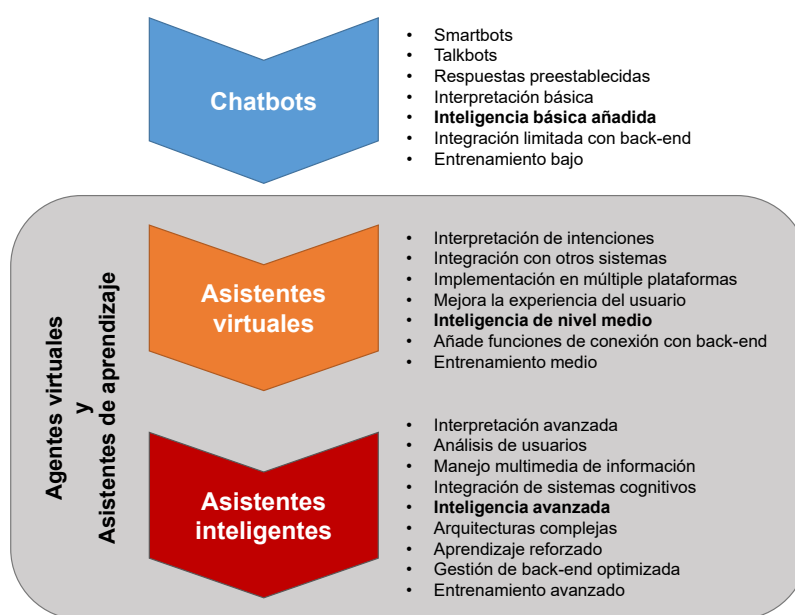


Fig. 1. Transición generacional de los chatbots y agentes virtuales.

Asistentes. Es posible considerar a los asistentes virtuales como parte de los agentes virtuales con un nivel de inteligencia medio. Las habilidades de los asistentes virtuales son mayores a las de los chatbots dado que utilizan un sistema de interpretación de intenciones. Con esto logran abordar diferentes tareas dentro de un abanico más amplio de opciones. Poseen la capacidad de integrarse con otros servicios y sistemas o correr en múltiples plataformas; en consecuencia mejoran considerablemente la experiencia al usuario.

Dentro de los asistentes virtuales, las entidades más sofisticadas son los asistentes inteligentes. Se busca crear asistentes con funciones y características más complejas que permitan interactuar de manera fluida con el usuario, como lo haría un ser humano. Incorporan capacidades de inteligencia mayor como interpretación y generación de lenguaje, análisis de sentimientos y son

capaces de definir un perfil del usuario y adaptarse a el, lo que se conoce como servicios cognitivos [2].

En este sentido los asistentes inteligentes son las entidades más difíciles de diseñar y su arquitectura roza la vanguardia tecnológica de desarrollo en IA, la cual sigue creciendo a pasos agigantados. El uso de datos masivos y limpios es un requisito importante dado que la capacidad de aprender por sus medios (aprendizaje por refuerzo) es una característica distintiva.

1.2. Asistentes de aprendizaje

En los últimos años, cada vez más organizaciones han comenzado a explotar los beneficios de IA más allá de la simple consulta de información seguida de una respuesta programada.

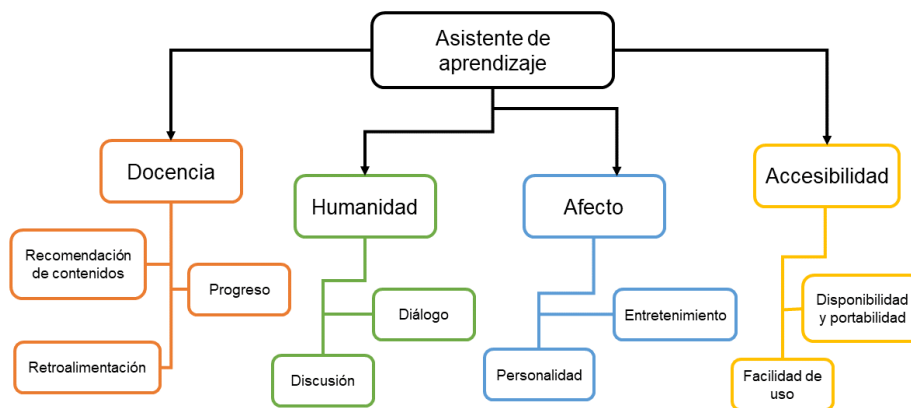


Fig. 2. Atributos y habilidades de un asistente de aprendizaje.

La incorporación de los chatbots y agentes al área educativa durante los últimos años implica un aumento en el interés por las formas en que estas entidades podrían implementarse para la enseñanza y el aprendizaje, lo que da lugar a la creación de asistentes de aprendizaje [19].

Los asistentes de aprendizaje pueden considerarse como parte de los agentes virtuales. Se enfocan en mejorar y personalizar la automatización en la enseñanza a través de la IA. El conocimiento de los modelos de IA es importante para desarrollar agentes pedagógicos útiles, interesantes y valiosos que no solo aprovechen al máximo los más recientes avances científicos, sino también identifiquen en cierta forma las preocupaciones emocionales, cognitivas y de educación, por lo que su diseño se convierte en un reto importante en la actualidad.

Los asistentes de aprendizaje pueden proporcionar los beneficios de la disponibilidad instantánea y la capacidad de responder de forma natural a través de una interfaz conversacional con ventajas similares a las de entablar un diálogo

con un profesor. Estos asistentes demuestran la capacidad de crear interacciones sencillas con los usuarios además de respaldar la participación, establecer objetivos, proponer estrategias y analizar los resultados de aprendizaje o capacitación [16].

Dentro de los métodos para evaluar la calidad de los asistentes de aprendizaje se encuentran los basados en el proceso jerárquico analítico [14], cuyo enfoque es el adecuado para resolver problemas de toma de decisiones multicriterio. Bajo este esquema y a través de múltiples estudios realizados en [11], se concluyó que los atributos sobresalientes para la medición de la calidad incluyen la eficacia (funcionalidad y humanidad), la eficiencia (rendimiento) y la satisfacción (accesibilidad, afecto, comportamiento y ética).

En [16] se sigue la misma línea de análisis obteniendo cuatro categorías principales: enseñanza, humanidad, afecto y accesibilidad. Finalmente, estos atributos son considerados en el desarrollo de asistentes de aprendizaje más complejos y avanzados, por lo que su diseño no debe omitir estos criterios de calidad. En la Figura 2 se observan las funcionalidades ligadas a los cuatro atributos de un asistente de aprendizaje.

El estudio realizado por [16] para diversos asistentes pedagógicos utilizados en la plataforma de Facebook messenger en diferentes idiomas indica que de los atributos existentes, humanidad es uno de los más relevantes ya que es el punto de partida de interacción con los usuarios y se encuentra íntimamente ligado al tipo de personalidad, por lo que las habilidades conversacionales de un asistente juegan un papel importante en los primeros pasos en el diseño e implementación.

2. Métodos y materiales

Esta sección detalla algunos conceptos útiles relacionados al procesamiento del lenguaje que son la base del desarrollo de aplicaciones que involucran texto. También se detallan varias de las técnicas comunes de pre-procesamiento de información que transforman los datos para su aprovechamiento por los sistemas computacionales.

2.1. Procesamiento del lenguaje

El procesamiento de lenguaje natural (NLP, por sus siglas en inglés) es el campo de estudio de la IA enfocado en la comprensión, manejo y generación de lenguaje humano por medio de algoritmos de Machine Learning (ML) [10]. En los últimos años ha permitido incrementar el número de posibilidades dentro del ámbito académico para obtener y entregar retroalimentación en tiempo real y así mejorar la calidad de la educación mediante la aplicación de técnicas de análisis de sentimientos [?].

Como parte de este procesamiento se encuentra la comprensión de lenguaje natural (NLU, por sus siglas en inglés) que se enfoca en procesar entradas no estructuradas y convertirlas a un formato estructurado, tarea que resulta difícil cuando se tienen las complejidades de lenguaje como la anáfora, elisión,

ambigüedad e incertidumbre [12]. Es aquí donde las redes neuronales recurrentes (RNN, por sus siglas en inglés) juegan un papel importante. Finalmente, la generación de lenguaje natural (NLG, por sus siglas en inglés) se encarga de transformar los datos estructurados en lenguaje natural, como voz o texto.

Debido a que parte de este procesamiento de lenguaje, mientras más grande sea el modelo tendrá una mayor certeza, como la adaptación de GPT-2 en DialoGPT [13] y sus 147 millones de diálogos, mientras que Menna, el modelo de Google AI, tiene 2,600 millones de parámetros y se entrena con 341 GB de texto, lo que lo coloca con una capacidad 1.7 veces mayor que los modelos de última generación como OpenAI GPT-2 y con 8.5 veces más datos.

2.2. Técnicas de procesamiento

La limpieza de datos es el proceso que permite eliminar datos incorrectos, duplicados o corruptos de un conjunto y es sumamente útil cuando se combinan múltiples fuentes de datos. Para realizar este proceso existen diferentes métodos que integran algoritmos de ML [6]. La tokenización es el primer paso en el proceso de NLP: divide los datos no estructurados y el texto en fragmentos que se consideran elementos discretos, con lo que se tienen datos numéricos adecuados para el aprendizaje automático, por ejemplo [7].

La tokenización puede separar oraciones, palabras, caracteres o subpalabras. Lemmatization y stemming son dos métodos empleados por los chatbots para analizar el significado detrás de una palabra. Por lo general, lemmatization se realiza mediante algoritmos Tree (árboles de palabras clave) fundamentados en la programación dinámica basada en la distancia de Levenshtein y en la estructura de los datos [15] y busca la mejor palabra de origen de las que se tienen.

Por otro lado, el proceso de stemming reduce la inflexión en las palabras a sus formas de raíz. Lo que ayuda en el pre-procesamiento de texto, palabras y documentos para la normalización del texto [17]. La derivación de cada idioma es diferente y está fuertemente afectada por el tipo de idioma del texto, lo que hace que existan múltiples variantes de los algoritmos para atender cada caso.

2.3. Transición de redes secuenciales a redes recurrentes

Los humanos no iniciamos nuestro razonamiento desde cero cada segundo que pasa, por ejemplo, mientras leemos un documento de texto, entendemos lo que está escrito con base en nuestro entendimiento de las palabras anteriores, lo que construye el contexto. En este sentido, las redes neuronales tradicionales (NNs) no pueden simular esta tarea por su características secuenciales y olvidan atributos previos.

Las redes secuenciales son un tipo de red neuronal en donde cada entrada se procesa de manera independiente sin considerar los datos procesados con anterioridad. En el caso de las redes recurrentes, se utiliza información de los procesos y datos anteriores para calcular una nueva salida. Las llamadas RNNs se utilizan ampliamente para realizar análisis de secuencias ya que están diseñadas

para extraer la información contextual definiendo las dependencias entre varios pasos de tiempo. De forma general, las RNNs se caracterizan por un flujo de retroalimentación de sus estados internos.

3. Arquitectura conversacional

Los modelos secuencia a secuencia (a menudo abreviados como *seq2seq*) son una clase especial de arquitecturas de redes neuronales recurrentes compuestas a su vez de dos arquitecturas en cascada nombradas como codificador y decodificador. Normalmente, la arquitectura *seq2seq* se utiliza para resolver problemas complejos relacionados al lenguaje como traducción automática, sistemas de pregunta y respuesta, resumen de texto, reconocimiento del habla, etc. [18].

El modelo *seq2seq* más simple consta de dos redes recurrentes LSTM (Long-Short Term Memory), una para el codificador y otro para el decodificador. Básicamente, el codificador procesa el texto de entrada produciendo un estado final el cual es utilizado por la etapa del decodificador. A través de este proceso se busca que el codificador capture toda la información sobre la fuente para que el decodificador pueda generar un texto de salida en función de estos estados obtenidos.

3.1. Proceso de entrenamiento

Similar a muchos procesos de entrenamiento de redes neuronales, los modelos *seq2seq* aplicados al lenguaje se entrenan para predecir distribuciones de probabilidad de una frase o 'token' dado un contexto (conjunto de tokens previos).

Tabla 1. Hiper-parámetros del modelo.

| Hiper-parámetro | Valor | Hiper-parámetro | Valor |
|---------------------|----------|--------------------------|---------|
| Dimensión LSTM | 256 | Longitud de la secuencia | 15 |
| Optimizador | RMSProp | Activación | Softmax |
| Función de pérdida | CCE | Tamaño de batch | 64 |
| Métrica | Accuracy | Dropout | 0.2 |
| Tasa de aprendizaje | 0.01 | Épocas | 50 |

A cada paso se intenta maximizar la probabilidad de asignar el token correcto reduciendo la función de pérdida (comúnmente entropía cruzada). Formalmente, se asume que tenemos una instancia de entrenamiento cuya fuente es una secuencia

$x = (x_1, \dots, x_m)$ y otra secuencia objetivo $y = (y_1, \dots, y_n)$, donde ambas secuencias pueden o no diferir de tamaño, el objetivo de *seq2seq* es estimar la probabilidad condicional $p^{(t)} = p(*|y_1, \dots, y_{t-1}, x_1, \dots, x_m)$, donde t representa

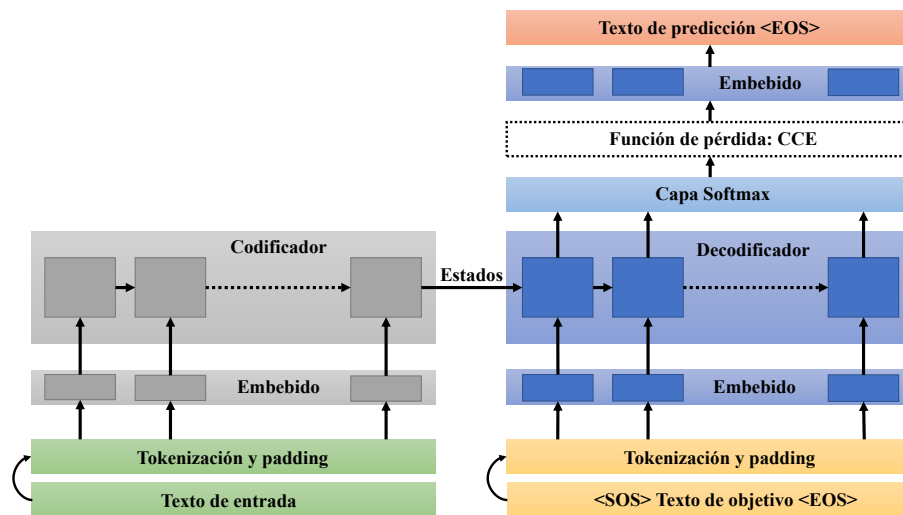


Fig. 3. Diagrama a bloques del entrenamiento *seq2seq*.

cada paso de tiempo. De forma general, el proceso completo puede modelarse como:

$$p(y_1, \dots, y_n | x_1, \dots, x_m) = \prod_{t=1}^n p(y_t | y_1, \dots, y_{t-1}, x_1, \dots, x_m), \quad (1)$$

donde n es la longitud de y , m es la longitud de x . En el caso práctico, dos arreglos de texto correspondientes a un diálogo de tipo pregunta-respuesta se procesan mediante técnicas de tokenización para la conversión de cada palabra, lo que produce una representación numérica de cada frase, lo que llamamos teóricamente x e y .

Para el caso del decodificador, se utilizan dos palabras reservadas $\langle \text{SOS} \rangle$ y $\langle \text{EOS} \rangle$ que indican respectivamente el inicio y el final de cada oración en la secuencia objetivo. Posteriormente, los arreglos x e y se embeben y utilizados como elemento de entrada de cada bloque LSTM, tanto del codificador, como del decodificador. Primero, el estado final resultante del codificador se utiliza como estado inicial del decodificador.

Posteriormente, el decodificador hará predicciones por cada palabra la cual se comparará con la palabra objetivo. Finalmente, mediante el cálculo del error a través de la función de pérdida, el sistema ajustará los pesos de la arquitectura a cada iteración. El proceso descrito anteriormente puede verse en la Figura 3 así como los hiper-parámetros utilizados en la Tabla 1, los cuales han sido ajustados heurísticamente.

3.2. Descripción de las bases de datos

Se ha realizado un proceso de búsqueda de datos en Internet para conseguir la información y acoplarla a la forma requerida por la arquitectura. Las características de diálogo estilo pregunta-respuesta en el idioma español no son tan comunes como se puede encontrar en otros idiomas, complicando más el hallar datos específicos con un contexto educativo. Se han realizado experimentos solo sobre tres bases de datos (BD) diferentes: dos BD son de acceso público y una fué elaborada localmente mediante técnicas de web scarping. Estas BD se detallan a continuación:

1. Los Sopranos. Extraída de Kaggle, consiste en la transcripción de los diálogos de una serie de televisión. En los experimentos descritos a continuación, se utilizan los archivos de la primera temporada con 13 episodios de 14k líneas de diálogo conteniendo 7,645 palabras. Ver la gráfica de nubes de palabras en Figura 4a.
2. OpenSubtitle. Corpus construido y estudiado en [9], es una BD masiva de 3.7M de archivos que recopila los diálogos de algunas películas en 62 idiomas diferentes. Para los experimentos llevados a cabo en este trabajo, se ocupó un extracto de 20k líneas de diálogo las cuales construyen un vocabulario de 9,680 palabras diferentes, lo que representa alrededor del 10 % del vocabulario del idioma español. Ver su gráfica de nubes de palabras en la Figura 4b.
3. Asistente. Es una BD la cual consta de una colección de datos generados localmente, así como diálogos extraídos de un conjunto público en [1]. Las oraciones han sido revisadas, limpiadas y pre-procesadas. Esta base de datos contiene 2k líneas de diálogo con un vocabulario de 1,857 palabras. Véase en la Figura 4c su gráfica de nube de palabras.

3.3. Recursos computacionales

Este trabajo se implementó en su totalidad en lenguaje Python versión 3.7. Una computadora con CPU Intel(R) Xeon(R) Silver 4216 CPU@2.10, así como una tarjeta gráfica NVIDIA GeForce RTX3090 se utilizó para entrenamiento e inferencia mediante las bibliotecas proporcionadas por Keras sobre Tensorflow versión 2.7.

4. Resultados

4.1. Proceso de inferencia

El proceso de inferencia es muy similar al entrenamiento y se define habitualmente como desarrollos independientes ya que las entradas y salidas son diferentes para ambos casos. En este proceso, la inferencia aprovechará todos los parámetros de la red aprendidos en la etapa de entrenamiento. La arquitectura



Fig. 4. Nube de palabras de las bases de datos extraídas de: Los Soprano a, Opensubtitle b y propia c.

del codificador no cambia, por lo que se alimenta la misma red con una nueva frase (conjunto de tokens).

El decodificador, por su parte, se alimentará tanto del estado final del codificador, como de la palabra reservada <SOS> que le indicará el inicio de iteración. El resultado de la predicción de la primera palabra será el elemento de entrada del segundo bloque y así, sucesivamente hasta alcanzar la longitud máxima de la secuencia o bien la palabra reservada que indica el final: <EOS>. Este proceso se conoce como greedy decoding y busca encontrar la máxima probabilidad de las palabras siguiendo un procedimiento en cadena, de la forma, donde y' representa la secuencia de predicción:

$$y' = \arg \max_y p(y|x) = \arg \max_y \prod_{t=1}^n p(y_t | y_{<t}, x). \quad (2)$$

Otras técnicas existentes en la literatura como beam search permiten mejorar los resultados del proceso de inferencia al producir varias hipótesis y calcular la mejor combinación, aunque este procedimiento puede incrementar considerablemente el tiempo de respuesta.

4.2. Gráficas de desempeño

El desempeño de la arquitectura puede visualizarse mediante las gráficas de pérdida y precisión para cada época $\in [0, 50]$. Dichas gráficas se muestran en la Figura 5 para las tres BD estudiadas. Se puede notar en la Figura 5(a) que la pérdida disminuye a cada época para todas las BD, lo que da un indicio de convergencia del modelo hacia los datos de entrenamiento. Se puede ver que los

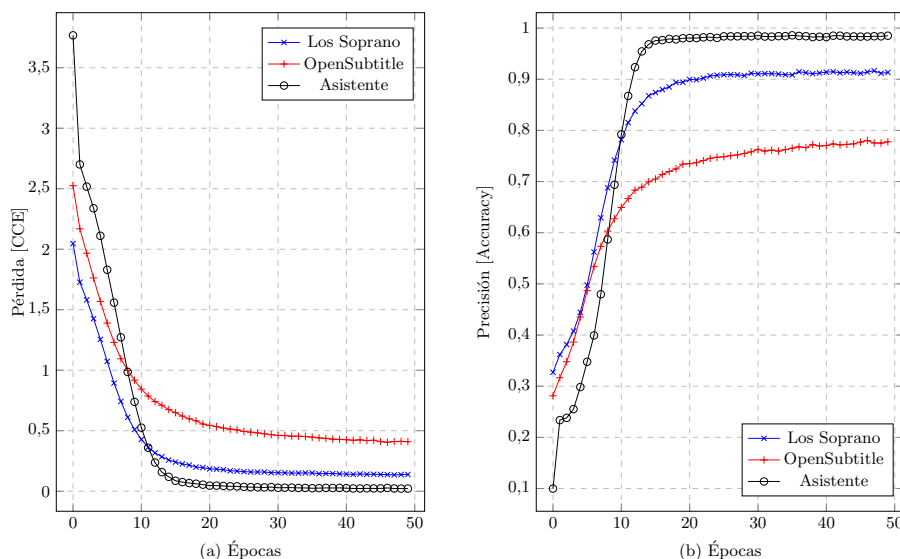


Fig. 5. Curvas de pérdida medida con CCE en (a), curvas de accuracy en (b).

valores mínimos de pérdida se alcanzan alrededor de la época 30, aunque para algunas BD la convergencia es más rápida debido a la cantidad de información utilizada y la limpieza de los datos.

Por otro lado, la métrica utilizada en términos de accuracy, ver Figura 5(b), indica la precisión del ajuste del modelo hacia los datos entrenados. Tanto las funciones de pérdida y la métrica de precisión funcionan de diferente manera cuando se analizan distintos datos (series de tiempo, texto, imágenes, video), por lo que posiblemente una función de pérdida más adecuada brindará mejor información del ajuste realizado sobre texto.

4.3. Comparativa de diálogo

El trabajo de [8] presenta un modelo de diálogo basado en *seq2seq* con el que se simulan diferentes comportamientos de personalidad. Sus resultados se comparan subjetivamente con el fin de medir la calidad de las respuestas y la cantidad de inconsistencias como complemento de las evaluaciones objetivas. Hemos extraído un conjunto de esas preguntas para estimar el desempeño subjetivamente del modelo con el entrenamiento de las 3 BD. En la Tabla 2 se muestran las respuestas obtenidas a las preguntas base que se han realizado, donde se pueden notar las inconsistencias entre algunas de las respuestas y la tendencia del modelo a adquirir una personalidad dada la información con la que ha sido entrenado.

Tabla 2. Tabla comparativa de las respuestas a un diálogo.

| Pregunta | Modelo | Respuesta |
|-----------------------|--------------|--|
| Hola | Los Soprano | Hola buscaba a tony soprano |
| | OpenSubtitle | Hola que tal estais |
| | Asistente | Hola estas de vuelta |
| ¿Donde vives? | Los Soprano | A que se refiere |
| | OpenSubtitle | Hay alguien mejor que podemos hablar de los insectos |
| | Asistente | Vivo en una computadora |
| ¿Cuántos años tienes? | Los Soprano | Feliz cumpleaños |
| | OpenSubtitle | Como es tu mejor |
| | Asistente | Puedo vivir tanto como mis padres quieran |
| ¿Que estudias? | Los Soprano | No se puedo hacerlo |
| | OpenSubtitle | Me gusta |
| | Asistente | Simplemente no puedo salir de la habitacion |
| ¿Tienes sentimientos? | Los Soprano | En serio |
| | OpenSubtitle | Siempre así que claro |
| | Asistente | No tengo sentimientos en este momento soy un chatbot |

4.4. Limitaciones

Básicamente, existen dos inconvenientes principales de la arquitectura *seq2seq*, ambos relacionados con la longitud de las secuencias procesadas. En primer lugar, se habla de una memoria limitada en donde frases más complejas o con un contexto mayor de entendimiento pueden insertar errores por la pérdida de información ya sea al forzar un número máximo de caracteres (operaciones de truncamiento y compresión), o bien los estados finales del codificador son insuficientes para alimentar el decodificador.

Segundo, por regla general, cuanto más profunda es una red neuronal, mayor incertidumbre de interpretabilidad. En el caso de redes neuronales recurrentes, a profundidades grandes y secuencias largas de texto los gradientes desaparecen al realizar el proceso de retro-propagación. Aunque en teoría las redes recurrentes ayudan a prevenir en cierta medida este problema, sigue estando presente dicha limitante.

5. Discusión: El reto de la personalidad coherente

A pesar de la tecnología disponible en PLN e IA, la cual ha demostrado alcanzar resultados muy cercanos e incluso rebasar a los que pueden lograrse mediante la interacción humana, existen algunas limitaciones debido a la naturaleza del razonamiento humano que para las máquinas les es imposible simular. Tales limitaciones emergen de la falta de conciencia e inteligencia emocional, confusión por complejidad de las intenciones o del contexto, así como la necesidad de un continuo entrenamiento y una base de datos robusta.

Tal es el caso que aborda este trabajo al exhibir una parte importante de las limitantes producidas al entrenar modelos con datos escasos y que muestran características de diálogo específicas, con lo que una tendencia de personalidad particular se hace presente. Lo ideal es que los chatbots y agentes produzcan respuestas coherentes a las entradas, esto parece simple pero incorporar un conocimiento establecido para crear una personalidad es uno de los mayores retos en investigación.

La personalidad de estas entidades se refiere básicamente al personaje que interpreta durante una interacción conversacional. Adoptar características como edad, género, idioma, forma de expresión, nivel de conocimiento o especialidad pueden resaltar mejor su carácter y con eso facilitar la comunicación con el usuario estableciendo vínculos de confianza. Como lo resalta el trabajo de [21] que sugiere el uso de agentes con memoria que puedan simular interés en el perfil del usuario al realizar preguntas básicas dentro de un escenario de conversación natural.

Muchos sistemas aprenden a generar respuestas lingüísticas apropiadas pero no están entrenados para generar respuestas semánticamente consistentes. A tales sistemas les es difícil asignar un perfil explícito para generar respuestas coherentes y una de las raíces de este problema se ubica en el tipo de datos con los que se entrenan algunos modelos. La necesidad de datos masivos requeridos por técnicas del aprendizaje profundo es un factor clave para su buen desempeño.

Sin embargo, un conjunto de datos regularmente consiste en una compilación de información de diferentes fuentes. Abordar el problema de escasez de datos limpios y un proceso particular de filtrado es uno de los primeros pasos en la construcción de modelos robustos que adquieran las características de personalidad que la aplicación necesite y que brinden de cierta forma sentido humano a una conversación fluida.

6. Conclusiones y trabajos a futuro

El presente trabajo aborda los fundamentos en el desarrollo e implementación de las habilidades conversacionales de un asistente de aprendizaje, partiendo de la necesidad actual de enfocar la tecnología existente en el sector educativo y de proponer herramientas pedagógicas robustas que se encuentren a la vanguardia. Describimos el proceso de construcción de un sistema basado en el aprendizaje profundo y técnicas del procesamiento de lenguaje natural que le permite entrenarse con tres diferentes bases de datos y formar oraciones en el idioma español en respuesta a una frase de entrada.

Los resultados indican la convergencia del modelo en épocas tempranas lo que da un indicio de rendimiento, aún teniendo bases de datos masivas con un vocabulario extenso. Los resultados subjetivos nos permiten abordar el tema de personalidad coherente que en la actualidad representa un reto importante para cualquier sistema conversacional. Mecanismos de atención o redes más complejas (arquitecturas de tipo Transformers) pueden ayudar a superar las limitaciones que los modelos *seq2seq* presentan por naturaleza.

Se sabe que la capacidad de un asistente de aprendizaje de manejarse dentro de un ambiente con temas simples de charla abre el panorama adecuado para la formación de vínculos de confianza con los estudiantes. De esta manera, se obtiene información útil que construya nuevas bases de datos limpias y completas.

Esto permite también recopilar la información suficiente para realizar tareas futuras relacionadas al analítica del aprendizaje y complementar las funciones empleadas por los sistemas actuales de tutoría inteligente.

Agradecimientos. Este proyecto recibió financiamiento de Santander Univer-

Referencias

1. Alblawi, A. S.: Big data and learning analytics in higher education: Demystifying variety, acquisition, storage, NLP and analytics. arXiv, (2018) doi: 10.1109/icbdaa.2017.8284118
2. Benke, I., Gnewuch, U., Maedche, A.: Understanding the impact of control levels over emotion-aware chatbots. *Computers in Human Behavior*, vol. 129, pp. 107122 (2022) doi: 10.1016/j.chb.2021.107122
3. Chen, X., Zou, D., Xie, H., Cheng, G., Liu, C.: Two decades of artificial intelligence in education: Contributors, collaborations, research topics, challenges, and future directions. *Educational Technology and Society*, vol. 25, no. 1, pp. 28–47 (2022)
4. Dale, R.: The return of the chatbots. *Natural Language Engineering*, vol. 22, no. 5, pp. 811–817 (2016) doi: 10.1017/s1351324916000243
5. Gartner: Chatbots will appeal to modern workers (2019)
6. Hirsch, T., Hofer, B.: Identifying non-natural language artifacts in bug reports. In: 2021 36th IEEE/ACM International Conference on Automated Software Engineering Workshops (ASEW). Ieee (2021) doi: 10.1109/asew52652.2021.00046
7. Li, B., Zhang, Y., Sainath, T., Wu, Y., Chan, W.: Bytes are all you need: End-to-end multilingual speech recognition and synthesis with bytes. arXiv, (2018)
8. Li, J., Galley, M., Brockett, C., Spithourakis, G. P., Gao, J., Dolan, B.: A persona-based neural conversation model. arXiv, (2016)
9. Lison, P., Tiedemann, J.: OpenSubtitles2016: Extracting large parallel corpora from movie and tv subtitles. In: Lrec (2016)
10. Madhuri, D., Prasad, R.: A ML and NLP based framework for sentiment analysis on bigdata. *International Journal of Recent Technology and Engineering*, vol. 8, no. 5, pp. 189–200 (2020) doi: 10.35940/ijitee.d9062.029420
11. Radziwill, N. M., Benton, M. C.: Evaluating quality of chatbots and intelligent conversational agents (2017)
12. Ram, A., Prasad, R., Khatri, C., Venkatesh, A., Gabriel, R., Liu, Q., Nunn, J., Hedayatnia, B., Cheng, M., Nagar, A., King, E., Bland, K., Wartick, A., Pan, Y., Song, H., Jayadevan, S., Hwang, G., Pettigru, A.: Conversational ai: The science behind the alexa prize. arXiv, (2018) doi: 10.48550/arxiv.1801.03604
13. Rastogi, A., Zang, X., Sunkara, S., Gupta, R., Khaitan, P.: Towards scalable multi-domain conversational agents: The schema-guided dialogue dataset. arXiv, (2019)
14. Saaty, T. L.: What is the analytic hierarchy process? *Mathematical Models for Decision Support*, pp. 109–121 (1988) doi: 10.1007/978-3-642-83555-1_5

15. Schmitt, M., Constant, M.: Neural lemmatization of multiword expressions. In: Proceedings of the Joint Workshop on Multiword Expressions and WordNet (MWE-WN 2019). Association for Computational Linguistics (2019) doi: 10.18653/v1/w19-5117
16. Smutny, P., Schreiberova, P.: Chatbots for learning: A review of educational chatbots for the facebook messenger. Computers & Education, vol. 151, pp. 103862 (2020) doi: 10.1016/j.compedu.2020.103862
17. Suci, F. W., Hayatin, N., Munarko, Y.: In-idris: Modification of idris stemming algorithm for indonesian text. IIUM, (2022)
18. Sutskever, I., Vinyals, O., Le, Q. V.: Sequence to sequence learning with neural networks. arXiv, (2014)
19. Tamayo, P. A., Herrero, A., Martín, J., Navarro, C., Tránchez, J. M.: Design of a chatbot as a distance learning assistant. Open Praxis, vol. 12, no. 1, pp. 145 (2020) doi: 10.5944/openpraxis.12.1.1063
20. Ubani, S., Nielsen, R.: Review of collaborative intelligent tutoring systems (CITS) 2009-2021. In: 2022 11th International Conference on Educational and Information Technology (ICEIT). Ieee (2022) doi: 10.1109/iceit54416.2022.9690733
21. Zhang, S., Dinan, E., Urbanek, J., Szlam, A., Kiela, D., Weston, J.: Personalizing dialogue agents: I have a dog, do you have pets too? arXiv, (2018)