

Modelos para la generación automática de diálogos: Una revisión

Andrés Vázquez, David Pinto, Darnes Vilariño, Mauricio Castro

Benemérita Universidad Autónoma de Puebla, Facultad de Ciencias de la Computación,
Puebla, México

{andrex, dpinto, darnes}@cs.buap.mx

Resumen. La generación automática de diálogos es una parte importante en la interacción humano-robot, los diálogos generados deben garantizar una conversación coherente entre el humano y el robot, se espera que la interacción sea lo más natural y eficaz que se pueda, es por eso por lo que en este trabajo se presenta una revisión de algunos modelos que se han utilizado para la generación de diálogos en la interacción humano-computadora.

Palabras clave: Sistema de diálogo, generación de lenguaje natural, interacción humano-robot, redes neuronales.

Models for Automatic Generation of Dialogues: A Review

Abstract. The automatic generation of dialogues is an important part of the human-robot interaction, the generated dialogues must guarantee a coherent conversation between the human and the robot, the interaction is expected to be as natural and effective as possible, that is why this paper presents a review of some models that have been used to generate dialogues in the human-computer interaction.

Keywords. Dialogue system, generation of natural language, human-robot interaction, neural networks.

1. Introducción

A medida que mejora la comprensión del lenguaje y la tecnología de generación automática de diálogos, existe un interés creciente en la construcción de sistemas de conversación de usuario, que pueden ser utilizados para una variedad de aplicaciones tales como planificación de viajes, sistemas tutoriales o soporte técnico basado en chat.

La mayoría de los trabajos en este contexto han enfatizado la comprensión o la generación de una secuencia de palabras asociadas con una sola oración o turno de un orador, potencialmente aprovechando el turno anterior. Más allá del contexto local, el uso del lenguaje en una conversación orientada a objetivos refleja el tema global de discusión, así como el rol respectivo de cada participante.

En los sistemas de diálogo hablado (SDS Spoken Dialogue Systems), la tarea de generación de lenguaje natural (NLG Natutal Language Generation) es convertir una representación con significado (MR Meaning Representation) producida por el gestor de diálogo en una o más oraciones en un lenguaje natural.

En este trabajo se realiza el análisis de algunos métodos utilizados en la generación de diálogos, con el objetivo de tener un punto de referencia en este contexto y en el futuro poder realizar trabajos que permitan proponer nuevas técnicas o mejoras en los modelos que ya existen de generación de diálogos.

2. Sistemas de diálogo

Los sistemas de diálogo hablado o sistemas conversacionales (SDS) son una tecnología concebida para facilitar la interacción natural mediante el habla entre una persona y una computadora. Constituyen como una interfaz o un intermediario entre un usuario y un sistema de cómputo, que tiene la ventaja de no requerir el uso de una pantalla, un teclado o un ratón y de recurrir, en cambio, al medio de comunicación propio de los seres humanos.

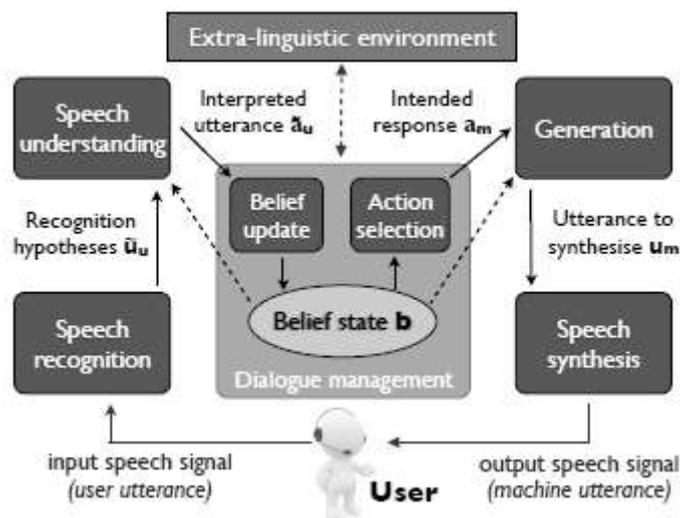


Fig. 1. Esquema de la arquitectura del sistema de diálogo, Fuente Lison [2].

Un sistema de diálogo consiste de una estructura modular en la que cada módulo se ocupa de unas determinadas tareas en interacción con todos los demás. La figura 1 describe una arquitectura de un sistema de diálogo, el cual consta de un módulo de

reconocimiento de habla que se ocupa de transformar la señal sonora en una representación que pueda procesar el módulo de análisis semántico; una vez realizado este análisis, interviene el módulo de gestión del diálogo, conectado a una interfaz que permite la consulta a una base de datos y que envía la información recogida al módulo de generación de respuestas; este último, con la intervención de un conversor de texto en habla, facilitaría al usuario la información deseada. Como elemento central se incluye el contexto del diálogo, que contiene información sobre el enunciado que se está procesando, la historia del diálogo y el estado actual de la interacción.

Puede observarse cómo cada uno de los módulos de un sistema de diálogo está asociado a uno o más modelos que reflejan los conocimientos necesarios para garantizar la interacción en el usuario y una aplicación computacional.

3. Algunas técnicas utilizadas en la generación de diálogos

Dentro de los trabajos que tienen que ver con la generación de diálogos podemos citar el trabajo de Konstantopoulos [1], donde se propone una arquitectura novedosa para los sistemas de diálogo afectivo y multimodal que permite un control explícito sobre los rasgos de personalidad que se quiere que el sistema exhiba. Más específicamente, acercarse a la personalidad como un medio de sintetizar diferentes y posiblemente conflictivos modelos de adaptación en un modelo general que se utilizará para impulsar los componentes de interacción del sistema. Además, esta síntesis se realiza en presencia de conocimiento de dominio, de manera que la estructura del dominio y las relaciones influyen en el cálculo de los resultados.

3.1. Modelos Bayesianos probabilísticos

En esta sección se describen dos trabajos relevantes con el enfoque bayesiano, el trabajo de Lison [2] muestra cómo representar la estructura subyacente de modelos probabilísticos para el diálogo utilizando reglas probabilísticas. Estas reglas se definen como asignaciones estructuradas sobre variables del estado de diálogo, especificadas usando condiciones y efectos de alto nivel. Estas reglas pueden incluir parámetros tales como probabilidades de efecto o utilidades de acción. Las reglas probabilísticas permiten al diseñador del sistema explotar poderosas generalizaciones en la especificación del dominio de diálogo sin sacrificar la naturaleza probabilística del modelo. El marco es muy general y puede expresar un amplio espectro de modelos, desde modelos clásicos totalmente estimados a partir de datos a aquellos que incorporan un rico conocimiento previo. La elección del modelo dentro de este espectro es, por lo tanto, esencialmente una decisión de diseño dependiente de las disponibilidades relativas de datos de entrenamiento y conocimiento de dominio. También se presentan algoritmos para construir redes bayesianas correspondientes a la aplicación de las reglas y para estimar sus parámetros a partir de datos utilizando la inferencia bayesiana como se muestra en la figura 2. El enfoque presentado se ha implementado en un sistema de diálogo oral para la interacción humano-robot y se ha validado en una tarea de aprendizaje de políticas basada en un conjunto de datos de Wizard-of-Oz.

Los resultados empíricos han demostrado que la estructura de reglas permite que el algoritmo de aprendizaje converja más rápido y con mejor rendimiento de generalización.

De igual forma en [3] presenta una evaluación de un sistema de diálogo oral que detecta y se adapta a la desconexión del usuario y a la incertidumbre en tiempo real. Los autores comparan esta versión de su sistema con una versión que se adapta sólo a la desconexión del usuario y a una versión que ignora por completo la desconexión del usuario y la incertidumbre. Encuentran un aumento significativo en el éxito de la tarea al comparar ambas versiones de adaptación de su sistema a su línea base no adaptativa, pero sólo para los usuarios masculinos. Su evaluación examina el impacto de la adaptación a diferentes números de estados afectivos en el éxito de la tarea, y también examina las interacciones con el género del usuario. Sin embargo, esta evaluación sólo se llevó a cabo en un escenario Wizard-of-Oz, donde un humano oculto reemplazó el reconocimiento de voz, el análisis semántico y los componentes de detección de afecto de su sistema de diálogo.

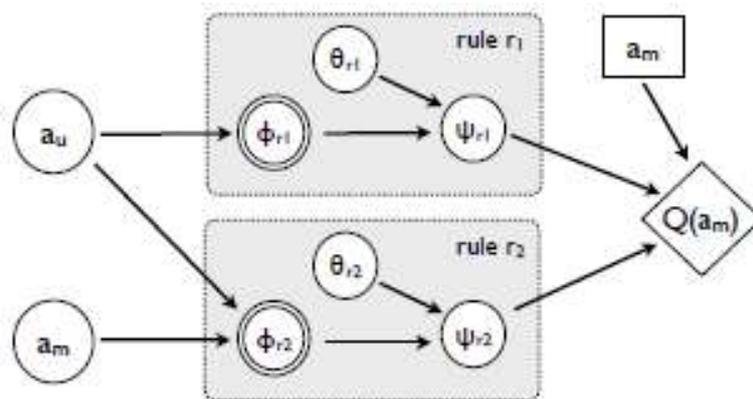


Fig. 2. Red bayesiana con las reglas r_1 y r_2 , fuente Lison [2].

3.2. Modelos estocásticos

En esta sección se describen algunos trabajos que utilizan el modelo basado en procesos de decisión de markov, tal es el caso de Barlier [4] que propone un marco original para modelar los diálogos hablados de humano-computadora para tratar la coadaptación entre usuarios y sistemas de diálogos hablados en tareas no cooperativas. La conversación es modelada como un juego estocástico: tanto el usuario como el sistema tienen sus propias preferencias, pero tienen que llegar a un acuerdo para resolver una tarea no cooperativa. Ellos son entrenados conjuntamente para que el gestor del diálogo aprenda la estrategia óptima contra el mejor usuario posible. Los resultados obtenidos por simulación muestran que se aprenden estrategias no triviales y que este marco es adecuado para el modelado del diálogo.

Ahora bien, en [5] se plantea que, en los últimos años, con la difusión de los robots domésticos, la necesidad de mejorar las capacidades de comunicación de esos robots

con las personas ha ido en aumento. El objetivo de este estudio es construir un marco para un sistema de diálogo que se ocupe de la información multimodal que un robot observa. Ellos aplican el proceso de decisión de Markov parcialmente observable (POMDP por sus siglas en inglés) para modelar la interacción multimodal entre un ser humano y un robot. A través de sus experimentos, han confirmado que el marco propuesto funciona correctamente y logran una interacción multimodal eficaz con un robot.

En los trabajos [6,7] utilizan el aprendizaje de refuerzo (RL) para aprender las políticas de diálogo de preguntas y respuestas para una aplicación del mundo real. En el trabajo de Mitsu y Georgila [6] se analiza un corpus de interacciones de visitantes del museo con dos personajes virtuales que sirven como guías en el Museo de Ciencias de Boston, con el fin de construir un modelo realista de comportamiento del usuario al interactuar con estos personajes. Se construye un usuario simulado sobre la base de este modelo y se utiliza para aprender la política de diálogo de los personajes virtuales usando RL. En este trabajo se sigue un enfoque basado en Procesos de Decisión de Markov Parcialmente Observable (POMDP).

En el trabajo de Lison [8] las contribuciones son dobles. En primer lugar, demostrar cómo aplicar el RL Bayesiano basado en modelos para aprender el modelo de transición de un dominio de diálogo. También se comparan dos enfoques de modelado en el contexto de un escenario humano-robot en el que se instruye a un robot Nao a moverse y recoger objetos. Los resultados empíricos muestran que el uso de representaciones estructuradas permite al algoritmo de aprendizaje converger más rápido y con mejor desempeño de generalización. La idea clave es modelar la gestión del diálogo como un proceso de decisión de Markov (MDP) o un proceso de toma de decisiones de Markov parcialmente observable (POMDP), y dejar que el sistema aprenda por sí mismo la mejor acción a realizar en cada posible situación conversacional mediante interacciones repetidas con un usuario (real o simulado). En [9, 10, 11] también se utiliza aprendizaje de refuerzo como se muestra en la figura 3 donde el modelo de diálogo y la política están parametrizados y dada una función de recompensa adecuada, pueden optimizarse utilizando el aprendizaje de refuerzo. Estudios empíricos han demostrado que las políticas optimizadas a través de RL son generalmente más robustas, flexibles y adaptables que sus homólogos hechos a mano.

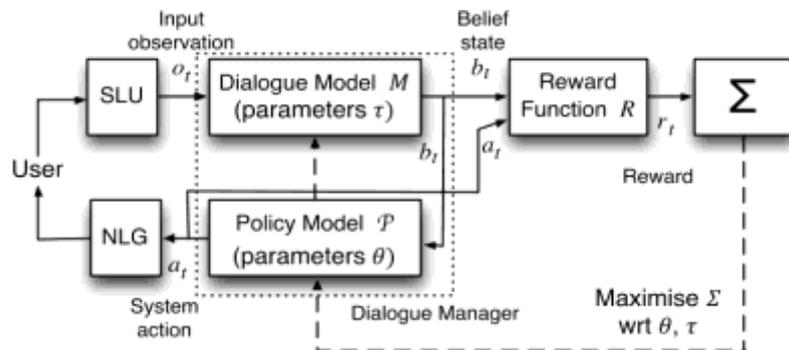


Fig. 3. Componentes de un sistema de diálogo hablado basado en POMDP, fuente Young [11].

3.2.1. Proceso de Decisión de Markov

Un Proceso de Decisión de Markov MDP [7] es formalmente una tupla (S, A, R, T, γ) donde S es el espacio de estado, A es el espacio de acción, $R: S \rightarrow \mathbb{R}$ es la función de recompensa, $T: S \times A \rightarrow P(S)$ es un conjunto de probabilidades de transición Markoviana y γ es un factor de descuento $0 \leq \gamma \leq 1$. La optimización del problema de la toma de decisiones consiste en encontrar una política $\pi: S \rightarrow P(A)$ que mapea estados a acciones de tal manera que las recompensas acumulativas obtenidas siguiendo esta política se maximiza. Para ello, la calidad de una política se mide en cada estado como la recompensa acumulativa esperada que puede obtenerse siguiendo la política que comienza en este estado. Esta medida se denomina función de valor $V^\pi: S \rightarrow \mathbb{R}$:

$$V^\pi(s) = E\left[\sum_{i=0}^{\infty} \gamma^i R(s_i) | s_0 = s, a_i = \pi(s_i)\right]. \quad (1)$$

Se puede definir un orden sobre funciones de valor tales como $V^1 > V^2$ si $\forall s V^1(s) > V^2(s)$. La política óptima π^* es el que maximiza la función de valor para cada estado: $\pi^* = \arg \max_{\pi} V^\pi$.

3.2.2. Proceso de Decisión de Markov Parcialmente Observable

Un POMDP [6] se define como una tupla $(S, A, P, R, O, Z, \gamma, b_0)$ donde S es el conjunto de estados (que representan contextos diferentes) en los que el sistema puede estar (el mundo del sistema), A es el conjunto de acciones del sistema, $P: S \times A \rightarrow P(S, A)$ es el conjunto de probabilidades de transición entre estados después de tomar una acción, $R: S \times A \rightarrow \mathbb{R}$ es la función de recompensa, O es un conjunto de observaciones que el sistema puede recibir sobre el mundo, Z es un conjunto de las probabilidades de observación $Z: S \times A \rightarrow Z(S, A)$ y γ es un factor de descuento que pondera las recompensas a largo plazo. En cualquier momento dado el paso i el mundo está en algún estado no observado $s_i \in S$. Porque si no se conoce exactamente se mantiene una distribución sobre estados llamados estado de creencia b , así $b(s_i)$ es la probabilidad de estar en estado s_i , con estado de creencia inicial b_0 . Cuando el sistema realiza una acción $\alpha_i \in A$ basada en b , siguiendo una política $\pi: S \rightarrow A$, recibe una recompensa $r_i(s_i, \alpha_i) \in \mathbb{R}$ y la transición al estado s_{i+1} de acuerdo con $P(s_{i+1} | s_i, \alpha_i) \in P$. El sistema recibe entonces una observación o_{i+1} según $P(o_{i+1} | s_{i+1}, \alpha_i)$. La calidad de la política π seguido por el agente se mide por la recompensa futura esperada también llamada Q-función, $Q^\pi: S \times A \rightarrow \mathbb{R}$.

3.3. Modelos de redes neuronales

En los trabajos [12, 13] presentan un nuevo sistema de generación de lenguaje natural para sistemas de diálogo hablado capaces de incorporar (adaptar) la forma de hablar de los usuarios, proporcionando respuestas contextualmente apropiadas. El generador se basa en redes neuronales recurrentes (RNN) y el enfoque de secuencia a secuencia. Es completamente entrenable a partir de datos que incluyen el contexto anterior junto con las respuestas que se generarán. Muestran que el generador de con-

texto genera mejoras significativas sobre la línea base tanto en métricas automáticas como en una prueba de preferencia por parejas.

En [14] se presenta una arquitectura de red neuronal para la generación de respuestas que son a la vez sensibles al contexto y basadas en datos. Como tal, puede ser entrenada de extremo a extremo en cantidades masivas de datos de medios sociales. Comentan que, esta es la primera aplicación de un modelo de red neuronal a la generación de respuesta de dominio abierto, y creen que el presente trabajo sentará las bases para modelos más complejos por venir. Además, introducen una nueva técnica de extracción de múltiples referencias que muestra una promesa para la evaluación automatizada.

Así también en [15] se presenta un enfoque simple del modelado conversacional. El modelo conversa al predecir la oración siguiente dada la oración u oraciones anteriores en una conversación. La fuerza de este modelo es que puede ser entrenado de extremo a extremo y por tanto, requiere mucho menos reglas hechas a mano. Además, este modelo sencillo puede generar conversaciones sencillas dado un gran conjunto de datos de entrenamiento de conversación. Los resultados preliminares sugieren que, a pesar de optimizar la función objetivo errónea, el modelo es capaz de conversar bien. Es capaz de extraer conocimiento de un conjunto de datos de dominio específico y de un conjunto de datos grande, ruidoso y de dominio general de subtítulos de películas. En un conjunto de datos ruidoso de transcripción de películas de dominio abierto, el modelo puede realizar formas sencillas de razonamiento de sentido común.

En [16] proponen un enfoque basado en redes neuronales que modela los procesos de atención e intención. Consiste esencialmente en tres redes neuronales recurrentes (RNN). La red de codificadores es un modelo de nivel de palabra que representa oraciones origen. La red de intenciones es una red recurrente que modela la dinámica del proceso de intención. La red decodificadora es una red recurrente que produce respuestas a la entrada desde el origen. Es un modelo de lenguaje que depende de la intención y tiene un mecanismo de atención para atender a las palabras de origen en particular, al predecir un símbolo en la respuesta. El modelo está entrenado de extremo a extremo sin datos etiquetados. Los experimentos muestran que este modelo genera respuestas naturales a las entradas de los usuarios. En la figura 4 se describe el modelo atención con intención y en la tabla 1 se da un ejemplo del proceso de diálogo.

Tabla 1. Un ejemplo de proceso de diálogo [16].

<i>user</i>	<i>computer is infected</i>
<i>agent</i>	do you want to retrieve the files that was deleted?
<i>user</i>	the ones that the virus deleted, yes.
<i>agent</i>	i can help you resolve the issue with our virus removal and protection service
<i>user</i>	ok
<i>agent</i>	here is a link how to run system restore
<i>user</i>	thank you.
<i>agent</i>	you welcome

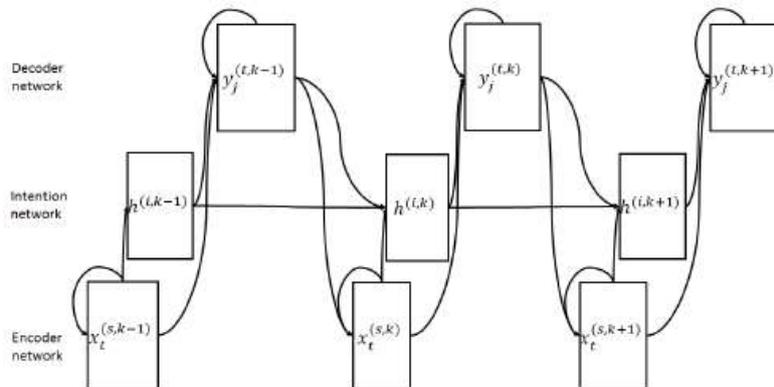


Fig. 4. El modelo de atención con intención (AWI). El modelo se desarrolla en tres turnos. En cada turno tiene RNNs para la red del codificador y la red del decodificador. Cada sesión está representada por un vector de dimensión fija, que es un estado oculto de una red de intención RNN, Fuente Yao [16].

En [17] se presenta un generador de lenguaje estadístico basado en una estructura de memoria a corto plazo (LSTM por sus siglas en inglés) controlada semánticamente. El generador de LSTM puede aprender de datos no alineados optimizando conjuntamente sus componentes de planificación de frases utilizando un criterio de entrenamiento de entropía cruzada simple sin ninguna heurística y se obtiene una variación de calidad de lenguaje simplemente mediante muestreo aleatorio de las salidas de la red. Con menos heurística, una evaluación objetiva en dos diferentes dominios de prueba muestran que el método propuesto mejoró el rendimiento en comparación con los métodos anteriores. Los jueces humanos anotaron el sistema LSTM más alto en informatividad y naturalidad y en general lo prefirieron a los otros sistemas.

Utilizando el mismo enfoque de estructura de memoria (LSTM) [18] presenta un modelo de conversación que incorpora tanto el contexto como el rol de participante en las conversaciones de dos partes. Se exploran diferentes arquitecturas para integrar la información de roles y contextos de los participantes en un modelo de lenguaje de memoria a corto plazo (LSTM). El modelo conversacional puede funcionar como un modelo de lenguaje o un modelo de generación de lenguaje. Los experimentos realizados en el corpus de diálogo de Ubuntu muestran que el modelo puede capturar la interacción de múltiples turnos entre los participantes. El método propuesto supera a un modelo LSTM tradicional medido por la perplejidad del modelo de lenguaje y la clasificación de respuestas. Las respuestas generadas muestran diferencias de características entre los dos roles de los participantes.

En [19] se comenta que los modelos de red neuronal de secuencia a secuencia para la generación de respuestas conversacionales tienden a generar respuestas habituales y seguras (por ejemplo, no sé) independientemente de la entrada. Sugieren que la función objetivo tradicional, es decir, la probabilidad de salida (respuesta) dada entrada (mensaje) no es adecuada para tareas de generación de respuesta. En su lugar, proponen utilizar Máxima Información Mutua (MMI por sus siglas en inglés) como la fun-

ción objetivo en los modelos neuronales. Los resultados experimentales demuestran que los modelos de MMI propuestos producen respuestas más diversas, interesantes y apropiadas, produciendo ganancias sustanciales en las puntuaciones de BLEU4 en dos conjuntos de datos conversacionales y en evaluaciones humanas.

Tabla 2. Análisis comparativo de los diferentes modelos utilizados en la generación automática de diálogos.

Modelo	Autor	Herramientas, técnicas y/o modelos	Dominio
<i>Reglas Probabilísticas</i>	Lison (2012)	Redes Bayesianas	Conjunto de datos de Wizard-of-Oz.
	Ritter (2013)	SMT , GIZA++	Twitter, WMT08, WMT09
<i>Procesos estocásticos</i>	Barlier et al., (2015)	MDP, POMDP y RL	100000,75000 diálogos
	Iijima y Kobayashi (2016)	POMDP, Q-learning	
	Lison (2013)	POMDP, Redes Bayesianas	Wizard-of-Oz y simulación de usuario
	Pietquin (2013)	MDP, IRL	Simulación del usuario
	Misu et al., (2012)	MDP, POMDP, RL	Corpus (Museo de Ciencias de Boston)
	Png (2011)	MDP, POMDP, BRL	Corpus SACTI1
	Mahadik (2017)	POMDP, Redes Bayesianas	
	Young y Thomson (2013)	POMDP y RL	
<i>Redes Neuronales</i>	Dušek y Jurčiček (2016)	LSTM y seq2seq	Crowdsourcing
	Sordoni et al., (2015)	RNN, RLM	Conversaciones de Twitter
	Vinyals (2015)	RNN (LSTM) y seq2seq	OpenSubtitles (dominio abierto). Conjunto de datos de solución de problemas (dominio cerrado)
	Yao et al., (2015)	RNN, seq-seq	Diálogos de servicio de chat
	Wen et al., (2015)	SC-LSTM	Ontologías (SF Restaurant SF Hotel)
	Luan et al., (2016)	LSTM	Corpus de diálogo UBUNTU
	Li et al., (2016)	NN, seq-seq y MMI	OpenSubtitles
	Su et al., (2016)	NN, SL y RL	Conjunto de datos de Wizard-of-Oz (Amazon Mechanical Turk)
	Li et al., (2016)	LSTM y seq2seq, RL	OpenSubtitles
	Shang et al., (2015)	NN, SMT , GIZA++	Microblog similar a Twitter en China
Sutskever (2014)	RNN, LSTM, seq-seq	WMT'14	

En la investigación [20] se describe un enfoque de dos pasos para la gestión del diálogo en sistemas de diálogo hablado orientados a tareas. Se propone un marco de red neuronal unificado para permitir al sistema aprender primero mediante supervisión a partir de un conjunto de datos de diálogo y luego mejorar continuamente su comportamiento a través del aprendizaje por refuerzo, todos utilizando algoritmos basados en gradiente en un solo modelo. Los experimentos demuestran la efectividad del modelo supervisado en la evaluación basada en el corpus, con simulación de usuarios y con sujetos humanos remunerados. El uso del aprendizaje por refuerzo mejora aún más el rendimiento del modelo en ambos entornos interactivos, especialmente bajo condiciones de mayor ruido.

Los recientes modelos neuronales de generación del diálogo [14, 16, 17, 18, 19, 21] ofrecen una gran promesa para generar respuestas para los agentes de conversación, pero tienden a ser miopes, prediciendo enunciados uno a la vez ignorando su influencia en los resultados futuros. Modelar la dirección futura de un diálogo es crucial para generar diálogos coherentes e interesantes, una necesidad que llevó a los modelos tradicionales de procesamiento de lenguaje natural (PNL por sus siglas en inglés) de diálogo a basarse en el aprendizaje por refuerzo. Como en el trabajo [22] que muestra cómo integrar estos objetivos, aplicando el aprendizaje por refuerzo profundo para modelar la recompensa futura en el diálogo chat-bot. El modelo simula diálogos entre dos agentes virtuales, utilizando métodos de políticas de gradiente para recompensar las secuencias [23] que muestran tres propiedades conversacionales útiles: informatividad, coherencia y facilidad de respuesta.

En la tabla 2 se hace un análisis de los modelos, autores, herramientas, técnicas y el dominio de datos empleados en la generación automática de diálogos de los artículos analizados en este trabajo.

4. Conclusión

En este trabajo se presentó una revisión de diferentes modelos para la generación de diálogos desde los modelos Bayesianos probabilísticos, los estocásticos hasta los modelos basados en redes neuronales, como los modelos que utilizan LSTM secuencia-secuencia (SEQ2SEQ) que son un tipo de modelo de generación neuronal que maximiza la probabilidad de generar una respuesta dado el turno anterior del diálogo y actualmente los modelos que buscan integrar SEQ2SEQ y los paradigmas de aprendizaje por refuerzo, aprovechando las ventajas de ambos. Esto da pauta para futuras investigaciones en donde se puedan generar respuestas cada vez más diversas e interactivas que fomenten una conversación más sostenida.

Agradecimientos. Queremos agradecer a Vicerrectoría de Investigación y Estudios de Posgrado de la Benemérita Universidad Autónoma de Puebla por apoyar este trabajo a través del proyecto Extracción de frases clave y sus relaciones en artículos científicos.

Referencias

1. Konstantopoulos, S.: An Embodied Dialogue System with Personality and Emotions. In: Proceedings of the 2010 Workshop on Companionable Dialogue Systems, ACL, 31–36 (2010)
2. Lison, P.: Probabilistic Dialogue Models with Prior Domain Knowledge. Association for Computational Linguistics. In: Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL), Seoul, South Korea, 179–188 (2012)
3. Litman, D., Forbes-Riley, K.: Evaluating a Spoken Dialogue System that Detects and Adapts to User Affective states. In: Proceedings of the SIGDIAL 2014 Conference, Philadelphia, U.S.A., ACL, 181–185 (2014)
4. Barlier, M., Perolat, J., Laroche, R., Pietquin, O.: Human-Machine Dialogue as a Stochastic Game. In: Proceedings of the SIGDIAL 2015 Conference, Prague, Czech Republic, 2–11 (2015)
5. Iijima, S., Kobayashi, I.: A POMDP-based Multimodal Interaction System Using a Humanoid Robot. In: 30th Pacific Asia Conference on Language, Information and Computation (PACLIC 30) Seoul, Republic of Korea, 519–525 (2016)
6. Misu, T., Georgila, K., Leuski, A., Traum, D.: Reinforcement Learning of Question-Answering Dialogue Policies for Virtual Museum Guides. In: Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL), Seoul, South Korea, 84–93 (2012)
7. Pietquin, O.: Inverse Reinforcement Learning for Interactive Systems. ACM, 978-1-4503-2019-1/13/08 MLIS'13, Beijing, China (2013)
8. Lison, P.: Model-based Bayesian Reinforcement Learning for Dialogue Management. arXiv:1304.1819v1 [cs.AI] (2013)
9. Png, S., Pineau, J.: Bayesian Reinforcement Learning for POMDP-Based Dialogue Systems. In: Conference Paper in Acoustics, Speech, and Signal Processing, 1988, ICASSP-88., 1988 International Conference on 2011 (2011). DOI: 10.1109/ICASSP.2011.5946754
10. Mahadik, S., Dwivedi, P., King, M., Zhu B., Tang S.: Dialogue Manager for Spoken Dialogue System: Review. In: International Conference on Emanations in Modern Technology and Engineering (ICEMTE-2017), 27–29 (2017)
11. Young, S., Gašić, M., Thomson, B., Williams, J. D.: POMDP-Based Statistical Spoken Dialog Systems: A Review. In: Proceedings of the IEEE, 101(5) (2013)
12. Dušek, O., Jurčiček, F.: A Context-aware Natural Language Generator for Dialogue Systems. In: Proceedings of the SIGDIAL 2016 Conference, Los Angeles, USA, 185–190 (2016)
13. Dušek, O., Jurčiček, F. 2016b.: A context-aware natural language generation dataset for dialogue systems. In Workshop on Collecting and Generating Resources for Chatbots and Conversational Agents-Development and Evaluation. 6–9 (2016)
14. Sordoni, A., Galley, M., Auli, M., Brockett, C., Ji, Y., Mitchell, M., Dolan, B.: A Neural Network Approach to Context-Sensitive Generation of Conversational Responses. In: Proceedings of NAACL-HLT (2015)
15. Vinyals, O., Le Q.: A Neural Conversational Model. In: Proceedings of the 31st International Conference on Machine Learning, Lille, France, JMLR: W&CP, 37 (2015)
16. Yao, K., Zweig, G.: Attention with Intention for a Neural Network Conversation Model. In: Presented at NIPS Workshop on Machine Learning for Spoken Language Understanding and Interaction 2015, arXiv:1510.08565v3 [cs.NE] (2015)
17. Wen, T., Gašić, M., Mrkšić, N., Su, P., Vandyke D., Young, S.: Semantically Conditioned LSTM-based Natural Language Generation for Spoken Dialogue Systems. In: Proceedings

- of the 2015 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Lisbon, Portugal, 1711–1721 (2015)
18. Luan, Yi., Ji, Y., Ostendorf, M.: LSTM based Conversation Models. arXiv:1603.09457v1 [cs.CL] (2016)
 19. Li, J., Galley, M., Brockett, C., Gao, J., Dolan, B.: A Diversity-Promoting Objective Function for Neural Conversation Models. In: Proceedings of NAACL-HLT 2016, Association for Computational Linguistics, San Diego, California, 110–119 (2016)
 20. Su, P., Gašić, M., Mrkšić, N., Rojas-Barahona, L., Ultes, S., Vandyke, D., Young S.: Continuously Learning Neural Dialogue Management. arXiv:1606.02689v1 [cs.CL] (2016)
 21. Shang, L., Lu, Z., Li, H.: Neural Responding Machine for Short-Text Conversation. arXiv.1503.02364v2[cs.CL] (2015)
 22. Yao, K., Zweig, G.: Attention with Intention for a Neural Network Conversation Model. In: Presented at NIPS Workshop on Machine Learning for Spoken Language Understanding and Interaction 2015. arXiv:1510.08565v3 [cs.NE] (2015)
 23. Li, J., Monroe, W., Ritter, A., Galley, M., Gao, J., Jurafsky, D.: Deep Reinforcement Learning for Dialogue Generation. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, Texas, 1192–1202 (2016)
 24. Sutskever, I., Vinyals, O., Le, Q.: Sequence to Sequence Learning with Neural Networks.arXiv:1409.3215v3 [cs.CL] (2014)