

# Jornadas de Automática

## Sistema de aprendizaje cooperativo para enjambre de robots sociales

Sofía Álvarez-Arias<sup>1,\*</sup>, Marcos Maroto-Gómez<sup>1</sup>, Sara Carrasco-Martínez<sup>1</sup>, María Malfaz<sup>1</sup>, Miguel Ángel Salichs<sup>1</sup>

*Departamento de Ingeniería de Sistemas y Automática, Universidad Carlos III de Madrid. Avenida de la Universidad, 30, 28911 Leganés, España.*

**To cite this article:** Álvarez-Arias, S., Maroto-Gómez, M., Carrasco-Martínez, S. Malfaz, M., Salichs, M.A. 2024. Social Robot Network for Collaborative Learning. Jornadas de Automática, 45. <https://doi.org/10.17979/ja-cea.2024.45.10865>

### Resumen

Cada vez son más los robots sociales dedicados a asistencia diaria en casas particulares y residencias, por lo que una adaptación al usuario y centralizar la información recogida por todos ellos resulta fundamental. Este artículo presenta un sistema de aprendizaje cooperativo para el robot social Mini. Este robot pretende asistir a usuarios de edad avanzada facilitando tareas mediante una interacción humano-robot adaptada a sus necesidades. El sistema propuesto permite integrar distintos métodos de aprendizaje por refuerzo de forma simultánea, permitiendo a los robots mejorar su comportamiento y aprender en entornos dinámicos de forma autónoma. Este sistema pretende superar limitaciones importantes del aprendizaje en robótica social como los amplios tiempos de entrenamientos necesarios para la recolección de datos y la gestión de grandes volúmenes de datos. La red permite compartir y centralizar la información de las interacciones de cada robot, agilizando el proceso de aprendizaje al disponer de un mayor volumen de datos en menos tiempo. El servidor almacena y procesa la información, mejorando la eficiencia y reduciendo la carga computacional en los robots. Este sistema permite a los robots adaptarse mejor a diferentes usuarios y entornos, mejorando sus respuestas y la personalización en la interacción humano-robot.

*Palabras clave:* Sistemas autónomos, Interacción multi-modal, Aprendizaje por Refuerzo, Coaprendizaje, Autoaprendizaje.

### Cooperative Learning System for Social Robots Swarm

#### Abstract

More and more social robots are dedicated to daily assistance in private homes and residences, so adapting to the user and centralising the information they collect is essential. This article presents a cooperative learning system for the Mini social robot. This robot aims to assist elderly users by facilitating tasks through human-robot interaction adapted to their needs. The proposed system simultaneously integrates different reinforcement learning methods, enabling robots to improve their behaviour and learn in dynamic environments autonomously. This system aims to overcome important learning limitations in social robotics, such as the extensive training times required for data collection and the management of large volumes of data. The network allows the sharing and centralization of information about each robot's interactions, speeding up the learning process by making more data available in less time. The server stores and processes the information, improving efficiency and reducing the computational load on the robots. This system allows robots to better adapt to different users and environments, improving their responses and personalization in human-robot interaction.

*Keywords:* Autonomous robotic systems, Multi-modal interaction, Reinforcement Learning, Co-Learning and self-Learning.

\* Autor para correspondencia: [sofalvar@pa.uc3m.es](mailto:sofalvar@pa.uc3m.es)

Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)

*Correos electrónicos:* [sofalvar@pa.uc3m.es](mailto:sofalvar@pa.uc3m.es) (Sofía Álvarez-Arias), [marmarot@ing.uc3m.es](mailto:marmarot@ing.uc3m.es) (Marcos Maroto-Gómez), [sacarras@ing.uc3m.es](mailto:sacarras@ing.uc3m.es) (Sara Carrasco-Martínez), [mmalfaz@ing.uc3m.es](mailto:mmalfaz@ing.uc3m.es) (María Malfaz), [salichs@ing.uc3m.es](mailto:salichs@ing.uc3m.es) (Miguel Ángel Salichs)

## 1. Introducción

El aprendizaje automático en máquinas permite a sistemas artificiales aprender a tomar decisiones más adecuadas en entornos dinámicos (Soori et al., 2023). A grandes rasgos el aprendizaje autónomo puede ser supervisado o no supervisado. El aprendizaje supervisado utiliza grandes volúmenes de datos etiquetados para clasificar datos no etiquetados mediante semejanza. De forma opuesta, el aprendizaje no supervisado trata de buscar patrones en conjuntos de datos que se encuentran sin etiquetar. Ambos sistemas han sido ampliamente utilizados para mejorar las capacidades de los robots, permitiéndoles adaptarse a las situaciones cambiantes del entorno y mejorar la experiencia de usuario personalizando sus acciones. Sin embargo, los algoritmos de aprendizaje automático presentan importantes limitaciones como la necesidad de utilizar grandes volúmenes de datos para realizar predicciones precisas en aprendizaje supervisado y de largos tiempos de entrenamiento en aprendizaje no supervisado (Smith and Lamprecht, 2024).

Hoy día uno de los algoritmos de aprendizaje más utilizados en robótica social e interacción humano-robot es el aprendizaje por refuerzo (Akalin and Loutfi, 2021). Esta metodología permite a un agente aprender a tomar decisiones mediante prueba y error en entornos dinámicos, maximizando una función de refuerzo que indica cómo de positivo o negativo resulta ejecutar cada acción en cada situación. El resultado del aprendizaje es una política de comportamiento que indica qué acción debe ejecutarse en cada situación (estado) que el robot experimenta Sutton and Barto (2018). Sin embargo, los algoritmos de aprendizaje por refuerzo utilizados en aplicaciones reales de interacción humano-robot presentan grandes limitaciones derivadas de la larga duración que puede tener cada acción, el gran número de acciones que se pueden considerar y la gran cantidad de situaciones que el robot puede experimentar. Estas limitaciones han sido parcialmente solucionadas por métodos de *Deep Reinforcement Learning* (Vouloutsi et al., 2023), donde redes neuronales sustituyen a las bases de datos, permitiendo generalizar de forma más eficaz el aprendizaje.

Este trabajo presenta un sistema de aprendizaje cooperativo para robots sociales basado en ROS (Quigley et al., 2009) dedicados a tareas asistenciales e interacción humano-robot. Este trabajo se desarrolla en el grupo de robótica social de la Universidad Carlos III de Madrid que cuenta con un total de 19 robots sociales Mini (Salichs et al., 2020). Mini es un robot de escritorio dedicado a estimulación cognitiva de mayores y entretenimiento. Estos robots se encuentran situados en diferentes localizaciones y son utilizados por distintos usuarios, por lo que se pretende centralizar la información obtenida por cada uno de ellos individualmente en un servidor que permita agrupar toda la información obtenida, almacenarla de forma eficiente teniendo un registro común de lo que hacen los robots y agilizar el aprendizaje conjunto en la medida de lo posible. De esta forma, el sistema propuesto pretende dar solución a algunas problemáticas de los algoritmos de aprendizaje por refuerzo. Cada robot o nodo entrenará sus propios algoritmos para adaptarse a cada uno de los usuarios que lo utilizan. Los robots se comunican con el servidor bidireccionalmente para entrenar los mismos algoritmos en el servidor (y poder generalizar para nuevos usuarios) y transferir esta información de nuevo a cada robot.

La principal novedad del sistema con respecto a otras aproximaciones disponibles en la literatura (Liu et al., 2020; Li et al., 2020; Gamboa-Montero et al., 2023) es la posibilidad de ejecutar varios algoritmos de aprendizaje de forma simultánea y la distribución del conocimiento a los nodos (robots). Además, ninguno de los sistemas desarrollados hasta la fecha en aprendizaje cooperativo se ha desarrollado para enjambres de robots sociales dedicados a interacción humano-robot. Los resultados del artículo muestran tres casos de uso de como el sistema permite agilizar y centralizar diferentes estrategias de aprendizaje y monitorización: el aprendizaje de las preferencias de cada usuario hacia las actividades del robot, el aprendizaje de la política de comportamiento general del robot para obtener un sistema autónomo, el almacenamiento de los registros de funcionamiento de cada robot.

La sección 2 ofrece una descripción del robot social utilizado para el enjambre y una descripción del aprendizaje del mismo. La sección 3 detalla el método de comunicación utilizado para el aprendizaje colaborativo. La sección 4 da una visión práctica de la colaboración entre robots. Por último, la sección 5 recoge las diferentes conclusiones de este trabajo, así como una señalización a los trabajos futuros que esta línea de investigación puede aportar.

## 2. El robot social Mini

Mini (Salichs et al., 2020) es un robot social de escritorio diseñado para asistir a personas mayores con bajo grado de deterioro cognitivo en situación de dependencia y a sus cuidadores realizando actividades de estimulación física, cognitiva y afectiva. Mini es capaz de percibir el entorno utilizando distintos sensores de tacto, presencia, audio y vídeo y actuar en consecuencia utilizando cinco motores situados en la cintura, brazos, cuello y cabeza. Además es capaz de mostrar expresividad con indicadores luminosos en las mejillas, corazón y boca y generar sonidos verbales y no verbales utilizando un altavoz. La interacción con el usuario se complementa con una tableta que sirve para mostrar información y obtener sus respuestas a las distintas actividades que presenta.

Mini tiene un comportamiento bioinspirado semejante al comportamiento humano con el fin de mostrarse natural y conseguir nuestra aceptación. Mini es capaz de adaptarse a las distintas situaciones que experimenta y personalizar la interacción con el usuario utilizando métodos de aprendizaje por refuerzo profundo (Sutton and Barto, 2018). Con esta metodología, el robot repite aquellas acciones que generan un refuerzo positivo y que favorecen una buena interacción con el usuario y evitar aquellas que proporcionan una mala experiencia. De esta forma, Mini es capaz de proporcionar una mejor experiencia de usuario y funcionar autónomamente reduciendo la intervención de personas externas.

Mini cuenta con un sistema de aprendizaje de preferencias que se aplica a las actividades de entretenimiento para conocer qué actividades son las favoritas del usuario y proponérselas de forma autónoma más frecuentemente. Este sistema utiliza métodos de aprendizaje por refuerzo clásicos basados en *Temporal Difference (TD)* que combinan la información explícita proporcionada por el usuario sobre cuánto le gusta cada actividad y métricas que infieren implícitamente si al usuario le ha gustado o no.

Por último, Mini tiene un sistema de monitorización y almacenamiento de eventos que permite a sus desarrolladores conocer qué actividades realiza el robot y las respuestas que ofrece el usuario durante la interacción. Así, es posible conocer si el usuario está haciendo un uso frecuente y adecuado del robot.

### 2.1. Aprendizaje de comportamiento

El sistema de aprendizaje de comportamiento tiene como objetivo permitir a Mini adaptarse al usuario de forma autónoma. Este sistema está basado en métodos de *Deep Reinforcement Learning (DRL)* (François-Lavet et al., 2018), en concreto, *Rainbow* (Hessel et al., 2018). *Rainbow* combina aprendizaje por refuerzo clásico (Sutton and Barto, 2018), que permite evaluar la bondad de la acción presentada al usuario en función de la recompensa obtenida, y el aprendizaje profundo (Goodfellow et al., 2016), que se caracteriza por utilizar redes neuronales para almacenar la información de las experiencias capacitando al robot para identificar patrones entre los datos recopilados y así generalizar comportamientos en experiencias similares. La combinación de ambas técnicas da lugar a un método de aprendizaje en el que se deben cumplir dos objetivos clave: minimizar la función de pérdidas representada en la ecuación 1 y maximizar la función de refuerzo acumulado representada en la ecuación 2.

$$\mathcal{L}(\theta) = E_{s,s',a,r}[(Q(s, a, \theta_i) - Y_i)^2] \quad (1)$$

$$R_t = \sum_{\tau=t}^{\infty} \gamma^{\tau-t} r_{\tau} \quad (2)$$

En la ecuación 1,  $\mathcal{L}$  es la pérdida de la acción realizada en el tiempo  $t$ . Esta pérdida depende de  $\theta_i$ , que representa las relaciones entre las neuronas de la red.  $E$  representa el valor esperado sobre los parámetros  $s, s', a$  y  $r$ , donde  $s$  es el estado del robot en el instante de tiempo  $t$ ,  $s'$  es el estado del robot tras realizar la acción  $a$  y  $r$  es el refuerzo obtenido tras realizar la acción  $a$ .  $Q$  es el valor aprendido para cada estado-acción y representa cómo de bueno es ejecutar la acción  $a$  en el estado  $s$ .  $Y_i$  representa el valor objetivo que se debe actualizar. En la ecuación 2,  $\gamma \in (0, 1)$  es la importancia de las experiencias futuras sobre las pasadas. Valores próximos a 1 consideran mejor las experiencias pasadas y próximos a 0 las nuevas. Finalmente,  $r_{\tau}$  es el valor de refuerzo numérico obtenido tras realizar una acción.

Mini aplica este aprendizaje para mejorar su comportamiento y ser un robot autónomo. El comportamiento se refiere a las secuencias de acciones que realiza para producir una respuesta adecuada a la situación que está experimentando sin alterar sus propios procesos. Para aprender qué acción es la más adecuada, es necesario conocer el estado  $s$  en el que se encuentra el robot. El estado del robot es una combinación de procesos internos bioinspirados, los eventos sociales que debe cumplir y la percepción que realiza del entorno. En total, lo componen 18 variables (continuas y discretas) como por ejemplo el nivel de energía o entretenimiento y si el usuario está presente o no. Mini es capaz de realizar un total de 14 acciones para satisfacer sus necesidades internas y también las del usuario. Por lo tanto, el objetivo de este sistema de aprendizaje es ligar cada estado o situación que vive el robot con la acción que maximiza su propio bienestar y el del usuario.

### 2.2. Aprendizaje de preferencias

El aprendizaje de preferencias consiste en la predicción (Maroto-Gómez et al., 2023) y el ajuste (Maroto-Gómez et al., 2024) de las preferencias del usuario hacia las actividades de entretenimiento del robot para proponerle actividades aquellas del interés del usuario más frecuentemente.

La predicción de preferencias necesita de una base de datos en la que conste una clasificación del usuario y las preferencias de cada uno de los usuarios. Mini dispone de un *dataset* en el que, para cada uno de los usuarios, se relacionan sus características (entrada) y sus preferencias hacia las actividades del robot (salida). Las características de los usuarios permite realizar relaciones entre sus características y preferencias para luego predecir las preferencias de los nuevos usuarios de los cuales se conocen sus características. De esta forma, el robot propone aquellas actividades que figuran en sus preferencias más frecuentemente. Este *dataset* se ha elaborado mediante un cuestionario que completaron aproximadamente 500 personas. La predicción genera un valor entre 0 y 5 de cuánto le gusta cada actividad al usuario. Cuanto más alto es el valor, más le gusta la actividad.

La predicción de preferencias puede no ajustarse a las preferencias reales del usuario. Por ello, tras la predicción, el robot permite al usuario realizar las distintas actividades para obtener *feedback* sobre si le gusta o no. Mini aprende a repetir aquellas acciones que obtienen un refuerzo positivo y evita aquellas que disgustan al usuario. Para determinar si la actividad propuesta gusta o no al usuario el algoritmo de aprendizaje actualiza el valor del refuerzo de cada una de las actividades a medida que se proponen siguiendo la ecuación 3. Esta ecuación se basa en el algoritmo TD.

$$V(a)' = V(a) + \alpha * r \quad (3)$$

$V(a)'$  es el valor de la acción entre 0 y 5 (preferencias) tras la experiencia con el nuevo usuario.  $V(a)$  es el valor de la actividad antes de realizarla. Al valor previo se le suma el refuerzo  $r$ , que toma valores entre 0 y 5 en función de cuanto le gusta la actividad propuesta al usuario, siendo 0 nada y 5 mucho. Al peso se le aplica un parámetro de ajuste  $\alpha$  que regula la influencia del refuerzo en el nuevo valor.

## 3. Metodología

El sistema propuesto tiene como objetivo conectar a un grupo dinámico de robots Mini con un servidor. La figura 1 muestra la comunicación en la red de aprendizaje colaborativo. La red comienza en los nodos. Cada nodo es un robot que tiene sus propios usuarios. El robot reconoce a los usuarios utilizando una cámara y tiene almacenados sus datos en local. Cada robot genera experiencias con cada uno de los usuarios. Durante estas experiencias o interacciones Mini aplica los dos tipos de aprendizajes de los que dispone para mejorar su adaptación y comportamiento. Además, cada robot guarda información sobre las actividades que el usuario realiza generando un sistema de eventos que tienen relevancia para los desarrolladores. Cada uno de los robots guarda sus propios datos en local y a la vez los comparte con el servidor para centralizar todo el conocimiento. Después, el servidor es capaz de devolver información más completa y actualizada de la agrupación de datos a los nodos para que estos mejoren su rendimiento

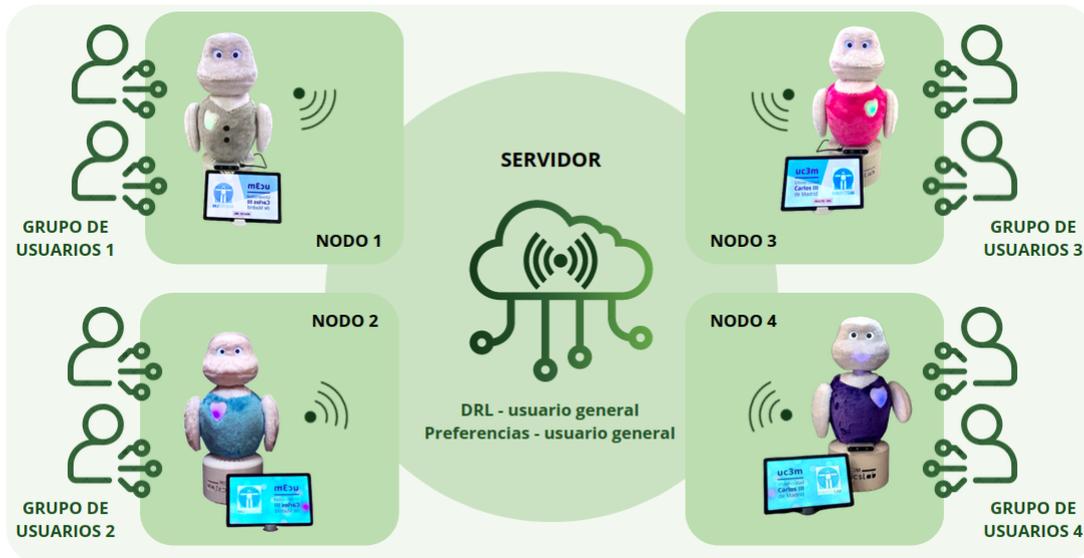


Figura 1: Representación de la red de robots Mini para el aprendizaje colaborativo a través de un servidor.

y conocimiento. Para compartir la información se ha desarrollado en ROS una arquitectura software que utiliza comunicación por *sockets*, comunicando cada uno de los robots con el servidor compartiendo datos de forma bidireccional. Esta información contiene datos relevantes de cada una de las interacciones y se almacena en el servidor hasta que alguno de los robots la necesita para sus procesos de aprendizaje. Los datos contienen el identificador del robot que envía los datos o los recibe, el usuario objetivo (si es un usuario genérico, el identificador será *default*), el algoritmo que genera los datos y los datos necesarios para construir una experiencia. Este sistema permite que cada robot aprenda más rápido al disponer de un mayor volumen de información en menos tiempo.

El sistema de *sockets* utiliza la propia librería nativa de *Python* con el mismo nombre para construir canales de comunicación entre dispositivos no relacionados, permitiendo el paso de información a través de *topics* de ROS. Los *topics* son canales de información que permiten a un nodo suscribirse y publicar en otros nodos. Suscribirse a un nodo permite recibir información concreta de él, mientras que publicar envía información deseada al nodo. El uso de *sockets* evita que la configuración nativa de ROS comparta toda la información de los nodos conectados entre sí, pudiendo dar problemas si hay más de dos nodos comunicándose. Compartir toda la información y no solo la deseada aumenta los tiempos de espera y el costo computacional.

La información obtenida de los robots se divide en aquella obtenida del aprendizaje de preferencias, la obtenida del DRL y el sistema de generación de eventos. Todo ello se almacena y transmite como cadenas de caracteres. La información generada por el DRL almacena datos sobre la fecha en la que ha tenido lugar la interacción, el estado del robot, la actividad realizada y el robot que transmite la información. En el caso del aprendizaje de preferencias, se recoge también la fecha de ejecución, la actividad propuesta y el robot en el que tiene lugar la interacción además del peso de la actividad propuesta.

El servidor sirve de nodo central de la cadena de comunicación en la red. Este nodo hace la función de almacén de datos y generación de conocimiento agrupado y general, per-

mitiendo que cualquier entrenamiento en cualquiera de los robots de la red aprenda de la información recogida en toda la red. Esto hace que el servidor sea una fuente de información bidireccional en la que se recoge toda la información sobre las experiencias y donde se desarrolla el grueso del aprendizaje para luego solo transmitir los resultados de esos entrenamientos. Mini tiene ciertos comportamientos que no necesitan partir de la inexperiencia y que se pueden preentrenar en el servidor aligerando el consumo de recursos en los robots. El uso del servidor tiene una ventaja adicional, ejecutar los algoritmos en los propios robots resta mucha potencia a los mismos por el costo computacional. Sin embargo, el servidor puede hacer lo mismo con la información focalizada sólo en el aprendizaje sin el resto de procesos del robot, agilizando el proceso y reduciendo ese coste computacional.

El sistema de aprendizaje permite que los robots de la red trabajen en entornos dinámicos muy diferentes de manera simultánea, mejorando no solo las respuestas al usuario sino también su adaptación al entorno donde tiene lugar la interacción. Compartir todos los datos permite a todos los robots que forman la red adaptarse a varios tipos de usuarios y dar respuestas apropiadas al entorno del que percibe información. Además, su diseño es escalable y permite incluir fácilmente nuevos robots y estrategias de aprendizaje.

#### 4. Casos de uso

Los siguientes casos de uso muestran el funcionamiento del sistema para cada uno de los dos métodos de aprendizaje que se consideran en la red de aprendizaje cooperativo propuesta y como se almacena la información generada.

##### 4.1. Aprendizaje de comportamiento

El estado del robot cuenta con 4 variables que simulan procesos biológicos (sueño, estrés, energía e interacción física), 4 procesos cognitivos (entrenamiento social, autoentrenamiento, interacción cognitiva y vínculo con el usuario), 4 emociones (alegría, tristeza, enfado y calma), 4 estímulos del entorno (presencia, emoción, satisfacción y acción del usuario) y

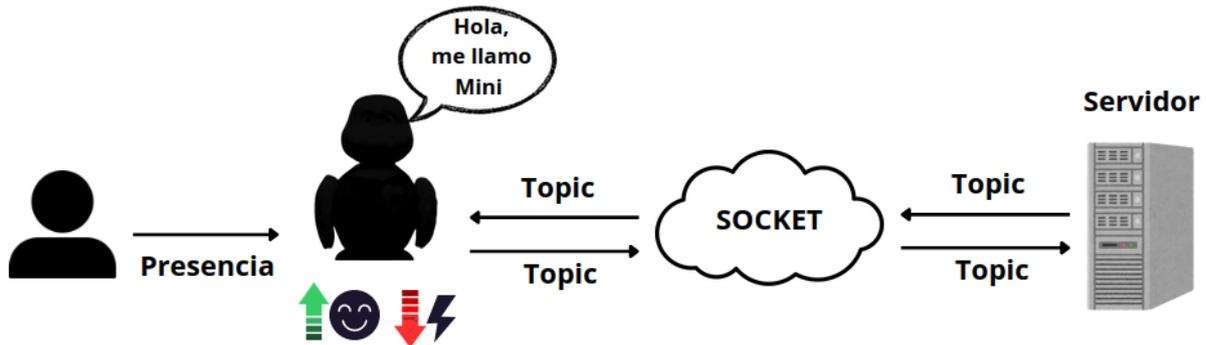


Figura 2: Representación de la comunicación entre humano, robot y servidor para el aprendizaje de comportamiento.

2 variables sobre eventos sociales que el usuario debe realizar. Cada una de estas variables toma distintos valores que generan el estado del robot. En base a él, debe elegir la acción que mejore los procesos internos que se encuentren en peor condición y ayudar en los eventos sociales del usuario teniendo en cuenta los estímulos que percibe como disponible.

La figura 2 muestra una interacción habitual con el robot donde interviene el aprendizaje de comportamiento. El robot detecta la presencia del usuario, lo que hace que las acciones que necesiten este estímulo tomen relevancia sobre las que no lo necesitan. Mini decide ejecutar la acción "hablar" para restaurar sus niveles de alegría, a su vez la acción consumirá energía. Cada vez que un robot toma una decisión y ejecuta una acción, obtiene una nueva experiencia que genera un valor de refuerzo y cambia el estado del robot a uno nuevo. Repitiendo este proceso muchas veces, el robot es capaz de aprender cómo regular sus procesos mientras mejora su comportamiento y la experiencia de usuario. Sin embargo, este proceso es muy lento y requiere de largos procesos de aprendizaje. Por ello, proponemos que cada experiencia obtenida por un robot sea enviada al servidor para que la agrupe con las experiencias de otros robots y así agilizar el aprendizaje. Después, el servidor devuelve a todos los robots activos el conocimiento generado agilizando el proceso de aprendizaje. De esta forma es posible no solo aprender más rápido si no combinar la información de varios usuarios, mejorando el comportamiento de los robots ya que no se sobreajustan a un usuario en concreto. La metodología propuesta permite que, a su vez, cada robot personalice la interacción ya que el algoritmo no solo se entrena en el servidor, si no también en los propios robots. De esta forma es posible combinar el aprendizaje generalizado de todos los usuarios con el específico de uno de ellos.

#### 4.2. Aprendizaje de preferencias

El robot Mini dispone de alrededor de 20 actividades de entretenimiento entre juegos, reproducción de contenido multimedia como vídeos o música e información del tiempo y noticias. Debido al alto número de actividades que no todos los usuarios conocen y a que algunas personas no tienen la proactividad necesaria para comenzar una interacción con el robot, se ha diseñado un sistema de predicción y ajuste de preferencias que permite al robot fomentar las actividades que más gustan al usuario a la vez que se exploran aquellas que no son las preferidas y se han ejecutado pocas veces.

Cada actividad del robot tiene un valor asociado entre 0

y 5 puntos que indica cuánto le gusta la actividad al usuario que está interactuando con el robot. Si el robot no reconoce al usuario, se utiliza un registro general para estos casos que engloba a todos los usuarios no reconocidos. Mini tiene la función de proponer actividades al usuario de forma proactiva. Para ello, genera una lista de probabilidades en base a los valores de preferencias de cada actividad. Después, escoge una actividad en base a las probabilidades y se la presenta al usuario. En base a cómo se realiza esta actividad y una pregunta sobre cuánto le ha gustado tras finalizar la misma, el robot obtiene una experiencia que le permite actualizar el valor de dicha actividad. Si le gustó al usuario, el valor aumentará y si no le gustó disminuirá. La figura 3 muestra cómo se obtiene esta información en una interacción cualquiera con Mini. El robot le propone la actividad chistes de robots.<sup>al</sup> usuario, este lleva a cabo la actividad y al final de la misma el Mini le pregunta al usuario por su nivel de satisfacción mostrándole las 5 opciones por pantalla. Cada vez que el robot obtiene una experiencia nueva, envía la información de la actividad realizada y el valor de refuerzo obtenido al servidor. En el servidor, se agrupan las experiencias recogidas por todos los robots y todos sus usuarios, entrenando el algoritmo para obtener un conocimiento generalizado de qué actividades del robot prefieren los usuarios. Al centralizar toda esta información, es posible generar un conocimiento más preciso y amplio sobre las preferencias de los usuarios. Por último, el servidor envía la información aprendida de nuevo a los nodos (robots) para que se adapten de forma más precisa a nuevos usuarios. Este proceso iterativo permite ajustar dinámicamente las preferencias del usuario hacia las actividades del robot y mejorar la experiencia de usuario.

#### 4.3. Sistema de almacenamiento de eventos

El sistema de almacenado de eventos tiene como objetivo almacenar detalles sobre la interacción de cada robot con cada usuario para que los desarrolladores y personal técnico puedan realizar tareas de postprocesado que les permitan concluir la bondad del proceso. A nivel práctico, cada robot envía una secuencia de datos cada vez que finaliza una actividad donde se recogen los datos de la actividad y las métricas de interacción del usuario. Entre las métricas que se almacenan, se encuentran la fecha y hora de la actividad, tiempo de ejecución, número de interacciones totales, modo de interacción (si usó la *tablet* o por voz), resultado de la actividad, y si es un juego con preguntas, número de aciertos y fallos.

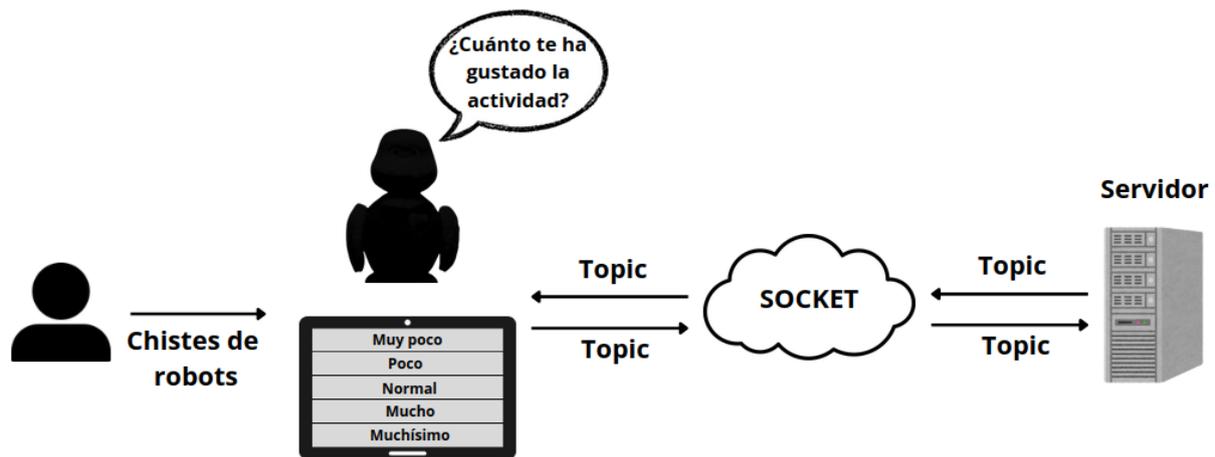


Figura 3: Representación de la comunicación entre humano, robot y servidor para el aprendizaje de preferencias.

El servidor recibe esta información y la guarda en un registro de datos junto con los datos del robot que proporcionó la información y el usuario que realizó la actividad. Este método es el único con un mecanismo de comunicación unidireccional desde los nodos o robots hacia el servidor, aunque se pretende en versiones futuras que el servidor sea capaz de encontrar patrones en las métricas de interacción para poder superar limitaciones de interacción de los usuarios como problemas de visión priorizando una comunicación por voz o problemas de audición fomentando el uso de la tableta.

## 5. Conclusiones

Este trabajo presenta un sistema de aprendizaje cooperativo para robots sociales usando un servidor como nodo central del sistema. La metodología propuesta trata de mejorar y agilizar el aprendizaje de robots sociales para obtener comportamientos autónomos personalizados adaptados al usuario. La propuesta muestra el diseño e implementación del sistema, así como tres casos de uso con distintos métodos de aprendizaje por refuerzo y monitorización de eventos. Como línea de actuación para futuras mejoras, es necesario la realización de pruebas con varios robots para comprobar la eficiencia del sistema y cuantificar los beneficios a nivel de aprendizaje y monitorización de errores y eventos.

## Agradecimientos

Estos resultados han sido financiados por los proyectos PID2021-123941OA-I00, financiado por MCIN/AEI/10.13039/501100011033 y por ERDF A way of making Europe; TED2021-132079B-I00 financiado por MCIN/AEI/10.13039/501100011033 y por la Unión Europea NextGenerationEU/PRTR; Mejora del nivel de madurez tecnológica del robot Mini (MeNiR) financiado por MCIN/AEI/10.13039/501100011033. 13039/501100011033 y por la Unión Europea NextGenerationEU/PRTR; Robot social portable con alto grado de vinculación (PoSoRo) PID2022-140345OB-I00 financiado por MCIN/AEI/10.13039/501100011033 y ERDF A way of making Europe.

## Referencias

- Akalin, N., Loutfi, A., 2021. Reinforcement learning approaches in social robotics. *Sensors* 21 (4), 1292.
- François-Lavet, V., Henderson, P., Islam, R., Bellemare, M. G., Pineau, J., et al., 2018. An introduction to deep reinforcement learning. *Foundations and Trends® in Machine Learning* 11 (3-4), 219–354.
- Gamboa-Montero, J. J., Alonso-Martín, F., Marques-Villarroya, S., Sequeira, J., Salichs, M. A., 2023. Asynchronous federated learning system for human–robot touch interaction. *Expert Systems with Applications* 211, 118510.
- Goodfellow, I., Bengio, Y., Courville, A., 2016. *Deep learning*. MIT press.
- Hessel, M., Modayil, J., Van Hasselt, H., Schaul, T., Ostrovski, G., Dabney, W., Horgan, D., Piot, B., Azar, M., Silver, D., 2018. Rainbow: Combining improvements in deep reinforcement learning. *Proceedings of the AAAI conference on artificial intelligence* 32 (1), 14.
- Li, L., Fan, Y., Tse, M., Lin, K.-Y., 2020. A review of applications in federated learning. *Computers & Industrial Engineering* 149, 106854.
- Liu, Y., Li, Z., Liu, H., Kan, Z., 2020. Skill transfer learning for autonomous robots and human–robot cooperation: A survey. *Robotics and Autonomous Systems* 128, 103515.
- Maroto-Gómez, M., Castro-González, Á., Castillo, J. C., Malfaz, M., Salichs, M. Á., 2023. An adaptive decision-making system supported on user preference predictions for human–robot interactive communication. *User Modeling and User-Adapted Interaction* 33 (2), 359–403.
- Maroto-Gómez, M., Malfaz, M., Castillo, J. C., Castro-González, Á., Salichs, M. Á., 2024. Personalizing activity selection in assistive social robots from explicit and implicit user feedback. *International Journal of Social Robotics*, 1–19.
- Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., Wheeler, R., Ng, A. Y., et al., 2009. Ros: an open-source robot operating system. *ICRA workshop on open source software* 3 (2), 5.
- Salichs, M. A., Castro-González, Á., Salichs, E., Fernández-Rodicio, E., Maroto-Gómez, M., Gamboa-Montero, J. J., Marques-Villarroya, S., Castillo, J. C., Alonso-Martín, F., Malfaz, M., 2020. Mini: a new social robot for the elderly. *International Journal of Social Robotics* 12 (6), 1231–1249.
- Smith, L., Lamprecht, C., 2024. Identifying the limitations associated with machine learning techniques in performing accounting tasks. *Journal of Financial Reporting and Accounting*.
- Soori, M., Arezoo, B., Dastres, R., 2023. Artificial intelligence, machine learning and deep learning in advanced robotics, a review. *Cognitive Robotics*.
- Sutton, R. S., Barto, A. G., 2018. *Reinforcement learning: An introduction*. MIT press.
- Vouloutsi, V., Cominelli, L., Dogar, M., Lepora, N., Zito, C., Martínez-Hernández, U., 2023. Towards living machines: current and future trends of tactile sensing, grasping, and social robotics. *Bioinspiration & biometrics* 18 (2), 025002.