
7 TÉCNICAS DE APRENDIZAJE

La capacidad de aprendizaje en un sistema multiagente representa un rasgo de indiscutible importancia en la inteligencia del mismo.

Una posible definición informal de aprendizaje, dada desde el punto de vista de la Inteligencia Artificial, describe a este como la adquisición de nuevo conocimiento, habilidades motrices y/o cognitivas e incorporación de los mismos en futuras actividades del sistema, habiéndose realizado esta adquisición e incorporación por el propio sistema en pos de mejorar su rendimiento (Sen y Weiss, 1999).

A lo largo del capítulo se presentan las ideas y clasificaciones típicas de las situaciones de aprendizaje. Tras esto, y a partir de la última clasificación considerada, se revisan algunas de las técnicas y tendencias actuales.

7.1 Aprendizaje de Máquinas e Inteligencia Artificial Distribuida

Las técnicas de Aprendizaje de Máquinas (*Machine Learning*, ML) constituyen uno de los campos centrales de Inteligencia Artificial. Estas técnicas tratan aspectos computacionales del aprendizaje en sistemas naturales y técnicos.

Hasta los años noventa, no se produce la confluencia entre la Inteligencia Artificial Distribuida y las técnicas de Aprendizaje de Máquinas. Las técnicas de ML habían sido ampliamente estudiadas anteriormente, pero de forma independiente de los MAS. A pesar de que el aprendizaje y la adaptabilidad sean características de gran importancia para la inteligencia en un sistema multiagente (Alonso et al., 2001).

Por su parte, la Inteligencia Artificial Distribuida estaba centrada en sistemas multiagente cuya estructura organizativa y comportamiento funcional estaban, típicamente, determinados a detalle y más o menos fijos.

La teoría del Aprendizaje de Máquinas, inicialmente, trataba el aprendizaje como un proceso centralizado y aislado que tenía lugar en una entidad inteligente. Las técnicas de aprendizaje de máquinas hacen que un agente máquina descubra por sí mismo, a menudo tras varios intentos, la forma de resolver una determinada tarea o problema.

Ante esta situación, y recordando que en los sistemas multiagente se enfatizan los comportamientos conjuntos de agentes con un cierto grado de autonomía y que interactúan entre sí, las principales razones que originaron esta confluencia fueron:

- Necesidad de dotar a los sistemas multiagente con capacidad de aprendizaje.
- Mejora en la comprensión del proceso de aprendizaje en sistemas técnicos y naturales a través de la extensión de ML desde el aprendizaje de agentes individuales al aprendizaje en sistemas multiagente.

La posibilidad de aprender cómo solucionar las situaciones o tareas por parte de los agentes tiene gran interés.

En la mayoría de los dominios dinámicos, el diseñador del sistema multiagente no puede prever ni identificar todas las situaciones potenciales con las que puede encontrarse un agente, así como tampoco los requerimientos ni necesidades de comportamiento que estas originan en el agente. A consecuencia, el agente necesita capacidad para adaptarse a nuevas situaciones y entornos. Esto es especialmente cierto para MAS, en el que el comportamiento global es emergente en lugar de predefinido. El aprendizaje es una parte crucial de la autonomía y por tanto, es de gran interés en MAS.

En los principios de esta confluencia, a un determinado nivel, los MAS pueden considerarse otro ámbito de aplicación para ML (Alonso et al., 2001). En ese caso, el estudio se reduce a la aplicación de algoritmos de aprendizaje para entidades individuales en agentes de un sistema multiagente.

Es conveniente señalar que el aprendizaje en un sistema multiagente va más allá de dotar a cada agente con capacidad de aprendizaje aislado. La inteligencia en un sistema multiagente es un concepto social y de igual modo lo es el aprendizaje. En consecuencia, los distintos agentes aprenden de forma distribuida e interactiva, viéndose afectados los unos por los otros en su proceso de aprendizaje.

Son numerosas las áreas de conocimiento involucradas en el aprendizaje en los sistemas multiagente. Estas van desde el Aprendizaje Reforzado (*Reinforcement Learning*, RL), la Computación Evolutiva (*Evolutionary Computation*, EC), la Teoría de Juegos (*Game Theory*, GT), el modelado de agentes o la Robótica.

7.2 Concepto de aprendizaje en MAS

El aprendizaje en MAS, básicamente, puede definirse como la aplicación de técnicas de aprendizaje de máquinas a problemas en los que están envueltos múltiples agentes. Este nuevo campo se caracteriza por la existencia de múltiples agentes situados en un gran espacio de búsqueda. Estos agentes interactúan entre sí, de forma que, pequeños cambios en los comportamientos ya aprendidos pueden dar lugar a comportamientos emergentes impredecibles. Al mismo tiempo, existen numerosos agentes que aprenden y que se adaptan entre sí y a los cambios en el contexto, con la gran dificultad que esto implica.

Las distintas clasificaciones que del aprendizaje en MAS pueden hacerse, parten del estudio y la valoración de las siguientes características (Sen y Weiss, 1999):

- Grado de descentralización.
- Características específicas de la interacción.

- Nivel de compromiso de los agentes.
- Rasgos relacionados con el objetivo a conseguir.
- Método de aprendizaje.
- Realimentación del aprendizaje.

7.3 Clasificaciones de los métodos y situaciones de aprendizaje

Atendiendo a la naturaleza de las características previamente enumeradas y enfatizando unas frente a otras, pueden obtenerse las cuatro clasificaciones siguientes (Tabla 1).

Tabla 1: Tabla de posibles clasificaciones de situaciones de aprendizaje en MAS (Fuente: elaboración propia).

Aprendizaje en sistemas multiagente			
CLASIFICACIÓN 1	CLASIFICACIÓN 2	CLASIFICACIÓN 3	CLASIFICACIÓN 4
Grado de centralización	Técnica de aprendizaje	Tipo/Nivel de interacción	Intencionalidad
Centralizado	On line	Social	Cooperativo
Descentralizado	Off line		Competitivo

7.3.1 Métodos de aprendizaje centralizados y descentralizados

Según el grado de descentralización, pueden darse dos tipos de aprendizaje:

- Centralizado: aislado, aprendizaje de un único agente de forma independiente del resto.
- Descentralizado: interactivo, varios agentes toman parte en un mismo proceso de aprendizaje, pudiendo participar en varios procesos a la vez.

La mayoría de los algoritmos de aprendizaje se han desarrollado a partir de una perspectiva individual, puesto que este era el enfoque de las técnicas heredadas de ML. En este caso, el agente se centra en mejorar sus rasgos individuales, independientemente del dominio en el que se encuentra. No obstante, esta situación no es realmente posible en un MAS, puesto que un agente percibe de forma indirecta al resto de agentes del sistema como parte del entorno. Así pues, parece posible conseguir un comportamiento de grupo coordinado usando estrategias de aprendizaje individuales. No obstante, el aprendizaje individual no siempre permite alcanzar la funcionalidad o el comportamiento óptimo. Existen dominios para los que el aprendizaje distribuido coordinado constituye una aproximación más natural y se consigue mejorar la efectividad.

7.3.2 Métodos de aprendizaje en línea y fuera de línea

El método o la técnica de aprendizaje empleada también permiten distinguir varias formas de aprendizaje.

Las técnicas de aprendizaje en línea (*on line*) y fuera de línea (*off line*) originan dos posibles modos de aprendizaje de igual nombre.

- El aprendizaje *on line* emplea algoritmos de aprendizaje incrementales, tales como Redes Neuronales (*Neural Network*, NN) o Aprendizaje Reforzado (*Reinforcement Learning*, RL). Estos algoritmos se emplean para obtener nuevas hipótesis incrementales tan pronto como un nuevo ejemplo de entrenamiento (parte de información básica para el proceso) está disponible.
- Por otra parte, las técnicas *off line* inducen hipótesis a partir de un conjunto de ejemplos de entrenamiento dados al algoritmo en instantes de tiempo concretos.

El aprendizaje *on line*, por su propia definición, parece más adecuado para MAS, ya que permite actualizar el conocimiento de forma constante. Sin embargo, convendría alcanzar un equilibrio entre ambas. Según Alonso et al., (2001) la mejor opción es mantener una base de ejemplos e hipótesis y cuando estos dejen de ser válidos, lanzar la creación de nuevas hipótesis a partir de los últimos ejemplos de entrenamientos recopilados. Una aproximación a este enfoque mixto es el empleado en (Guerra-Hernández et al., 2004).

7.3.3 Métodos basados en procesos lógicos o reactivos

Otra posible distinción, que emana de las técnicas de aprendizaje empleadas, es la que distingue entre un aprendizaje reactivo y otro basado en sistemas lógicos.

- Sistemas reactivos: en estos sistemas el comportamiento global emerge de la interacción de los comportamientos de los agentes sin que existan protocolos de coordinación específicos ni se dote a los agentes de complejos modelos de conocimiento.

Se asume un comportamiento de los individuos basado en los valores de la información, a partir de los cuales se origina el comportamiento social (por ejemplo, distancia a mantener con los vecinos).

El aprendizaje se reduce a un mapeo de situaciones en acciones. Al evitarse el procesamiento interno en los agentes, este método permite ofrecer respuestas en tiempo real. Como contrapartida, la falta de capacidad para abstraer conceptos y transformar la experiencia vivida en conocimiento impide el despliegue de comportamientos de alto nivel. Es decir, no es posible simular interacciones sociales complejas.

- Sistemas basados en procesos lógicos: permiten incorporar conocimiento del dominio en el proceso de aprendizaje de forma natural. Las herramientas EBL (*Explanation Based Learning*) e ILP (*Inductive Logic Programming*) son técnicas *offline* adecuadas para el aprendizaje basado en conocimiento. Permiten además superar las limitaciones de los sistemas reactivos.

EBL (Carbonell et al., 1990) extrae reglas generales a partir de ejemplos sencillos. Para ello, genera explicaciones sobre el éxito o fracaso del sistema en un problema y generaliza el resultado. El conocimiento se convierte, mediante una ley deductiva, en pericia, experiencia.

ILP (Muggleton y Raedt, 1994), por su parte, se apoya en procesos inductivos. Este método formula hipótesis basadas en circunstancias externas e inicialmente desconocidas.

Pese a lo usado y difundido de la arquitectura BDI para los agentes, esta no se ha tenido muy en cuenta a la hora de estudiar el aprendizaje en MAS. Esto es debido a que la mayoría de las aproximaciones al aprendizaje en MAS usaban métodos de aprendizaje orientados a acción y con representaciones menos expresivas que las empleadas por los agentes BDI. Los métodos de aprendizaje que emplean representaciones más expresivas se usan habitualmente para aprendizaje aislado.

Una arquitectura BDI expandida para suplir las limitaciones de los agentes BDI, es decir, su falta de competencia para el aprendizaje es la expuesta en (Guerra-Hernández et al., 2004). Los métodos ILP, concretamente, inducción de árboles de decisión lógicos permiten dotar a los agentes con propiedades para el aprendizaje.

Para la consecución de agentes BDI aprendices se considera una arquitectura genérica para el agente. ACE (*Action Estimation Algorithm*) es un sistema de aprendizaje que, partiendo de interpretaciones, construye árboles de decisión lógicos, es decir, árboles de decisión en los que los nodos internos son conjunciones de sentencias de primer orden (Blockeel et al., 2000).

La arquitectura propuesta, junto con la autonomía e intencionalidad del agente, determina cuándo este debe aprender, configura su módulo de aprendizaje y ejecuta ACE.

Los ejemplos desarrollados en (Guerra-Hernández et al., 2004) muestran cómo se incrementan las posibilidades de aprendizaje cuando se pueden compartir ejemplos de entrenamiento.

La combinación de estas dos técnicas permite alcanzar resultados óptimos en entornos dinámicos y complejos característicos de MAS. Puesto que una puede suplir las dificultades o problemas de la otra. Por ejemplo, apoyarse únicamente en las reglas generadas por EBL puede ser impráctico en entornos reales, ya que la observación y percepción que se tiene del entorno por un agente nunca es completa. ILP puede computar hipótesis a partir de ejemplos de entrenamiento y del conocimiento de fondo o base (Alonso et al., 2001).

7.3.4 Aprendizaje social y métodos

La naturaleza de la interacción que tiene lugar entre los agentes y el papel que estos juegan en ella da lugar a lo que se conoce como aprendizaje social. Las raíces del aprendizaje social se encuentran en la confluencia entre la Biología y la Inteligencia Artificial.

El aprendizaje social engloba situaciones como la que se produce cuando un nuevo agente, que puede tener algunos comportamientos codificados en tiempo de diseño, entra en un entorno en el ya existen otros agentes. En este caso, no existe necesidad de que el último agente que acaba de entrar aprenda todas las condiciones y posibilidades del entorno por sí

mismo, sino que puede beneficiarse del aprendizaje acumulado por la población de agentes ya existente.

En el caso de aprendizaje social cooperativo, lo que es bueno para un agente es bueno para todo el resto, por lo que se asume la buena voluntad de todos los agentes en el proceso de aprendizaje. No obstante, en entornos no cooperativos pueden existir agentes maliciosos que distorsionen el proceso de aprendizaje, por ejemplo, transmitiendo enseñanzas erróneas con la intención de engañar.

Si en un sistema multiagente es posible este tipo de aprendizaje, deben determinarse los momentos en los que es conveniente usarlo y cómo hacer uso del mismo. El aprendizaje social se desarrolla o evoluciona con mayor probabilidad si los costes del aprendizaje individual son altos. Debe tenerse en cuenta que el aprendizaje social es adecuado cuando el ritmo de cambio espacial o temporal del entorno es medio/bajo (Alonso et al., 2001).

Existen numerosas formas por las que un agente puede aprender comportamiento de otros. Una posible forma es la imitación, más compleja de lo que aparenta, pues no sólo hay que percibir y reproducir, sino entender los cambios originados en el entorno y por qué llevan a la consecución de la meta.

Algunos de los mecanismos más sencillos a partir de los cuales pueden construirse habilidades de aprendizaje que den lugar a un fenómeno global complejo son, entre otros, el comportamiento contagioso o el aprendizaje por observación.

Una realización concreta de estas normas sencillas es, por ejemplo, la norma de permanecer cerca de la madre. Esta regla, combinada con otras normas sencillas, permite originar un aprendizaje social sofisticado.

Una arquitectura de agentes basada en componentes con apariencia visual software y que permite al agente aprender vocabulario sencillo mediante observación es presentada en (Machado y Botelho, 2006).

El proceso de aprendizaje considerado implica observación visual e interacción social (intercambio de mensajes para deshacer confusiones y errores). Una vez el aprendizaje es validado, cuando el observador está seguro del éxito del proceso, crea una entrada en su mapa de vocabulario e integra el nuevo concepto.

Este modo de aprendizaje pone de manifiesto la importancia de la interacción social para el aprendizaje por observación, y se muestra como un procedimiento válido para diversos dominios de aplicación.

7.3.5 Aprendizaje Cooperativo y Competitivo

Desde el punto de vista del comportamiento genérico del propio sistema multiagente, estos pueden describirse como cooperativos o competitivos, si bien, en muchos casos un mismo sistema se clasifica de un modo u otro según las circunstancias consideradas.

Esta última clasificación presentada no es disjunta, sino que es posible el solapamiento entre los dos tipos establecidos.

Desde el prisma del aprendizaje y con esta clasificación, puede verse cómo un agente aprende usando técnicas de ML, ya sea para actuar individualmente o para coordinarse con otros agentes en pos de un objetivo común. Si bien, esta última situación constituye aún un desafío abierto para sistemas con entornos en tiempo real, ruidosos, colaborativos y posiblemente también competitivos.

- **Sistemas cooperativos:** Los agentes de un sistema cooperativo persiguen un objetivo común.
- **Sistemas competitivos:** Los agentes en sistemas competitivos no tienen objetivos alineados, sino que el interés de los agentes es maximizar su propia ganancia.

Los trabajos recientes en este ámbito tratan de extrapolar las técnicas de ML de dominios estacionarios y de un único agente a los entornos multiagente.

La principal diferencia entre agentes cooperativos y competitivos radica en el objetivo de cada uno de ellos. Los agentes cooperativos tienen como meta la maximización de la utilidad del grupo mientras que los competitivos sólo quieren maximizar su propia utilidad. Esta clasificación de Hoen et al., (2005) enfatiza la intención del agente, el objetivo de su diseño frente a su comportamiento. Esto es debido a que, por ejemplo, un agente competitivo (también llamado egoísta) puede cooperar con otros porque espera conseguir una mayor recompensa de este modo. Así pues, el comportamiento aparente no constituye la base de la clasificación elegida, sino que esta nace de la intencionalidad del agente.

El carácter cooperativo o competitivo de un MAS se refleja ya en aspectos relativos al diseño del mismo. Mientras que el primero suele ser diseñado por un mismo diseñador que ya crea en la estructura las características necesarias para que se produzcan complejos fenómenos de cooperación, en el segundo caso, los agentes pueden provenir de fuentes bien distintas, por lo que las tareas de coordinación se complican.

La elección de unas estrategias u otras y la implementación de ciertas técnicas son aspectos que deben tenerse en cuenta al diseñar un sistema multiagente cooperativo o competitivo.

7.4 Aprendizaje en sistemas multiagente Cooperativos

Los sistemas multiagente cooperativos se componen de una serie de agentes que, mediante su interacción, tratan de resolver tareas conjuntamente o de maximizar la utilidad.

En este contexto, cobra importancia el concepto de aprendizaje, con un interés inicial en las técnicas de ML como vía para la automatización del proceso inductivo, de los procesos de búsqueda y optimización de problemas (Panait y Luke, 2005).

Existen dos categorías básicas de aprendizaje dentro de los MAS cooperativos:

- **Aprendizaje en equipo:** un único agente busca los comportamientos para todo el equipo. Este enfoque mantiene una estrecha relación con las técnicas tradicionales de ML. Puede tener problemas en cuanto a escalabilidad se refiere a medida que aumenta el tamaño del

equipo. Una posibilidad para mantener el espacio de búsqueda manejable es asignar comportamientos idénticos a múltiples miembros del equipo.

- Aprendizaje concurrente o simultáneo: en este caso, existen varios procesos aprendiendo al mismo tiempo. El hecho de que cada agente de un conjunto que coopera aprenda de forma independiente implica la reducción del espacio de búsqueda de soluciones por división del mismo. No obstante, la presencia de múltiples aprendices hace que el entorno no pueda ser considerado estacionario, premisa básica para la aplicación de técnicas de ML. A consecuencia, se requieren nuevos enfoques y métodos.

7.4.1 Aprendizaje en equipo

Esta modalidad es una de las dos grandes categorías existentes en los modelos de aprendizaje en MAS cooperativos y se sitúa muy cercana a la línea de aprendizaje de máquinas.

El modelo de aprendizaje en equipo constituye una aproximación sencilla al aprendizaje en sistemas multiagente, pues permite aplicar las técnicas estándar de ML al agente aprendiz del equipo. Consiste, básicamente, en que un único aprendiz busca los modelos de comportamiento para todo el equipo de agentes. Esta característica hace que se eviten problemas relacionados con la adaptación del conocimiento adquirido por distintos agentes, lo que constituye una ventaja en cuanto a facilidad de implementación se refiere.

No obstante, aparecen problemas de escala para dominios complejos. La explosión del espacio de soluciones que se produce para grandes espacios de búsqueda puede saturar o disminuir la eficiencia de técnicas como la de Aprendizaje Reforzado o Computación Evolutiva (*Evolutionary Computation*, EC). Este tipo de problemas son difícilmente tratables mediante técnicas de aprendizaje reforzado, que exploran el espacio de utilidades. Las técnicas de computación evolutiva, que exploran el espacio de comportamientos, son más adecuadas en estas situaciones por permitir una mejor escalabilidad (Salustowicz et al., 1998). En cualquier caso, no se puede olvidar el hecho de que este método es centralizado, con los problemas que esto puede acarrear.

Dentro de esta misma categoría, pueden distinguirse otras tres, en función de que el aprendizaje en equipo sea homogéneo, heterogéneo o un híbrido de los dos anteriores.

- Aprendizaje homogéneo: en este caso, el comportamiento de todos los agentes será igual al desarrollado por el o los aprendices.

Asumir que todos los agentes tienen el mismo comportamiento reduce drásticamente el espacio de búsqueda de aprendizaje. La aplicación de distintas técnicas de aprendizaje, la inclusión de capacidad de comunicación, etc. hacen que varíe la eficiencia de esta técnica.

- Aprendizaje heterogéneo: se caracteriza porque cada agente puede tener un comportamiento distinto gracias a la especialización en las tareas.

En esta ocasión, el equipo está formado por agentes con distintos comportamientos y un único aprendiz que trata de mejorar el equipo como un todo. Permite la existencia de diversidad en el equipo a costa de un incremento en el espacio de búsqueda.

- Aprendizaje híbrido: representa el punto medio entre las dos concepciones anteriores y consiste en adoptar distintas pautas de comportamiento para grupos de agentes.

La adecuación de estas técnicas a un problema concreto depende de las necesidades de especialización para la resolución del mismo.

7.4.2 Aprendizaje concurrente

El segundo modo más empleado para el aprendizaje en MAS cooperativos es el aprendizaje concurrente, en el que numerosos procesos de aprendizaje tratan de mejorar, al mismo tiempo, partes del equipo.

Típicamente, cada agente se constituye como un proceso para modificar su propio comportamiento. De aquí surge una de las mayores dificultades de este enfoque: mientras que un agente aprende para adaptarse a su entorno, al mismo tiempo, y sin que este lo pueda controlar, otros agentes hacen lo mismo, con las modificaciones en el entorno que esto conlleva.

La adecuación de una técnica de aprendizaje u otra a un sistema multiagente depende de la situación a la que este se enfrente. Canssen y Wiegand (2003) establecen que el aprendizaje concurrente resulta más efectivo para dominios en los que se puede realizar una cierta descomposición útil del mismo. O bien cuando es adecuado centrarse, hasta cierto punto, en un subproblema de manera independiente del resto.

El mayor desafío o la mayor dificultad de esta técnica se debe a que un agente adapta su comportamiento a un entorno al que otros aprendices también se están adaptando al mismo tiempo, originando así unos cambios sobre los que el primer agente no tiene ningún control. La adaptación de un agente al entorno provoca cambios en el mismo que pueden hacer inútiles las tareas de adaptación de otros agentes. Esta situación hace que se viole la premisa básica sobre la que se apoya ML.

La investigación en aprendizaje concurrente sigue tres líneas: asignación de crédito, dinámica de aprendizaje (entender el impacto de la coadaptación en los procesos de aprendizaje) y modelado de otros agentes para mejorar las interacciones (colaboración).

- Problema de la asignación de crédito (*Credit Assignment Problem, CAP*): el objetivo de esta línea es dar respuesta a la pregunta de cómo dividir la recompensa entre los agentes. Esto puede hacerse a partes iguales con el método de recompensa global (*global reward*) o valorando, según distintas políticas, el comportamiento individual de cada agente, es decir, con la técnica de recompensa local (*local reward*).

La recompensa global no es escalable para problemas de dificultad creciente porque los aprendices no reciben la suficiente realimentación compartida para sus acciones específicas (Wolpert y Tumer, 2001). Esta técnica también presenta dificultades para su aplicación a entornos distribuidos en los que no se puede computar la recompensa global eficientemente (Hoen et al., 2005).

Por su parte, la recompensa local penaliza la pereza en los agentes y se favorece la aparición de comportamientos voraces, pues los agentes no tienen ningún incentivo para ayudar a los otros.

Otra posibilidad para la asignación de la recompensa es la Utilidad de la Vida Maravillosa. Según esta filosofía (*Wonderful Life Utility*), en lugar de calcular la recompensa que le corresponde a un agente según su contribución a la recompensa global, el cálculo se basa en el comportamiento que habría tenido el equipo si este agente no hubiera estado presente (Wolpert y Tumer, 2001).

Este problema de Asignación de Crédito puede descomponerse en la asignación de crédito intra-agente (asignación de crédito a las decisiones e inferencias subyacentes al agente) e inter-agente (a las acciones externas de los agentes) (Sen y Weiss, 1999).

La asignación inter-agente resulta muy compleja para MAS, pues los agentes que han contribuido a la producción de un cambio en el entorno pueden estar lógicamente y espacialmente distribuidos, además de ser de naturaleza heterogénea. El problema de asignación de crédito intra-agente es igual de complejo para el ámbito del aprendizaje individual o el de los sistemas multiagente.

- Dinámica del aprendizaje: se enfrenta al problema de conseguir, con el limitado número de herramientas de estudio y análisis disponibles, un comportamiento óptimo en un entorno continuamente cambiante.

Cuando se aplica aprendizaje individual a un entorno estacionario, el agente experimenta con múltiples comportamientos hasta encontrar el comportamiento óptimo global.

En los entornos dinámicos, el agente trata, en el mejor de los casos, de mantenerse al día de los cambios que se producen en el entorno y perseguir el comportamiento óptimo (óptimo móvil). La situación es aún más complicada en los MAS, en los que el proceso de aprendizaje de los propios agentes provoca cambios en el entorno de aprendizaje de los otros agentes.

Según Hoen et al. (2005), dos son las herramientas capaces de ayudar a modelar y analizar la dinámica de aprendices concurrentes que siguen distintas técnicas de aprendizaje. Estas dos técnicas son la Teoría de Juegos Evolutiva (*Evolutionary Game Theory*, EGT) y por otro lado, una herramienta que combina información sobre la velocidad de cambio del comportamiento de cada agente, ritmo de aprendizaje y retención y, por último, ritmo en el que los otros agentes aprenden, para modelar y predecir así el comportamiento de los aprendices concurrentes existentes.

Muchos de los métodos de aprendizaje concurrente convergen al punto de equilibrio de Nash, aunque este no sea óptimo. Esto se debe a que los agentes, normalmente, no tienen control sobre lo que hacen los otros y es difícil crear alianzas para escapar de este equilibrio.

- Modelado de los compañeros de equipos: se centra en el aprendizaje de un agente sobre los otros. Consiste en que un agente sea capaz de realizar hipótesis válidas de comportamiento para los demás y adaptar así las decisiones propias en pos de una mejor aproximación a la solución óptima.

Para ello, pueden usarse Redes Bayesianas con el fin de estimar las creencias, capacidades y preferencias de los otros. El modelado no implica una mejor coordinación necesariamente. La experiencia ha demostrado que las asunciones iniciales condicionan fuertemente el comportamiento final desarrollado.

A modo de resumen, la Figura 1 muestra las divisiones y líneas principales de investigación dentro de esta corriente.

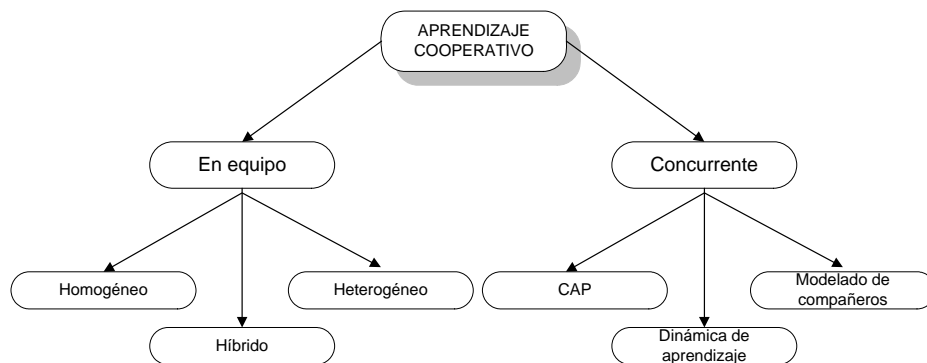


Figura 1: Posibles líneas dentro del Aprendizaje Cooperativo en MAS (Fuente: elaboración propia).

7.5 Aprendizaje en sistemas multiagente Competitivos

Los sistemas multiagente tratados en el punto anterior estaban caracterizados porque los agentes implícita o explícitamente tenían como meta común el trabajar juntos. Los agentes se consideraban benévolo y elegían sus acciones con el fin de promover la utilidad común o global del sistema.

Una nueva perspectiva que contribuye a mejorar la comprensión del aprendizaje en MAS y a la creación de nuevos algoritmos es propuesta en (Tuyls et al., 2004).

En primer lugar, se señala la falta de un marco teórico en el que el aprendizaje y la adaptabilidad de los MAS sean entendidos profundamente. Se propone como base para la construcción de este marco la Teoría de Juegos Evolutiva.

Para ello, se estudia la relación existente entre los sistemas multiagente y la Teoría de Juegos Evolutiva. Entre ambas disciplinas existen numerosas similitudes, pues la propia Teoría de Juegos tradicional modela interacciones entre agentes racionales en juegos de dos o más jugadores que eligen entre estrategias conocidas las recompensas y preferencias. La Teoría de Juegos Evolutiva permite tener en cuenta la falta de conocimiento completo por parte de los agentes, así como repetir el juego varias veces. De ahí la necesidad de aprendizaje y adaptación.

La relación existente entre la Teoría de Juegos Evolutiva y el Aprendizaje Reforzado es fundamentalmente matemática: la extensión de métodos clásicos de RL para obtener nuevos métodos en EGT. Por último, se tiene en cuenta también la estrecha relación existente entre MAS y RL, ya que esta técnica de aprendizaje es una de las de mayor difusión en MAS.

Así pues, se identifican las necesidades de aprendizaje en MAS y se las relaciona con EGT para establecer el marco teórico.

Una serie de principios generales para diseñar grandes sistemas multiagente complejos y descentralizados de manera automática es propuesta en (Wolpert et al., 1999). Se considera que cada agente emplea RL, lo que convierte el problema en uno de determinación y actualización de las funciones de recompensa para alcanzar el objetivo global.

El marco matemático propuesto se denomina *Collective Intelligence*, COIN y es aplicable a problemas tales como el diseño y control para el rutado en una red de comunicaciones, la construcción de algoritmos paralelos para solución numérica de problemas de optimización, etc.

COIN está relacionado con mecanismos estadísticos, GT, EGT y se apoya en conceptos como los submundos, la alineación de restricciones o la utilidad de la vida maravillosa para conseguir un funcionamiento óptimo de forma rápida, pese a la naturaleza local de las informaciones disponibles para los agentes. Esto se consigue con los algoritmos de microaprendizaje (*microlearning*) y macroaprendizaje (*macrolearning*).

El primero de ellos proporciona una recompensa en valores reales para los agentes en el problema tratado. El estudio del aprendizaje desde el punto de vista macro, permite mejorar un sistema no alineado en restricciones inicialmente, mejorando su utilidad del mundo. Para conseguirlo, varía la pertenencia de los agentes a los submundos.

Los experimentos estudiados en (Wolpert et al., 1999) muestran la adecuación de los principios de COIN para considerarlos un marco genérico de aprendizaje reforzado en sistemas multiagente (*Multiagent Reinforcement Learning*, MARL).

Son numerosos los distintos enfoques y las técnicas novedosas propuestas para enfrentar el problema MARL.

Aprender a actuar en un sistema multiagente es un problema difícil, puesto que la definición de política óptima deja de ser válida. La política óptima depende en cada momento de las políticas seguidas por los otros agentes. Se origina así una situación de objetivo de aprendizaje móvil. Los dominios multiagente considerados constan de agentes que están obligados a interactuar entre sí. Estos agentes pueden tener objetivos, asunciones, algoritmos y convenciones independientes.

Según Bowling y Veloso (2002) los primeros algoritmos de aprendizaje solían tener uno de estos dos inconvenientes:

- Falta de convergencia a la política óptima frente a políticas específicas de sus adversarios.
- Falta de convergencia genérica.

Para suplir estos problemas, los mismos autores proponen el método WoLF (*Win or Learn Fast*) que se basa en un ritmo de aprendizaje variable según la situación en la que se encuentre el agente. Mediante el aprendizaje, el agente puede descubrir y sacar partido de la dinámica del entorno, así como adaptarse a dificultades imprevistas en la realización de tareas.

La filosofía de WoLF consiste en aprender rápido si el agente está perdiendo (está en una situación de desventaja) y aprender cautelosamente cuando está ganando.

El método propuesto verificaba los principios de racionalidad y convergencia, considerados en este caso como criterios básicos de evaluación de los métodos de aprendizaje.

Un aspecto importante en el aprendizaje en sistemas multiagente que no había recibido mucha atención con anterioridad es resaltado en (Hoen et al., 2006). Aquí se propone el empleo de técnicas que tengan en cuenta el efecto de las acciones propias en las políticas desarrolladas por los otros agentes. De este modo, se intenta que un agente pueda aprender una buena política enfrentado a oponentes adaptativos y competitivos.

La técnica presentada permite al agente aprender con la posibilidad de modelar su entorno y sus adversarios, frente a los cuales quiere optimizar su recompensa jugando a Mejor Respuesta (*Best Response*, BR). La política que maximiza esta recompensa depende de las acciones cambiantes o variables de los adversarios. La consideración de los cambios en las acciones de los otros como respuesta a las acciones propias de un agente en cuestión, permite a este último obtener una visión panorámica y una cierta ventaja en situaciones en las que interactúa repetidamente con los mismos adversarios.

El nuevo marco MARL propuesto se denomina StrOPM (*Strategic Opponent Policy Modeling*). Este supera a anteriores aproximaciones a la filosofía central del método como MOPM (*Myopic Opponent Policy Modeling*), ya que tiene en cuenta tres factores: la política actual estimada de los oponentes, las recompensas a recibir por las acciones y el impacto de las acciones elegidas en la política futura del adversario.

Este marco novedoso resulta válido para juegos de matrices iterados representativos, como el iPD (*iterated Prisoner Dilemma*).

No obstante, Nachbar y Zame (1996) probaron que, en general, es imposible aprender perfectamente a jugar de forma óptima frente a un adversario adaptativo y, al mismo tiempo, estimar la política de este perfectamente (de forma exacta).

El aprendizaje frente a múltiples oponentes ha sido tratado en (Vuc et al., 2006). El objeto básico de estudio son los denominados problemas de aprendizaje en juegos *general sum* repetidos y con n jugadores, que pueden ser cooperativos o adversarios.

Para esta situación se proponen criterios que se centran en la garantía de recompensas o pagos frente a distintas clases de adversarios. Estos criterios refuerzan algoritmos de aprendizaje que permiten cooperación entre agentes al mismo tiempo que garantizan una recompensa mínima básica para cada agente.

Se proponen algoritmos que alcanzan los criterios propuestos para situaciones de adversarios estacionarios y adaptativos. Es reseñable que la implementación del algoritmo para estrategias estacionarias mejora los resultados obtenidos con otras técnicas en entornos similares.

7.6 Aprendizaje y otras dimensiones del sistema multiagente

La relación e influencia que el aprendizaje tiene con aspectos del sistema multiagente, tales como la organización, cooperación o la comunicación, ya se ha hecho notar y se puede intuir de los puntos anteriores.

Los agentes pueden aprender a coordinar sus actividades adecuadamente, evitando posibles fallos y beneficiándose, cuando sea posible, de las condiciones del entorno. Para ello, los agentes deben aprender sobre su entorno y los otros agentes en el sistema, para adaptar así sus comportamientos locales.

El aprendizaje también tiene un papel importante en la definición de la estructura organizativa del sistema o grupo de agentes que acometen una serie de tareas. Los agentes en grupos han de aprender la asignación de roles para complementarse de forma efectiva. La adaptación de la estructura de un grupo y las actividades de sus miembros a las distintas situaciones a enfrentarse permite ampliar el rendimiento del sistema y resolver desafíos imprevistos. Los agentes aprenden roles según la situación concreta en un dominio de resolución de problemas cooperativo (Sen y Weiss, 1999).

Según el problema al que se enfrente el sistema multiagente la comunicación puede ser una necesidad o una herramienta para mejorar el rendimiento y la eficacia del proceso.

Los agentes pueden comunicarse para coordinarse de forma más efectiva, distribuir modelos más exactos del entorno y aprender las soluciones de subtareas ya resueltas por otros agentes.

No obstante, la comunicación también puede hacer que aumente el espacio de búsqueda y se dificulte o impida el aprendizaje del comportamiento óptimo.

Según Stone y Veloso, (2000) para que un sistema multiagente comunicativo sea considerado sistema multiagente necesita restricciones.

Sin centrarse en ella, Nunes y Oliveira (2004) emplean la comunicación como herramienta para permitir que los agentes aprendan a partir de múltiples de fuentes de información.

En los sistemas multiagente se producen situaciones en las que distintos agentes necesitan aprender conceptos similares en paralelo. El enfoque propuesto, amplía las posibilidades del Aprendizaje de Máquinas, puesto que el intercambio de información puede aumentar la exactitud y velocidad de resolución del problema a costa de la comunicación de los agentes.

Los agentes pueden tener distintas estructuras y algoritmos de aprendizaje. Esta aproximación al aprendizaje favorece la emergencia de nuevas técnicas a partir de reglas simples de comunicación. Los resultados medios obtenidos no permitieron concluir que el entorno de aprendizaje social mejorara significativamente los resultados frente al aprendizaje individual, pero sí mostró que la funcionalidad media de los agentes con peores características individuales mejoraba.