

Aprendizaje en SMA

**Jose Aguilar
CEMISID-ULA**

Características

- La capacidad de aprendizaje permite a los agentes adaptarse a las nuevas situaciones que aparecen en el entorno.

cada agente en estos sistemas deben aprender a adaptarse a la comportamiento dinámico y desconocido de los otros agentes y/o ambiente para competir o colaborar eficazmente.

- El aprendizaje, como la inteligencia, es un fenómeno social en los SMA.

Los agentes aprenden de forma distribuida e interactiva, afectándose los unos a los otros.

Aprendizaje

- Adquisición de nuevo conocimientos y habilidades cognitivas, y la incorporación de los mismos en las actividades del SMA
- El proceso de adquisición de nuevo conocimientos y habilidades cognitivas es guiado por el mismo sistema, coadyuvando a mejorar su rendimiento

Proceso de aprendizaje se refiere a toda actividad (planificación, inferencia, toma de decisión) que es ejecutada con la intención de alcanzar un *objetivo de aprendizaje*

Aprendizaje de un Agente vs Aprendizaje SMA

- Casi todos los algoritmos de *aprendizajes han sido hechos para un agente*

Como usarlos en SMA?

- Algoritmos de aprendizajes para un agente se focaliza en como un agente mejora sus habilidades individuales.
- No podemos hablar de aprendizaje SMA, si un agente no afecta ni es afectado por otros agentes
si un agente no es explícitamente consciente de otros agentes, lo percibe como parte del medio ambiente y su comportamiento será parte de la hipótesis a aprender.
- Es posible lograr un comportamiento coordinado del grupo usando aprendizaje para un solo agente

Aprendizaje de un Agente vs Aprendizaje SMA

- Investigaciones anteriores han demostrado que ciertos niveles de conocimiento de los agentes pueden perjudicar el rendimiento.
- El aprendizaje para un agente no siempre produce un rendimiento óptimo en SMA y pueden existir dominios donde *un aprendizaje coordinada multi-agente* es una metáfora más natural y mejora la eficacia.

es una pregunta abierta si niveles altos de conocimiento en un agente producen un mejor desempeño.

Aprendizaje multi-agente

- Definición en sentido amplio:
es la aplicación de aprendizaje de máquina a problemas que afectan a múltiples agentes
- Características del aprendizaje multi-agentes.
 - **Involucran a múltiples agentes**, los espacio de búsqueda pueden ser inusualmente grande, y debido a la interacción de los agentes pequeños cambios en los comportamientos a menudo pueden dar lugar a cambios impredecibles en el resultado a nivel macro ("emergente"), es decir el SMA como un todo.
 - **Pueden participar múltiples alumnos**, c/u aprendiendo y adaptándose en el contexto de los demás, lo que presenta problemas de la teoría de juegos en el proceso de aprendizaje

¿Cuál es el objetivo en el aprendizaje multi-agente?

- Debido a que no conoce su entorno
 - ¿Cómo se comportan los otros agentes?
 - ¿Cuál es la función de recompensa?
- Aprender una mejor respuesta
 - Convergencia de políticas.
 - El minimax óptimo.
- Información completa: resuelto
 - Soluciones exactas o aproximadas
- Información Incompleta: **se requiere aprender**
 - El entorno no es markoviano
 - La convergencia no está garantizada
 - Comportamientos impredecibles

Aprendizaje Social

- Considere un SMA, donde los nuevos agentes entran en un mundo ya poblado con agentes experimentados.
 - Los nuevos agentes se inicia con una pizarra en blanco, ya que no ha tenido todavía la oportunidad de aprender sobre su entorno (aunque pueden tener programado comportamientos).
 - Sin embargo, un nuevo agente no tiene que saber todo lo relacionado con el medio ambiente por sí mismo: puede beneficiarse del aprendizaje acumulado de la población de agentes experimentados.

Esta situación podría caracterizar los agentes de software altamente autónomos que operan en Internet

Aprendizaje Social

La situación descrita también coincide con el problema de aprendizaje en un animal recién nacido, sobre todo en especies sociales como la nuestra.

- Una diferencia importante entre los agentes artificiales y los animales es que en un SMA a menudo hablamos de escenarios completamente cooperativos: lo que es bueno para un agente es bueno para todos (función de utilidad común).
- Aunque la cooperación se produce en muchas especies animales, el potencial conflicto no está ausente debido a la competencia en el corazón del proceso evolutivo.

Aprendizaje Social

- Los conflictos de interés son relevantes en SMA si los agentes están operando en un entorno con competidores malintencionados, como es el caso en Internet.
- El aprendizaje social, en tal caso, podría implicar la complicada tarea de asegurarse de que su "maestro" no este tratando de engañarlo con el fin de sus propios intereses.

¿ Cuando se deben incluir las habilidades de aprendizaje social en un SMA, y cómo debe hacerse?

¿ Cuales son las condiciones en las que será ventajoso para un agente aprender de los demás en lugar de por sí mismos?

Aprendizaje Social

- La conclusión es sencilla:
el **aprendizaje social** es mejor cuando los costos de un agente por un **aprendizaje por ensayo y error** es alto.
- Un ejemplo:
 - situaciones en que un error de un animal podría significar su muerte: comer una planta venenosa o no correr al ver un depredador
 - Muchos depredadores basan su lógica en la probabilidad de encontrar animales pequeños que están aprendiendo a partir del comportamiento de los demás

Hacer equivalencias en agentes de software

Aprendizaje Social

- El aprendizaje social será seleccionado cuando *las tasas de cambio en el medio ambiente (espacial o temporal) se encuentran en niveles intermedios*.
- La lógica es la siguiente:
 - en un entorno que cambia muy lentamente, las estrategias de *lógica cableada* (es decir, la información transmitida genéticamente) permitirán a los animales responder adecuadamente.
 - *Si el entorno cambia muy rápidamente*, el animal debe aprender por sí mismo basado en las condiciones locales. El **aprendizaje social será insuficiente** porque el animal inocente estaría tratando de aprender de otro cuya experiencia en el mundo ya no es pertinente.

La capacidad de aprendizaje social en un grupo de agentes de software depende de la velocidad de los cambios en el entorno

Mecanismos de aprendizaje social

Comportamiento contagioso

- Ejemplificado por:

"Si otros están huyendo, yo huyo también."
- Los estímulos producidos por un comportamiento particular sirven como disparo para que otros se comportan de la misma manera.
- Por ejemplo, especies de animales donde un movimiento rápido de uno de ellos hace que el grupo de animales se mueva. Así, cualquier de uno de ellos al huir dará lugar a una reacción en cadena de movimientos rápidos.
- No implica un aprendizaje real, es más reactivo, sin embargo es una especie de **comportamiento social adaptativo**.
- Ejemplos: el movimiento de los rebaños de animales, de bancos de peces, la risa y el bostezo en los seres humanos

Mecanismos de aprendizaje social

Seguir una conducta o comportamiento

- Ejemplificado por:
"seguir a alguien mayor, y luego aprender de lo que sucede"
- Por ejemplo, si usted sigue a sus padres, y ellos a veces comen chocolate, podríamos desarrollar un gusto por el chocolate. Eventualmente aprenderemos que comer chocolate es bueno.
- Capacidad de aprender para generar una conducta adquirida.
- Es un estímulo para la adquisición de una conducta de alimentación en ciertas especies (p.ej. Las ratas negras)

Mecanismos de aprendizaje social

Aprendizaje por observación

- Ejemplificado por:
 - "Preste atención a lo que otros hacen o experimentan, y si los resultados para ellos son buenos o malos, entonces aprenda de eso"
 - El aprendizaje por observación puede existir sin la evaluación *explícita* de la experiencia como buena o mala.
 - La adquisición del miedo en monos ilustra esa idea:
 1. monos criados en laboratorio se les permitió observar a otros de la misma especie y su reacción de miedo ante la presencia de una serpiente.
 2. Los observadores, que antes eran indiferentes a las serpientes, adquirieron rápidamente un miedo.

Mecanismos de aprendizaje social

Comportamiento dependiente de Mapeos:

Capacidad de discriminar

- Este tipo de aprendizaje permite generar un estímulo discriminativo,
- Por ejemplo entrenadas para seguir a un líder.
No hay indicios de que el seguidor entiende las intenciones del líder , ni siquiera que el seguidor es consciente de la coincidencia entre el comportamiento del líder y el suyo.
- Las ratas y palomas pueden ser fácilmente entrenados para discriminar,
- Por ejemplo, una paloma silvestre podría aprender por ensayo y error que rascarse es probable que sea gratificante (es decir, que puedan dar lugar a la ingestión de alimentos). Se podría aprender de las demás la correspondencia entre la comida oculta y la evidencia de la alimentación,

Mecanismos de aprendizaje social

Mapeo Cross Modal: Mímica vocal de las aves

- Caso especial de aprendizaje social debido a que el estímulo original y la respuesta del animal se encuentran en la misma modalidad sensorial,
- Copiar los movimientos de otro animal requiere coincidencia intermodal: el observador debe ser capaz de traducir la información visual asociada a otros a sus movimientos.
- La idea es imaginarse un animal capaz de identificar los movimientos de los demás, y asignarlos a los movimientos de sus propios músculos.
- El trabajo sobre "**las neuronas espejo**" en los monos y los seres humanos (capacidad innata para realizar cross-modal en los seres humanos) es altamente sugestivo.

Clasificación del Aprendizaje

- **Aprendizaje Centralizado**: todo el proceso de adquisición de conocimiento es ejecutado por un solo agente. Es posible que el agente se encuentre situado en un MAS, sin embargo, el proceso de aprendizaje se lleva a cabo como si este estuviera solo.
- **Aprendizaje Distribuido**: varios agentes se encuentran implicados en el proceso de aprendizaje.

Clasificación

Tipo	Técnica de Aprendizaje	Intencionalidad	Cada Agente aprende	Participantes
Centralizado	On-line	Cooperativo	Aislado	Un solo agente
Descentralizado	Off-line	Competitivo	Interactivo	Varios Agentes

- Cooperativo
 - En equipo
 - Concurrente

Un agente puede estar inmerso en varios procesos de aprendizaje (centralizados o descentralizados) en un mismo momento

Propiedades de las clases (primera valida solo caso descentralizado)

Propiedades	Valores
Grado de descentralizado	Distribuido o paralelo
Característica de las interacciones	Nivel, frecuencia, persistencia, patrón y variabilidad
Compromiso de los Agentes	Relevancia y rol
Características de los objetivos de aprendizaje	Tipo de mejora esperada, compatibilidad, etc
Método de aprendizaje	De memoria, por instrucciones/asesoramiento, Desde ejemplos,/practica por analogias, por descubrimiento,
Retroalimentación del Aprendizaje	Supervisado, no supervisado, reforzado

Clasificación del Aprendizaje

- Otra clasificación basada en qué forma c/agente modela a su entorno social, y el comportamiento de los otros agentes:
 - Nivel 1: los agentes aprenden a partir de sus propias interacciones con el ambiente, sin intervención directa con otro agente. Los cambios hechos en el ambiente pueden ser vistos por cualquier agente.
 - Nivel 2: interacción directa entre los agentes por medio del intercambio de mensajes.
 - Nivel 3: aprendizaje a partir de la observación de las acciones tomadas por otros agentes.

Interés del Aprendizaje en SMA

Relacionada con la Organización

Aprendizaje de roles

Aprendizaje sobre otros agentes

Aprender a jugar mejor contra un oponente

Aprender a beneficiarse de las condiciones del entorno

Relacionada con Coordinación

Aprender a coordinarse evitando fallos

Adaptación a distintas situaciones

(

Aprendizaje y Coordinación

Cómo los agentes pueden aprender a coordinar sus actividades?

- *Basado en Aprendizaje reforzado*
 - Agentes reactivos y adaptativos
 - Acciones que maximicen la retroalimentación o reforzamiento
 - Proceso de decisión Markoviano (S, A, P, r) donde S: conjunto de estados, A, conjunto de acciones, P es la probabilidad de ir de; estado e1 a e2 a través de la acción a1, r es la función de recompensa.
 - Cada agente tiene una política T para decidir que acción tomar, así su esperada recompensa es:

$$V(T, \gamma) = E(\sum_t \gamma^t r(T, s, t)) \quad \text{ep: tasa descuento}$$

Aprendizaje y Coordinación

- *Q-learning*: se escoge acción a en estado s tal que se maximice la recompensa
 - $V(T, \gamma, s) = \max_{a \in A} Q(T, \gamma, s, a)$ para todo $s \in S$
 - Si una acción a en estado s produce un reforzamiento de R y una transición a s' , Q ahora es (β es la tasa de aprendizaje y R reforzamiento):
$$Q(s,a) = (1-\beta)Q(s,a) + \beta(R + \gamma \max_{a' \in A} Q'(s', a'))$$
- *Sistema de Clasificación*
 - Sistema basado en reglas, cada regla es (c_i, a_i) :
condición c_i genera acción a_i
 - Aprendizaje usando AG (reglas), donde $S(c_i, a_i)$ calidad de la regla i en el tiempo t (R viene dado por la recompensa por la acción a_i)
$$S(c_i, a_i, t+1) = (1-\beta) S(t, c_i, a_i) + \beta(R + S(t+1, c_i, a_i))$$

Algoritmos Multiagente de Aprendizaje por refuerzo

- En los SMA otros agentes que se adaptan al entorno lo hacen no estacionario, violando la propiedad de Markov que el aprendizaje tradicional hace al establecer que solo se basa en el comportamiento del agente.
 - En el aprendizaje de un robot individual, Q-learning tradicional es bueno.
 - También se puede aplica Q-learning a cada agente en un SMA.
 - Sin embargo, el hecho de que el ambiente ya no es estacionario en el SMA es generalmente descuidada.
- *Minimax-Q learning algorithm for games*: el jugador maximiza sus beneficios de aprendizaje en la peor situación. En esencia, en Minimax-Q el jugador siempre trata de maximizar su valor esperado ante la peor posible escogencia de su oponente.

Q-learning Coordinado

- La idea principal de este método consiste en descomponer la función-Q en una combinación lineal de Q-funciones de agentes locales:

$$Q(s, a) = \sum_{i=1, n} Q_i(s_i, a_i).$$

- Cada función Q-local (Q_i) de un agente i se basa en s_i y a_i que, respectivamente, representan el subconjunto de todos los estados y las variables medidas del agente i .
- Un grafo de relación se construye mediante la adición de un arco entre el agente i y j , cuando la acción del agente j está incluido en las variables de acción del agente i ,
- Una función Q local se actualiza (para un a' vecinos en el grafo)

$$Q_i(s_i, a_i) := Q_i(s_i, a_i) + \alpha [R(s, a) + \gamma \max_{a'} Q(s', a') - Q(s, a)].$$

Aprendizaje y Coordinación

Aprendizaje reforzado Interactivo

- Estimación de acción:
 1. Dada percepciones Si del estado actual S ,
 2. C/agente a_i calcula el grupo de acciones $A_i(S)$ que puede ejecutar, y su relevancia $E(i,j,S)$.
 3. Se calcula lo que ofrece cada acción como $B(i,j,S) = (\alpha + \beta) E(i,j,S)$ donde α es factor de riesgo y β un termino de ruido,
 4. Se selecciona la acción con mas B y se actualiza $E(i,j,S)$ como $E(i,j,S) = E(i,j,S) - B(i,j,S) + R$ donde R es la recompensa externa

Aprendizaje de roles

- Suponer S_k y R_k conjunto de situaciones y roles para el agente k
- Estimación de roles para diferentes situaciones = $|S_k| |R_k|$
- Fase aprendizaje, la probabilidad de seleccionar un rol r en situación s es:

$$Pr(s, r) = \frac{f(U(r,s), P(r,s), C(r,s), \text{Potencial}(r,s))}{\sum_{j \in R_k} (f(U(j,s), P(j,s), C(j,s), \text{Potencial}(j,s)))}$$

donde U es utilidad, P es la probabilidad y C es el costo medidos.

- Para escoger el rol a ser jugado en un momento dado

$$\max_{j \in R_k} (f(U(j,s), P(j,s), C(j,s), \text{Potencial}(j,s)))$$

y $U, P, C, \text{Potencial}$ son actualizados: $U(r, s_{n+1}) = (1 - \beta) U(r, s_n) + \beta U_f$ U edo. Final
 $P(r, s_{n+1}) = (1 - \beta) P(r, s_n) + \beta P_f$ $P_f = 1$ si f es ok

Aprendizaje a beneficiarse de las condiciones del entorno

- Agentes que no modelan a otros (Caso 1):

- un comprador escoge a vendedor s^* tal que

$$s^* = \max_{s \in S} f(g, p(g, s))$$

donde g es el producto deseado, S conjunto de vendedores, f es la función con el valor esperado por el comprador al comprar g al precio p .

- Esa función se aprende como

$$f(g, t+1) = (1-\beta) f(g, t, p) + \beta V(g, b, p, q) \quad \beta \text{ decrece con el tiempo}$$

- Un vendedor s venderá un bien g a precio $p(s^*)$ tal que

$$p(s^*) = \max_{p \in P} p > c(g, s) h(g, s, p)$$

donde h es la función que da el esperado beneficio por la venta

- Esa función es aprendida como

$$h(g, p, t+1) = (1-\beta) h(g, p, t) + \beta \text{Prof}(g, p, s)$$

Aprendizaje a beneficiarse de las condiciones del entorno

- Agentes que analizan comportamientos pasados de otros (Caso 2)
 - un comprador escoge a vendedor s^* tal que
$$s^* = \max_{s \in S} \frac{1}{|Q|} \sum_{x \in Q} q(g,s,x) V(g,b, p(g,s), q(g,s,x))$$
donde Q es el conjunto de posibles **niveles de calidad**.
 - Un vendedor s venderá un bien g a precio
$$p^* = \max_{p \in P} (p - c(g,s)) \prod_{s' \in S} \sum_{p'} N(g,b,s,p, s', p')$$
donde N es la función de la **densidad probabilística**

Aprender a jugar mejor contra un oponente

- Dado el conjunto de estados posibles de juego S , una función sucesora σ en S , d es la profundidad de la búsqueda, un modelo del oponente de que jugará $\varphi: S \times S$ viene dado por la función

$$M: f(s) \quad d \leq 0$$

$$M(s, d, f, \varphi) = \max_{s' \in \sigma(s)} f(s') \quad d = 1$$

$$\max_{s' \in \sigma(s)} M(\varphi(s'), d-2, f, \varphi) \quad d > 1$$

- Otro enfoque basado en la máxima utilidad: suponer que se debe escoger del conjunto de posibilidades $\alpha = \{\alpha_1, \dots, \alpha_i\}$ y las del oponente $\beta = \{\beta_1, \dots, \beta_n\}$ y la utilidad del movimiento es $u(\alpha_i, \beta_i)$

$$MEU = \max_{\alpha_i \in \alpha} \sum_{\beta_j \in \beta} p(\beta_j / \alpha_i) u(\alpha_i, \beta_j)$$

Donde $p(\beta_j / \alpha_i)$ es la prob. Cond. que oponente escoja β_j dado que agente escogió α_i

Aprendizaje y Comunicaciones

- Aprender a Comunicarse
- Comunicación como aprendizaje
- Temas:
 - ¿ Qué comunicar?
 - ¿ Cuando comunicar?
 - ¿ Con quien comunicarse?
 - ¿ Cómo comunicarse?

Comunicación como aprendizaje

- ¿ Qué tan rápido se hayan los resultados del aprendizaje con comunicación?
- ¿ Los resultados de la comunicación son mejores o no con la comunicación?
- ¿ Qué tan complejo es el proceso de aprendizaje con la comunicación?
- *Bajo nivel* de comunicación: se intercambian piezas de información en simples consultas
 - Data Sensada (percepción), Decisiones y Políticas (por ejemplo valor de $Q(s,a)$ en el caso del agente y Q-learning para decidir que hacer)
- *Alto nivel*: hay negociaciones y explicaciones sintetizando información.
 - Mas complejo (caso humano)

Aprender a Comunicarse

Por ejemplo. para decidir quién hace qué

- Una tarea se especifica como $T_i = \{A_{i1}V_{i1}, \dots, A_{im}, V_{im}\}$ donde A_{ij} es un atributo de la tarea y V_{ij} su valor

$$\text{SIMILAR}(T_i, T_j) = \sum_r \sum_s \text{DIST}(A_{ir}, A_{js}) \text{ y}$$
$$\text{DIST}(A_{ir}, A_{js}) = \text{SIMIL_ATR}(A_{ir}, A_{js}) \text{ SIMIL_VA}(A_{ir}, A_{js})$$

- Conjunto de similares tareas a T_i ($S(T_i)$)

$$S(T_i) = \{T_j; \text{SIMILAR}(T_i, T_j) > 0.85\}$$

- Agentes no difunden hacer una tarea, sino que preselecciona c/agente según lo que hace usando la expresión

$$\text{SUIT}(M, T_i) = 1 / |S(T_i)| \sum_{T_j \in S(T_i)} \text{PERFORM}(M, T_j)$$

Donde $\text{PERFORM}(M, T_j)$ indica que tan bueno agente M ha hecho tarea T_j en el pasado

Comunicación de Alto nivel como aprendizaje

- Lenguaje comunicación (ejemplo):
 - Hipótesis: Introducir(h), propone(h,c), Negar(h) donde c es la confianza en dicha hipótesis
 - Evaluación Hipótesis: Confirmar(h,c), Desacuerdo(h,c), SInOpinion(h.c), Modifica(h,g,c)-genera una versión modificada de h
 - Modificación status hipótesis: actualiza(h,t)-cambia status de estar de acuerdo con el valor de confianza t, acepta(h)
t= soporte(h)(1-contra(h)) donde

- Ejemplo: 3 agentes con tres hipótesis

1. A1: propone(h1, 0.6)
2. A2: modifica(h1, h2, 0.5)
3. A3: modifica(h1,h3, 0.55)
4. A1: modifica (h2,h3,0.45)
5. A3: confirma(h2, 0.6)
6. A3: confirma(h3, 0.6)
7. A2: confirma(h3, 0.5)
8. A3: niega(h1)
9. A2: niega(h2)
10. A1: Actualiza(h3, 0.8)
11. A3: acepta(h3)
12. A2: acepta(h3)

Aprendizaje en equipo simultaneo

Múltiples procesos de aprendizaje tratan de mejorar las partes del equipo.

- Normalmente, cada agente tiene su proceso de aprendizaje propio y único para modificar su comportamiento.
- Hay diferentes grados de granularidad: el equipo se puede dividir en "Escuadrones", cada uno con su propio proceso de aprendizaje
- Desafío central: cada alumno adapta sus comportamientos en el contexto de otros que se van co-adaptando también y no hay ningún control de ese proceso.
- El problema es que a medida que los agentes aprenden, modifican sus comportamientos, a su vez pueden arruinar las conductas aprendidas de otros agentes, haciendo obsoletos los supuestos en los que se basan

Aprendizaje en equipo simultaneo

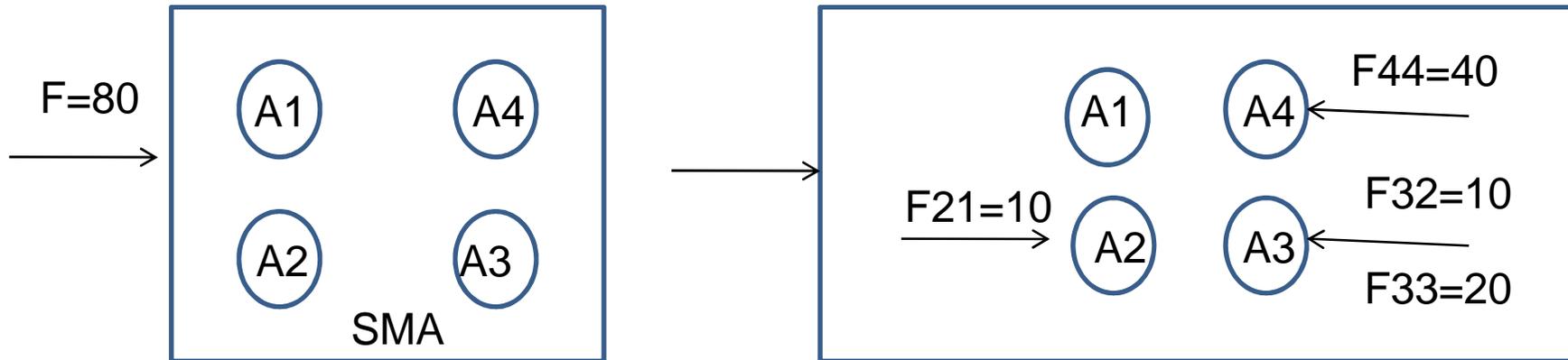
- Un enfoque simplista para hacer frente a la co-adaptación es ver a los otros alumnos como parte de un ambiente dinámico al cual alumno debe adaptarse.
- El problema es que la adaptación de los agentes al medio ambiente puede cambiar el propio medio ambiente. Esto es una violación significativa de los supuestos básicos de las tradicionales técnicas de aprendizaje automático.
- Tres ejes principales de investigación
 - **Problema de asignación de créditos**, que se ocupa de cómo repartir el premio obtenido en un equipo de alto nivel a los alumnos individuales.
 - problemas en la dinámica del aprendizaje: pretende entender el impacto de la co-adaptación en los procesos de aprendizaje.
 - **modelado de otros agentes** con el fin de mejorar las interacciones (y colaboración) con ellos

Problema de Asignación de Crédito

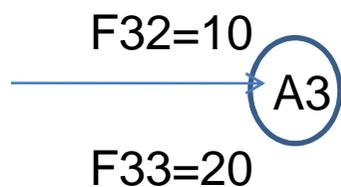
- **Objetivo:** asignación de la retroalimentación (crédito o penalización) para c/elemento del sistema
 - Inter-agente: asignación de la retroalimentación a c/u de los agentes (qué acción de qué agente contribuye a mejorar que rendimiento?)
 - Intra-agente: asignación de retroalimentación a los componentes internos de las acciones de los agentes (qué conocimiento, inferencia o decisión de la acción del agente mejora el rendimiento del sistema)
- **Típico problema de aprendizaje en SMA que ataca a ambos niveles del sistema con técnicas distintas**

Problema de Asignación de Crédito

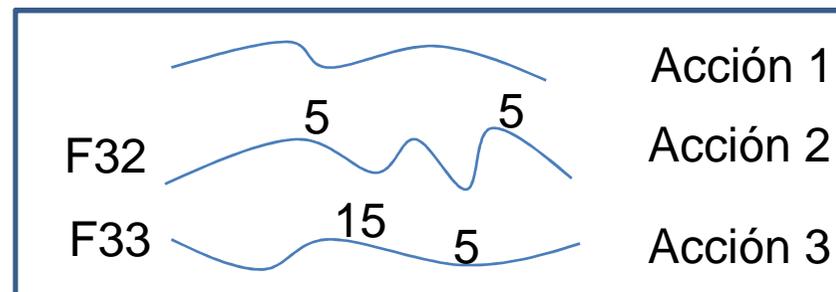
- Inter-agente:**



- Intra-agente:**



A3



Ejemplos de Mecanismo de Aprendizaje en SMA

Aprendizaje Distribuido: SMILE

- El protocolo de SMILE (Sound Multiagent Incremental LEarning) supone que cada agente puede aprender de manera incremental en función de la información que va obteniendo.

Agentes Sapientes

- aprendizaje incremental se refiere a la capacidad para inferir relaciones causa-efecto a partir de la experiencia.

Por ejemplo, aprendizaje por inducción

- Cada agente debe contar con un mecanismo de revisión de creencias que mantenga la consistencia en toda actualización realizada a la base de creencias de un agente.
- El aprendizaje surge mientras se llevan a cabo una serie de interacciones entre los agentes del SMA que tienen en común ciertas creencias, con el propósito de mantener la consistencia global.

Aprendizaje Distribuido: SMILE

- Un agente r es capaz de actualizar su estado B para mantener la consistencia después de que cierta información K ha sido percibida (a-consistencia).
- Existe un conjunto BC de creencias que es común a los agentes del SMA. Si el agente r_i actualiza la parte común BC entonces cada agente debe actualizar su estado para mantener la sma-consistencia).
- Con la sma-consistencia un agente juega el rol de aprendiz, e implica una comunicación con otro agente que toma el papel de crítico.
- Los agentes no son especializados y pueden tomar roles diferentes y la información esta distribuida entre los agentes sin una memoria centralizada, ni de agentes que juegan un papel particular.

Aprendizaje Distribuido: SMILE

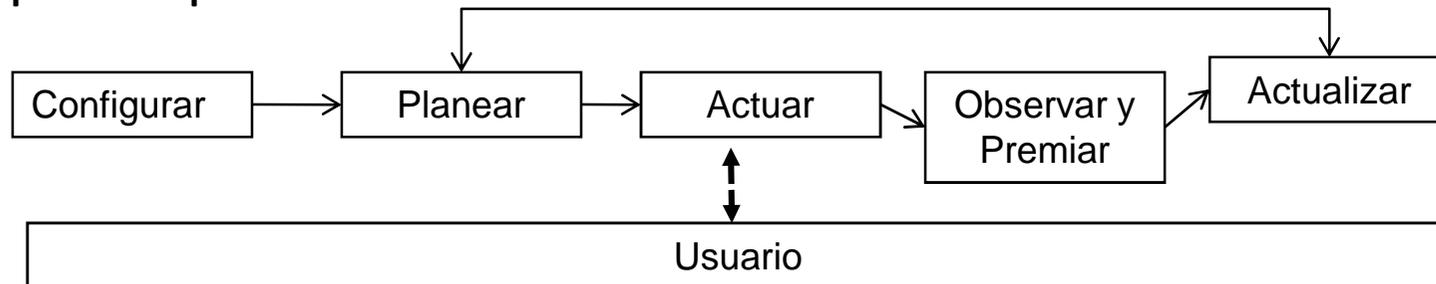
- Idea general es que para que un mecanismo M_s sea sma-consistente, se supone que exista una interacción entre el agente aprendiz r_i y los otros agentes.
- Con el propósito de lograr la sma-consistencia de M_s , el agente aprendiz r_i tiene un mecanismo interno M encargado de mantener la consistencia del agente. El proceso es:
 - El mecanismo M_s es disparado por un agente r_i que recibe una información k (necesita actualización de estado para restablecer su a-consistencia).
 - El estado del agente aprendiz se actualiza a $M(B_i)$;
 - B_0C es la parte común modificada por el agente r_i ; y B_0j el estado de cualquier otro agente r_j inducido por la modificación de B_0C .
 - Se da una interacción $I(r_i; r_j)$ entre el agente aprendiz r_i y el agente r_j jugando el papel de crítico,
 - El mecanismo M_s termina hasta que ningún agente puede proporcionar mas información k_0 (la sma-consistencia del agente aprendiz r_j es restaurada y todos los agentes adoptan la actualización de estado BC).

Aprendizaje Distribuido: SMILE

- Dos situaciones por las cuales un agente podrá considerar comunicarse mientras se encuentra en el proceso de aprendizaje:
 - Cuando el agente no es capaz de iniciar el proceso de aprendizaje, e.g., no tiene suficientes ejemplos;
 - El agente no puede hallar una hipótesis para explicar la falla del plan en cuestión.
- En ambos casos el agente aprendiz puede preguntar a los demás agentes en el MAS por mas ejemplos de entrenamiento.
- Bosquejo en el caso de agentes con planes (BDI):
 - El elemento de aprendizaje retroalimenta al sistema (función inversa), utilizando las creencias relevantes al plan ejecutado que fallo.
 - La función es planteada como un meta-plan que recolecta información sobre los planes ejecutados y detecta fallo de planes
- Como resultado, el agente BDI, aprende de forma incremental

Interactive Artificial Learning (IAL).

- Un usuario no experto interactúa con una máquina de aprendizaje para ayudar al agente a aprender comportamientos exitosos autónomos que satisfacen las necesidades y los objetivos del usuario
- Ejemplo etapas en IAL



- Primer paso: el algoritmo se configura.
- En los pasos del dos al cinco, el algoritmo varias veces planea su comportamiento, realiza acciones, observa las consecuencias, y actualizaciones sus representaciones internas según su experiencias.
- En IAL, los agentes pueden potencialmente interactuar con los humanos en cualquier etapa del proceso.

Interactive Artificial Learning (IAL).

- Este enfoque tiene éxito cuando el usuario final conoce el comportamiento autónomo deseado y el dominio se entiende bien.
- Las investigaciones actuales se centran en el desarrollo de algoritmos IAL y tecnologías de interfaz que permiten a los usuarios colaborar con el aprendizaje de agentes en todo el proceso de aprendizaje.
- Algunas áreas de aplicación:
 - redes eléctricas inteligentes,
 - programación de tareas dinámicas en sistemas de transporte
 - robótica asistencial para el tratamiento del autismo.
- Algunos métodos
 - **Aprendizaje por demostración** (LBD): comportamiento autónomo de un agente se deriva mediante la observación de las acciones de los usuario,
 - **Aprendizaje evolutivo interactivo**, en donde la intervención humana se le proporciona a un algoritmo genético.

Interactive Artificial Learning (IAL).

- J. W. Crandall, M. A. Goodrich, and L. Lin. Encoding Intelligent Agents for Uncertain, Unknown, and Dynamic Tasks: From Programming to Interactive Artificial Learning. AAI Spring Symposium: Agents that Learn from Human Teachers, 2009.
- B. Argall, B. Browning, and M. Veloso, Learning by demonstration with critique of a human teacher, In Proceedings of the Second ACM/IEEE International Conference on Human Robot Interaction, 2007.
- A. L. Thomaz, and C. Breazeal, Teachable robots: Understanding human teaching behavior to build more effective robot learners, Artificial Intelligence, 2008.
- J. W. Crandall, M. H. Altkrori, and Y. M. Hassan, Learning by Demonstration in Repeated Stochastic Games, Submitted to European Conference on Machine Learning, 2010.
- G. Dozier, B. Carnahan, C. Seals, L. Kuntz , S. Fu, An Interactive Distributed Evolutionary Algorithm (IDEA) for Design, IEEE International Conference on Systems, Man and Cybernetics , 2005.
- W. Ma, K. Wang, On-Line Taxi Problem on the Benefit-Cost Graphs, In Proceedings of the Fifth International Conference on Machine Learning and Cybernetics, 2006.
- A. Duquette, F. Michaud, and H. Mercier, Exploring the use of a mobile robot as an imitation agent with children with low-functioning autism, Autonomous Robots, Vol. 24, No. 2, 2008.
- M. B. Colten et. Al, Toward therapist-in-the-loop assistive robotics for children with autism and specific language impairment, In Proceedings of Adaptive & Emergent Behavior & Complex Systems, 2009.

La teoría CLRI

Método formal de aprendizaje de agentes: determina cómo el aprendizaje de un agente afecta el aprendizaje de otros agentes.

- Se supone un sistema en el que cada agente tiene una función de decisión que rige su comportamiento, y una función objetivo que describe el mejor comportamiento posible del agente.

La función objetivo es desconocido por el agente.

- Parámetros de un número inmenso de algoritmos de aprendizaje: tasa de Cambio, tasa de **A**prendizaje, tasa de **R**etención y tasa de **I**mpacto
- El objetivo del aprendizaje del agente es tener su función de decisión como un duplicado exacto de su función objetivo

la función objetivo va cambiando como resultado del aprendizaje.

- Se suponen N agentes, con un mundo visto como un conjunto de estados discretos $w \in W$ que se le presentan al agente según una probabilidad con distribución $D(W)$.
- Cada agente i tiene un conjunto de posibles acciones A_i $|A_i| > 2$.
- En cada tiempo t todos los agentes se les presenta una nueva w , toman una acción simultáneamente, y reciben algún pago.

La teoría CLRI

- Comportamiento de cada agente i se define por una función de decisión $d(i, t, w)$:
- En cualquier momento t hay una función óptima de i dado por su función objetivo $O(t, i, w)$.
- Algoritmo de aprendizaje trata de reducir la discrepancia entre d y O usando pagos que recibe por cada acción.
- Como otros agentes aprenden y cambian su función de decisión, la función objetivo de i también va a cambiar,

