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

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

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

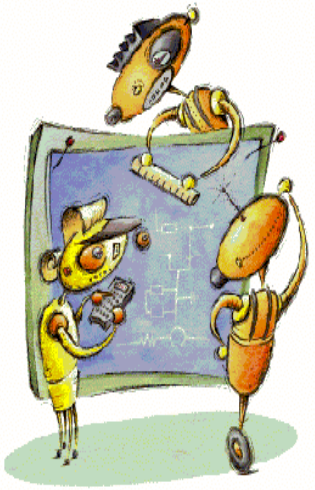
¿Cuál es el objetivo en el aprendizaje multi-agente?

- **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 en Sistemas Multi-Agente

Aprendizaje colectivo. Es el aprendizaje que se lleva a cabo por los agentes como grupo, e. g., mediante el intercambio de conocimientos o la observación de otros agentes



Algunas técnicas del aprendizaje colectivo:

- Aprendizaje reactivo
- Aprendizaje basado en la lógica
- Aprendizaje social

- Comportamiento contagioso
- Seguir conducta o comportamiento
- Aprendizaje por observación

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
Según Boesch basada en sociedades de chimpancés

- **Individualismo:** sólo requiere un chimpancé para atrapar a su presa.

Esta estrategia es para entornos en los que depredadores tienen la ventaja.

- **Similitud:** varios individuos realizan las mismas acciones para cazar a sus presas, pero sin coordinación aparente.
- **Sincronía:** similar a la similitud, pero con la adición de la coordinación temporal entre los cazadores.
- **Coordinación:** los cazadores se coordinan en el espacio, además de sincronía temporal, para sincronizar posiciones y velocidades entre ellos.
- **Colaboración:** es la más compleja estrategia y tiene un mayor nivel de cooperación, ya que requiere especialización de roles.

Esta estrategia es para entornos en los que los depredadores tienen la desventaja.

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.

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, se podría aprender 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
Caracreristíca de las interacciones	Nivel, frecuencia, persistencia, patrón y variabilidad
Compromiso de los Agentes	Relevancia y rol
Caracteristícas 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.

Clasificación del Aprendizaje

Aprendizaje en equipo

un solo alumno que decide el comportamiento de un equipo de agentes. Se puede categorizar en dos modelos.

- **Modelo homogéneo**: todos los agentes reciben el mismo comportamiento del alumno, incluso si los agentes no son idénticos.
- **Modelo heterogéneo**: proporciona a cada agente un comportamiento único, que permite la especialización de cada uno.
- **Modelo híbrido** combina ambos aprendizajes, formando escuadrones especializados cuyos miembros tienen el mismo comportamiento.

Clasificación del Aprendizaje

Aprendizaje concurrente

- **múltiples procesos de aprendizaje** que se ejecutan en paralelo, típicamente uno para cada agente del sistema.
- Esta estrategia es útil para **problemas en los que la descomposición es posible** y ventajoso, y cuando es posible centrarse en cada sub-problema independientemente,
- Con el fin de obtener la cooperación real, **cada agente debe aprender y converger a un comportamiento que mejor se adapta** al comportamiento aprendido por los otros agentes.
- Esto implica que **el comportamiento de cada agente está cambiando constantemente** al tratar de adaptarse a otros agentes, que serán diferentes en cada instante de tiempo, lo que le obliga a cambiar su comportamiento nuevamente.
- Esta segmentación dinámica **puede conducir el proceso de aprendizaje a no converger a un comportamiento óptimo** del grupo, que es la razón principal por la que en la literatura existen diferentes consideraciones para abordar el problema.

Aprendizaje por refuerzo, aprendizaje dinámico, aprendizaje cooperativo

Clasificación 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

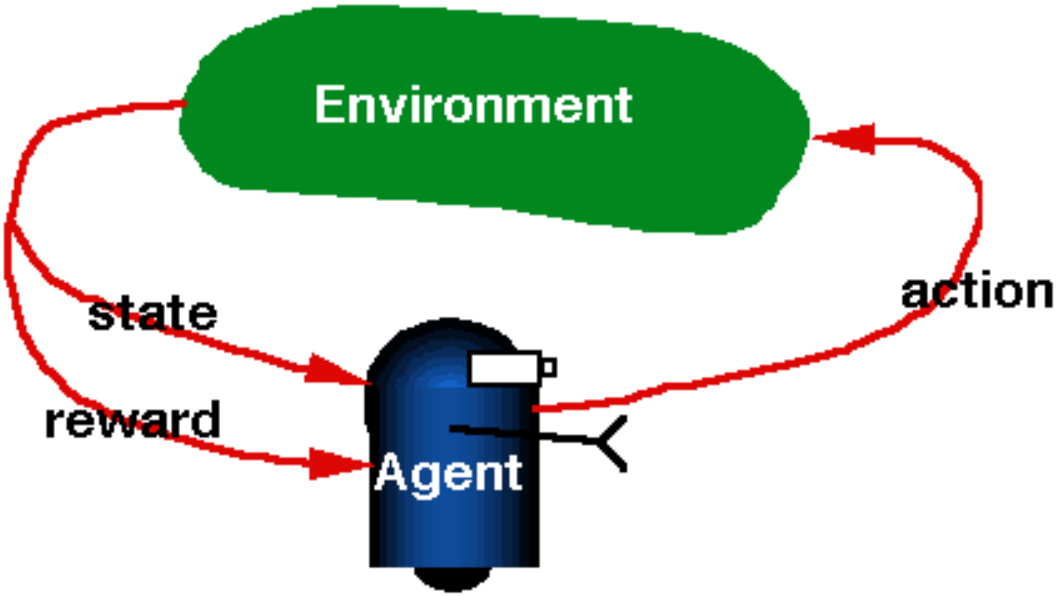
(

Basados en la Teoría de Juegos

Postulado Fundamental de la Teoría de Juegos: "racionalidad"

Un agente racional toma decisiones que maximizen su utilidad

- .



Teoría de Juegos

- Juego especificado por matriz de pago: jugadores (1…N), acciones,)

acción A

Acción B

	R	P	S
R	0	-1	$+1$
P	$+1$	0	-1
S	-1	$+1$	0

pago A

	R	P	S
R	0	$+1$	-1
P	-1	0	$+1$
S	$+1$	-1	0

pago B

- Si matrices son identicas, juego es cooperativo, sino no-cooperativo (sum-cero = puro competitivo)

Vent and Desv TJ

- Provee unas bases para el proceso de aprendizaje en SMA.

- Ya que:
  - Juegos son estacionarioros y bien especificado; **X**
  - Hay poder de calculo en los PV; **X**
  - Se puede asumir otros agentes la usan; **X**
  - Resuelve pb. coordinación. **X**

- Esas condiciones raramente estan en situaciones reales

Aprendizaje SMA

Basica idea: agente se adapta, ignorando no-estationariedad de las estrategias de los agentes

Juego Fictioso: Agente observa frecuencia de escogencias de acciones de los otros agentes en el tiempo:

$$prob(action\ k) = \frac{\#times\ k\ observed}{total\ \ \#\ observations}$$

Agente juega mejor (mas usado)

Aprendizaje SMA

TJ Evolucionaria: muchos agentes usando diferentes estrategias.

- x= vector de población de estrategias,
- x_k = fracción de pobl. Jugando estrategia k. Evolución:

$$\frac{dx_k}{dt} = x_k\left(u(e^k, x) - u(x, x)\right)$$

- Se pueden llegar a puntos atractores (estables)

Iterativo Gradiente Ascendente: (Singh, Kearns and Mansour): adaptación a estrategias de otros jugadores.

$$\frac{dx_i}{dt} = \varepsilon \frac{\partial}{\partial x_i}\left(u(x_i, x_{-i})\right)$$

- u is lineal entre x_i y x_{-i}

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 s1 a s2 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:

$$V(T, \gamma) = E(\Sigma_t \gamma(t) r(T, s, t)) \qquad \gamma : \text{tasa descuento}$$

Aprendizaje y Coordinación

- *Q-learning:* se escoge acción a en estado s tal que se **maximice la recompensa**
  - V(T, γ, s)= max $a \in A$ Q(T, γ, s, a) para todo s $\in$S
  - Q ahora es (β es la tasa de aprendizaje y R reforzamiento):

$$Q(s,a)=(1-\beta)Q(s,a)+ \beta(R+\gamma \max_{a' \in A} Q'(s', a'))$$

Si una acción a en estado s produce una transición a s',

- *Sistema de Clasificación*
  - **Sistema basado en reglas**, cada regla es (ci,ai):

condición ci genera acción ai

  - Aprendizaje usando AG (reglas), donde S(ci,ai) calidad de la regla i en el tiempo t (R viene dado por la recompensa por la acción ai)

$$S(ci,ai, t+1)=(1- \beta) S(t,ci,ai)+ \beta(R+ S(t+1,ci,ai))$$

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 para juegos:*** 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 funcion-Q en una combinación lineal de Q-funciones** de agentes locales:

$$Q(s, a) = \sum_{i=1,n} Qi(si, ai).$$

- Cada función Q-local (Qi) de un agente i se basa en si y ai 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 a' vecinos en el grafo)**

$$Qi(si, ai) := Qi(si, ai) + \alpha[R(s, a) + \gamma \max_{a'} Q(s', a') - Q(s, a)].$$

Juegos estocásticos

- Para tareas de aprendizaje concurrentes con múltiples agentes,
- Modela las distribuciones de probabilidad de las acciones de cada agente para varios estados.

extensión de los procesos de decisión de Markov y la teoría de juegos de matriz.

- Un juego estocástico es una tupla (N, S, A1 ... n, T, R1 ... n), donde n es el número de agentes, S es el conjunto de estados, Ai es el conjunto de acciones a disposición del agente i, T es la función de transición, y Ri es la función de recompensa del agente i.

El objetivo es encontrar una política π determinista o estocástico : S → Ai, que asigna los estados a una distribución de probabilidad sobre las acciones del agente de manera que se maximiza la recompensa futura.

Juegos estocásticos

Dos conceptos importantes:

- Para un juego, la función de mejor respuesta para un agente i, BRI (ζ-i), es el conjunto de todas las estrategias que son óptimas dado que los otros agentes utilizan la estrategia conjunta ζ-i
- Una estrategia para el agente i se define como ζi ∈ PD (Ai) y A-i es el conjunto de acciones para todos los agentes, sin agente i.
- Un equilibrio de Nash es un conjunto de estrategias para todos los jugadores ζi, de tal manera que ζi ∈ BRI (ζ-i). Por lo tanto, ningún agente puede obtener una recompensa mayor cambiando su estrategia, a condición de que todos los demás agentes continúan utilizando la estrategia de equilibrio.

Propiedades en un algoritmo de aprendizaje por refuerzo concurrente:

- Racionalidad: Si las políticas de otros agentes convergen, el algoritmo de aprendizaje convergerá a una mejor respuesta.
- Convergencia: El agente de aprendizaje necesariamente converge a una política π.

Aprendizaje y Coordinación

Aprendizaje reforzado Interactivo

- **Estimación de acción:**
  1. Dada Si percepciones del estado actual S,
  2. C/agente ai calcula el grupo de acciones Ai(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 SMA

Variando tasas de aprendizaje

WoLF: “Win or Learn Fast” (Bowling): agente reduce su tasa de aprendizaje cuando lo hace bien, y aumenta cuando lo hace mal.

- tiene las propiedades de racionalidad y convergencia
- dos ritmos de aprendizaje $\delta l > \delta w$, dependiendo de si el agente está ganando o perdiendo
- Eso se determina comparando la recompensa esperada utilizando la política actual contra la recompensa esperada de la política actual promedio.
- Si el valor esperado actual es menor, el agente está perdiendo.
- El objetivo de heurística es aprender rápidamente a escapar de situaciones peligrosas cuando el agente está perdiendo, y estimular la convergencia cuando el agente está utilizando una buena política,

Aprendizaje SMA

Variando tasas de aprendizaje

- **Multi-timescale** Q-Learning (Leslie): diferentes agentes usan diferentes leyes para descender tasa de aprendizaje

"Enseñando Estrategias": reconoce que otros jugadores tienen estrategias adaptativas

"agente puede usar una estrategia que no sea optima (caso del dilema de prisionero) esperando asi inducir al otro jugador adaptar su estrategia en el futuro",

Aprendizaje de roles

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

$$Pr(s, r)= f(U(r,s),P(r,s),C(r,s)/$$

$$\sum_{j \in Rk} (f(U(r,s),P(j,s),C(j,s))$$

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

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

$$\mathbf{max}\ j \in Rk\ (f(U(j,s),P(j,s),C(j,s))$$

y U,P, C son actualizados:

$U(r,sn+1)=(1- \beta)\ U(r,s,n)+\beta Uf$ U edo. Final

$P(r,sn+1)=(1- \beta)\ P(r,s,n)+\beta Pf$ Pf=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 es 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

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

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

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

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

  - Esa función es aprendida como

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

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 φ: SxS viene dado por la función M(s,d,f, φ):

$$M = f(s) \qquad d<=0$$

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

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

- Otro enfoque basado en la máxima utilidad: suponer que se debe escoger del conjunto de posibilidades $\alpha = \{\alpha 1, \ldots \alpha i\}$ y las del oponente $\beta=\{\beta 1, \ldots \beta n\}$ y la utilidad del movimiento es $u(\alpha i, \beta i)$

$$\mathbf{MEU = \max_{\alpha i \in \alpha} \Sigma_{\beta j \in \beta} p(\beta j / \alpha i) u(\alpha i, \beta j)}$$

Donde $p(\beta j / \alpha i)$ es la prob. Cond. que oponente escoja βj dado que agente escogió αi

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 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 Ti={Ai1Vi1, …, Aim,Vim} donde Aij es un atributo de la tarea y Vij su valor

$$SIMILAR\ (Ti, Tj)=\sum_r \sum_s DIST(Air,Ajs)$$

y DIST(Air,Ajs)=SIMIL_ATR(Air,Ajs) SIMIL_VA(Air,Ajs)

- Conjunto de similares tareas a Ti (S(Ti))

$$S(Ti)=\{Tj;\ SIMILAR(Ti,Tj)> 0.85\}$$

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

$$SUIT(M,Ti)=1/|S(Ti)|\ \sum_{Tj \in S(Ti)} PERFORM(M,Tj)$$

Donde PERFORM(M,Tj) indica que tan bueno agente M ha hecho tarea Tj 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,

$$acepta(h, t) = soporte(h)(1-contra(h))$$

- 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: confirma(h3,0.55)
5. A3: confirma(h2, 0.5)
6. A2: confirma(h3, 0.55)
7. A3: niega(h1)
8. A2: niega(h2)
9. A1: acepta(h3)
10. A2: acepta(h3)
11. A1: Actualiza(h3, 0.55)

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 agente adapta sus comportamientos en el contexto de otros que se van también co-adaptando, sin 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 agentes como parte de un ambiente dinámico** al cual agente 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

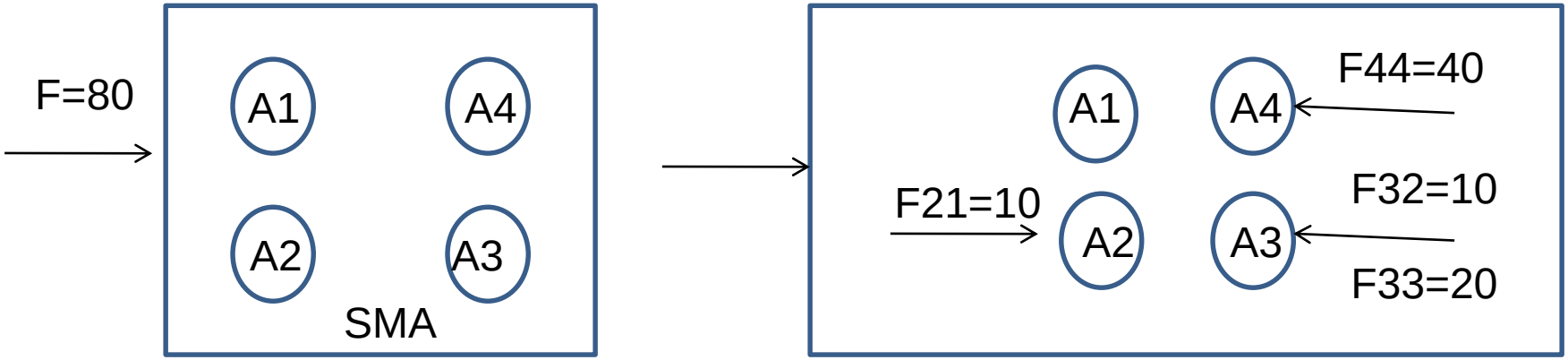
Problema de Asignación de Crédito

- **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 qué 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)

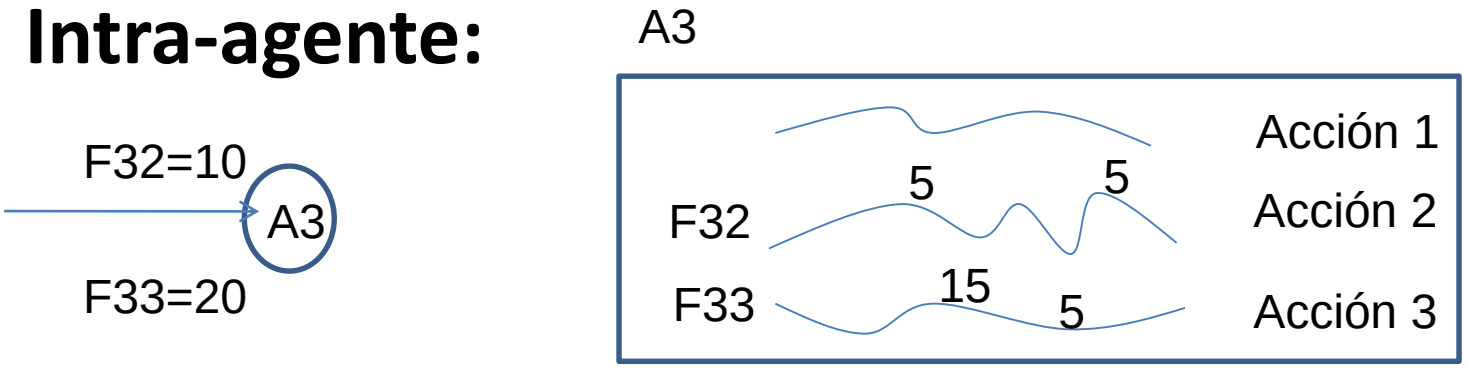
Típico problema de aprendizaje en SMA que ataca a ambos niveles del sistema con técnicas distintas

Problema de Asignación de Crédito

- **Inter-agente:**



- **Intra-agente:**



Aprendizaje Colectivo para Coordinarse en SMA basado en Algoritmos Culturales

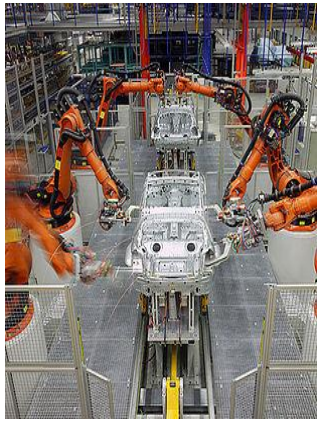
Caso de Estudio: Problema de Integración en Automatización

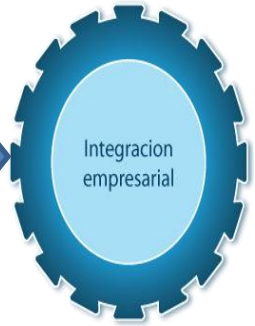
Motivación

Experiencias

Conocimientos

Aprendizaje colectivo





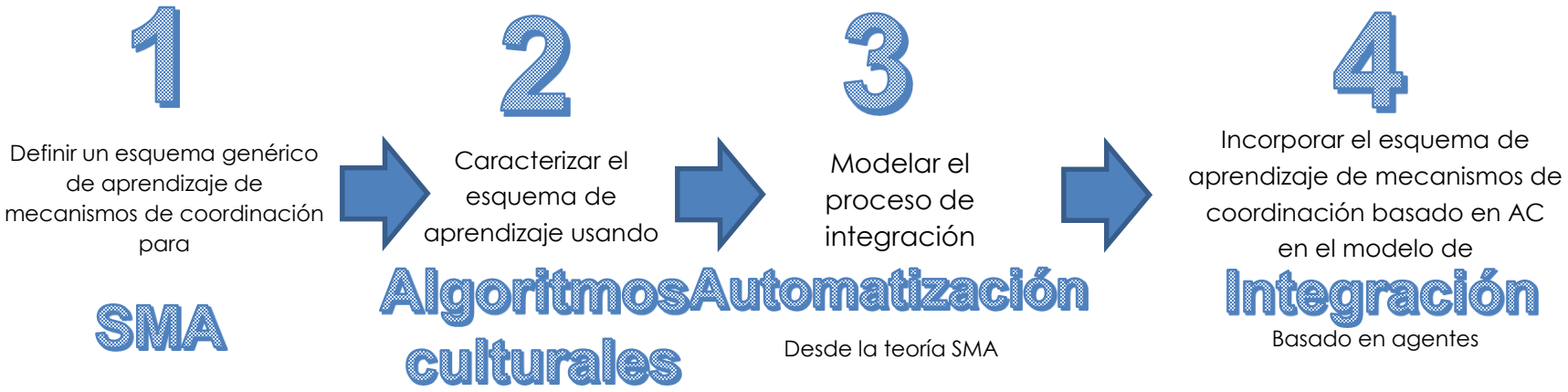
Automatización Industrial basadas en Comunidades de Agentes

Resolver el problema de la integración a través de la coordinación

- Modelando el problema de automatización usando SMA
- Analizando las estrategias de coordinación y comunicación
- Proveyendo capacidad de aprendizaje colectivo
- Usando la Computación Evolutiva

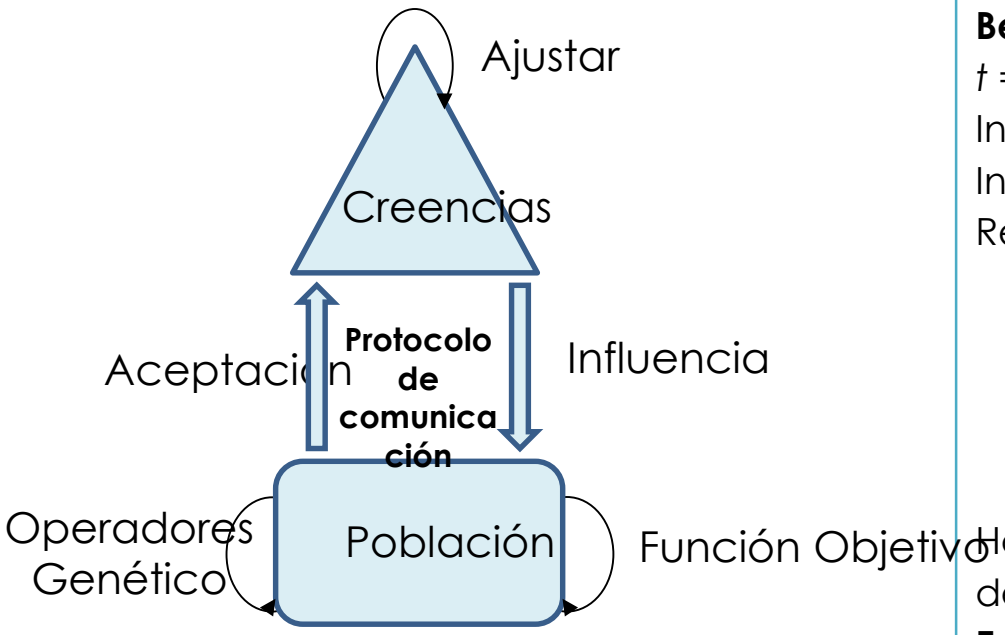
Objetivos general y específicos

Proponer un esquema de aprendizaje de mecanismos de coordinación en una sociedad de agentes (sistema multi-agente) usando algoritmos culturales, y aplicar dicho esquema en el problema de integración en automatización industrial.



Algoritmos culturales

- Algoritmos culturales (AC): es una de las diversas técnicas de la computación evolutiva



Begin
$t = 0$;
Iniciar P^t
Iniciar B^t
Repetir
    Evaluar P^t
    Ajuste (B^t, Acepte (P^t))
    Variación (P^t, influencia (B^t))
    $t = t + 1$;
    Seleccionar P^t de P^{t-1}
Hasta (haber alcanzado la condición de finalizar)
End

Pseudo-código de un AC

Integración en automatización

es un proceso que incluye la producción de datos, la infraestructura de comunicación, los mecanismos de procesamiento de datos, la comunicación interna en cada nivel y la comunicación entre niveles, con el fin de lograr sistemas que permitan ejecutar las diferentes tareas de control y gestión existentes en una empresa

Arquitecturas de integración

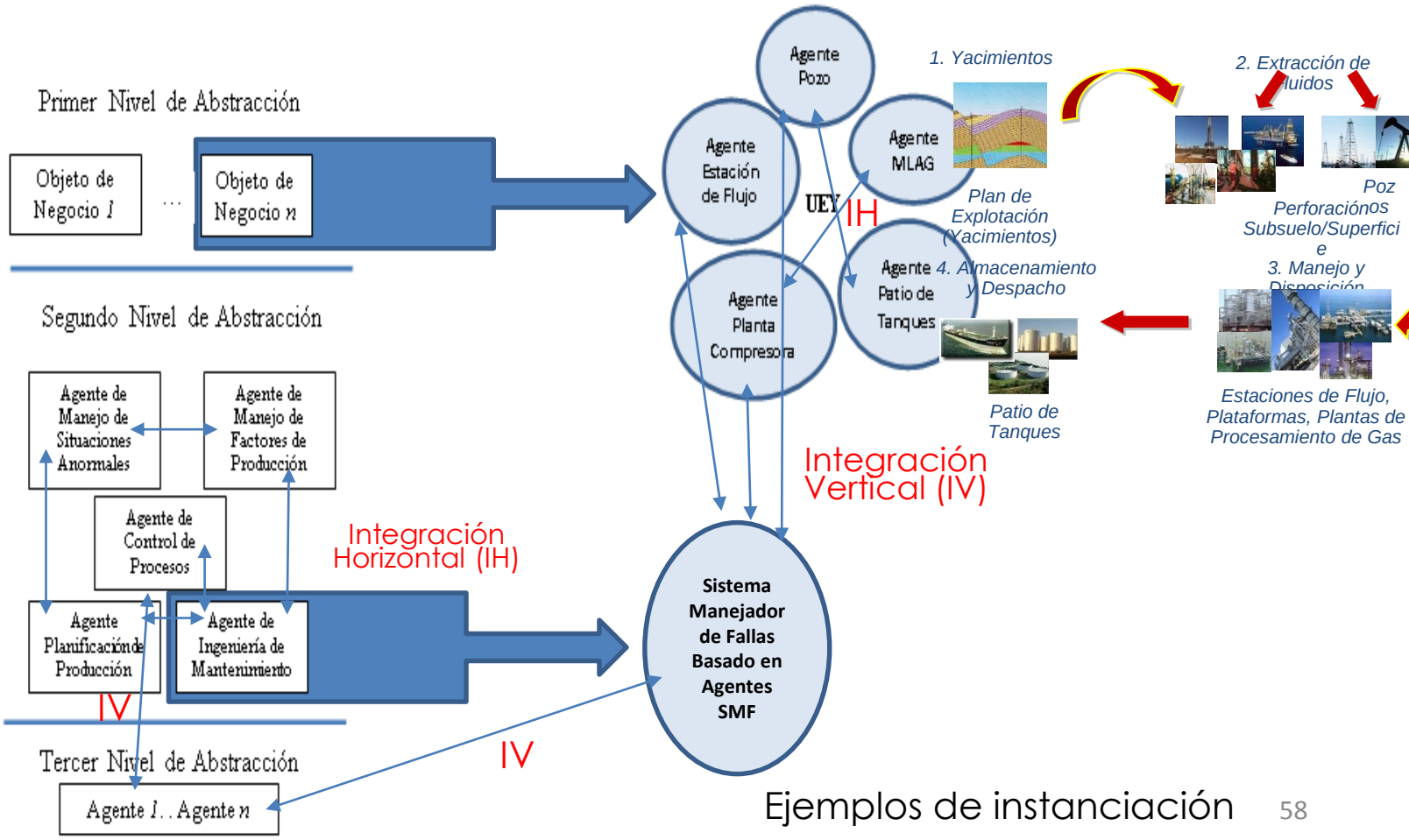
- Datos (DOI)
- Servicios (SOI)

Tipos de integración

- Horizontal (IH)
- Vertical (IV)

Integración en automatización basada en la coordinación de los SMA

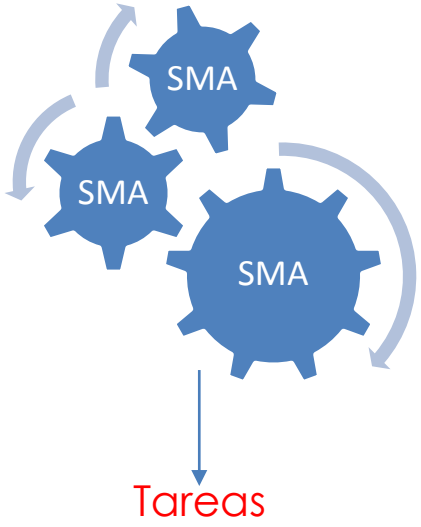
SADIA



Ejemplos de instanciación

Hipótesis:

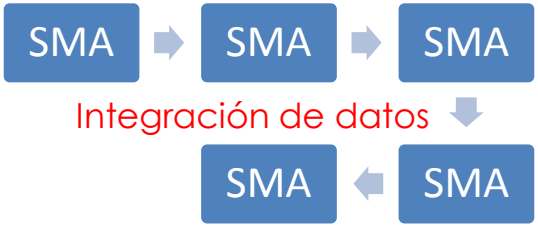
Primera



Los agentes al coordinase en sus diferentes conversaciones, logran integrarse horizontal y verticalmente

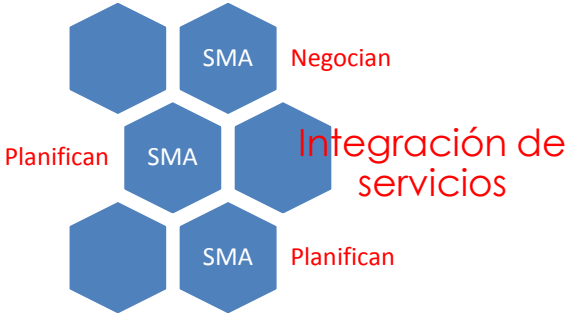
Segunda

Al garantizar la comunicación entre los agentes, permite que los agentes hablen entre si, enviando y recibiendo datos



Tercera

Cualquier modelo de SMA orientado a servicios al obtener los mecanismos adecuados, permite la ejecución de servicios con un mínimo de conflictos



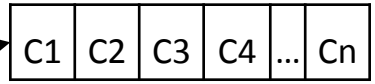
Cuarta

Al trabajar conjuntamente los agentes del sistema automatizado, realizando sus tareas

Bases de la formalización

Muchas Conversaciones y sub-conversaciones

SMA



- SMA
- Conversaciones
- mecanismos coordinación

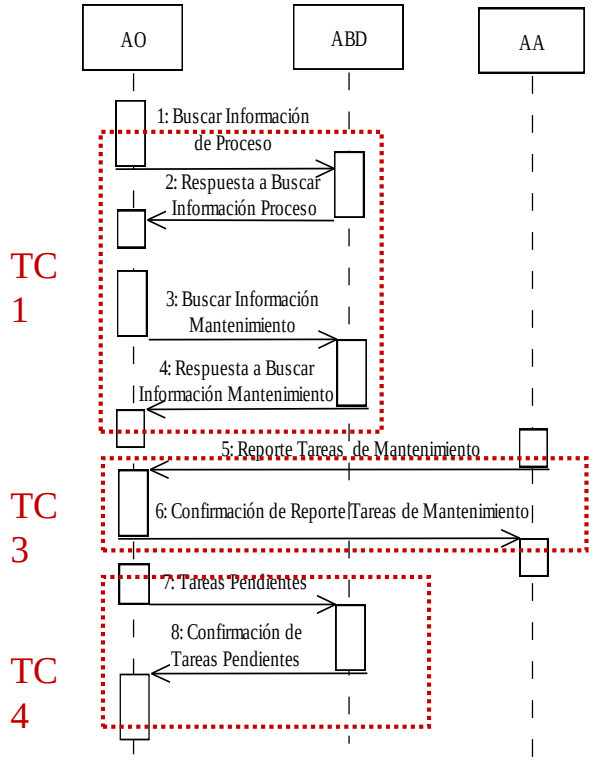
Usa muchos mecanismos de coordinación

Pueden clasificase por tipos de interacción

Objetivo: aprender mecanismo de coordinación

Tipos de interacción (TC)

Tipos de interacción	Protocolos de interacción FIPA
TC1: Consultar	Brokering, Query, Request,
TC2: Asignar	Brokering, Contract Net, English and Dutch Auction
TC3: Informar	Brokering, propose
TC4: Solicitar	Request, Query



Describen que se desea alcanzar

Mecanismo de coordinación

Formalización de los mecanismos de coordinación : **subasta**

$$S = \langle C_0, Of_i^j, \vec{\varepsilon}_i, \alpha_i^j, C_p, C \rangle$$

- C_0 es el precio inicial.
- Of_i^j es una matriz de ofertas
- $\vec{\varepsilon}_i$ es la máxima cantidad que un agente puede ofertar.
- α_i^j especifica una propuesta dada
- $C_P(j, t, x_r)$ es la condición de parada
- C es el precio final del recurso

Formalización de los mecanismos de coordinación:

planificación

$$PL = \langle AP, AA, AE, T, G, P \rangle$$

- $AP, AA\ y\ AE \subset A$ son los conjuntos de agentes que planifican, asignan y ejecutan un plan, o sub-planes, respectivamente.
- T es la matriz de asignación (conformada por agentes y servicios)
- G es el objetivo general,
- P es el conjunto de sub-planes sp_e , $\forall\, e = 1..q$, donde q es el número de sub-planes

Matriz de Planificación Centralizada para Planes Distribuidos

Agentes vs Servicios

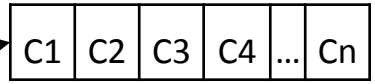
T(PCpD)	S_1	S_2	S_3
AP_1	1	1	0
AE_1	0	0	1
AE_2	0	0	1
AE_3	0	0	1
AE_4	0	0	1
AE_5	0	0	1

- S1 planificación
- S2 Asignación
- S3 Ejecución

Caracterizar los componentes del AC

- Población
- Espacio de creencias
- Protocolo de comunicación

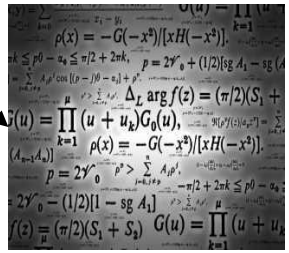
Individuo



Conversaciones y sub-conversaciones

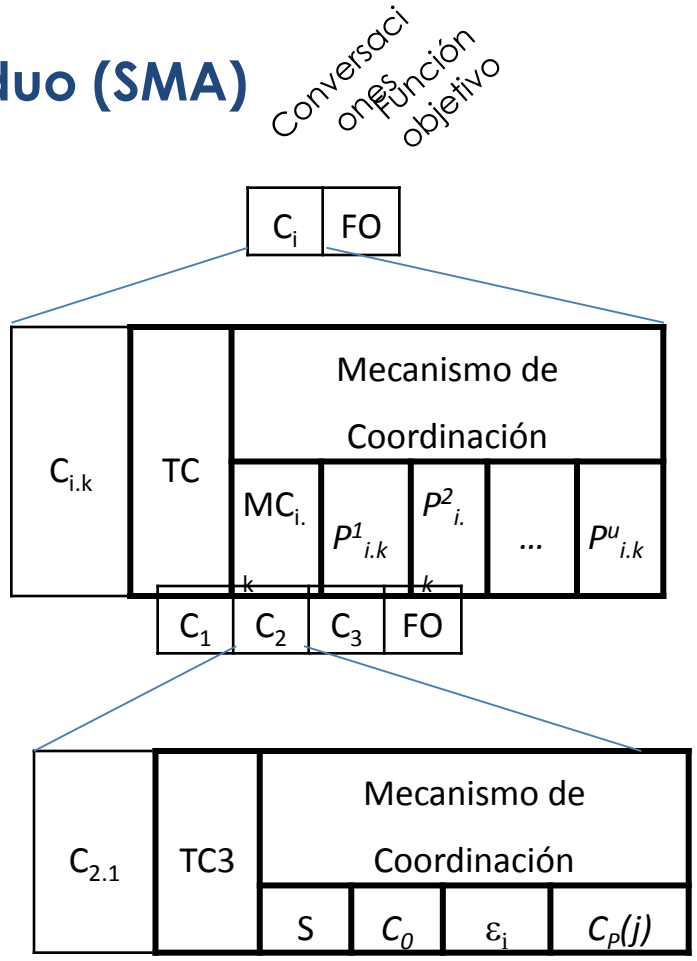
Conocimiento

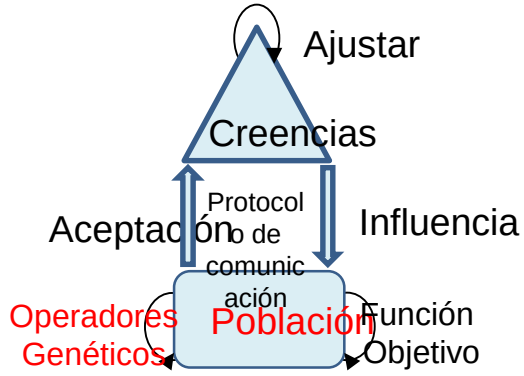
Tipos de conversación y mecanismos de coordinación



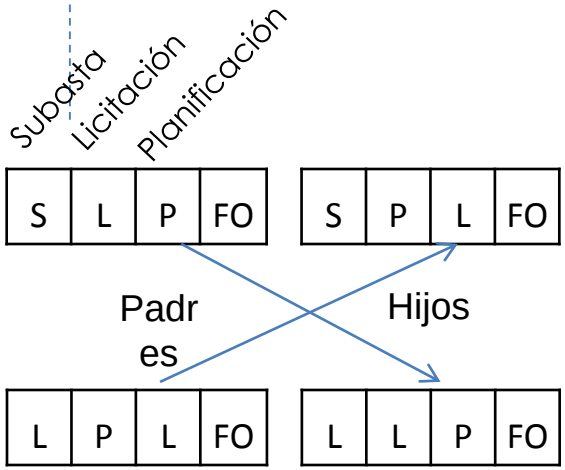
Ecuaciones

Individuo (SMA)

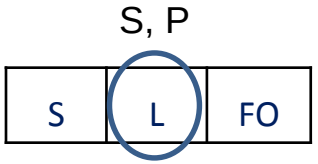




Operadores genéticos



Cruce (punto sencillo)



Mutación

• Función objetivo

Permite evaluar el performance o desempeño de los individuos, esta basada en el costo de procesamiento (CP) y en el costo comunicacional (CC) de cada mecanismo de coordinación usado por el individuo (la idea es minimizar la función objetivo)

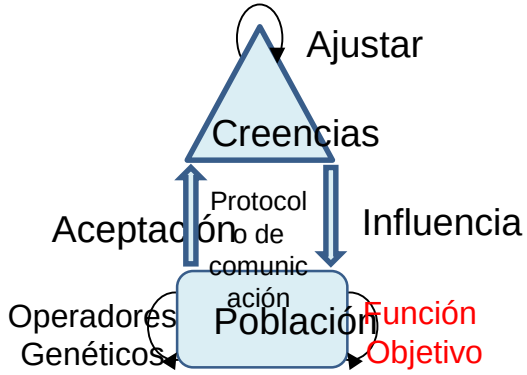
$$FO = \Sigma^{n}_{i=1} \Sigma^{m}_{i\ k=1} (a * CP_{i.k} + b * CC_{i.k})$$

- Donde a y b son constantes definidas por el usuario y permiten normalizar las unidades de la función
- Conversaciones i = {1 … n}
- Sub-conversaciones k = {1 … m_i}

Costo de Procesamiento

$$CP_{i.k} = PI_k + PE_k + \Sigma^{j}_{l=1} \Sigma^{nj}_{q=1} A_{l,q}$$

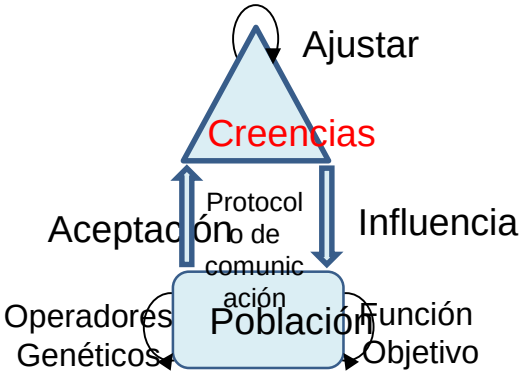
- *PI* fijación del precio inicial; especificación de condiciones en las que se requiere un servicio, generación de planes
- *PE* proceso de selección del agente ganador, asignación de planes
- $A_{l,q}$ tiempo para preparar propuestas, tiempo para generar planes parciales



Costo de Comunicación

$$CC_{i.k} = \Sigma^{j}_{l=1} (\Sigma^{N-1}_{r=1} CEP_{l,r} + \Sigma^{nj}_{s=1} CEO_{l,s}) + \Sigma^{N-1}_{r=1} CS_r$$

- *CEP* costo de envío de propuesta
- *CEO* costo de envío de ofertas, envío de planes parciales
- *CS* costo de informar al ganador, de enviar el plan global



Espacio de creencias

Conocimiento Circunstancial

Ejemplos específicos de eventos importantes, e. g., experiencias exitosas

TC	*MC*	$IO_{(TC, MC, t-1)}$	$TO_{(TC)}$

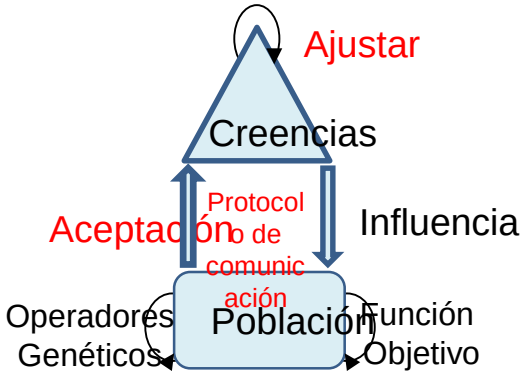
- *IO* es el índice de ocurrencias
- *TO* es el total de ocurrencias

Conocimiento Normativo

Son rangos de valores idóneos de cada una de las variables que integran al mecanismo de coordinación usado por los individuos

MC	P^1		P^2		...	P^u	
	LI	*LS*	*LI*	*LS*		*LI*	*LS*

- *LI, LS* son los límites inferiores y superiores de cada variable P^u



Protocolo de comunicación: función aceptación

Conocimiento Circunstancial

$$IO_{(TC,MC,t)} = IO_{(TC,MC,t-1)} + \left(\frac{NO_{(TC,MC,t)}}{TO_{(TC)}} \right)$$

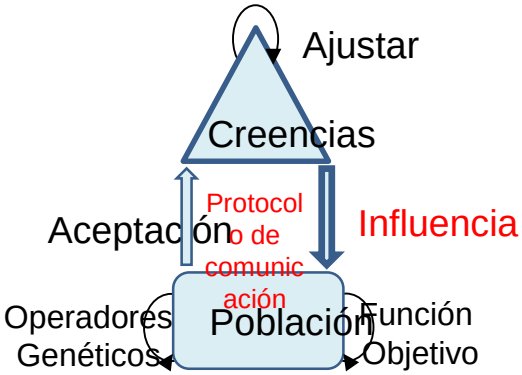
$$TO_{(TC)} = TO_{(TC)} + \sum_{i=1}^{k} NO_{(TC,MC,t)i}$$

TC, tipo de conversación
MC, mecanismo de coordinación
t, estado actual

Conocimiento Normativo

$$Lac(P^u) = \left[\frac{\left(lv + \overline{P} \right)}{2} \right]$$

Lac, limite actual
Lv, limite anterior



Protocolo de comunicación: función influencia

Conocimiento Circunstancial en el Individuo

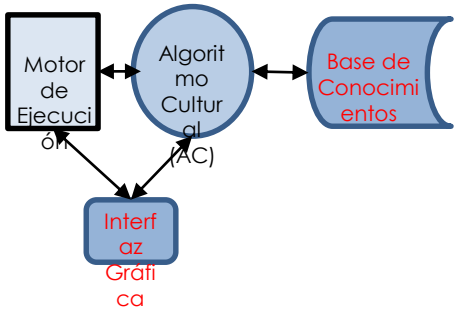
$C_{2.1}$	TC3	Mecanismo de Coordinación			
		SI	C_0	ε_i	$C_P(j)$

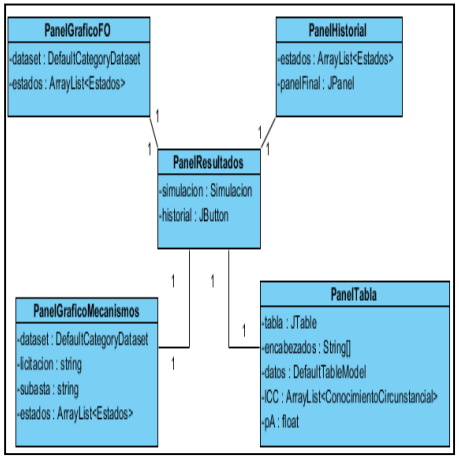
Conocimiento Normativo en el Individuo

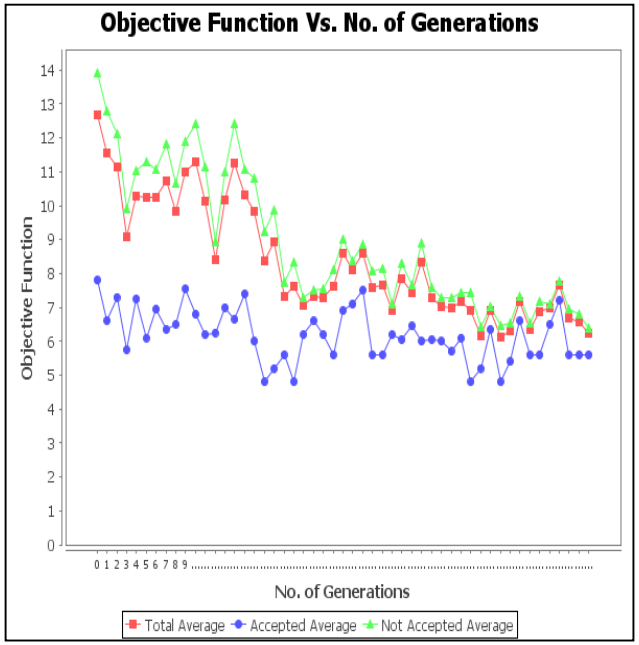
$C_{2.1}$	TC2	Mecanismo de Coordinación			
		SI	C_0 (LI, LS)	ε_i (LI, LS)	$C_P(j)$ (LI, LS)

Mutación dirigida

Arquitectura de CLEMAS (Cultural LEarning for Multi-Agent Systems)







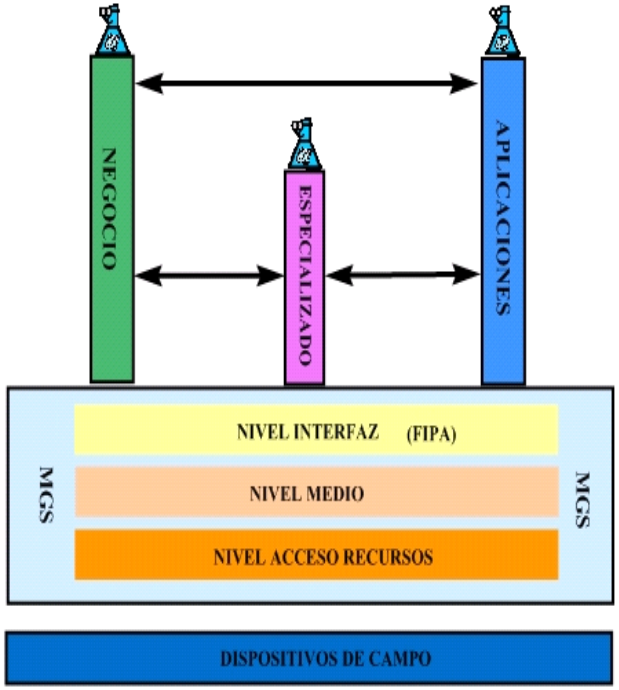
Resultados de simulacion
Licitacion: 120
Subasta Inglesa: 95
Subasta Holandesa: 145

Condiciones Iniciales:
Tamano de la Poblacion:20
Maximo de Generaciones:8
Porcentaje de aceptacion: 20.0%
Probabilidad de Cruce: 0.7%
Probabilidad de Mutacion Directa: 0.5%

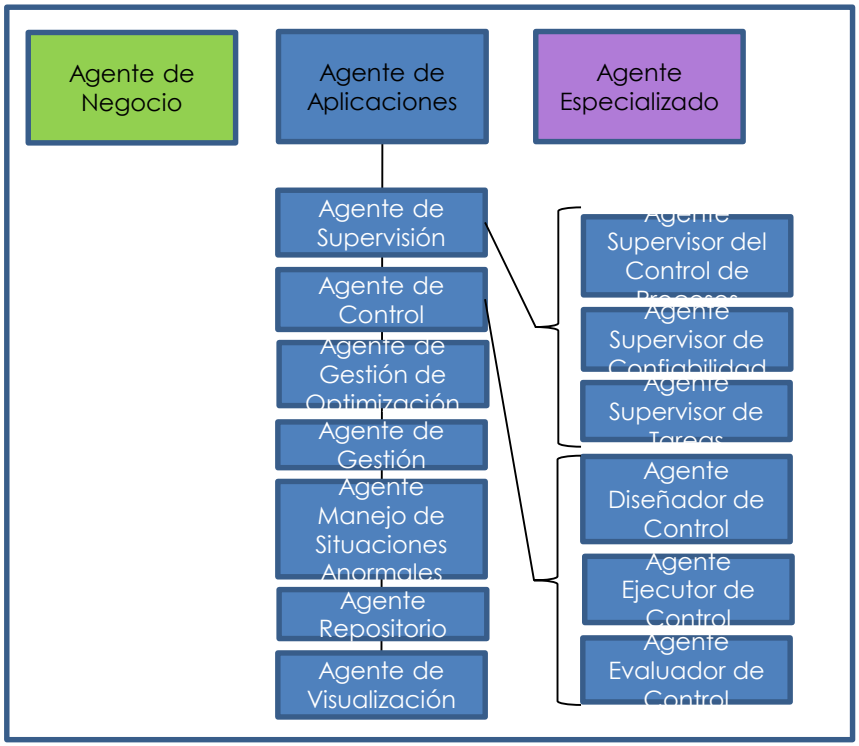
Resultados (En base al 20.0% de la poblacion)

Tipo Conversacion	Licitacion	Subasta Inglesa	Subasta Holandesa
Consulta	53.125%	46.875%	0.0%
Asignacion	0.0%	0.0%	0.0%
Infomacion	15.625%	3.125%	81.25%
Solicitud	0.0%	0.0%	0.0%
Total de Ocurrencia	34.375%	25.0%	40.625%

SMA de una explotación de yacimientos de producción petrolera

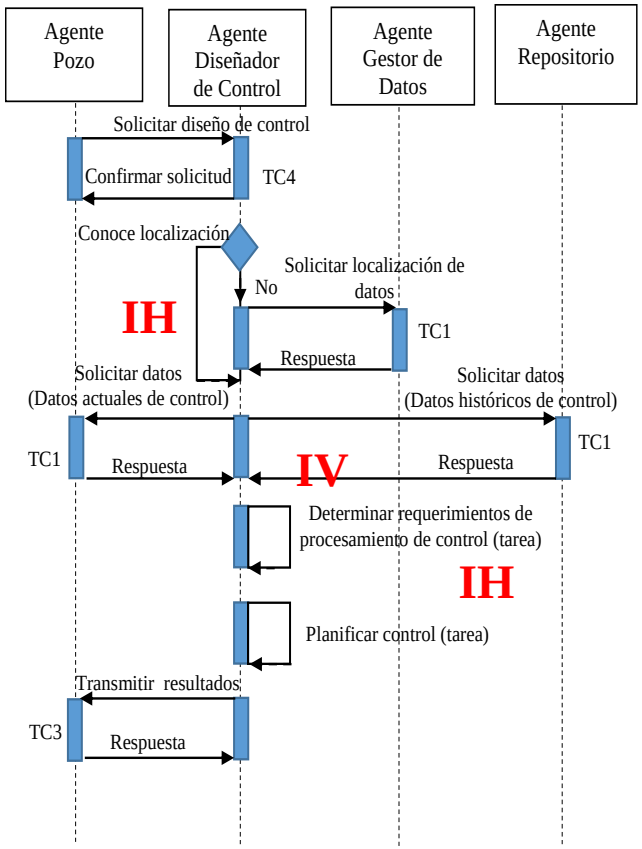


Modelo



SMA

Diseñar el control de un agente pozo



Diseño de experimentos para la verificación de la capacidad de aprendizaje en los sistemas de automatización basados en SMA usando CLEMAS

C_i	FO

Individuo Para i = {1, 2, 3}

- Los valores iniciales de los parámetros de los mecanismos son, para subasta: C_0 = [4...10] ε = [5...20], $C_P(j)$ = [1...5]. Para licitación $M(F)$ = [4...10], y $f(T)$ = [5 ... 20]. Para planificación: AP = 1, AA = 1, AE = 3, $PCpD$, con 6 sub-planes

  - Población: 50 individuos
  - Numero de generaciones: 65
  - Probabilidad de cruce: 0.7
  - Probabilidad de mutación: 0.07

Resultados de los experimentos

TC	Licitación	Subasta Inglesa	Subasta Holandesa	Planificación
Consulta (TC1)	55,36%	5,20%	37,90%	1,53%
Asignación (TC2)	0,0%	0,0%	0,0%	0,0%
Informa (TC3)	30,22%	33,18%	34,80%	1,78%
Solicitud (TC4)	90,39%	5,21%	1,32%	3,07%
Total de ocurrencias	55,01%	12,55%	30,32%	2,12%

En la tabla se aprecia que TC1 usa licitación, TC3 Subasta holandesa y TC4 licitación

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 (consistencia).
- Existe un conjunto BC de creencias que es común a los agentes del SMA.
- Si el agente ri actualiza la parte común BC entonces cada agente debe actualizar su estado para mantener la consistencia.
- Para garantizar consistencia un agente juega el rol de aprendiz, e implica una comunicación con otro agente que toma el papel de crítico.

Aprendizaje Distribuido: SMILE

- Idea general es que para que un mecanismo Ms sea consistente, se supone que exista una **interacción entre el agente aprendiz ri y los otros agentes**.
- Con el propósito de lograr la consistencia de Ms, el agente aprendiz ri **tiene un mecanismo interno M encargado de mantener la consistencia del agente**. El proceso es:
  - El mecanismo Ms es disparado por un agente ri que recibe una información k (necesita actualización de estado para restablecer su a-consistencia).
  - El estado del agente aprendiz se actualiza a M(Bi);
  - B0C es la parte común modificada por el agente ri; y B0j el estado de cualquier otro agente rj inducido por la modificación de B0C.
  - Se da una interacción I(ri; rj) entre el agente aprendiz ri y el agente rj jugando el papel de crítico,
  - El mecanismo Ms termina hasta que ningún agente puede proporcionar mas información k0 (la sma-consistencia del agente aprendiz rj es restaurada y todos los agentes adoptan la actualización de estado BC).

Aprendizaje Distribuido: SMILE

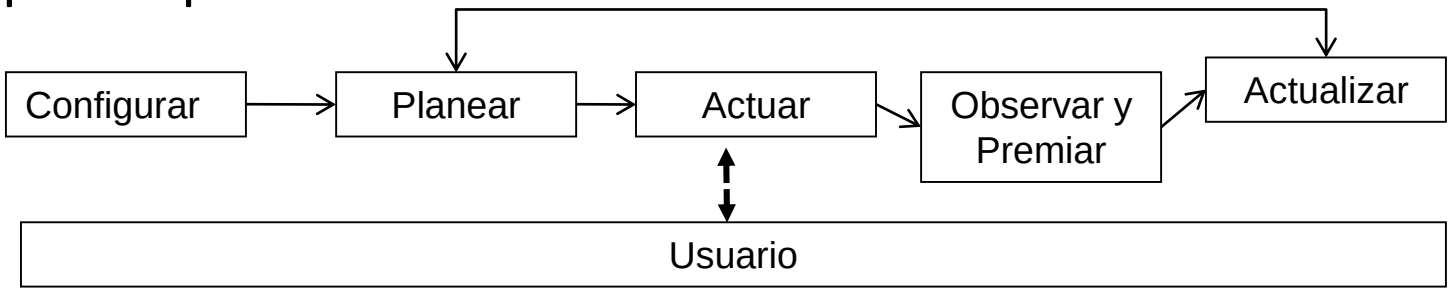
- Dos situaciones por las cuales un agente podrá considerar comunicarse mientras se encuentra en el proceso de aprendizaje:
  - Cuando el agente no es capaz de iniciar el proceso de aprendizaje, e.g., no tiene suficientes ejemplos;
  - El agente no puede hallar una hipótesis para explicar la falla del 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 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.

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.

- 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 Ai $|Ai| > 2$.
- 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,