

Aprendizaje en agentes

Jose Aguilar

CEMISID, Facultad de Ingeniería

Universidad de los Andes

Mérida, Venezuela

aguilar@ula.ve

Aprendizaje

Por mas formalismos de representación y razonamiento, nunca se abarcara todo el conocimiento posible

- Comportamiento inteligente: capacidad de adaptarse o aprender de experiencias

Un sistema **aprende** cuando es *capaz de experimentar modificaciones estructurales y/o funcionales*, de acuerdo con la *experiencia*, en vista a conseguir *mayor eficacia* en su interacción con el medio

G. Pajare, M. Santos

Aprendizaje

- Se utiliza para :
 - Memorización de algo
 - Aprendizaje de hechos con la observación o la exploración
 - Desarrollo de habilidades cognoscitivas a través de la práctica
 - Organización de nuevo conocimiento

El aprendizaje denota *cambios adaptativos* en el sistema en el sentido que permiten al sistema *hacer la(s) tarea(s)* más eficientemente y eficazmente la próxima vez.

H. Simon

Aprendizaje

- Esencial para **ambientes desconocidos**
- Es un método de **construcción de sistemas**
- **Modifica los mecanismos de decisión** de los agentes para mejorar su comportamiento
- Única forma para que **agente adquiera por si solo lo que requiere**
- **Provee autonomía**
- Forma automática de **adquisición del conocimiento**

¿Qué se deriva del Aprendizaje?

- **Incrementa el conocimiento** que se posea
 - Adquisición del conocimiento
 - Capacidad de memorizar
 - Adaptación al entorno
- **Mejora la exactitud de la solución**
 - El objetivo es la síntesis (definición) de una estructura nueva
 - Supone realizar una inferencia (inductiva, abductiva)
- **Mejora el tiempo de calculo de una solución**
 - El objetivo es el análisis (explicación) de una observación
 - Supone realizar una inferencia deductiva

Aprendizaje

Sentencia clásica: $P \vee BC \Rightarrow C$

- P: Premisa
- BC: información previa
- C: consecuente, observación, datos de entrada
- Aprendizaje tiene dos elementos:
 - Objetivo de Aprendizaje (O)
 - Información inicial (I)
- Inferencia inductiva en el aprendizaje:
 - $O = P$
 - $I = C + BC$

Descubro desde lo que observo
- Inferencia deductiva en el aprendizaje :
 - $I = P + BC$
 - $O = C$

Construyo desde experiencias previas

Tipos de Aprendizaje

Aprendizaje inductivo

- Basados en observaciones o ejemplos
- Ejemplo:
 - El pedazo de pan número 1 me alimento cuando yo lo comí.
 - El pedazo de pan número 2 alimentaba cuando yo lo comí.
 - El pedazo de pan número 3 alimentaba cuando yo lo comí.
 - El pedazo de pan número 100 alimentaba cuando yo lo comí.
- Por lo tanto,
todos los pedazos de pan me alimentan cuando los como (P)

Tipos de Aprendizaje

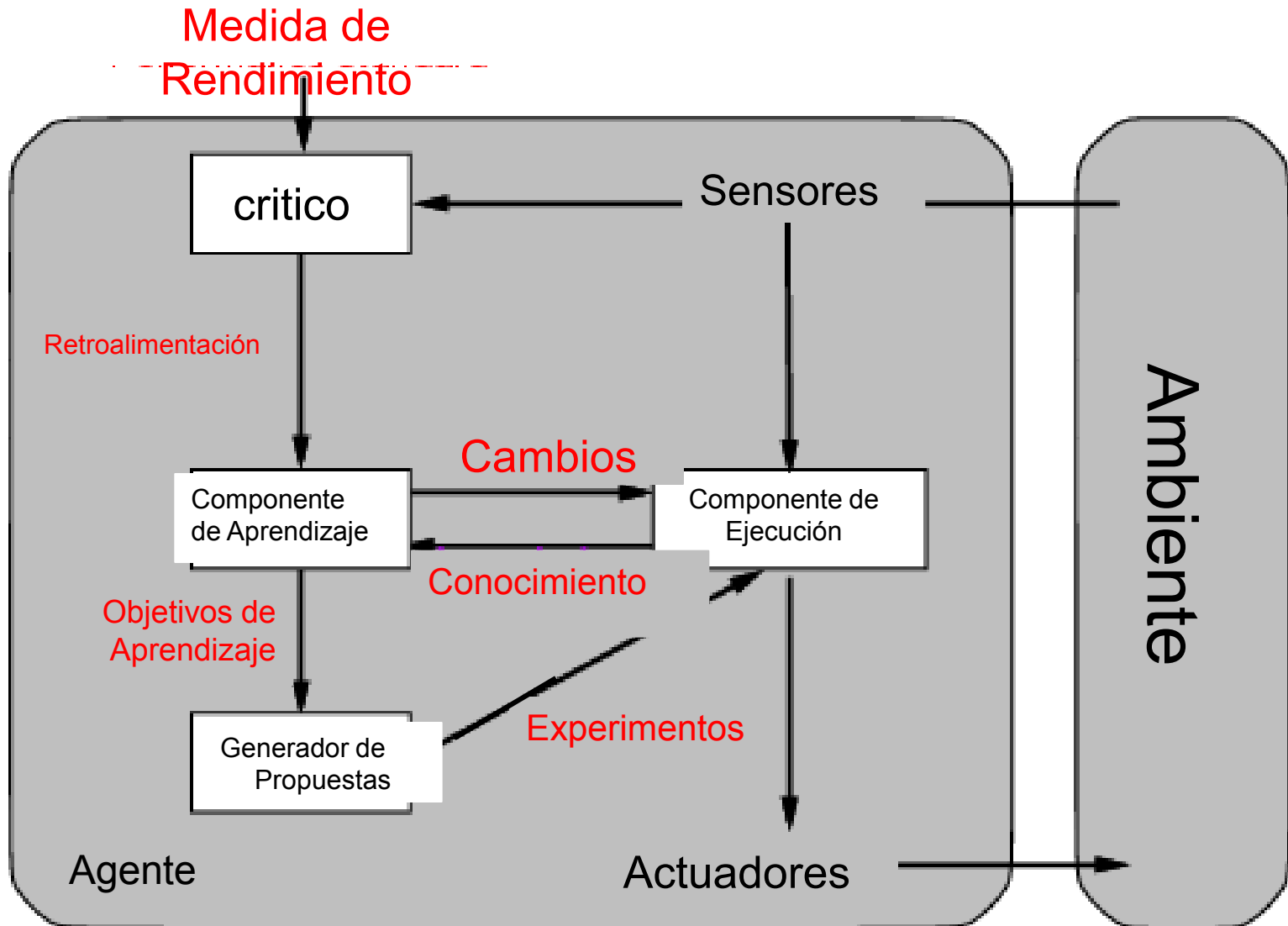
Aprendizaje deductivo

- Basado en explicaciones, hipótesis, premisas
- Ejemplo: dada una regla sintáctica en un lenguaje de una oración ($O=S+V$)=>**Premisa**, y conociendo otras reglas (por ejemplo la concordancia de genero, la posibilidad de varios sujetos, etc.)=>**BC**, el objetivo es encontrar una mejor definición de la oración (**O(C)**) que tenga más sentido en el contexto dado

• Otros tipos de Aprendizaje

- Por analogía
- Basado en casos
- Por imitación
- Multiestrategias

Agente con aprendizaje



Un *componente de ejecución* puede incluir

- Un mapeo de las condiciones actuales a acciones
- Un medio para inferir propiedades relevantes del mundo según lo que se percibe
- Usa
 - Información de cómo evoluciona el mundo
 - Información sobre resultados de las posibles acciones
 - Información indicando lo deseable de cada estado
 - Información de que tan deseable es una acción en un estado en particular
- Tiene objetivos a optimizar

Diseño del *componente de aprendizaje*

- Cuales **componentes** del componente de ejecución deben ser aprendidos
- Cual es la **representación** usada para esos elementos
- Cual es el **mecanismo** de retro-alimentación disponible
 - Aprendizaje Supervisado
 - Aprendizaje No supervisado => agrupamiento
 - Aprendizaje Reforzado
- Que **conocimiento-a-priori** esta disponible

Clases de aprendizaje en Agentes

- **Aprendizaje Supervisado**

Dado un conjunto de ejemplos de pares de entrada/salida, encontrar una regla que prediga la salida asociada con una nueva entrada

- **Aprendizaje No supervisado => agrupamiento**

Dado un conjunto de ejemplos, pero sin etiqueta, agruparlos o caracterizarlos naturalmente,

- **Aprendizaje Reforzado**

Un agente interactúa con su ambiente haciendo observaciones y tomando acciones, y es recompensado o penalizado por ellas. El aprende a escoger las acciones que lo recompensen mas

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.

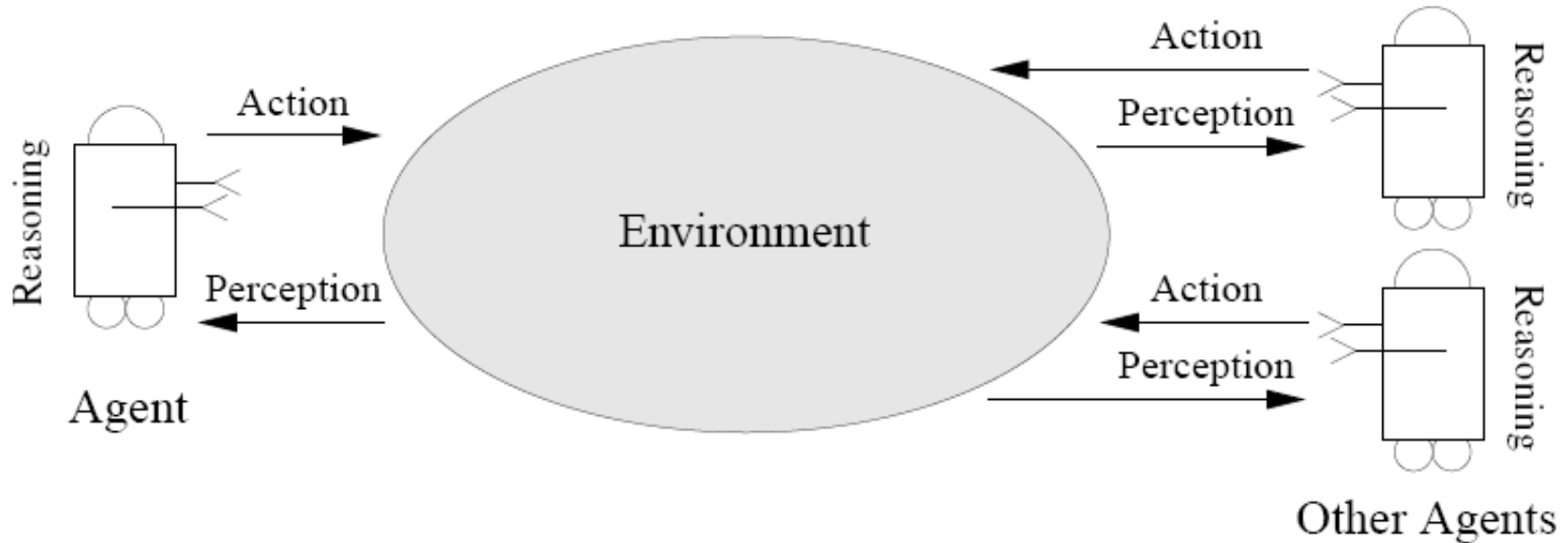
Los agentes aprenden de forma distribuida e interactiva, afectándose los unos a los otros.

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*

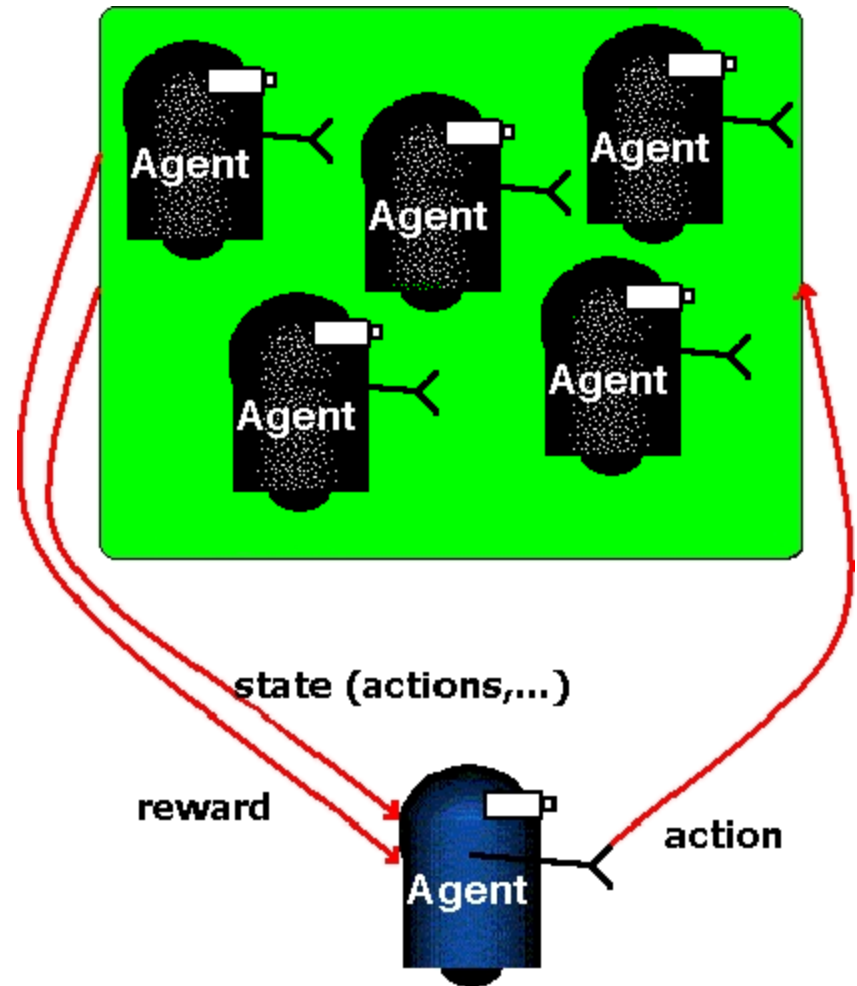
Aprendizaje Multi-agente



- Aprender en un entorno multi-agente
 - todos los agentes simultáneamente aprenden
 - medio ambiente no estacionario (otros agentes están evolucionando)
 - problema de un "blanco móvil"

Aprendizaje Multi-agente

- Agente trata de resolver su problema de aprendizaje, mientras que otros agentes en el medio ambiente también están tratando de resolver sus propios problemas de aprendizaje: desafiante no estacionario.
- Escenarios principales: (1) cooperativa, (2) interés propio
- Agente puede saber muy poco sobre otros agentes: recompensas y algoritmos de aprendizaje desconocidos
- Método tradicional de solución: la teoría de juegos (usa varias suposiciones)



Aprendizaje Multi-agente

- los agentes aprenden porque:
 - no saben todo sobre el medio ambiente
 - no saben cómo los otros agentes se comportan
- Otros agentes están aprendiendo también. ¿Podemos tener un marco unificador en el que este aprendizaje se puede entender.
- Aplicaciones aprendizaje en SMA
 - Equipos de robot de futbol
 - tráfico
 - rescate con Robot
 - Agentes para comercio electrónico
 - conducción automatizada

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

Aprendizaje de un Agente vs Aprendizaje SMA

- Investigaciones 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 coordinado multi-agente* es una metáfora más natural y mejora la eficacia.

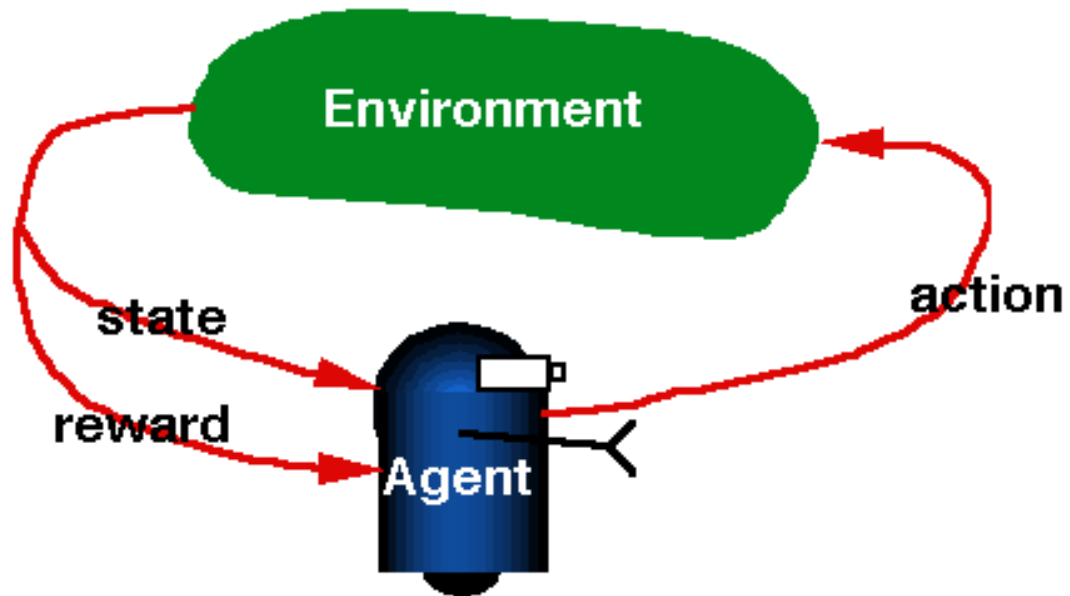
es una pregunta abierta si niveles altos de conocimiento en un agente producen un mejor desempeño.

Algunos (discutibles) Objetivos de aprendizaje en SMA

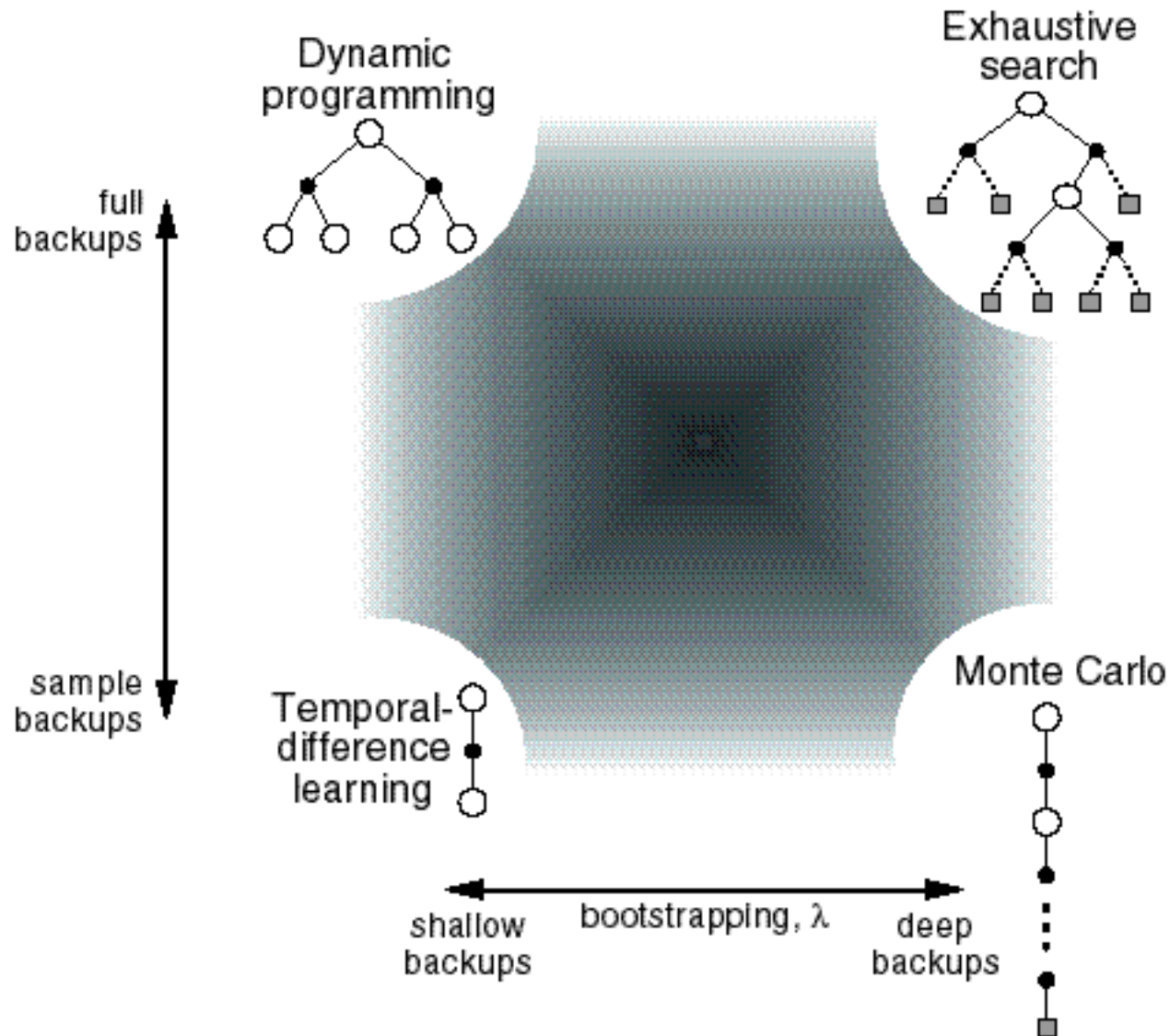
- El aprendizaje debe converger a una estrategia estacionaria
- En aprendizaje "auto-play" (todos los agentes utilizan mismo algoritmo de aprendizaje), los agentes deben converger en una estrategia
- El aprendizaje debe lograr rentabilidades tan buenas como la mejor respuesta a las estrategias de otros agentes
- (Peor de los casos) El aprendizaje debe garantizar una rentabilidad mínima

Normal aprendizaje en un agente

- Supongamos que el medio ambiente tiene estados observables, las recompensas esperadas y las transiciones de estado caracterizables , y todo lo anterior es estacionario (MDP-ish)
- Aprendizaje sin una especificación completa: RL + exploración, Monte Carlo, ...



Dimensiones de aprendizaje por Reforzamiento



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 en 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

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 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

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.

Clasificación del Aprendizaje

- **Aprendizaje cooperativo** - por ejemplo, cada agente tiene su propio mapa, y junto con otros agentes agregar una visión global
- **Aprendizaje competitivo** - por ejemplo, cada agente egoísta trata de maximizar propia utilidad al aprender acerca de los comportamientos y las debilidades de los otros agentes

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:

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

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)

$$S(c_i, a_i, t+1) = (1 - \beta) S(t, c_i, a_i) + \beta(R + S(t+1, c_i, a_i))$$

Aprendizaje RL en Multi-agente

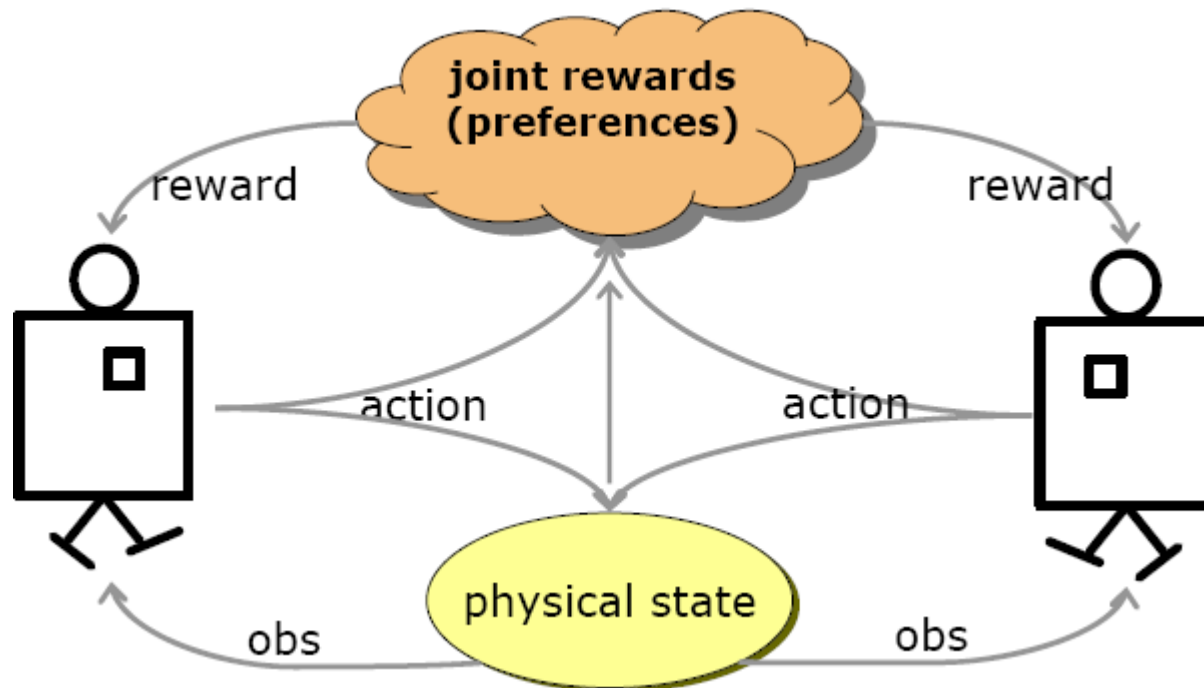
Espacio de estado más grande

- Podría tener que incluir el estado de otros robots en el propio estado

Problemas de aprendizaje RL en SMA

- Todos los problemas de un solo agente
- Otros agentes impredecible o no estacionario
- En caso de refuerzo ser local o global?
- Robot trata lograr la meta o reaccionar a otros robots cuando se hace una buena acción?

Aprendizaje RL en Multi-agente



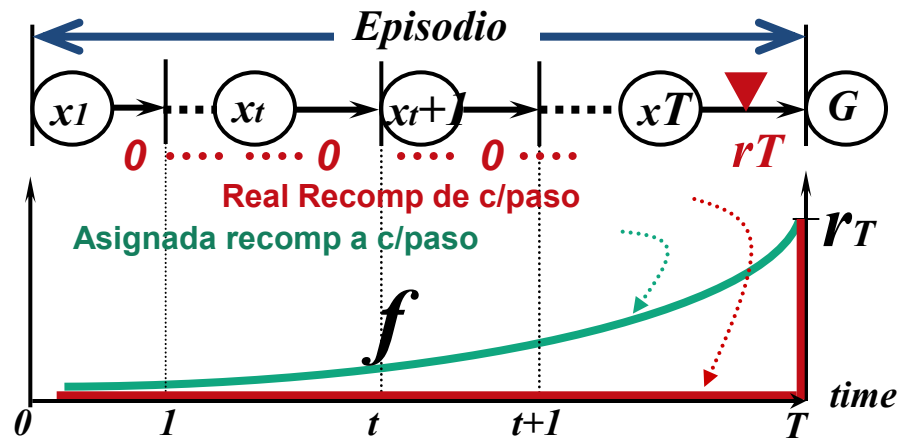
Aprendizaje RL en Multi-agente

Ambiente Multi-agente es no-Markoviano.

Porque : probabilidad de transición entre S_t a S_{t+1} puede variar.

Porque : agentes' hacen un proceso “*concurrente de aprendizaje*”

$$W_{n+1}(x_t, a_t) \leftarrow W_n(x_t, a_t) + f(r_T, t)$$



r_T : refuerzo tiempo T (objetivo)

W_n : peso par estado-acción despues de
 n episodios,

(x_t, a_t) : estado y acción en el tiempo t de
 n -vo episodio.

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 no es estacionario generalmente no es considerado.
- *Minimax-Q learning 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.

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

- Fase aprendizaje, la probabilidad de seleccionar un rol r en situación s es:

$$Pr(s, r) = f(U(r,s), P(r,s), C(r,s), \text{Potencial}(r,s)) /$$

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

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

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

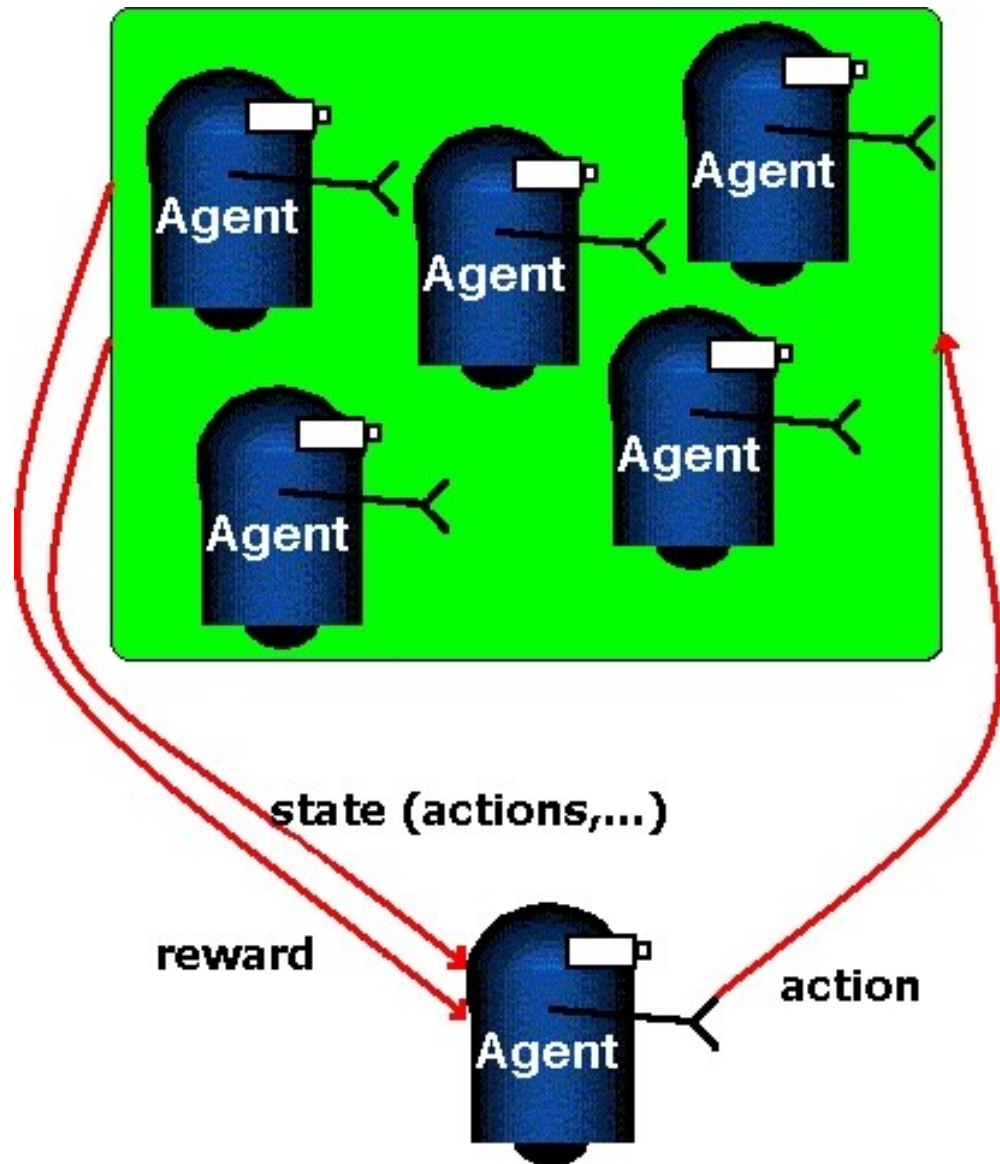
y $U, P, C, \text{Potencial}$ son actualizados: $U(r, s_{n+1}) = (1 - \beta) U(r, s_n) + \beta U_f$ U edo. Final

$P(r, s_{n+1}) = (1 - \beta) P(r, s_n) + \beta P_f$ $P_f = 1$ si f es ok

Problema aprendizaje en SMA

- Agente trata de resolver su problema de aprendizaje, mientras que otros agentes en el medio ambiente también están tratando de resolver sus propios problemas de aprendizaje.
- Solución teórica basada en no aprendizaje para problema completamente especificado:

teoría de juegos



Teoría de Juego

- Juego definido por: **jugadores** (1...N), **acciones**, y **matriz de pago**

		acciones B								
		<i>R</i>	<i>P</i>	<i>S</i>			<i>R</i>	<i>P</i>	<i>S</i>	
acciones A	<i>R</i>	0	−1	+1			<i>R</i>	0	+1	−1
	<i>P</i>	+1	0	−1			<i>P</i>	−1	0	+1
	<i>S</i>	−1	+1	0			<i>S</i>	+1	−1	0
		pago A					pago B			

- Si matriz son idénticas, juego es **cooperativo**, sino es **no-cooperativo**

Teoría de Juego

- Juegos sin estados: (bi)-matriz de juego
- Juegos con estados : juegos Markovianos, estocásticos, etc.
- Juegos con movimientos simultáneos: forma normal
- Juegos con turnos alternativos: forma extensiva
- No. of rondas = 1: juego de una vez
- No. of rondas > 1: juego repetido
- Escogencia de la acción determinística: estrategia pura
- Escogencia de la acción no determinística : estrategia mixta

Teoría de Juego

- Una estrategia conjunta x es un **Pareto-óptimo** si x mejora pagos de todo el mundo
- un x_i de un agente es una **estrategia dominante** si siempre es mejor, independientemente de las otras acciones
- x_i es la **mejor respuesta** si se maximiza la rentabilidad dado x_{-i}
- Una estrategia conjunta x es un **equilibrio** si la estrategia de cada agente es al mismo tiempo su mejor respuesta, es decir, no hay incentivos para desviarse (Nash, correlacionados)

Un equilibrio de Nash siempre existe, pero no es fácil de calcular

Juegos repetidos

- Equilibrio de Nash - Elijo lo que es mejor para mí cuando usted está haciendo lo que está haciendo y usted elige lo que es mejor para usted cuando estoy haciendo lo que estoy haciendo
- En los juegos repetidos tenemos dos jugadores se enfrentan entre sí, como por ejemplo, El dilema del prisionero
- Equilibrio de Nash se basa en la asunción de jugadores perfectamente racionales,
- la teoría determina la estrategia de equilibrio que se llega por diversos mecanismos de aprendizaje

Juego dilema del prisionero



Prisionero 2

Prisionero
1

		Confiesa (Defecto)	resiste (Coopera)
Confiesa (Defecto) resiste (Coopera)	Confiesa (Defecto)	-8 -8	0 -10
	resiste (Coopera)	-10 0	-1 -1

Juego dilema del prisionero



		Prisionero 2	
		↓ Confiesa (Defecto)	↓ resiste (Coopera)
Prisionero 1	Confiesa (Defecto)	-8 -8	0 -10
	resiste (Coopera)	-10 0	-1 -1

Sea cual sea Prisionero 2, lo mejor que Prisionero 1 puede hacer es confesar.

Juego dilema del prisionero



Prisionero 2

		Confiesa (Defecto)	resiste (Coopera)
Prisoner. 1	Confiesa (Defecto)	-8 -8	0 -10
	resiste (Coopera)	-10 0	-1 -1

Sea cual sea Prisionero 1, lo mejor que Prisionero 2 puede hacer es confesar.

Juego dilema del prisionero



Prisionero 2

		Confiesa (Defecto)	resiste (Coopera)
Prisionero 1	Confiesa (Defecto)	-8 -8	0 -10
	resiste (Coopera)	-10 0	-1 -1

Cada jugador tiene una estrategia dominante: confesar.

La estrategia dominante en El equilibrio (confesar, confesar)

Una estrategia es una estrategia dominante si es estrictamente mejor respuesta de un jugador para cualquier estrategia que los otros jugadores pueden escoger.

Una estrategia dominante en equilibrio es una combinación de estrategias que consiste de cada estrategia dominante de c/jugador.

Juego dilema del prisionero



Prisionero 2

		Confiesa (Defecto)	resiste (Coopera)
Prisionero 1	Confiesa (Defecto)	-8 -8	0 -10
	resiste (Coopera)	-10 0	-1 -1

La rentabilidad de la estrategia de equilibrio dominante $(-8, -8)$ es peor para ambos jugadores que $(-1, -1)$, el pago en el caso de que ambos jugadores resistan. Por lo tanto, el dilema del juego del prisionero es un juego de conflicto social.

Oportunidad para el aprendizaje de múltiples agentes: aprendiendo durante el juego repetido, la solución óptima de Pareto $(-1, -1)$ puede surgir como resultado de aprendizaje (también puede surgir en la teoría de juegos evolutiva).

Enfoques de aprendizaje SMA

- Idea básica: el agente se adapta, haciendo caso omiso de la no estacionariedad de las estrategias de otros agentes
- Juego ficticio: Agente observa la frecuencia de tiempo promedio de opciones de acción de otros jugadores, y lo modela:

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

agente luego juega mejor respuesta a este modelo

Las variantes de juego ficticio: ponderación exponencial "suavizada" mejor respuesta, pequeño ajuste hacia la mejor respuesta, ...

Enfoques de aprendizaje SMA

- Evolutiva teoría de juegos: modelos "Replicator Dynamics": gran población de agentes, con distintas estrategias, los agentes más aptos reproducen más copias.

Sea x = vectores estrategia poblacional y x_k = fracción de la población jugando estrategia k . La Tasa de crecimiento a continuación:

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

Replicator dynamics

- Este modelo supone que la fracción de agentes que juegan un particular crecerá en proporción a lo bien que la estrategia es en la población.
- Se supone una población homogénea de agentes.
- Los agentes se emparejan al azar con el fin de jugar a un juego simétrico (mismas estrategias y pagos).
- Se inspira en la evolución biológica.

$\phi^t(s)$ es número de agentes utilizando la estrategia s en el tiempo t , $u^t(s)$ la utilidad esperada para un agente que juega la estrategia s en el tiempo t y $u(s, s')$ utilidad que s recibe contra s' .

Podemos definir:

$$\theta^t(s) = \frac{\phi^t(s)}{\sum_{s' \in S} \phi^t(s')}$$

$$u^t(s) = \sum_{s' \in S} \theta^t(s') u(s, s')$$

$$\phi^{t+1}(s) = \phi^t(s)(1 + u^t(s))$$

Aprendizaje y Comunicaciones

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

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

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

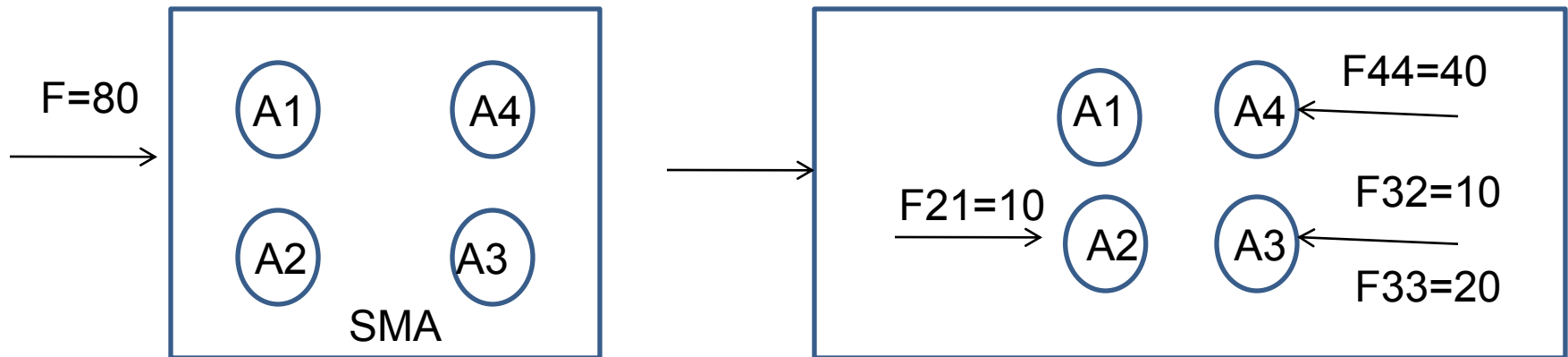
Donde $\text{PERFORM}(M, T_j)$ indica que tan bueno agente M ha hecho tarea T_j en el pasado

Problema de Asignación de Crédito

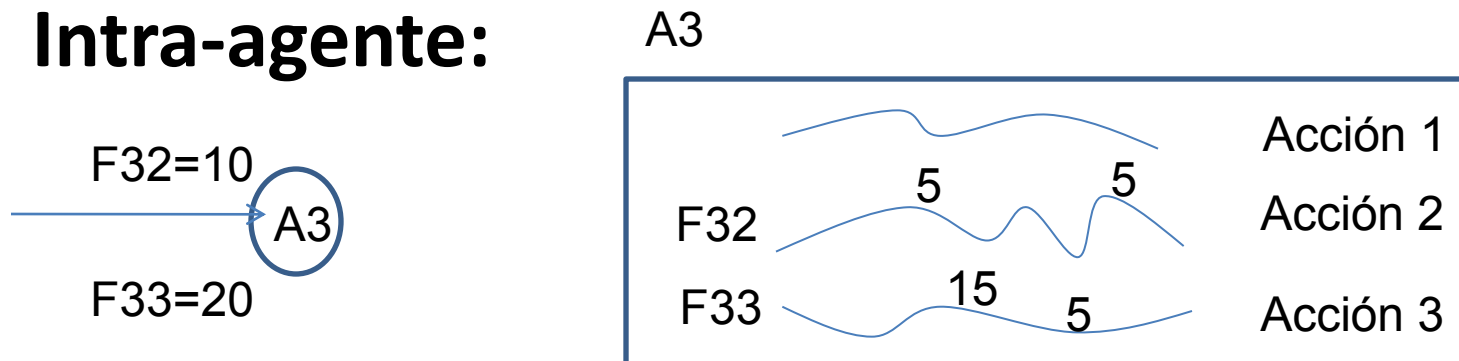
- **Objetivo:** asignación de la retroalimentación (crédito o penalización) para c/elemento del sistema
 - Inter-agente: asignación de la retroalimentación a c/u de los agentes (qué acción de qué agente contribuye a mejorar que rendimiento?)
 - Intra-agente: asignación de retroalimentación a los componentes internos de las acciones de los agentes (qué conocimiento, inferencia o decisión de 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:**



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



UNIVERSIDAD
DE LOS ANDES
MERIDA VENEZUELA

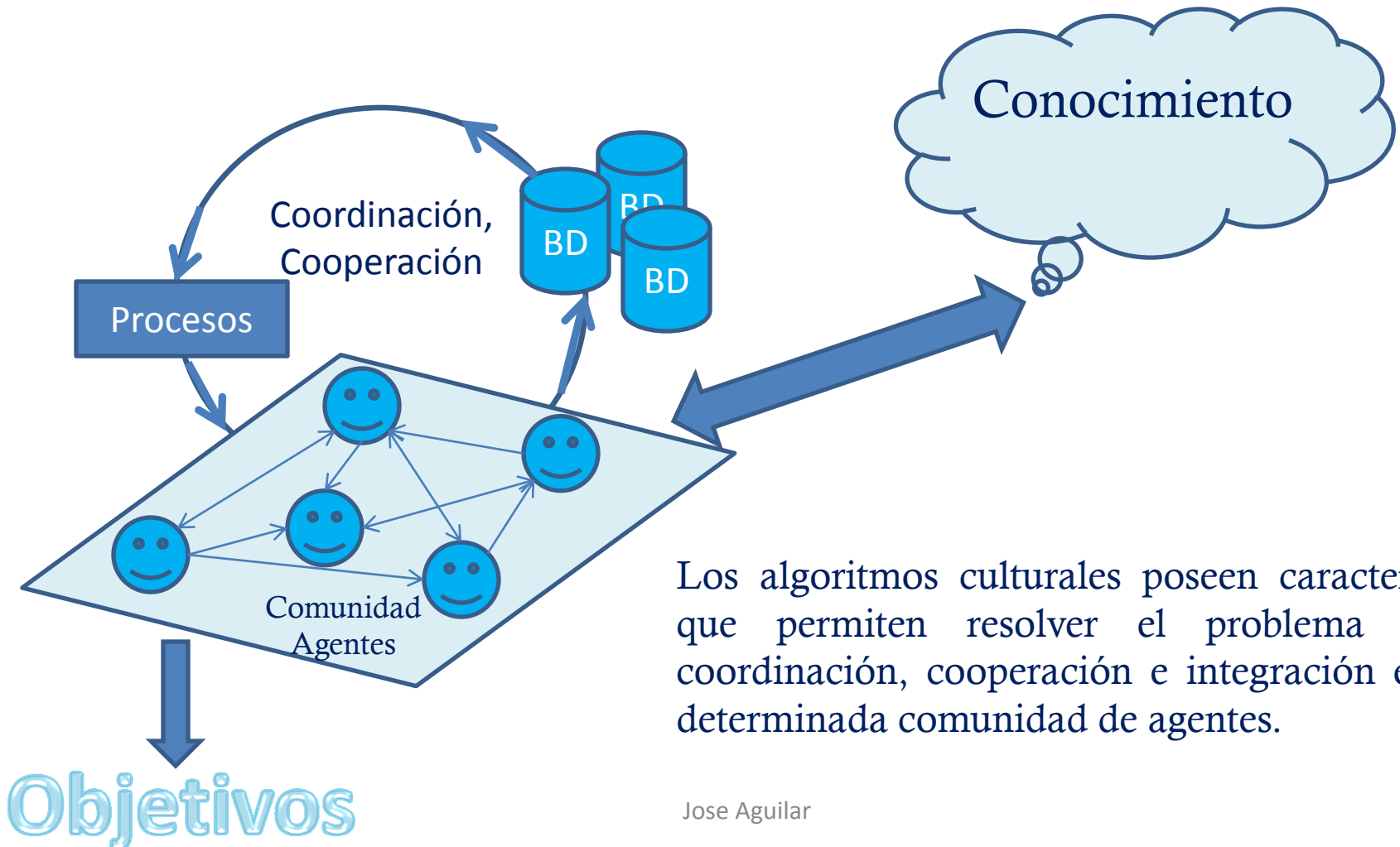
CEMISID

Centro de Estudios en Microelectrónica
y Sistemas Distribuidos

Cultural Algorithms-Based Learning Model for Multi-Agent Systems



Introducción



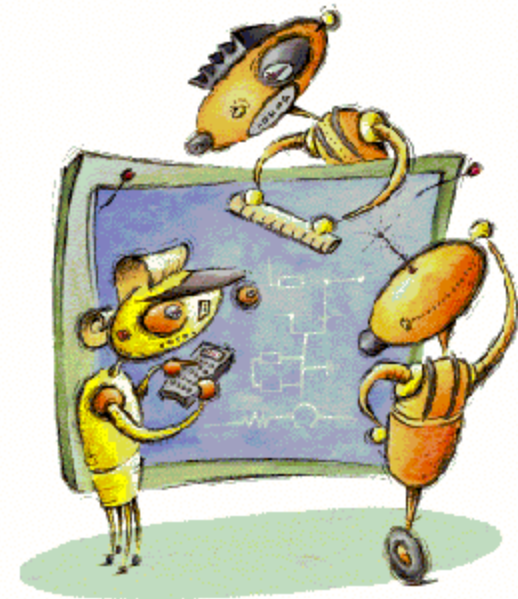
Los algoritmos culturales poseen características que permiten resolver el problema de la coordinación, cooperación e integración en una determinada comunidad de agentes.

Sistemas MultiAgentes (SMA)



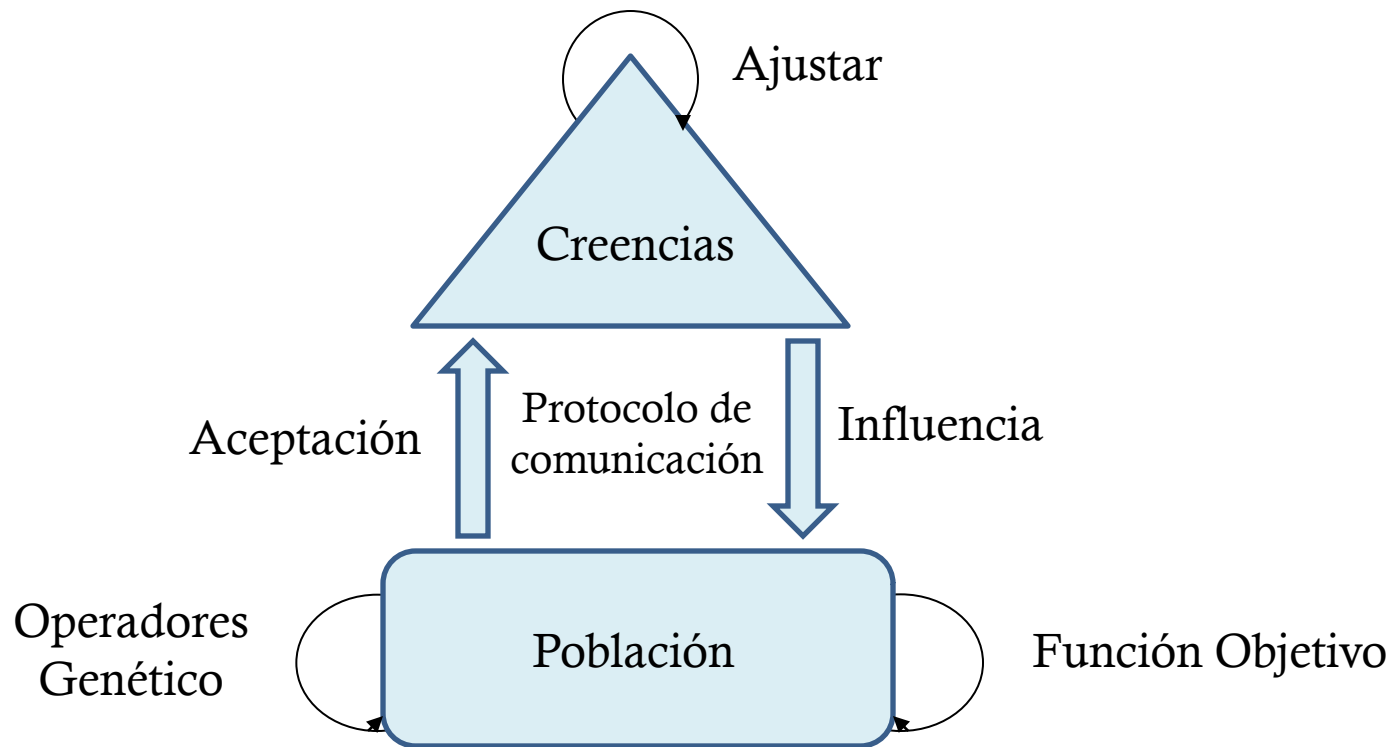
Coordinación en SMA. Conjunto de actividades suplementarias necesarias a realizar en una comunidad de agentes para poder actuar colectivamente

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



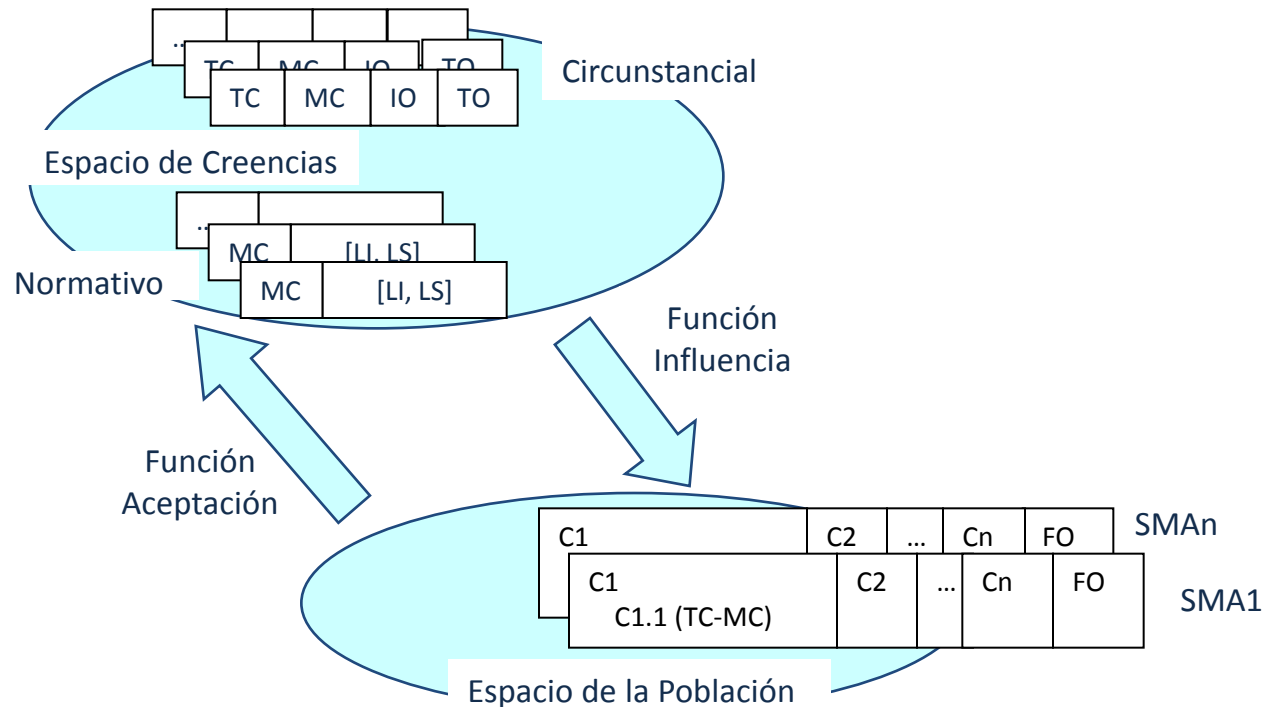


Algoritmos Culturales (AC)



Objetivo General

Proponer un modelo de optimización de aprendizaje de esquemas de coordinación para sociedades de agentes (SMA) usando algoritmos culturales, y aplicar dicho modelo en el problema de integración en automatización



Tipos de Conversación

De las interacciones entre los agentes de una comunidad (Sistemas Multiagentes) nacen conversaciones

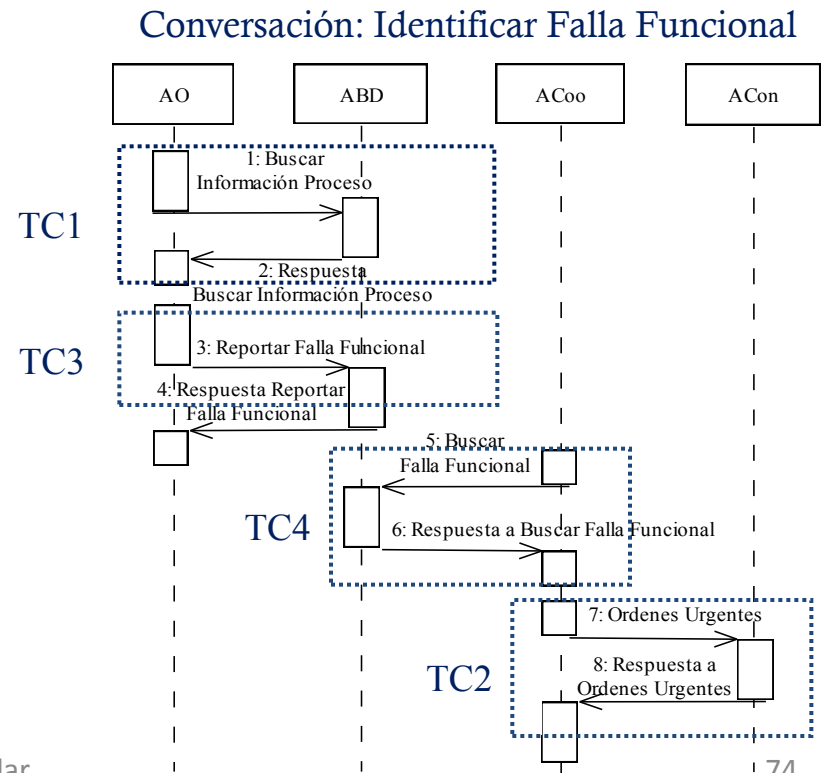
Basadas en FIPA

TC1: Consulta

TC2: Asignación

TC3: Información

TC4: Solicitud



Parámetros de los Mecanismos de Coordinación

Mecanismos de Coordinación (MC)	Parámetros (P)		
Subasta (SI/SH)	C_0	ε_i	$C_P(j, t, x_r)$
Licitación (L)	$M(f)$	$f(T_h)$	-

- ❑ C_0 es el precio inicial.
- ❑ ε_i es la máxima cantidad que un agente puede ofertar.
- ❑ $C_P(j, t, x_r)$ es la condición de parada y puede ser por número de rondas, tiempo, o un valor umbral respectivamente.
- ❑ $M(f)$ es el mensaje inicial que contiene la fecha de expiración para el recibimiento de ofertas.
- ❑ $f(T_h)$ es una función que permite a los potenciales contratistas, evaluar sus capacidades para responder a la notificación de la solicitud de realización de una tarea T_h .



CLEMAS

(Sistema Cultural de Aprendizaje para SMA)

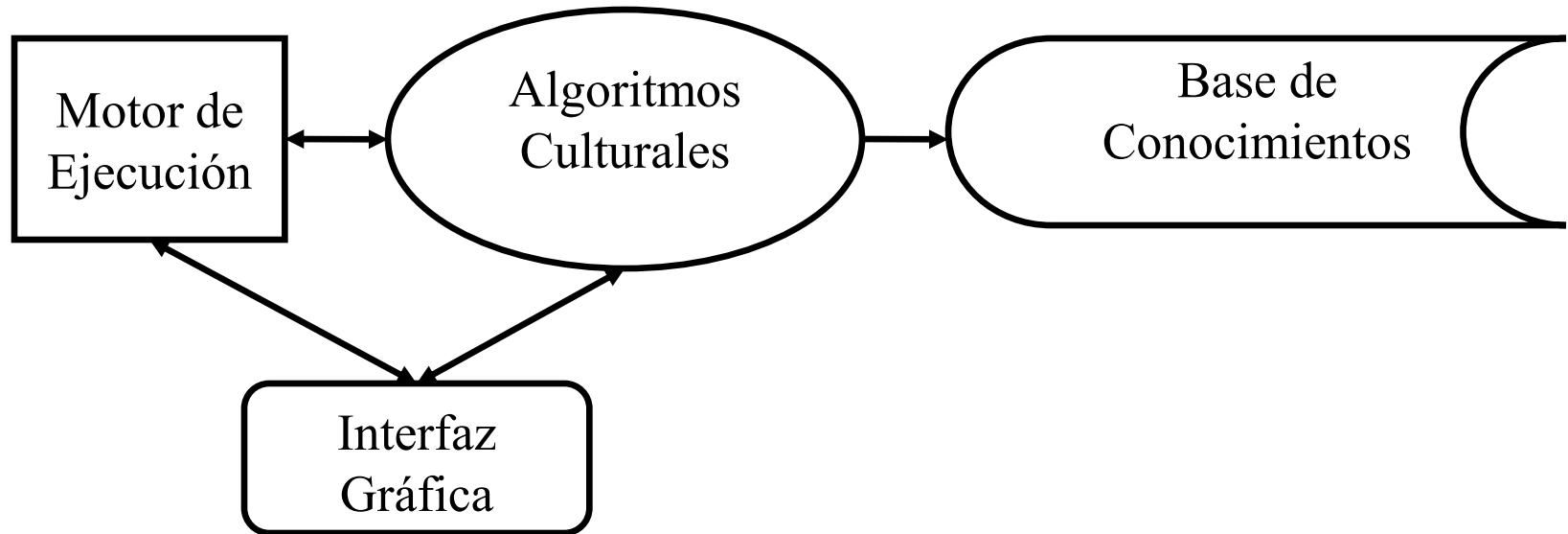
Es un sistema computacional que permite ejecutar el modelo de aprendizaje de esquemas de coordinación para un SMA dado.

CLEMAS requiere que se le configuren distintos parámetros y una vez culminada la ejecución, se puede visualizar cual MC propone para cada conversación/sub-conversación del SMA

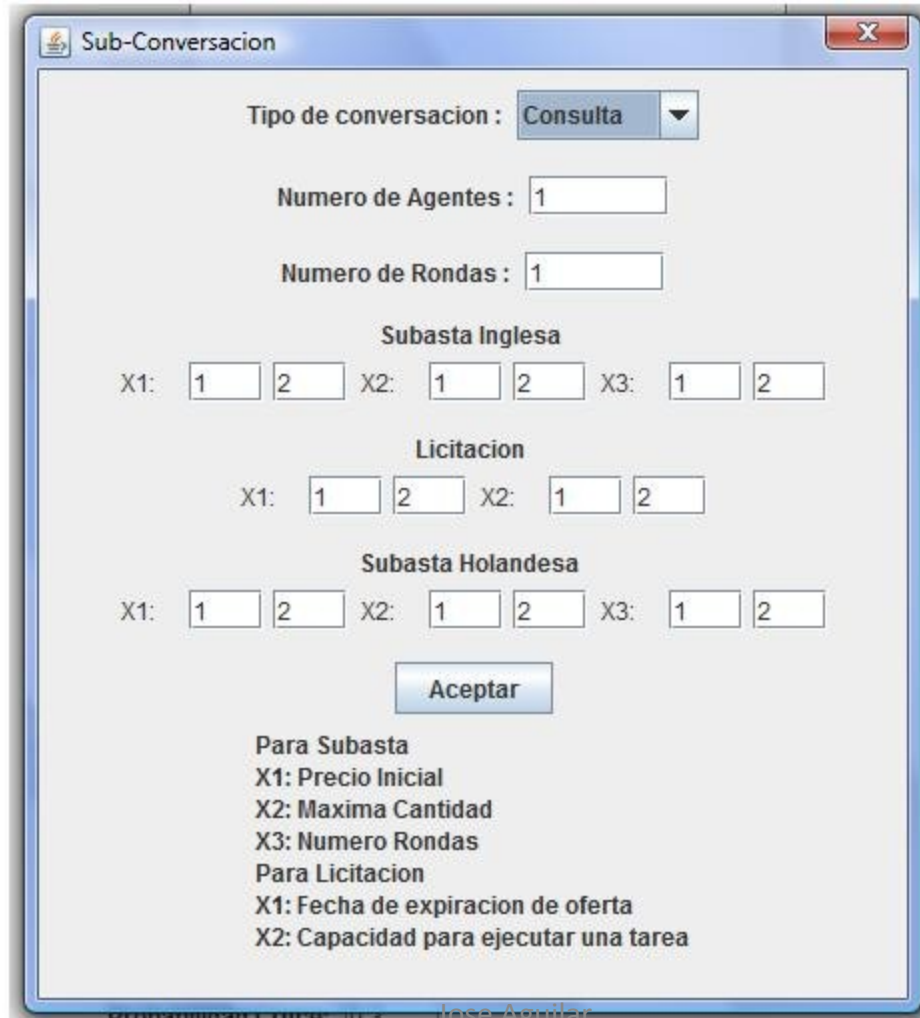
Permite visualizar el progreso del proceso de aprendizaje



Estructura de CLEMAS



Parámetros de los MC en CLEMAS



Sub-Conversion

Tipo de conversacion : Consulta

Numero de Agentes : 1

Numero de Rondas : 1

Subasta Inglesa

X1: 1 2 X2: 1 2 X3: 1 2

Licitacion

X1: 1 2 X2: 1 2

Subasta Holandesa

X1: 1 2 X2: 1 2 X3: 1 2

Aceptar

Para Subasta
X1: Precio Inicial
X2: Maxima Cantidad
X3: Numero Rondas
Para Licitacion
X1: Fecha de expiracion de oferta
X2: Capacidad para ejecutar una tarea

Resultados en CLEMAS

Condiciones Iniciales:
Tamaño de la Poblacion:10
Maximo de Generaciones:100
Porcentaje de aceptacion: 20.0%
Probabilidad de Cruce: 0.6%
Probabilidad de Mutación Directa: 0.3%

Resultados (En base al 20.0% de la poblacion)

Tipo Conversacion	Licitacion	Subasta Inglesa	Subasta Holandesa
Consulta	72.16667%	19.833334%	8.0%
Asignacion	99.5%	0.5%	0.0%
Informacion	15.000001%	22.333332%	62.666668%
Solicitud	0.0%	0.0%	0.0%
Total de Ocurrencia	51.57143%	18.142857%	30.285713%

Graficas de Simulacion

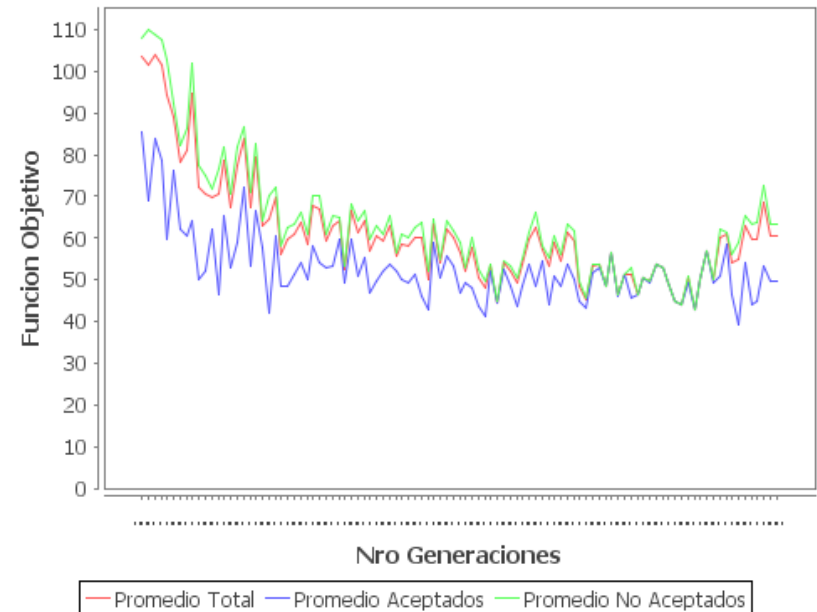
Comportamiento de los mecanismos de coordinación(del 20.0% de Aceptacion)

Frecuencia de Uso Vs Nro Generaciones



Comportamiento del mejor Individuo de la Poblacion

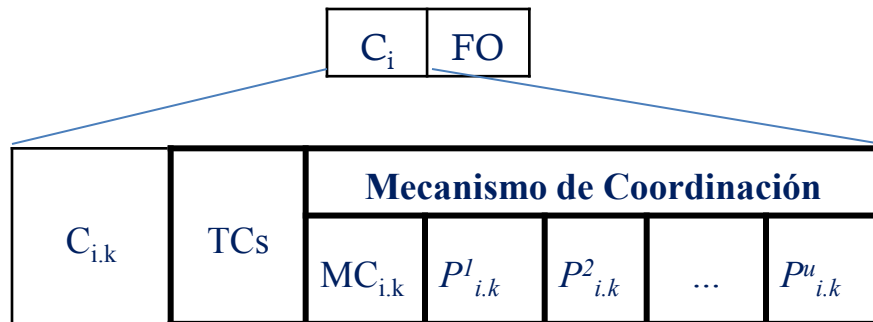
Funcion Objetivo Vs Nro Generaciones



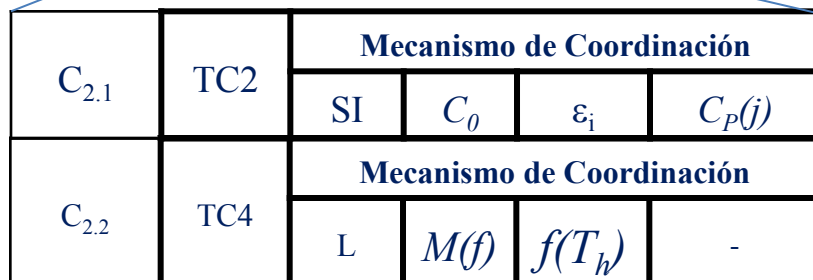
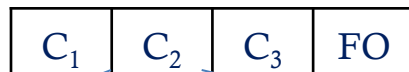


Población

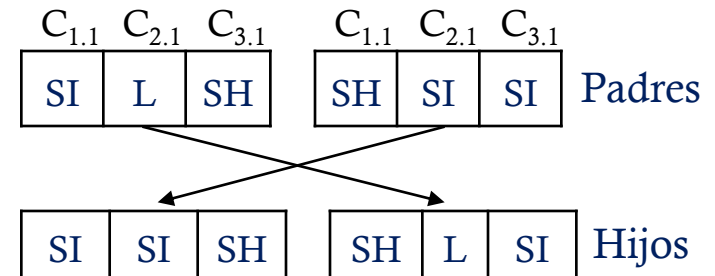
Individuo (SMA)



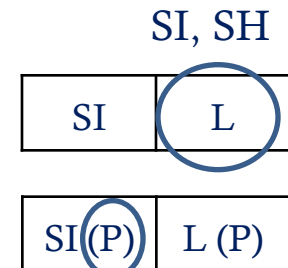
Ejemplo:



Operadores genéticos



Cruce punto sencillo



Mutación

Función Objetivo

Permite evaluar el desempeño de los individuos, esta basada en el costo de procesamiento (CP) y en el costo de comunicación (CC) de cada mecanismo de coordinación usado por el individuo

$$FO = \sum_{i=1}^n \sum_{k=1}^{m_i} (a * CP_{i,k} + b * CC_{i,k})$$

- ❑ Donde a y b son constantes definidas por el usuario y permiten normalizar las unidades de la función
- ❑ Conversaciones $i = \{1, \dots, n\}$
- ❑ Subconversaciones $k = \{1, \dots, m_i\}$
- ❑ Esta función se debe minimizar

CP y CC

Costo de Procesamiento

$$CP_{i,k} = PI_k + \sum_{l=1}^j \sum_{q=1}^{n_j} A_{l,q} + PE_k$$

- ❑ PI fijación del precio inicial (subasta); especificación de condiciones en las que se requiere un servicio (Licitación)
- ❑ A tiempo para preparar propuestas
- ❑ PE proceso de selección del agente ganador

Mec. De Coord.	PI	PE	$A_{l,q}$
Licitación (L)	Medio	Medio	Medio
Subasta Inglesa u Holandesa (SI,SH)	Bajo	Medio	Bajo

Costo de Comunicación

$$CC_{i,k} = \sum_{l=1}^j \left(\sum_{r=1}^{N-1} CEP_{l,r} + \sum_s^{n_j} CEO_{l,s} \right) + \sum_{r=1}^{N-1} CS_r$$

- ❑ CEP costo de envío de propuesta
- ❑ CEO costo de envío de ofertas
- ❑ CS costo de informar al ganador

Mec. De Coord.	CEP	CEO	CS
Licitación (L)	Bajo	Medio	Bajo
Subasta Inglesa u Holandesa (SI,SH)	Bajo	Medio	Bajo

Alto = 1; Medio = 0.6; Bajo = 0.2



Espacio de creencias

Conocimiento Circunstancial

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

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

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

Espacio de creencias

Conocimiento Normativo

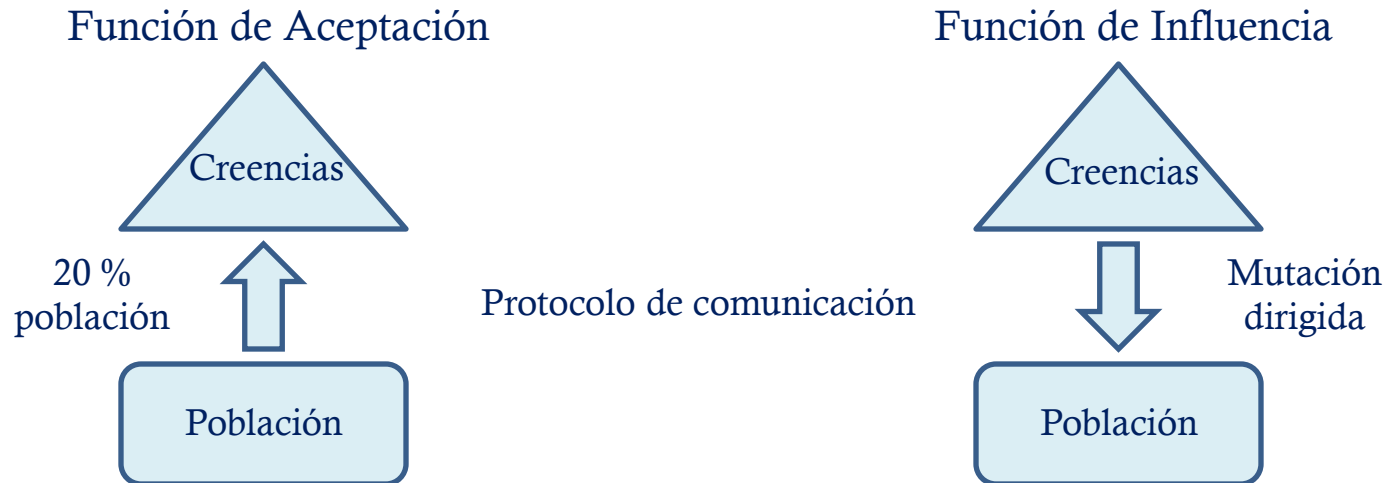
Son rangos de valores idóneos de cada una de las variables que integran al mecanismo de coordinación usado por los individuos

MC	P^1		P^2		$\dots$	P^u	
	LI	LS	LI	LS		LI	LS

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

Protocolo de comunicación

Permite la interacción entre los dos espacios (creencias y población), y esta compuesto por dos funciones: de aceptación y de influencia



Protocolo de comunicación

Función de Aceptación

Conocimiento Circunstancial

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

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

Conocimiento Normativo

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

$$\bar{m} = (1 - m)$$

$$m = \frac{\mu}{t}$$

$$\mu = \{0, \dots, 1\} \neq \{1, 2, \dots, \}$$



Protocolo de comunicación

Función de Influencia

Conocimiento Circunstancial en el
Individuo

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

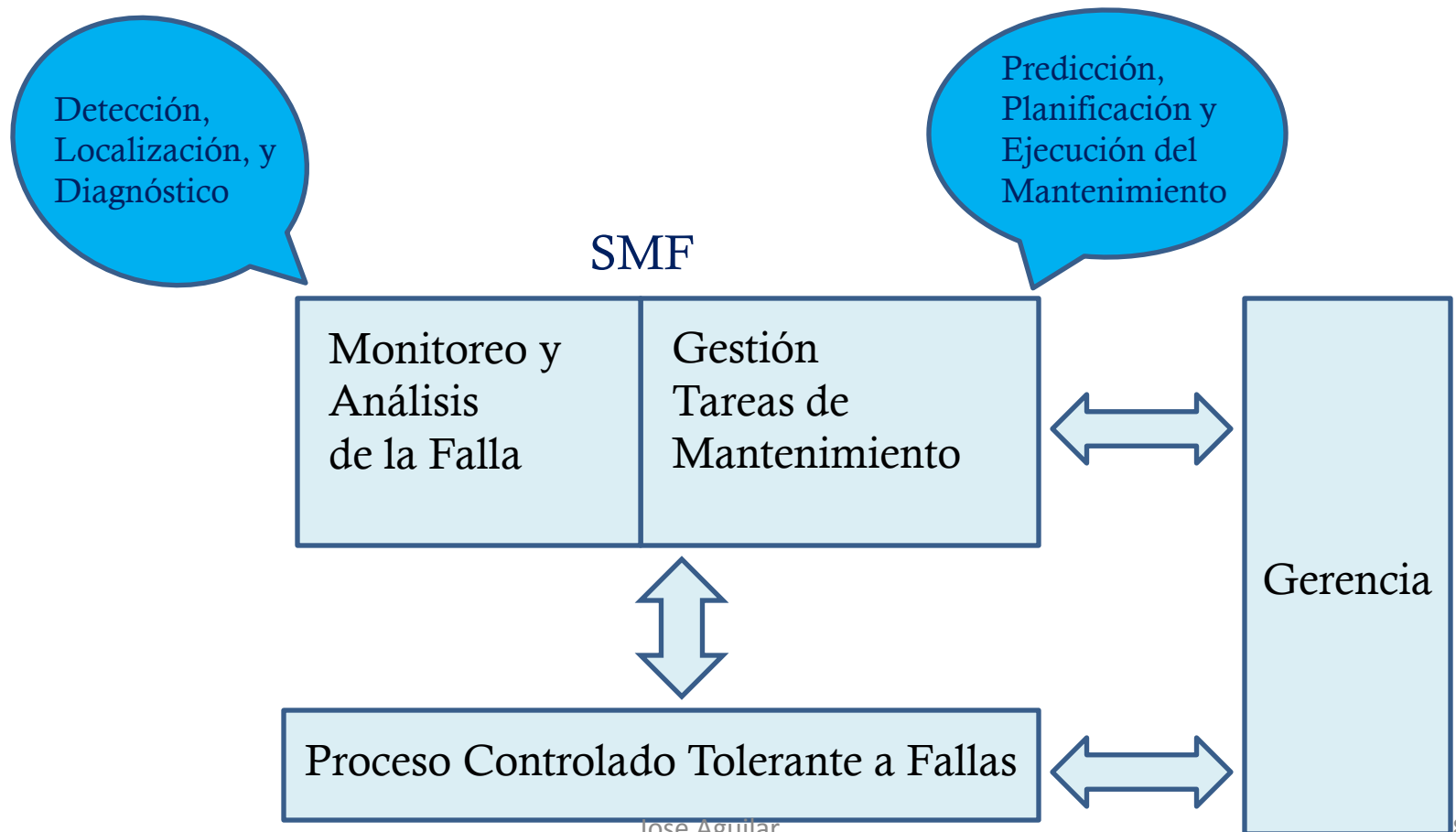
Conocimiento Normativo en el
Individuo

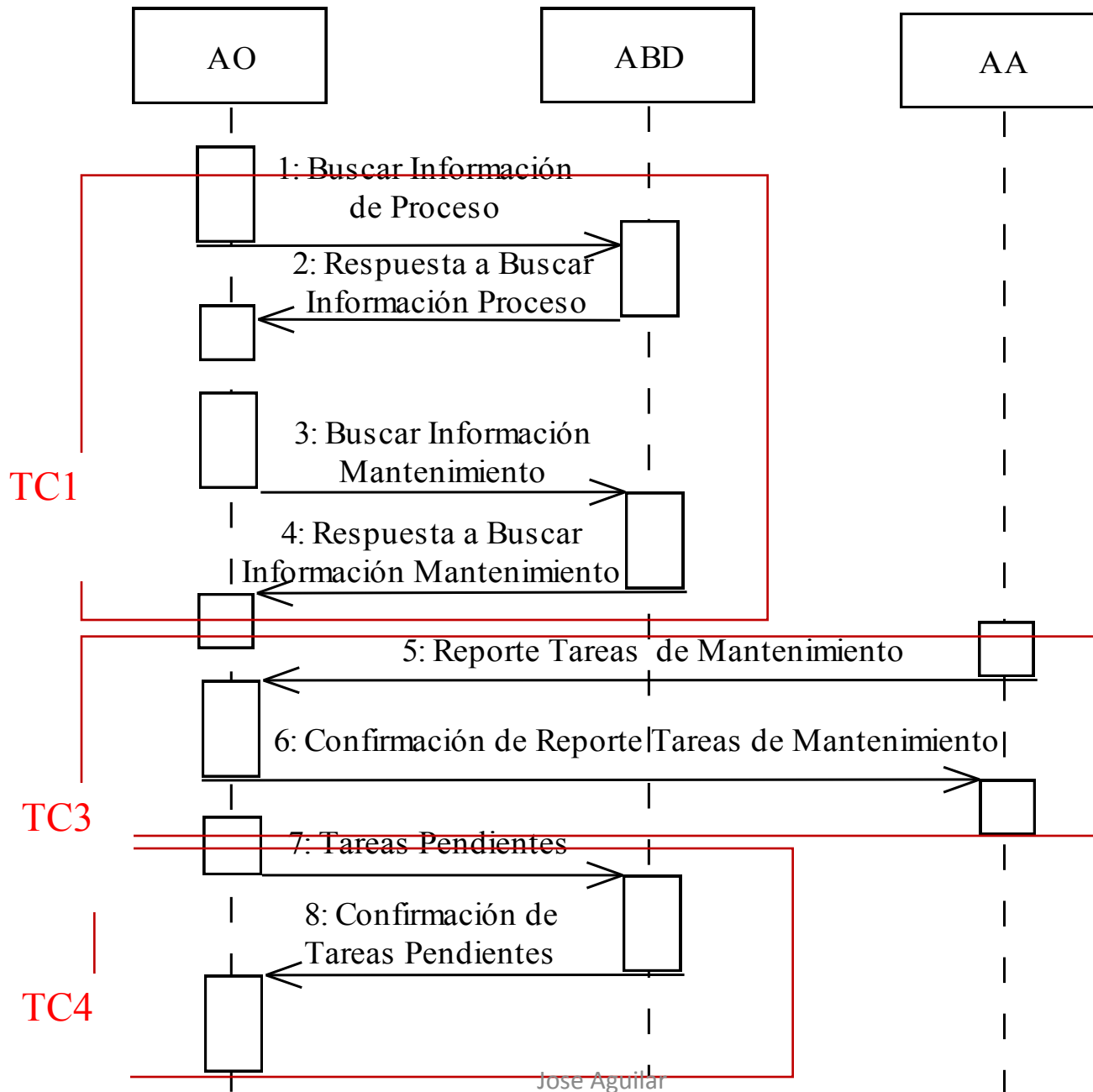
$C_{2.1}$	TC2	Mecanismo de Coordinación			
		SI	C_0 (LI, LS)	ε_i (LI, LS)	$C_P(j)$ (LI, LS)



Caso de Estudio

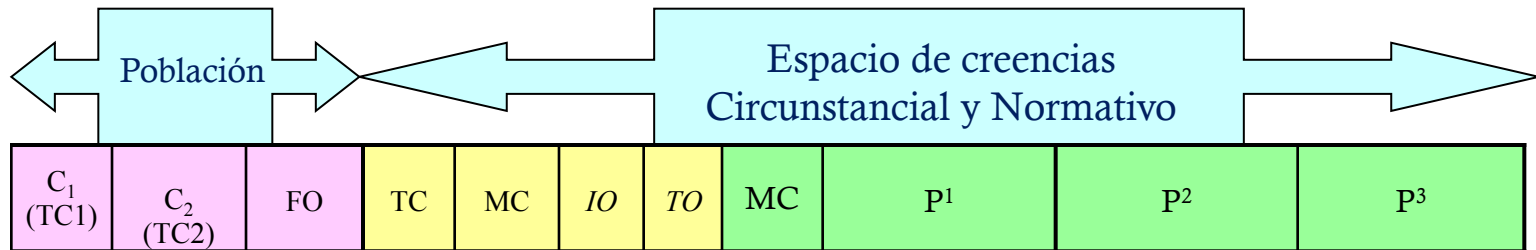
Sistema Manejador de Fallas (SMF) basado en un Sistema Multi-Agente







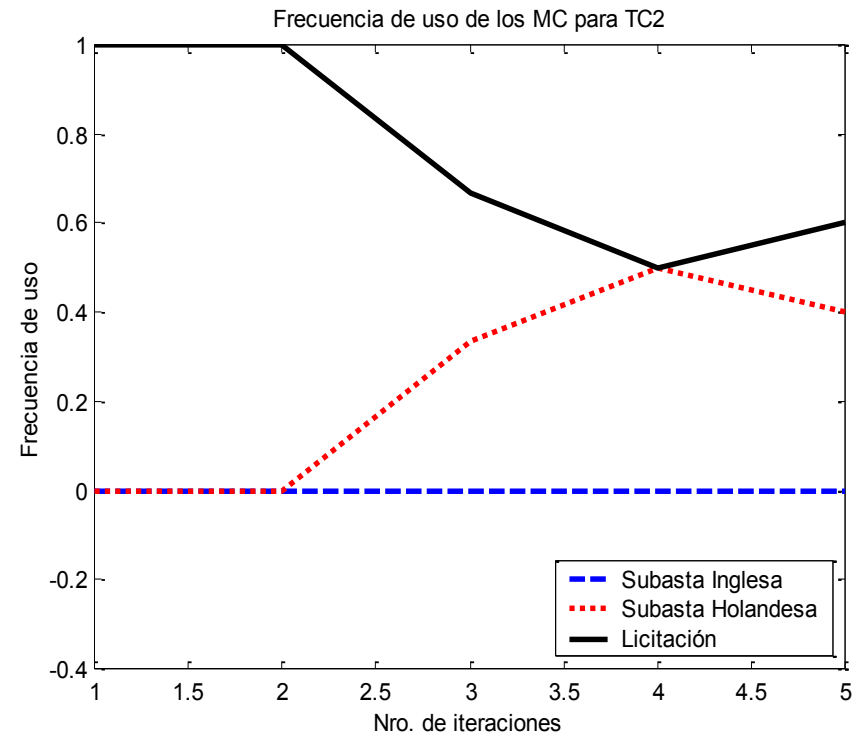
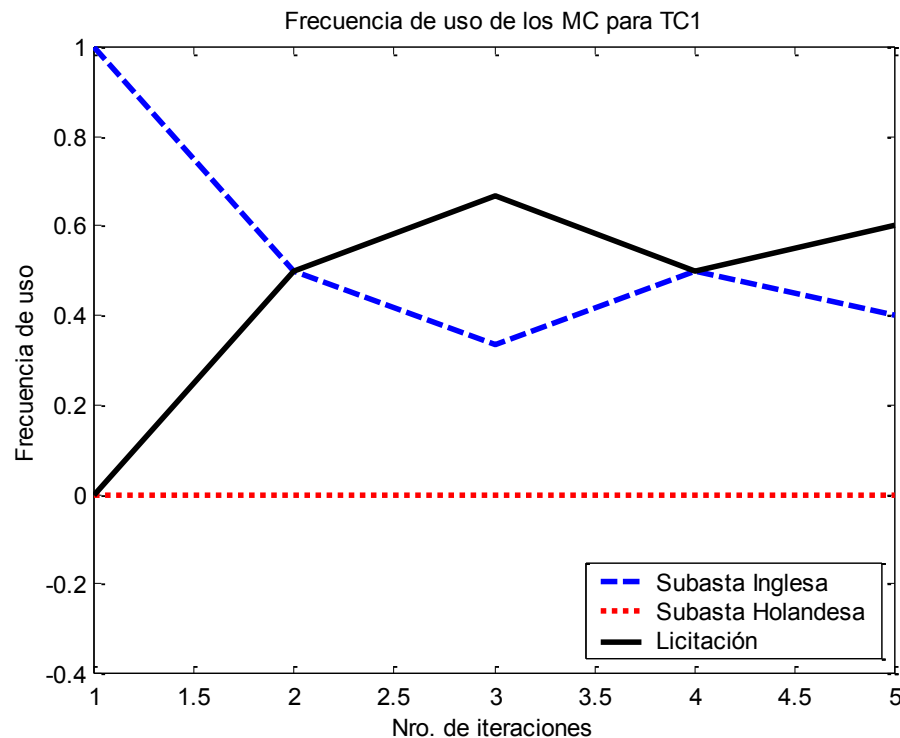
Resultados



		C_1 (TC1)	C_2 (TC2)	FO	TC	MC	IO	TO	MC	P ¹	P ²	P ³
2da	1	L	L	12,0	TC1	L	1	1	L	$M(f)$	$f(Th)$	-
	2	SI	SI	FO						[0,5...1,5]	[2,5...35]	
	3	L	SH	FO	TC2	L	1	2	L	$M(f)$	$f(Th)$	-
	4	C_1	C_2	FO						[0,5...1,5]	[2,5...35]	
	5	C_1	C_2	FO						[0,5...1,5]	[2,5...35]	
3ra	1	L	SH	25,2	TC1	L	1	2	L	$M(f)$	$f(Th)$	-
	2	L	L	FO						[0,25...0,75]	[1,25...17,5]	
	3	C_1	C_2	FO	TC2	SH	1	1	SH	C_0	ε_i	$C_P(j)$
	4	C_1	C_2	FO						[2,5...7,5]	[2,5...35]	[0,5...2]
	5	C_1	C_2	FO						[2,5...7,5]	[2,5...35]	[0,5...2]

Segunda y tercera iteración

Resultados



Gráficas para TC1 y TC2

Jose Aguilar