

# CICLO DE VIDA PARA EL APRENDIZAJE POR COMPARTICIÓN DE CONOCIMIENTOS ENTRE SISTEMAS INTELIGENTES AUTÓNOMOS

Ierache, J.<sup>1,2</sup>, Garcia Martinez, R.<sup>3,2</sup>

<sup>1</sup> Facultad de Informática, Ciencias de la Comunicación y Técnicas Especiales. UM

<sup>2</sup> Laboratorio de Sistemas Inteligentes. Facultad de Ingeniería. UBA

<sup>3</sup> Centro de Ingeniería de Software e Ingeniería del Conocimiento. Escuela de Postgrado. ITBA

jierache@unimoron.edu.ar , rgm@itba.edu.ar

## Resumen

En este artículo se describe el trabajo de investigación que en la actualidad se está desarrollando dentro del área de "Sistemas Inteligentes Autónomos". El objetivo principal de este trabajo es la continuación y profundización de la arquitectura LOPE (Learning by Observation in Planning Environments) con la incorporación de un ciclo de vida de aprendizaje compuesto de distintos layers que acompañan el aprendizaje del agente desde las teorías del creador, las teorías que se generan como producto de su entrenamiento y las que surgen de la interacción del agente con otros agentes y con el mundo de actuación.

## 1. Introducción

Dados ambientes desconocidos, los sistemas autónomos reales deben generar teorías de cómo sus ambientes reaccionan a sus acciones, y cómo las acciones afectan al ambiente. Usualmente estas teorías de aprendizaje son parciales, incompletas e incorrectas, pero pueden ser usadas para modificar esas teorías o para crear nuevas. Trabajos anteriores basados en aprendizaje de máquinas eran aplicados a resolución de problemas que principalmente estaban enfocados en aprendizaje de conocimientos cuya meta era mejorar la eficiencia de la tarea de resolución de problemas [Borrajo y Veloso 1997, Laier et al, 1986, ; Minton, 1988; Veloso, 1994].

Hay también, un actual interés en aprendizaje de estados de transición probabilísticos en el contexto de aprendizaje de refuerzo [Sutton, 1990; Watkins and Dayan, 1992]. Sin embargo pocos investigadores han abordado el problema de adquisición de operadores de manera generalizada [Carbonell and Gil, 1990; Wang, 1996], describen técnicas para la adquisición automática de descripciones generalizada de la teoría de un dominio.

Este problema es crucial cuando se está tratando con sistemas que deben "autonómicamente" adaptarse a un ambiente dinámico y desconocido. LOPE (Learning by Observation in Planning Environments) es una arquitectura de agente que integra planificación, aprendizaje y ejecución en un reiteración cerrada, mostrando cómo funciona el comportamiento de inteligencia autónoma [García-Martínez and Borrajo, 1997, 2000].

En nuestras investigaciones anteriores hemos presentado una arquitectura de agente simple. Este trabajo se concentra en el comportamiento de múltiples agentes bajo un ciclo de vida de aprendizaje sobre la base inicial de LOPE. Más precisamente, trabajaremos en el mecanismo de aprendizaje, generalización del mismo y una extensión de este a través de la demostración de cómo el conocimiento debe compartirse a través de muchos agentes.

## 2. Objetivo y Metodología de la Propuesta

Uno de los principales objetivos de cada agente LOPE [García-Martínez and Borrajo, 1997, 2000], conciste en determinar cómo los operadores aprenden automáticamente (modelos de acción), que predice los efectos de acciones en el ambiente de actuación del agente por observación de las consecuencias de sus acciones.

Para poder aprender estas descripciones, éste agente es capaz de lograr metas auto-definidas, ejecutar los planes, encontrar los comportamientos correctos o incorrectos, y aprender de la interacción con el ambiente y otros agentes.

Cada agente recibe la percepción desde el ambiente, llamado situaciones, aplica acciones y aprende desde su interacción con el mundo externo (el ambiente y otros agentes). Al principio, el agente percibe la situación inicial, y selecciona una acción al azar para ejecutar en el ambiente. Entonces, entra en un loop para la ejecución de una acción, percibiendo las situaciones resultantes y la utilidad de la situación, aprendiendo de la observación los efectos de aplicar las acciones en el ambiente y la planificación de próximas interacciones con el ambiente cuando el plan anterior ha finalizado su ejecución, o el sistema observa una diferencia entre la situación predicha por el operador del agente y la situación que es percibida desde el ambiente.

La Figura 1 muestra una vista esquemática de la arquitectura, donde puede haber n agentes LOPE. Cada uno de estos agentes recibe un input: percepciones desde el ambiente (situaciones y utilidades); un conjunto de acciones que pueden ejecutarse, y operadores aprendidos de otros agentes. La salida de cada agente es una secuencia de acciones sobre el tiempo (para el ambiente) y regularmente ese conjunto de operadores que él aprende (a través de otros agentes).

Sobre la base de LOPE (Learning by Observation in Planning Environments) [García-Martínez y Borrajo, 2000; García-Martínez et al., 2006], se trabaja en el ciclo de vida de aprendizaje del agente como aportación principal del presente trabajo, el que considera tres Layers: [a] el layer de BI (Built-In Operators) sobre la que se desarrolla el aprendizaje del agente a partir de los operadores implantados por el programador o creador, [b] el layer TB (Trained Base Operators) es la capa de aprendizaje en los que los operadores evolucionan en el contexto de entrenamiento del agente sobre la base inicial de los operadores implantados previamente y [c] el layer WI (World Interaction Operators) es la capa de aprendizaje en la que los operadores se aprenden por la interacción con el mundo y otros agentes. El ciclo de vida de aprendizaje del agente se muestra en la Figura 1.

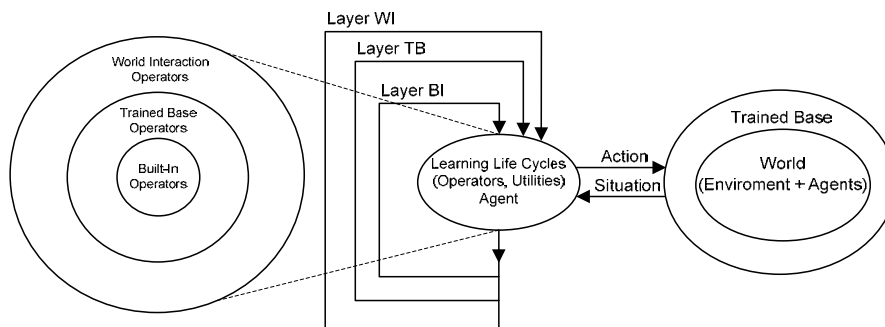


Figura 1. Ciclo de vida del aprendizaje del Agente

El sistema inteligente autónomo (AIS) "nace" con los operadores implantados por su programador. Estos representan a los operadores que conforman el conocimiento basal que permiten el comportamiento inicial de los agentes. La formación adquirida por los operadores facilita la evolución de éstos, mediante un mecanismo de refuerzo de los exitosos y de castigo sobre aquellos operadores que tuvieron mal funcionamiento.

Uno de los principales objetivos de la propuesta LOPE-LLC (LOPE-Learning Life Cycle) es que el agente aprenda autónomamente sobre la base de los operadores (modelos de acción) que predicen los efectos de las acciones en el medio ambiente mediante la observación de las consecuencias de esas acciones, conocer el comportamiento correcto o incorrecto, y aprender de los operadores [a] BI (Built-In Operators) iniciales, reforzados por la generación de los operadores de [b] TB (Trained Base Operators) y enriquecidos por el intercambio de operadores [c] WI (World Interaction Operators), basados en la interacción del agente con el mundo de actuación y otros agentes. La

Figura 2 muestra una vista esquemática de la arquitectura para la actuación de n-LOPE LLC agentes, que aprenden.

El ciclo de vida en estudio considera el aprendizaje de los agentes sobre los layers correspondientes a los operadores (BI, TB, WI) los que reciben como entrada: la percepción del medio ambiente, el conjunto de acciones que pueden ejecutar, y los operadores obtenidos de otros agentes por compartición de conocimientos. La salida de cada agente es una secuencia de acciones en el tiempo (que se ejecutan en el ambiente de actuación del agente) y con regularidad el conjunto de los operadores que ha aprendido (para compartir con los demás agentes).

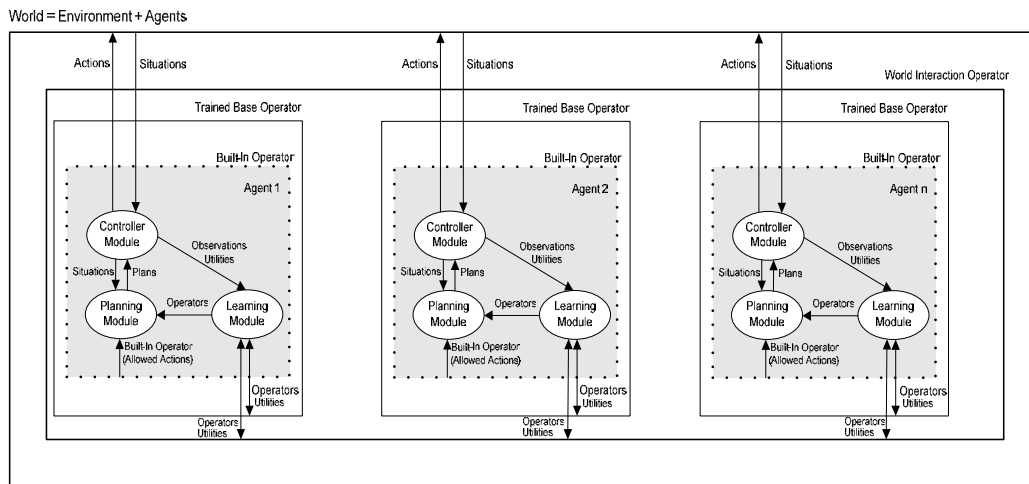


Figura. 2. Arquitectura de agentes LOPE-LLC

### 3. Resultados Esperados

Se llevará a cabo la experimentación del ciclo de vida propuesto que se muestra en la Figura 3 sobre la base de LOPE-LLC, considerando la compartición de operadores entre agentes y la valoración de éstos como eje principal del trabajo en el contexto de actuación de agentes simulados y reales. Para poder mejorar la convergencia de aprendizaje y para probar la generalización del conocimiento compartido, se llevarán a cabo experimentos en los cuales los sistemas se acuerdan de los operadores aprendidos con estos experimentos; mostraremos cómo el conocimiento previo brinda una más rápida convergencia de aprendizaje que cuando no se lo usa.

En la configuración de múltiples agentes, los agentes compartirían su conocimiento entre ellos. Se espera comparar aquí distintos experimentos, partiendo con un agente solo aprendiendo en un ambiente (grilla single), en la cual los operadores son generalizados, para luego experimentar con un agente solo LOPE-LLC aprendiendo de una grilla single donde un estimador probabilístico es asignado a cada operador. Las decisiones del agente estarán basadas en entrada de sensores solamente cuando no hay plan de ejecución. Se continuará con la experimentación sobre la base de la actuación de grupos de agentes LOPE-LLC con la misma configuración de grilla con una estrategia de compartimiento completa y donde un estimador de probabilidad es asignado a cada operador para asignar un grado de confianza a los planes generados, y los planes con baja confianza son descartados. Se realizarán experimentos conformados por grupos de agentes LOPE-LLC aprendiendo al mismo tiempo con una estrategia de compartición de operadores completa.

La experimentación se llevará a cabo con la aplicación de los primeros resultados sobre la base del simulador de Khepera, para extenderlo en ambientes de actuación real con robots Khepera, y en otros ambientes de robots y bípedos de bajo costo [Ierache, J., Bruno, M., Mazza, N., Dittler, M., 2008]. En estas aplicaciones se trabajará en conjunto con el Instituto de Investigaciones en Informática LIDI de la Universidad Nacional de La Plata.

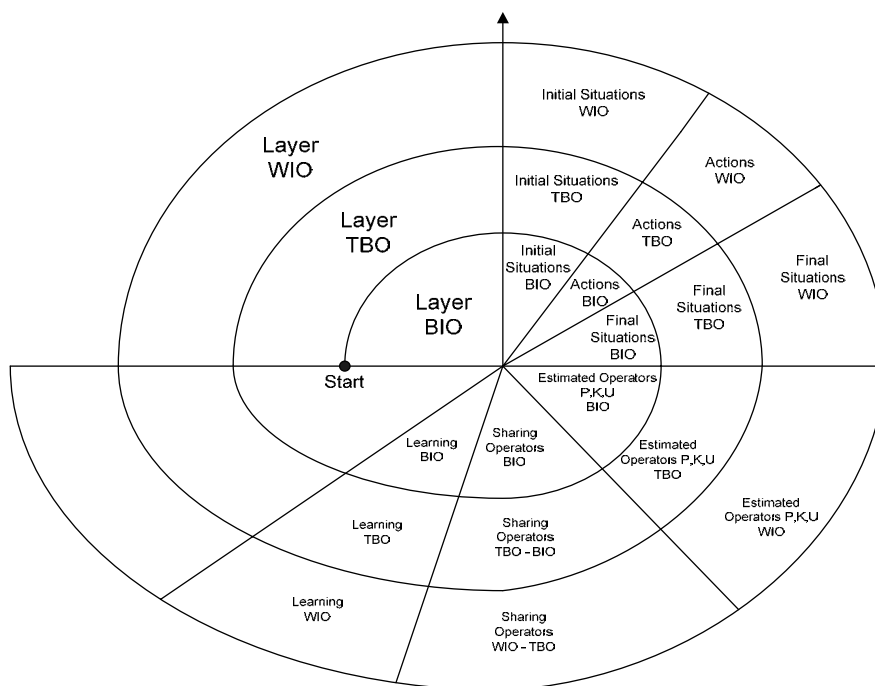


Figura 3. Ciclo de vida de aprendizaje propuesto

Para llevar adelante el proyecto se desarrollará un framework basado en la propuesta para el aprendizaje y compartición de conocimiento entre sistemas inteligentes autónomos distribuidos [Ierache, J., Naiouf, M., García Martínez, R., De Giusti A., 2008], para experimentar con la inclusión de nuevos agentes de un mismo tipo, aprendiendo y compartiendo lo que ellos aprendieron en la misma configuración de grilla y prueba, a fin de evaluar cómo esto afecta el comportamiento de aprendizaje y planificación. Sobre el framework cada agente continuamente aprende, planifica y ejecuta, como también se podrán configurar distintas estrategias de comunicación entre los agentes, para poder intercambiar lo que han aprendido, y las descripciones del operador. Conforme al desarrollo del trabajo se espera realizar publicaciones en congresos y revistas de la especialidad con los resultados de investigación obtenidos.

#### 4. Formación de Recursos Humanos

En la línea de investigación cuyos resultados parciales se reportan en esta comunicación, se encuentran trabajando: un tesista de doctorado.

#### 5. Referencias

Borrajo, D. and Veloso, M. (1997). Lazy incremental learning of control knowledge for efficiently obtaining quality plans. *AI Review Journal*, 11, 371-405.

Carbonell, J. G. and Gil, Y. (1990). Learning by experimentation: The operator refinement method. In (Michalski and Kodratoff, eds) *Machine Learning: An AI Approach*, Vol. III (pp. 191-213). San Francisco: Morgan Kaufmann.

García Martínez, R. y Borrajo, D. (2000). An Integrated Approach of Learning, Planning and Executing. *Journal of Intelligent and Robotic Systems*, 29, 47-78.

García-Martínez, R. and Borrajo, D. (1997). Planning, learning, and executing in autonomous systems. *Lecture Notes in Artificial Intelligence*, 1348, 208-220.

García-Martínez, R., Borrajo, D., Britos, P. y Maceri, P. Learning by Knowledge Sharing in Autonomous Intelligent Systems *Lecture Notes in Artificial Intelligence* Vol. 4140. Pág. 128-137. Springer-Verlag. 2006

- Ierache, J., Bruno, M., Mazza, N., Dittler, M., “Robots y Juguetes Autónomos una Oportunidad en el Contexto de las Nuevas Tecnologías en Educación”, VII Jornadas Iberoamericanas de Ingeniería de Software e Ingeniería del Conocimiento IIISIC Febrero 2008 Guayaquil Ecuador, , Proceedings VII Ibero-American Symposium on Software Engineering. Pag 371-379.
- Ierache, J., Naiouf, M., García Martínez, R., De Giusti A., “Un modelo de arquitectura para el aprendizaje y compartición de conocimiento entre sistemas inteligentes autónomos distribuidos” VII Jornadas Iberoamericanas de Ingeniería de Software e Ingeniería del Conocimiento IIISIC 2008 Guayaquil Ecuador. Proceedings VII Ibero-American Symposium on Software Engineering Pag 179-187
- Sutton, R. (1990). Integrated architectures for learning, planning, and reacting based on approximating dynamic programming. Proc. 7th Int. Conf. on ML: 216-224. Kaufmann.