Agentes y aprendizaje de máquina en optimización, control y entornos de enseñanza

Guillermo Aguirre Marcelo Luis Errecalde Guillermo Leguizamón

LIDIC *
Departamento de Informática
Universidad Nacional de San Luis(UNSL)
Ejército de los Andes 950 - Local 106
5700 - San Luis. Argentina
e-mail:{gaguirre,merreca,legui}@unsl.edu.ar

Resumen

Este trabajo describe el estado actual de las tareas de investigación realizadas en forma conjunta por un grupo de investigadores pertenecientes al LIDIC.

Las problemáticas abordadas por los integrantes de este grupo son disímiles y abarcan temas como optimización, control y entornos de enseñanza. Sin embargo, tienen como elementos en común la utilización de agentes y el aprendizaje de máquina.

Un aspecto interesante de este trabajo conjunto es que permitió detectar problemáticas comunes que son abordadas con técnicas diferentes. Estas técnicas son usualmente investigadas en forma independiente y existen pocos trabajos que analizan sus similitudes y diferencias. A nuestro entender esto constituye una seria limitación ya que un mejor entendimiento de estos aspectos permitiría que cada una de estas áreas se nutriera con conceptos y resultados desarrollados en las restantes.

Con este objetivo en mente, estamos actualmente trabajando en forma conjunta en tres líneas principales de investigación que podemos referenciar en términos generales como: a) agentes, aprendizaje y optimización, b) control de agentes mediante técnicas de aprendizaje híbridas y c) aprendizaje en agentes de interfaz para entornos de enseñanza. Cada una de estas líneas de investigación se describen en forma general en este trabajo.

Palabras Claves: Agentes Inteligentes, Aprendizaje de Máquina, Optimización, Control, Entornos de Enseñanza-Aprendizaje.

1 Agentes, aprendizaje y optimización

La técnica ACO (Ant Colony Optimization) ha surgido recientemente como una nueva meta-heurística para atacar problemas duros de optimización combinatoria. Los algoritmos ACO se definen como instancias de la meta-heurística ACO, los cuales son básicamente un sistema multiagente donde la interacción a muy bajo nivel de sus agentes componentes (llamados hormigas

*El laboratorio es dirigido por el Dr. Raúl Gallard y subvencionado por la UNSL y la ANPCYT (Agencia Nacional para la Promoción de la Ciencia y la Tecnología)

artificiales) resulta en un comportamiento complejo del sistema en su totalidad. Dichos algoritmos han sido inspirados por las colonias de hormigas reales [5] las cuales depositan una substancia química llamada feromona sobre el terreno por donde caminan. Esta substancia influye en la elección que ellas hacen al momento de decidir que paso o camino seguir en su ruta de exploración (generalmente hacia y desde la fuente de alimento), así, mientras mayor sea la acumulación de feromona sobre un paso particular, mayor será la probabilidad de elegir dicho paso. De esta manera, las hormigas artificiales de un algoritmo ACO se comportan de manera similar.

Las primeras aplicaciones de la meta-heurística ACO están vinculadas con el Problema del Viajante de Comercio [5, 7, 6, 8], siendo tal vez el problema más estudiado en el contexto de estos algoritmos. Sin embargo, existe un importante conjunto de estudios experimentales que incluye los siguientes problemas: Problema de Asignación Cuadrática, Ruteo, Coloreo de grafos, Redes de Telecomunicación, Scheduling [5], como así también, el Problema de Múltiples Mochilas, Máximo Conjunto Independiente [15, 16], entre otros.

Es importante destacar que los algoritmos ACO pueden fácilmente aplicarse a problemas de optimización discreta que puedan ser caracterizados como grafos (aunque hay excepciones demostradas para otro tipo de problemas [15, 16]). Así, la solución a un problema de optimización puede ser expresada en término de pasos factibles sobre dicho grafo.

Uno de los principios más importantes que guían la búsqueda de estos algoritmos está directamente vinculado con la cooperación indirecta, denominada stigmergy, de los agentes que conforman el sistema. Dicha cooperación se refleja a través del depósito de una cierta cantidad de feromona (refuerzo) sobre los pasos transitados por los agentes en su recorrido por el grafo proporcional a la calidad de la solución encontrada por cada uno de los agentes. La cantidad acumulada de feromona durante los sucesivos ciclos de ejecución del algoritmo conformará el bloque de construcción a partir del cual nuevas soluciones puedan ser generadas. En consecuencia, dicho bloque de construcción determinará en subsecuentes iteraciones la región del espacio de búsqueda a ser considerada. Más allá que un algoritmo ACO incluya información específica del problema e inclusive alguna forma de hibridización para mejorar su performance, nuestro interés está principalmente centrado en los aspectos relacionados a la forma de actualización del refuerzo y la manera en que dicha información es usada al momento de decidir cual es el paso a seguir. Recientes trabajos de algoritmos ACO aplicados a problemas de scheduling [17, 18] están dirigidos principalmente a formas alternativas de usar la información proveniente de la cooperación (feromona sobre los pasos) para mejorar la performance del algoritmo, aunque aplicado a un problema de scheduling particular.

Nuestra propuesta de trabajo actual pretende enfocar y fundamentar el comportamiento de los algoritmos ACO desde la perspectiva del aprendizaje por refuerzo. Aun más, dada ciertas semejanzas entre los distintos enfoques, un estudio posible está relacionado con la transferencia de resultados desde el área de aprendizaje por refuerzo que puedan mejorar y/o entender con mayor profundidad el funcionamiento de los algoritmos ACO.

Con respecto a este último punto, ya hemos realizado una experiencia similar estudiando las similitudes entre Sistemas Clasificadores (SC) y el método de AR denominado Q-Learning. A partir de este estudio logramos adaptar otros métodos de AR (p. ej. SARSA) al contexto de los SC. Esto permitió derivar un nuevo esquema dentro de los SC, que además de lograr un buen comportamiento tiene garantías de convergencia al óptimo que ya han sido demostradas con SARSA en el área del AR.

2 Control de agentes mediante técnicas de aprendizaje híbridas

En el grupo de investigación se han realizado distintos trabajos que utilizan el Aprendizaje por Refuerzo (AR) como mecanismo principal para resolver el problema de controlar agentes autónomos [10, 12, 9, 11, 20].

El AR se basa en la idea central del aprendizaje *por prueba y error* [21], un concepto originalmente analizado en estudios relacionados con el comportamiento animal.

En su concepción más pura el AR plantea, como solución al problema del control de agentes autónomos, distintas técnicas de aprendizaje que se basan exclusivamente en la señal de recompensa que recibe desde el ambiente. Dentro de este esquema, no se asume ningún tipo de conocimiento previo sobre el ambiente o la tarea a resolver. Estas características llevaron a que muchos investigadores plantearan al AR como una solución general para aquellos dominios desconocidos o poco entendidos para el diseñador de un agente artificial.

En este contexto, una pregunta interesante que se plantea es: ¿puede un agente de AR, con tan poca información disponible, escalar a problemas generales y complejos del mundo real?

En este sentido, trabajos relacionados a la complejidad del AR elemental, han demostrado que el problema de AR es en general intratable si no se recurre a variantes que permiten incorporar conocimiento en el proceso de exploración del ambiente.

Por otro lado, no es realista pensar que todos los ambientes en que se desempeñará el agente deban ser completamente desconocidos. En algunos casos puede existir información previa disponible que sirva para un mejor comportamiento inicial del agente. Esta información cumplirá un rol similar al de los *reflejos*, que permiten mantener orientado al agente en la dirección correcta mientras intenta aprender.

En otros casos, el agente podrá convivir con otros agentes y el proceso de aprendizaje puede involucrar el descubrimiento por prueba y error, pero también cumple un rol fundamental la información obtenida desde otros agentes con quienes convive. En estas situaciones, y en concordancia con lo expresado en [22], *el aprendizaje es más a menudo una cuestión de transferencia que de descubrimiento.* Dentro de este contexto social, el agente puede mejorar su comportamiento mediante distintos procesos de aprendizaje. Una forma de hacerlo es observando e *imitando* el comportamiento de otros agentes más experimentados. Los otros agentes también pueden actuar como *consejeros* o *instructores* quienes proveen consejos, órdenes de ejecución obligatoria o simplemente ejemplos del comportamiento deseado. Es interesante notar que estas formas de aprendizaje, ya han sido analizadas en distintos trabajos como una forma de extender y superar las limitaciones del AR. Sin embargo, estos enfoques han sido incorporados por separado y sin un marco general que permita integrarlos en forma coherente a todos ellos.

En este sentido, nuestra hipótesis actual de trabajo, plantea que para construir agentes artificiales basados en aprendizaje, que se desempeñen en forma efectiva en la resolución de problemas complejos, es necesario extender el modelo del AR con nuevas formas de experiencia que puedan servir para mejorar el proceso de aprendizaje. Nuesto punto de partida, consiste en considerar al aprendiz como parte de un contexto social donde puede aprender no sólo desde las recompensas recibidas por sus acciones, sino también desde las múltiples fuentes de conocimiento que puedan estar socialmente disponibles, incluyendo agentes humanos y artificiales.

Nuestro objetivo es poder integrar en forma coherente en el proceso de aprendizaje estas múltiples fuentes de información a las que denominaremos *fuentes de experiencia.* En el contexto del aprendizaje de políticas de control, es decir, el problema de aprender a mapear situaciones en

acciones, consideramos como *fuente de experiencia* a *cualquier dispositivo computacional que sirva para expresar sus preferencias sobre qué acción tomar en distintas situaciones.* Este dispositivo podrá tomar la forma de un sistema de aprendizaje por refuerzo, de aprendizaje supervisado, de aprendizaje por consejos, un planner deliberativo u otro cualquiera que sirva en la toma de decisiones sobre la acción a tomar.

Para poder integrar las diversas fuentes de experiencia en el proceso de decisión del agente, se pueden identificar al menos 3 tareas fundamentales a realizar:

1. **Definición de un lenguaje común de preferencias de acciones:** La salida de cada fuente, deberá ser traducida a este lenguaje común de preferencias antes de poder ser utilizadas en la toma de decisiones del agente.

2. **Definición de los traductores para cada fuente de experiencia:** Para cada fuente de experiencia se deberá asociar un traductor particular que convierta la salida producida por la fuente en el lenguaje común de preferencias de acciones.

3. **Definición de un mecanismo de consenso:** Las preferencias expresadas por las distintas fuentes en el lenguaje común deberán ser integradas en forma coherente para lograr como resultado una única política de control del agente. Será necesario por lo tanto, definir un mecanismo que establezca las reglas mediante las cuales se realiza esta integración.

Actualmente se ha definido un lenguaje común de preferencias de acciones que permite expresar políticas de control estocásticas. También se ha definido un mecanismo de consenso híbrido que combina un esquema de votación ponderado con un sistema jerárquico dictatorial para aquellas situaciones que requieren un tratamiento prioritario.

3 Aprendizaje en agentes de interfaz para entornos de enseñanza

La investigación sobre esta temática gira principalmente en torno de la problemática del diseño y la realización de una interfaz que facilite y haga más efectiva la interacción Hombre-Máquina, logrando comprender y controlar relaciones establecidas para resolver problemas en forma colaborativa sobre entornos virtuales basados en texto (MOO's). Los principales tópicos de investigación abarcan los agentes de interfaz, trabajo colaborativo soportado por computadoras y aprendizaje automático [4, 2]. El trabajo es abordado en el marco de conocimiento distribuido considerando varias aproximaciones que otorgan a los aspectos cognitivos y sociales una ponderación particular considerando que el medio de comunicación utilizado es la computadora. Esta ponderación particular implica emplear enfoques especiales para los términos *conocimiento* y *colaboración*: no se restringue a un conocimiento *en la cabeza* sino que se incluyen varios agentes y los elementos que estos emplean para hacer su tarea. En relación al trabajo colaborativo se lo considera como una actividad sincrónica y coordinada que es el resultado de un intento contínuo de construir y mantener una concepción compartida de un problema. Esto se consigue creando una estructura de conocimiento compartida que integre metas, descripciones del estado actual del problema, conocimiento de acciones disponibles para la resolución de problemas y asociaciones que vinculan estos tres elementos. Esta estructura de conocimiento es una pieza central para conseguir un comportamiento como el esperado. El trabajo colaborativo sólo puede progresar en la medida que se logre

mantener y aumentar una plataforma de conocimiento compartida (ground) construida a partir de coincidencias establecidas entre los pares que intervienen en la actividad colaborativa [19].

El medio de comunicación condiciona el volúmen del ground requerido; así en el caso de la comunicación a través de computadoras, se requiere un ground más grande que en una comunicación cara a cara. En nuestro caso los recursos de software que los usuarios pueden emplear para desarrollar su tarea son: un MOO en el cual se desarrollan comunicaciones basadas en texto y alguna aplicación específica de acuerdo a la problemática sobre la que se está trabajando [3]. A partir de la observación de las actividades que los pares realizan mientras usan estos recursos de software una interfaz inteligente puede construir el ground necesario para permitir una labor colaborativa. La interfaz, por medio de agentes computacionales irá monitoreando a los usuarios y recolectando información.

La aproximación socio-cultural al conocimiento humano ha ganado influencia en la tecnología educativa siguiendo la corriente encabezada por Vigotsky. Es así que hoy en día se otorga bastante importancia al papel que juegan los compañeros y los docentes en el aprendizaje. En particular la posibilidad de trabajar con ellos colaborativamente, usando computadoras como soporte, es una manera de lograr aprendizajes con un mínimo de enseñanza [1]. El aspecto social es muy importante en este tipo de entornos de aprendizaje por lo que nuestro trabajo le concede un lugar primordial. Usando técnicas de aprendizaje automático se pretende conseguir conocimiento sobre la forma en la cual los grupos interactúan y sobre el desempeño de cada usuario individualmente. El centro de la actividad social será el MOO, desde allí se puede conseguir información sobre los encuentros entre pares, los horarios y lugares dentro del MOO en que han coincidido, etc. Nuestro trabajo toma en consideración los logros conseguidos por el agente **Cobot**, que recolecta información social dentro de *LambdaMOO* [14, 13] para conseguir aprender cual es el comportamiento que prefieren los usuarios. Las acciones que Cobot realiza por su propia iniciativa resultan de lo aprendido sobre las preferencias de los usuarios que en determinado momento estén conectados al MOO. La técnica de aprendizaje por refuerzo fue escogida por las particulares características que presenta el entorno. En el MOO, las fuentes de aprendizaje son múltiples, ocasionalmente los usuarios se comportan de manera inconsistente o contradictoria, los experimentos son irreproducibles ya que se consideran entre otras cosas quienes son los usuarios conectados, la mayoría de los usuarios pierde interés en proveer rewards a Cobot por lo que las chances de aprendizaje son escazas y surge la posibilidad que un grupo reducido de usuarios condicione el comportamiento de Cobot. Por estas razones el algoritmo de aprendizaje considera para cada usuario una función de valuación individual basada sobre su feedback particular, así dentro del espacio de estado considerado se combinan las funciones de los usuarios presentes.

4 Conclusiones

Esta presentación describe los trabajos que actualmente se están realizando en tres temáticas en apariencia diferentes pero que, sin embargo, tienen características en común. Estas incluyen el uso de agentes, aprendizaje de máquina y los aspectos sociales y colaborativos de la sociedad de agentes como un todo. A partir de algunas experiencias que hemos realizado en base a este trabajo interdisciplinario hemos podido transferir resultados teóricos y prácticos entre las áreas que han llevado a un mejor entendimiento y aprovechamiento de trabajos realizados en forma aislada.

Es nuestro objetivo seguir aplicando la misma metodología de trabajo para poder justificar técnicas existentes u obtener nuevas propuestas tanto en las áreas de optimización, control como de entornos de enseñanza.

Referencias

[1] G. Aguirre and M. Lucero. Aprendizaje en entornos virtuales basados en texto. In *Congreso de Informatica en las universidades InfoUni'2001*, 2001.

[2] Guillermo Aguirre. Creacion de entornos de aprendizaje en mundos virtuales. In *CACiC 2001*, pages 15–26, 2001.

[3] Guillermo Aguirre. Interfaces inteligentes para resolución de problemas en entornos virtuales. In *WICC 2001*, pages 80–82, 2001.

[4] C. Buiu and G. Aguirre. Learning interface for collaborative problem solving. In *Advanced Research in computers and comunications in education*, pages 301–303, 1999.

[5] D. Cornea, M. Dorigo, and F. Glover, editors. *New Ideas in Optimization*. McGraw-Hill International, 1999.

[6] M. Dorigo, G. Di Caro, and L.M. Gambardella. Ant Algorithms for Discrete Optimization. *Artificial Life*, 5(2):137–172, 1996. Also available as Tech. Rep. IRIDIA/98-10, Universitè Libre de Bruxelles, Belgium.

[7] M. Dorigo and L.M. Gambardella. Ant Colony System: A Cooperative Learning Approach to the Traveling Salesman Problem. *IEEE Transactions on Evolutionary Computation*, 1(1):53–66, 1997.

[8] M. Dorigo, V. Maniezzo, and A. Colorni. Positive feedback as a search strategy. Technical Report Tech. Rep. No. 91-016, Politecnico di Milano, Italy, 1991.

[9] Marcelo Errecalde. Marcos teóricos del aprendizaje por refuerzo multiagente. In *Proceedings del Workshop de Investigadores en Ciencias de la Computación 2001 - WICC 2001*, pages 77–79, San Luis, Argentina, 2001.

[10] Marcelo Errecalde, Maria Liz Crespo, and Cecilia Montoya. Aprendizaje por refuerzo: Un estudio comparativo de sus principales métodos. In *Proceedings del Segundo Encuentro Nacional de Computación*, México, 1999.

[11] Marcelo Errecalde and Alfredo Muchut. Exploración dirigida por el objetivo en aprendizaje por refuerzo basado en modelo para ambientes no estacionarios. In *Proceedings del Congreso Argentino de Ciencias de la Computación 2001 - CACIC 2001*, pages 1117–1129, El Calafate, Argentina, 2001.

[12] Marcelo Errecalde, Alfredo Muchut, Guillermo Aguirre, and Montoya Cecilia. Aprendizaje por refuerzo aplicado a la resolución de problemas no triviales. In *Proceedings del Workshop de Investigadores en Ciencias de la Computación 2000 - WICC 2000*, pages 49–51, La Plata, Argentina, 2000.

[13] Charles Lee Isbell, Chistian Shelton, and Peter Stone Michael Kearns, Satinder Singh. Cobot in lambdamoo: A social statistics agent. In *Fifth International Conference on Autonomous Agents*, 2000.

[14] Charles Lee Isbell, Chistian Shelton, and Peter Stone Michael Kearns, Satinder Singh. A social reinforcement learning agent. In *Fifth International Conference on Autonomous Agents*, 2000.

[15] G. Leguizamón and Z. Michalewicz. A New Version of Ant System for Subset Problems. In *Proceedings of the 1999 Congress on Evolutionary Computation*, pages 1459–1464. IEEE Press, Piscataway, NJ, 1999.

[16] G. Leguizamón and M. Schütz. An ant system for the maximum independent set problem. In *VII Congreso Argentino de Ciencias de la Computación*, El Calafate, Santa Cruz, Argentina, 2001.

[17] D. Merkle and M. Middendorf. An ant algorithm with a new pheromone evaluation rule for total tardiness problems. In *Proceeding of the EvoWorkshops 2000*, number 1803 in Lectures Notes in Computer Science, pages 287–296. Springer Verlag, 2000.

[18] D. Merkle, M. Middendorf, and H. Schmeck. Ant colony optimization for resource constrained project scheduling. In D. Whiley et al., editor, *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2000)*, pages 893–900, Las Vegas, Nevada, USA, 10-12 July 2000. Morgan kaufmann.

[19] D. Traum P. Dillenbourg and D. Shneider. Grounding in multi-modal task-oriented collaboration. In *EuroAI and Education Conference*, 1996.

[20] Marcela Printista, Marcelo Errecalde, and Cecilia Montoya. Una implementación paralela del algoritmo de q-learning basada en un esquema de comunicación con caché. In *Proceedings del Congreso Argentino de Ciencias de la Computación 2000 - CACIC 2000*, Ushuaia, Argentina, 2000.

[21] Richard S. Sutton and Andrew G. Barto. *Reinforcement learning: an introduction*. The MIT Press, 1998.

[22] Steven D. Whitehead. A complexity analysis of cooperative mechanisms in reinforcement learning. In *Proceedings of the AAAI*, pages 607–613, 1991.