Modelos de decisión y optimización en presencia de adversarios: estudios y resultados hasta el momento

Doctorando: Pablo J. Villacorta
Director: Dr. D. Jose Luis Verdegay
Co-direcor: Dr. David A. Pelta

Grupo de Modelos de Decisión y Optimización (MODO),
Departamento de Ciencias de la Computación e IA,
ETSIIT - Universidad de Granada, 18071, Granada
{pjvi, verdegay, dpelta}@decsai.ugr.es

Abstract. La toma de decisiones, la optimización, y el razonamiento en presencia de adversarios dan lugar a una serie de problemas donde los comportamientos y las maneras de decidir que se deben contemplar son muy diferentes a las que se utilizarían cuando estos no existen. Esto es así puesto que en ocasiones se vuelve necesario recurrir a decisiones sub-óptimas solo con objeto de confundir a los adversarios. Se han realizado ya estudios sobre un modelo simple en el que se investigan métodos automáticos para la obtención de nuevas estrategias y aproximaciones analíticas sobre la eficacia de las mismas basadas en Teoría de la probabilidad.

1 Introducción

La toma de decisiones en presencia de adversarios plantea dificultades propias de un contexto que a veces exige recurrir a decisiones sub-óptimas solo con objeto de confundir a los adversarios. Este tipo de situaciones surgen claramente en el ámbito militar, pero también en áreas como la vigilancia de perímetros, desarrollo de juegos de ordenador, diseño de sistemas inteligentes para el entrenamiento de personal, ciber crimen, etc. Todos estos escenarios son de actualidad, pero además son interesantes desde un punto de vista científico y relevantes desde un punto de vista práctico.
En términos generales, un adversario es una entidad cuyos beneficios (en algún sentido) son inversamente proporcionales a los nuestros. Este adversario es capaz de alterar nuestros beneficios tomando ciertas acciones y además, puede observar nuestras acciones/decisiones teniendo la oportunidad de aprender nuestro patrón de comportamiento. Este aprendizaje le llevará a ser más efectivo en su intento de maximizar sus beneficios y minimizar los nuestros.

2 Modelo básico

El modelo [5] consta de dos agentes S y T (el adversario), un conjunto de posibles entradas o eventos $E = \{e_1, e_2, \ldots, e_N\}$ generados por un tercer agente R, y un conjunto de posibles respuestas o acciones $A = \{a_1, a_2, \ldots, a_M\}$ asociadas a cada evento. Existe también una matriz de pagos P de dimensiones $N \times M$

Algorithm 1.1 Secuencia de pasos que forman el modelo

for l = 1 hasta L **do**
  Llega un nuevo estímulo e_i.
  El agente T hace una predicción a_g
  El agente S escoge una acción a_j
  Se calcula el pago para S
  El agente T registra el par e_i, a_j
end for

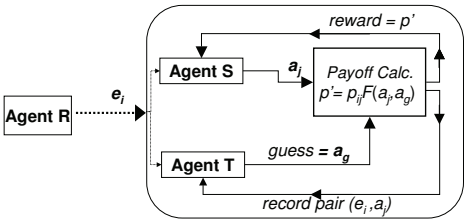

Fig. 1. Representación gráfica del modelo

tal que p_{ij} representa el pago obtenido por S al responder con la acción a_j ante el estímulo e_i. El agente S debe decidir qué acción elegir dada una entrada particular e_i y con un perfecto conocimiento de la matriz de pagos P. Su objetivo es maximizar la suma de los pagos obtenidos tras una secuencia de entradas o estímulos. El agente T no conoce la matriz de pagos P, pero está observando el comportamiento de S con el fin de aprender de sus acciones. Su objetivo es reducir el beneficio del agente S adivinando qué acción elegirá como respuesta al estímulo de la secuencia recibido en cada instante. El algoritmo 1.1 y la Fig. 1 describen estos pasos. L representa la longitud de la secuencia de estímulos.
Ante un estímulo e_i, S elige una acción y al mismo tiempo el agente T da una predicción sobre la acción que escogerá S. Además, T mantiene su propia matriz de observaciones, O, de dimensiones $N \times M$. O_{ij} representa el número de veces que, hasta el momento actual, el agente S decidió tomar la acción a_j cuando el estímulo era e_i. El cálculo de la recompensa para S se define como:

$$p' = p_{ij} \cdot F(a_g, a_j) \qquad \text{donde} \qquad F(a,b) = \begin{cases} 0 & \text{si a = b} \\ 1 & \text{en otro caso} \end{cases}$$

Esto significa que el agente S no obtiene ninguna recompensa cada vez que T consiga adivinar correctamente su respuesta. El agente S puede comportarse de manera totalmente determinista, de manera totalmente aleatoria, o siguiendo cualquier patrón intermedio. Lo mismo ocurre con T, que basa su comportamiento en su matriz de observaciones, seguramente también de un modo estocástico.

3 Estudios realizados hasta el momento

Diseño automático. Asumiendo una determinada estrategia para el adversario T, es posible plantear la búsqueda de buenas estrategias para S como un problema de optimización combinatoria en un espacio de estados adecuados. Dicho problema puede ser resuelto mediante metodos heurísticos, en los cuales la función de

fitness conlleva la ejecución de una simulación (algoritmo 1.1). Esta simulación es no determinista ya que las estrategias contienen un componente de aleatoriedad, lo que significa que podemos obtener distintos valores de fitness al evaluar varias veces una misma estrategia. Ello da lugar al problema de la comparación fiable de dos individuos, tarea necesaria en cualquier proceso de optimización. Se han propuesto técnicas basadas en intervalos de confianza para reducir el número de simulaciones necesarias en una comparación, y se han utilizado con éxito en un Algoritmo Genético [8], en Búsqueda Local [7] y en Algoritmos basados en Colonias de Hormigas [9].

Estudios teóricos del pago esperado. Otra línea diferente ha consistido en calcular el pago esperado para S mediante expresiones matemáticas basadas en conceptos de Teoría de la probabilidad y que no requieren la ejecución de una simulación. Dichas expresiones han sido validadas empíricamente mediante simulaciones que demuestran que son correctas ya que la predicción coincide con el pago obtenido mediante la experimentación. Tienen la ventaja de que pueden ser sometidas a un proceso de optimización, bien analítico (programación no lineal) o bien heurístico pero mucho más rápido al ser utilizadas directamente como función fitness[1], en lugar de recurrir a simulaciones. Este enfoque se ha seguido en [12, 13, 10].

Estudios sobre manipulación estratégica. Estrechamente relacionado con lo anterior, la optimización del pago esperado ha hecho uso del concepto de manipulación del oponente, beneficiándose del hecho de que T está observando el comportamiento de S. Más concretamente, S va variando su comportamiento en ciertos momentos, lo cual en general no es detectable por T excepto que éste olvide las observaciones más antiguas. Así, S combina períodos donde su comportamiento está más orientado a acumular confusión para T, con otros en los elige las mejores acciones con la seguridad de que la manipulación llevada a cabo previamente le permitirán no ser adivinado correctamente. Este enfoque se ha utilizado en [10, 11, 13].

4 Conclusiones y trabajos futuros

La conexión de este modelo con otros similares empleados en patrullaje de perímetros mediante robots autónomos [1, 3] será investigada próximamente para eliminar en la medida de lo posible suposiciones poco realistas [2], como el perfecto conocimiento de la estrategia del guardián por parte del ladrón que previamente lo ha estado observando. La incertidumbre inherente a estos dominios puede ser abordada mediante conceptos de Lógica Difusa. La aplicación del modelo a otros campos también es actualmente objeto de estudio [6, 4].

Agradecimientos

Los trabajos mencionados se llevaron a cabo con el apoyo de los proyectos TIN2008 - 06872 - C04 - 04 y TIN2008 - 01948 del Ministerio de Ciencia e Innovación, y P07 - TIC2970 de la Junta de Andalucía.

[1] Esta idea se ha aplicado con éxito en un trabajo actualmente en preparación que empleará optimización mediante *Evolución diferencial*

References

1. F. Amigoni, N. Basilico, and N. Gatti. Finding the optimal strategies for robotic patrolling with adversaries in topologically-represented environments. In *Proceedings of the 26th International Conference on Robotics and Automation (ICRA'09)*, pages 819–824, 2009.
2. F. Amigoni, N. Basilico, N. Gatti, A. Saporiti, and S. Troiani. Moving game theoretical patrolling strategies from theory to practice: An USARSim simulation. In *Proceedings of the 27th International Conference on Robotics and Automation (ICRA'10)*, pages 426–431, 2010.
3. N. Basilico, N. Gatti, T. Rossi, S. Ceppi, and F. Amigoni. Extending algorithms for mobile robot patrolling in the presence of adversaries to more realistic settings. In *Proceedings of the International Conference on Web Intelligence and Intelligent Agent Technology (IAT'09)*, pages 557–564, 2009.
4. N. Basilico, D. Rossignoli, N. Gatti, and F. Amigoni. A game-theoretical model applied to an active patrolling camera. In *Proceedings of the 1st International Conference on Emerging Security Technologies (EST)*, 2010.
5. D. Pelta and R. Yager. On the conflict between inducing confusion and attaining payoff in adversarial decision making. *Information Sciences*, 179:33–40, 2009.
6. J. Pita, M. Jain, J. Marecki, F. Ordonez, C. Portway, M. Tambe, C. Western, P. Paruchuri, and S. Kraus. Deployed armor protection: The application of a game theoretic model for security at the los angeles international airport. In *Proceedings of Industry Track in the Seventh International Conference on Autonomous Agents and Multiagent Systems (AAMAS '08)*, 2008.
7. P. Villacorta and D. Pelta. Comparación estadística de estrategias de decisión en presencia de adversarios: análisis y aplicación en búsqueda local. In *Proceedings of the Congreso Español sobre Metaheuristicas, Algoritmos Evolutivos y Bioinspirados (MAEB'10)*, pages 127–134, 2010.
8. P. Villacorta and D. Pelta. Evolutionary design and statistical assessment of strategies in an adversarial domain. In *Proceedings of the IEEE Conference on Evolutionary Computation (CEC'10)*, pages 2250–2256, 2010.
9. P. Villacorta and D. Pelta. Ant colony optimization for automatic design of strategies in an adversarial model. In *Proceedings of the 5th International Workshop on Nature Inspired Cooperative Strategies for Optimization (NICSO 2011)*, 2011. In press.
10. P. Villacorta and D. Pelta. Expected payoff analysis of dynamic mixed strategies in an adversarial domain. In *Proceedings of the 2011 Symposium on Intelligent Agents. IEEE Symposium Series on Computational Intelligence (SSCI 2011)*, pages 116 – 122, 2011.
11. P. Villacorta, D. Pelta, and M. T. Lamata. Formulation of an OWA-based dynamic mixed strategy for an adversarial model. In *Proceedings of the 2011 World Conference on Soft Computing (WCSC'11)*, pages 203.1 – 203.7, 2011.
12. P. J. Villacorta and D. Pelta. Theoretical analysis of expected payoff in an adversarial domain. *Information Sciences*. In press, 2011.
13. P. J. Villacorta, D. Pelta, and M. T. Lamata. Forgetting as a way to avoid deception in a repeated imitation game. *Autonomous Agents and Multi-Agent Systems*. Sometido a revisión.