

Toma de Decisiones basadas en Argumentación

Edgardo Ferretti,[†] Marcelo Errecalde,[†] Alejandro García,[‡] Guillermo Simari[‡]

[†]Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (LIDIC)¹
Departamento de Informática
Universidad Nacional de San Luis
Ejército de los Andes 950 - Local 106
(D5700HHW) - San Luis - Argentina
Tel: (02652) 420823 / Fax: (02652) 430224
e-mail: {ferretti, merreca}@unsl.edu.ar

[‡]Laboratorio de Investigación y Desarrollo en Inteligencia Artificial (LIDIA)²
Departamento de Ciencias e Ingeniería de la Computación
Universidad Nacional del Sur
Av. Alem 1253, (B8000CPB) Bahía Blanca, Argentina
Tel: (0291) 459-5135 / Fax: (0291) 459-5136
e-mail: {ajg, grs}@cs.uns.edu.ar

Resumen

Este artículo describe, en forma resumida, parte de los trabajos de investigación y desarrollo que se están llevando a cabo en la línea “Agentes y Sistemas Multi-agente” del LIDIC, en conjunto con investigadores del LIDIA. El objetivo de este trabajo es presentar las principales temáticas que están siendo abordadas actualmente en el área de agentes cognitivos.

Uno de los objetivos principales de esta línea, es el estudio y desarrollo de modelos de coordinación para agentes que forman parte de un sistema multi-agente; asimismo, uno de los objetivos parciales del grupo de trabajo, es analizar la utilización de técnicas de argumentación en modelos de coordinación de alto nivel.

Un paso previo necesario para lograr la coordinación de múltiples agentes, es que cada agente perteneciente al sistema tome una decisión acerca de cual será su próxima acción a seguir. De esta manera, cuando un agente toma decisiones, se produce normalmente la presentación de argumentos a favor o en contra de diferentes alternativas. En consecuencia, el estudio y desarrollo de modelos de toma de decisiones que usan argumentación para seleccionar las alternativas candidatas, en sistemas multi-agente, es otra área de estudio relevante para nuestra línea de trabajo. Actualmente, este estudio se está llevando a cabo en el ámbito de la toma de decisiones individuales (sistemas con un único agente). El estudio se está abordando con un enfoque teórico/práctico que establece similitudes con modelos teóricos clásicos de toma de decisiones, y prevé aplicaciones en problemas complejos del mundo real. En particular, el énfasis estará puesto en problemas que involucren coordinación y toma de decisiones en dominios con un único y con múltiples robots.

¹Las investigaciones realizadas en el LIDIC son financiadas por la Universidad Nacional de San Luis y por la Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT).

²Las investigaciones realizadas en el LIDIA son financiadas por la Universidad Nacional del Sur, por la Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT) y por el Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET).

1. Introducción

La toma de decisiones, a menudo vista como una forma de razonamiento orientada a la acción, durante mucho tiempo ha sido objeto de estudio de filósofos, economistas, psicólogos y científicos de la computación, entre otros. Dada cierta información acerca del estado actual del mundo y las consecuencias de las acciones potenciales, cualquier problema de decisión equivale a seleccionar las “mejores” acciones o aquellas suficientemente “buenas” que son posibles entre las diferentes alternativas. Es importante destacar, que la información disponible puede ser incompleta o incierta. Además, la bondad de una acción es juzgada estimando, quizás por medio de múltiples criterios, cuantas de sus consecuencias posibles concuerdan con las intenciones del tomador de decisiones o satisfacen sus preferencias. Asimismo, se asume que el agente se comporta de manera racional [23], al menos en el sentido que sus decisiones deben ser consistentes con sus preferencias. Sin embargo, se podría tener un punto de vista de la racionalidad, que requiriera que el comportamiento del tomador de decisiones estuviera conforme a postulados que establecen como debe comportarse un agente racional [17].

Los problemas de decisión han sido considerados desde diferentes puntos de vista, de los cuales se pueden distinguir dos tendencias principales que están influenciando actualmente la investigación en *Inteligencia Artificial (IA)*; estas son: la *teoría de decisión clásica* y los enfoques de decisión *BDI (Beliefs-Desires-Intentions)* y de *razonamiento práctico*.

1.1. Teoría de Decisión Clásica

La teoría de decisión clásica, desarrollada mayormente por economistas, se ha centrado principalmente en explicitar cuándo un tomador de decisiones es racional. En consecuencia, se han elaborado principios para comparar distintas alternativas. Un principio de decisión particular como el de la *utilidad esperada clásica*, debe ser justificado sobre la base de un conjunto de postulados de racionalidad, que deben ser satisfechos por la relación de preferencia existente entre las acciones. Esto significa, que en este enfoque la *racionalidad* es capturada a través de un conjunto de postulados que describen, qué es un comportamiento de decisión racional.

Este enfoque tiene como entradas un conjunto de acciones candidatas y una función que calcula el valor de sus consecuencias, cuando las acciones son realizadas en un estado dado junto con información parcial o completa acerca del estado actual del mundo. La salida, es una relación de preferencia entre acciones. Es importante destacar, que tal enfoque tiene como objetivo hacer un ranking del grupo de acciones candidatas, en vez de centrarse en una acción individual. Además, se presupone que las acciones candidatas son factibles de llevarse a cabo.

Realizando una sobre-simplificación, se pueden distinguir entonces dos grupos de trabajo en IA que siguen este enfoque de toma de decisiones. El primer grupo estaría representado por investigadores que usan *Redes Bayesianas* [19] y que trabajan en planeamiento bajo incertidumbre, por ejemplo [8]. Además, algunos trabajos en IA han tenido como objetivo desarrollar frameworks de decisión cualitativos, pero aún siguiendo la línea de pensamiento de la teoría de decisión clásica, entre ellos pueden citarse [9, 22].

1.2. Enfoques de Decisión BDI y de Razonamiento Práctico

Otros investigadores en IA que trabajan en razonamiento práctico, comenzando con la pregunta “¿cuál es la acción correcta que un agente debe realizar en una situación determinada?” [20, 21], han propuesto un proceso de dos pasos para responder a esta pregunta: el primer paso, a menudo llamado *deliberación* [23] consiste en identificar las metas del agente; en el segundo paso, se buscan maneras de lograr esas metas, es decir planes, y en consecuencia metas intermedias y sub-planes. Este

tipo de enfoque trae a colación aspectos tales como: ¿cómo se generan las metas?, ¿son las acciones posibles?, ¿tienen las acciones efectos indeseables?, ¿son los sub-planes compatibles?, ¿existen planes alternativos para lograr una meta dada?, etc.

En [5, 6], se ha argumentado que esto puede hacerse representando los estados cognitivos del agente; a saber, sus creencias, deseos e intenciones (arquitectura BDI). Esto requiere un enfoque expresivo en cuanto a la representación de conocimiento y preferencias, que contrasta con el enfoque de decisión clásica que usa directamente una distribución de incertidumbre (una distribución de probabilidad en el caso de la utilidad esperada) y una función de utilidad (valor). Además, el paso de deliberación es meramente un problema de inferencia, dado que equivale a encontrar un conjunto de deseos que están justificados sobre la base del estado actual del mundo y sobre los deseos condicionales. De hecho, chequear si un plan es factible y que no nos guiará a consecuencias indeseadas es aún un problema de inferencia. Un problema de decisión ocurre solamente cuando se tienen varios planes o sub-planes posibles y uno de ellos debe ser elegido.

2. Tareas en Progreso y Trabajos Futuros

La *argumentación* es un modelo de razonamiento basado en la construcción y evaluación de argumentos que interactúan entre sí. Estos argumentos tienen como fin, soportar, explicar o atacar enunciados que pueden ser opiniones, decisiones, etc. La argumentación ha sido utilizada en diferentes dominios; como razonamiento no-monótono [7, 10], manejo de inconsistencias en bases de conocimiento [1, 2] y modelado de diferentes tipos de diálogos, en particular persuasión [3] y negociación [4]. Un enfoque de negociación basado en argumentación tiene la ventaja de que además de intercambiar ofertas, se intercambian razones que las soportan y que eventualmente podrían llevar al receptor a cambiar sus preferencias. En consecuencia, se puede llegar a un acuerdo entre las partes de manera más conveniente, pues en otros enfoques donde las preferencias de los agentes son fijas la negociación puede fallar.

De esta manera, adoptar un enfoque de este estilo en los problemas de decisión, tendría el beneficio de que el tomador de decisiones además de tener una buena elección, tendría las razones subyacentes que la soportan de una manera fácil de entender. La toma de decisiones basada en argumentación, es más afín a la forma en que los seres humanos deliberan y finalmente toman o entienden una elección.

Es por eso que desde nuestra perspectiva consideramos la toma de decisiones individuales usando un enfoque argumentativo, donde se hace énfasis en la justificación de cual es la mejor decisión a tomar en una situación dada, y donde se dejan de lado otros aspectos del razonamiento práctico tales como la generación de metas, la factibilidad de las mismas y su planeamiento. De esta forma, el modelo de decisión rebatible en el que se está trabajando, sigue los lineamientos clásicos de la toma de decisiones, pero justificado desde el punto de vista de la argumentación. En particular, en [12], se presentó un modelo para la toma de decisiones rebatibles que combina reglas de decisión y argumentos. En el framework de decisión allí propuesto, la política de decisión del agente se puede modificar de manera flexible, realizando pequeños cambios en los criterios que influyen en las preferencias del agente y en la comparación de argumentos. Nuestro enfoque incluye una metodología simple para desarrollar los componentes de decisión del agente. De esta manera, un framework de decisión desarrollado siguiendo esta metodología, exhibe algunas propiedades interesantes con respecto a la consistencia en las decisiones requerida por el enfoque clásico de teoría de decisión [17]. Si el agente (tomador de decisiones) tiene disponible todo el conocimiento relevante acerca de sus preferencias con respecto a las posibles alternativas candidatas que se le pueden presentar, entonces nuestra propuesta implementa una *relación de preferencia racional*. No obstante, si el agente posee conocimiento parcial acerca de sus preferencias, aún exhibirá un comportamiento de elección consistente, que formalmente se relaciona con el *axioma débil de la preferencia revelada* del enfoque

de *reglas de elección*, un enfoque clásico de teoría de *Toma de Decisiones Individuales* [17] mucho más flexible que el enfoque *basado en preferencias*. Los principios presentados en [12], fueron ejemplificados en un dominio de un único robot Khepera 2 simulado [16, 18], que debe tomar decisiones acerca de que caja debe transportar a continuación, mientras realiza tareas de limpieza en un ambiente estático (donde los únicos cambios que se producen es por el accionar del mismo). Para ello, hemos hecho uso del framework *Khe-DeLP* [11, 13] que posibilita a los robots Khepera razonar usando *Defeasible Logic Programming* (DeLP) [15].

Como trabajo futuro, una primer extensión de nuestro trabajo consistirá en considerar la presencia de múltiples agentes en el ambiente. En [14], se presentó un primer enfoque a la toma de decisiones rebatibles en un sistema con múltiples robots. Sin embargo, en este nuevo escenario, se deberán considerar formalismos relacionados con la coordinación de agentes y también tendremos que establecer las conexiones formales con nuestro enfoque.

3. Consideraciones finales

Con respecto a los enfoques tradicionales de la teoría de decisión clásica, nuestro trabajo difiere principalmente en que el análisis se centra directamente en la relación de preferencia del agente, y no en una función de utilidad que representa a esta relación. Este aspecto nos permite establecer de manera directa, conexiones formales entre nuestro enfoque de toma de decisiones basadas en argumentación y enfoques más esenciales para modelar el comportamiento de elección individual.

El hecho de poder experimentar con dominios de un único y de múltiples robots, reales y simulados, por un lado permite transferir a la práctica los resultados teóricos formales que se obtienen y por otro, presenta dificultades propias del dominio, cuya solución práctica permite la abstracción de los conceptos utilizados para formular dicha solución, logrando una retroalimentación que permite perfeccionar los modelos teóricos obtenidos.

Referencias

- [1] T. Alsinet, C. Chesñevar, L. Godo, S. Sandri, and G. Simari. On the computation of warranted arguments within a possibilistic logic framework with fuzzy unification. In *Proceedings of the 11th International Workshop on Non-Monotonic Reasoning (NMR)*, 2006.
- [2] L. Amgoud and C. Cayrol. Inferring from inconsistency in preference-based argumentation frameworks. *Journal of Automated Reasoning*, 29(2):125–169, 2002.
- [3] L. Amgoud, N. Maudet, and S. Parsons. Modeling dialogues using argumentation. In *Proceedings of the Fourth International Conference on MultiAgent Systems (ICMAS)*, pages 31–38. IEEE Computer Society, 2000.
- [4] L. Amgoud, S. Parsons, and N. Maudet. Arguments, dialogue and negotiation. In *Proceedings of the 14th European Conference on Artificial Intelligence (ECAI)*, pages 338–342, 2000.
- [5] M. Bratman. *Intentions, Plans and Practical Reason*. Harvard University Press, 1987.
- [6] M. Bratman, D. Israel, and M. Pollack. Plans and resource bounded reasoning. *Computational Intelligence*, 4, 1988.
- [7] C. Chesñevar, A. Maguitman, and R. P. Loui. Logical models of argument. *ACM Computing Surveys*, 32(4):337–383, 2000.

- [8] T. Dean, L. Kaelbling, J. Kirman, and A. Nicholson. Planning under time constraints in stochastic domains. *Artificial Intelligence*, 76(1-2):35–74, 1995.
- [9] J. Doyle and R. Thomason. Background to qualitative decision theory. *AI Magazine*, 20(2), 1999.
- [10] P. M. Dung. On the acceptability of arguments and its fundamental role in non-monotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–357, 1995.
- [11] E. Ferretti, M. Errecalde, A. García, and G. Simari. Khedelp: A framework to support defeasible logic programming for the khepera robots. In *International Symposium on Robotics and Automation (ISRA)*, pages 98–103, San Miguel Regla, Hidalgo, México, August 2006.
- [12] E. Ferretti, M. Errecalde, A. García, and G. Simari. Decision rules and arguments in defeasible decision making. In *Proceedings of the 2nd International Conference on Computational Models of Argument (COMMA)*, Frontiers in Artificial Intelligence and Applications. IOS Press, Mayo 2008. (ACEPTADO).
- [13] E. Ferretti, M. L. Errecalde, A. J. García, and G. R. Simari. Khepera robots with argumentative reasoning. In *Proceedings of the 4th International AMIRE Symposium*, pages 199–206, Buenos Aires, Argentina, October 2007.
- [14] E. Ferretti, N. Rotstein, M. Errecalde, A. García, and G. Simari. Defeasible decision making in a multi-robot environment. *Research in Computing Science*, 32:150–160, 2007. Special issue “Advances in Artificial Intelligence and Applications”.
- [15] A. J. García and G. R. Simari. Defeasible logic programming: an argumentative approach. *Theory and Practice of Logic Programming*, 4(2):95–138, 2004.
- [16] K-Team. Khepera 2. <http://www.k-team.com>, 2002.
- [17] A. Mas-Collel, M. D. Whinston, and J. R. Green. *Microeconomic Theory*. Oxford University Press, 1995.
- [18] O. Michel. Webots: Professional mobile robot simulation. *Journal of Advanced Robotics Systems*, 1(1):39–42, 2004.
- [19] J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan-Kaufmann, 1988.
- [20] J. Pollock. The logical foundations of goal-regression planning in autonomous agents. *Artificial Intelligence*, 106(2):267–334, 1998.
- [21] J. Raz. *Practical Reasoning*. Oxford University Press, 1978.
- [22] S. W. Tan and J. Pearl. Qualitative decision theory. In *Proceedings of the 11th National Conference on Artificial Intelligence*, pages 928–933, 1994.
- [23] M. J. Wooldridge. *Reasoning about Rational Agents*. MIT Press, Cambridge Massachusetts, London England, 2000.