

Agentes inteligentes con razonamiento dirigido por factores emocionales

Santiago E. Fulladoza Dalibón, Diego C. Martínez, Guillermo R. Simari

Laboratorio de Investigación y Desarrollo en Inteligencia Artificial
Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur,
Av. Alem 1253, Bahía Blanca, Argentina
Tel: (0291) 459-5135
e-mail: santiagofulladoza@gmail.com, dcm@cs.uns.edu.ar, grs@cs.uns.edu.ar

Resumen

Esta investigación está centrada en la integración de formalismos de argumentación rebatible en agentes inteligentes creíbles. Los agentes creíbles son agentes autónomos en los que aspectos tales como la personalidad, las actitudes y las emociones juegan un rol preponderante para definir su comportamiento. Por lo tanto, se considera a estos aspectos como factores cuya influencia resulta de utilidad tanto para servir como guía en la construcción de argumentos de forma oportuna, como para definir criterios de evaluación y comparación de argumentos.

Palabras clave: Agentes creíbles, Argumentación, Razonamiento dirigido, Factores emocionales

1. Contexto

Esta línea de investigación se llevará a cabo dentro del ámbito de colaboración entre el Laboratorio de Investigación y Desarrollo en Inteligencia Artificial (LIDIA) del Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur y el Área de In-

teligencia Artificial de la Facultad de Ciencias de la Administración, Universidad Nacional de Entre Ríos.

2. Introducción

Una parte significativa de las arquitecturas de agentes con habilidades cognitivas tienen en común que, para estas, un agente está básicamente definido por un conjunto de metas, un conjunto de creencias, y un conjunto de reglas de razonamiento. Su modelo de funcionamiento se basa en la selección de metas según algún criterio y la búsqueda de reglas de razonamiento que “expliquen” cómo conseguir estas metas. Las reglas brindan información acerca de las condiciones que deben cumplirse para la verificación de la meta. Cada una de estas condiciones constituye una submeta que debe ser verificada a su vez mediante el cumplimiento ciertas otras condiciones (algunas de ellas satisficibles mediante la realización de una acción). De esta forma, el modo en que un agente razona y en consecuencia se comporta es descubierto mediante la búsqueda descendente a través de una jerarquía de reglas tendientes a la satisfacción de una meta de orden superior. La estrategia de razonamiento utilizada es conocida como razonamiento por encadena-

miento hacia atrás (backward-chaining reasoning), también conocida como razonamiento dirigido por metas (goal-driven reasoning), y está soportado por la visión de las reglas de razonamiento como una estructura en que la conclusión (la cabeza de la regla) puede ser garantizada a partir de la validación de las condiciones listadas en el cuerpo de la regla.

Existen, sin embargo, diversas aplicaciones (tales como determinados casos de simulación, entretenimiento digital, creación de compañeros electrónicos, entre otros) en que es preferible que el desempeño de un agente inteligente sea más flexible; dominios en donde un agente debe poder razonar partiendo de la información de la situación en la que se encuentra, guiado por factores que no determinen estrictamente su comportamiento, sino que den lugar a una evaluación de alternativas que no esté comprometida al cumplimiento de fines preacordados. Es además de interés, en muchos de estos casos, que el comportamiento de los agentes sea diferenciado. No sería realista, por ejemplo, en la simulación computacional de una catástrofe, que todos los agentes involucrados reaccionen de igual forma ante el evento. Es en este dominio de problemas en el cual deseamos aportar mejoras tendientes a un incremento en la credibilidad de los agentes. De acuerdo a [Ort03], convertir un agente inteligente en creíble requiere que se garanticen respuestas internas (cognitivas) y externas (comportamientos) situacional e individualmente apropiadas, y que se gestione una coordinación sensible entre respuestas internas y externas. De este modo, identificamos como dos principales características deseables que el comportamiento del agente sea:

- contextual, es decir, con un alto grado de relevancia respecto a las condiciones de cada situación en la que se desenvuelve el agente
- particular, es decir, que no sea determinado exclusivamente a partir de la información de la situación, sino que tenga en

consideración además factores particulares del agente que también varíen con el contexto

Respecto a la primer característica mencionada, resulta interesante una visión de las reglas de razonamiento diferente a la primeramente descrita. No es tan natural en este contexto ver a las reglas de modo en que los literales del cuerpo representan condiciones necesarias para garantizar una conclusión que requiere ser probada; sino de forma tal que los literales del cuerpo representan condiciones que favorecen la consideración de la cabeza de la regla, de modo que esta es una consecuencia circunstancial de las premisas. Siguiendo este punto de vista, consideramos de utilidad un método de resolución en que el razonamiento se realice haciendo un encadenamiento iterativo a partir de los datos actuales, de modo que las conclusiones consiguientes sean relevantes a la situación presentada al agente. Esta forma de razonamiento es conocida como encadenamiento hacia adelante (forward chaining) o razonamiento dirigido por los datos (data-driven reasoning). A la vez, resulta práctico representar el conjunto de reglas que soporta cada conclusión como un argumento progresivamente construido a través del encadenamiento de reglas, debido a que los argumentos proveen un medio que hace explícita la derivación de las conclusiones de una forma clara, y constituyen en si mismos una justificación del razonamiento llevado a cabo por el agente.

Este método representa una ventaja tendiente a un incremento en la contextualidad del razonamiento; sin embargo, si no es controlado, produce la derivación de todas las conclusiones inferibles iterativamente desde los datos, algo que no es deseable tanto debido a su elevado costo computacional, como a que carecería de sentido "práctico". Se requiere entonces un modo de guiar la búsqueda que responda en cierto modo al "interés" del agente respecto a la situación y a las reglas de razonamiento susceptibles de ser utilizadas en la misma.

Encontramos aquí un punto de conexión entre ambos requerimientos (la contextualidad y la particularidad del razonamiento y el comportamiento): podemos identificar en la naturaleza factores que los satisfacen simultáneamente. Los seres humanos contamos con mecanismos naturales para desenvolvernos a partir de la información que nos brindan las situaciones que se nos presentan a cada momento. No derivamos toda las conclusiones posibles sino que poseemos ciertos “elementos de información de fondo” que nos permiten “decidir” en qué aspectos centraremos nuestra atención y representan nuestro criterio de relevancia, y que son además, en conjunto, particulares a cada individuo. Algunos de estos factores que podemos identificar son nuestras emociones, personalidad y actitudes. Existen actualmente múltiples aproximaciones para la utilización de estos factores como influencia en el razonamiento y comportamiento de los agentes [HJC03, POM06]. Estos factores emocionales (nombre con el que englobaremos los tres conceptos) pueden ser modelados en el modelo por nosotros propuesto para agentes artificiales creíbles, y constituirían a efectos prácticos una heurística que guíe el razonamiento de los agentes hacia conclusiones que resulten de interés a la situación particular considerando el perfil emocional del agente.

Resulta natural que, bajo este modelo de razonamiento, un agente pueda derivar conclusiones contradictorias a partir de una situación. Teniendo en cuenta los principios en los que el modelo está fundado, no resulta deseable establecer restricciones al respecto, sino que es preferible brindar mecanismos para conseguir soluciones prácticas a cada contradicción que se presente. Teniendo en cuenta lo sostenido en [GS04] para common-sense reasoning, consideramos que las reglas de razonamiento de un agente guiado por factores emocionales deben ser rebatibles, de modo que su uso en una situación particular pueda ser “dejado de lado” por la aplicación de otra regla que justifique ser más relevante a la situación y el perfil

emocional particular.

Consideramos que la detección de conflictos entre argumentos, así como la resolución de los mismos (lo que puede involucrar búsqueda de nuevos soportes y derrotadores para los argumentos) es mucho más directa mediante el razonamiento backward-chaining. Por lo tanto, se propone una estrategia de razonamiento híbrida: se utilizará un método forward-chaining para la construcción progresiva de argumentos a partir de los datos de la situación, y un método backward-chaining para la identificación y resolución de conflictos entre argumentos. Un enfoque similar es presentado en [JH11] para su uso en sistemas de soporte a las decisiones.

Observando nuevamente el trabajo presentado en [GS04], encontramos a la argumentación rebatible como un mecanismo adecuado para la representación de los argumentos construidos, así como de las contradicciones que puedan darse entre conclusiones y como un marco para la resolución de las mismas.

Nuestra propuesta concreta consiste en lo siguiente:

- la formalización de un marco que permita la definición de un perfil emocional del agente, compuesto por un conjunto de factores de personalidad, un conjunto de actitudes y un conjunto de emociones.
- el enriquecimiento de las reglas de razonamiento con factores emocionales que cumplirían dos fines
 - actuar como un estímulo emocional a la consideración de una regla de razonamiento (típicamente, en una situación particular, las reglas con mayor grado de factores emocionales compatibles con el estado emocional actual del agente serán más susceptibles de ser recuperadas y utilizadas para el razonamiento en la situación actual)
 - proporcionar fuerza emocional a la regla a la que están ligados basada en

la positividad/negatividad e intensidad de las emociones que se derivan del uso de una regla

- la adaptación de un formalismo de argumentación rebatible, de modo que se automatice la construcción de argumentos y la detección y resolución de conflictos entre estos, siguiendo los lineamientos de lo previamente establecido respecto a una estrategia de razonamiento híbrida guiada por factores emocionales.
- la definición de un criterio de comparación de argumentos basado en la fuerza emocional de las reglas que los conforman, de modo que, en caso de que dos argumentos estén en conflicto (es decir, si sostienen conclusiones contradictorias) la fuerza emocional sea utilizada para decidir que argumento prevalece sobre el otro.

3. Líneas de investigación

Ésta línea de investigación se centrará en el problema de desarrollar una arquitectura de agente basada en sistemas de argumentación que considere, además de elementos racionales, aspectos emocionales para definir su comportamiento; de modo que proporcionen un mayor grado de credibilidad a los agentes que la implementen.

3.1. Argumentación

La argumentación es una forma de razonamiento en la cual, para una afirmación dada, se presta atención explícita a las justificaciones presentadas y a la resolución de los posibles conflictos entre ellas; de modo que la afirmación es aceptada o no según el análisis de los argumentos a su favor y en su contra. La forma en que los argumentos y las justificaciones son considerados permite definir un tipo de razonamiento automático, en el cual puede haber

información contradictoria, incompleta, e incierta. El estudio de la argumentación ha sido abordado desde diferentes enfoques en años recientes. A nivel lógico puede considerarse como una forma de modelar inferencia rebatible, y a nivel dialógico como una forma de interacción entre agentes. Actualmente es un atractivo paradigma para conceptualizar el razonamiento de sentido común [CML00, PV98]. Esto produjo como resultado la formalización de diferentes frameworks de argumentación abstracta como [Dun95], entre otros; y de Sistemas de Argumentación Basados en Reglas (SABR) como [AK07, GS04]; lo que permitió el desarrollo de diversas aplicaciones del mundo real basadas en argumentación.

3.2. Agentes inteligentes creíbles

Un agente inteligente es una entidad autónoma que es capaz de percibir su entorno y de actuar en él acordemente, posiblemente alterándolo en diversos grados. El estudio y desarrollo de agentes inteligentes ha sido abordado desde diferentes perspectivas; pero esta línea de investigación particularmente está relacionada con (y parte de) el concepto fuerte de agencia (en inglés, *strong agency*), según el cual se atribuye a los agentes, además de las características propias de agente, propiedades mentales antropomórficas tales como creencias, deseos, intenciones, emociones, actitudes, personalidad. Así llegamos al concepto de agente creíble, cuya diferencia fundamental con el espectro tradicional de agentes inteligentes es que el énfasis es puesto en la creación de agentes que poseen personalidades propias y distintivas [Loy97].

4. Resultados y Objetivos

Esta línea de investigación tiene por objetivo desarrollar una arquitectura de agente que integre los sistemas argumentativos rebatibles

con características emocionales, actitudinales y de personalidad, a modo de producir un aumento en la credibilidad de los agentes inteligentes artificiales.

Por otro lado, el enriquecimiento de los sistemas argumentativos con factores y criterios que establezcan preferencias entre argumentos basadas en información dinámica (tal como lo son, en este contexto, los factores emocionales) representará un avance significativo dentro del área de sistemas argumentativos en Inteligencia Artificial y Ciencias de la Computación.

5. Formación de Recursos Humanos

Esta investigación se lleva a cabo en el contexto de una beca de postgrado otorgada por el CONICET; por lo que tiene previsto la formación de un Doctor en Ciencias de la Computación.

Referencias

- [AK07] L. Amgoud and S. Kaci. An argumentation framework for merging conflicting knowledge bases. *Int. J. Approx. Reasoning*, 45:321–340, July 2007.
- [CML00] C. I. Chesnevar, A. G. Maguitman, and R. P. Loui. Logical models of argument. *ACM COMPUTING SURVEYS*, 32:337–383, 2000.
- [Dun95] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77:321–357, 1995.
- [GS04] A. J. García and G. R. Simari. Defeasible logic programming: an argumentative approach. *Theory Pract. Log. Program.*, 4:95–138, January 2004.
- [HJC03] A. E. Henninger, R. M. Jones, and E. Chown. Behaviors that emerge from emotion and cognition: implementation and evaluation of a symbolic-connectionist architecture. In *Proceedings of the second international joint conference on Autonomous agents and multiagent systems, AAMAS '03*, pages 321–328, New York, NY, USA, 2003. ACM.
- [JH11] N. Khalid Janjua and F. Khadeer Hussain. Defeasible reasoning based argumentative web-idss for virtual teams (vts). *Web Intelligence and Intelligent Agent Technology, IEEE/WIC/ACM International Conference on*, 3:330–334, 2011.
- [Loy97] A. B. Loyall. *Believable agents: building interactive personalities*. PhD thesis, 1997.
- [Ort03] A. Ortony. On making believable emotional agents believable. In R. P. Trappl, P. Petta, and S. Payer, editors, *Emotions in Humans and Artifacts*, chapter 6, pages 189–212. MIT Press, Cambridge, Massachusetts London, England, 2003.
- [POM06] David Pereira, Eugénio Oliveira, and Nelma Moreira. N.: Modelling emotional bdi agents. In *In: Workshop on Formal Approaches to Multi-Agent Systems (FAMAS 2006)*, Riva Del Garda, 2006.
- [PV98] H. Prakken and G. Vreeswijk. Logical systems for defeasible argumentation. In *Handbook of Philosophical Logic*. D. Gabbay, second edition, 1998.