

Sistemas de Argumentación Masiva sobre Bases de Datos Federadas

Cristhian A. D. Deagustini Marcelo A. Falappa
Guillermo R. Simari

Laboratorio de Investigación y Desarrollo en Inteligencia Artificial
Departamento de Ciencias e Ingenierías de la Computación
Universidad Nacional del Sur
Alem 1253 - Bahía Blanca - Prov. Buenos Aires - Argentina
(0291) 459- 5135
cadd@cs.uns.edu.ar, mfalappa@cs.uns.edu.ar, grs@cs.uns.edu.ar

Resumen

En la presente investigación se propone la integración de un Sistema de Argumentación Rebatible Basado en Reglas, en particular DeLP, con tecnologías de Bases de Datos Federadas como soporte de los procesos argumentativos. El objetivo final es poder definir tecnologías que nos permitan el desarrollo automático (o semi-asistido) de sistemas de argumentación capaces de soportar sus conclusiones en grandes volúmenes de datos, sin perder las ventajas de los actuales como manejo de información contradictoria e incompleta.

Una ventaja de un sistema de estas características respecto de los sistemas basados en argumentación actuales es que el mismo podrá acceder a un gran número de fuentes de conocimiento sin tener que conocer cómo acceder a cada una de ellas y obtener conclusiones a partir de grandes cantidades de datos, de manera que estas sean más objetivas e imparciales al estar basadas en un conocimiento mucho más rico, amplio y diversificado. A su vez, el conocimiento de un sistema como el expuesto puede *evolucionar*, tanto en el hecho de que los datos en que se soportan las conclusiones cambian a medida que las bases de datos usadas cambian (por la influencia de sistemas externos) como en la actualización de las reglas en las que se basa el sistema para la construcción de argumentos si realizamos minería de reglas en la base de dato actualizada. Esto permitirá producir sistemas capaces de entregar respuestas más precisas y justas a las consultas realizadas a un agente autónomo con estas capacidades.

Para lograr lo propuesto, debemos atacar varios problemas. Primeramente, hemos de definir métodos eficientes de integración de DeLP en su forma actual con Bases de Datos Relacionales que nos permitan identificar y proveer al proceso argumentativo de información relevante dentro del universo de datos almacenados por las Bases de Datos. Además, para lograr un desarrollo de sistemas mas objetivos, debemos establecer métodos que nos permitan la obtención de reglas de conocien-

to basados en datos o documentos externos, de manera que el conocimiento reflejado en las reglas no represente la visión de un tema particular de unos pocos individuos. También estudiaremos el problema de la creación de una vista unificada de datos basada en el conocimiento almacenado por varias Bases de Datos (tanto datos en sí como en el esquema de las mismas), resolviendo de manera escéptica los conflictos potenciales debido a inconsistencias e incompletitudes, y finalmente se pasará a la investigación acerca de la eficiencia de los procesos argumentativos basados en cantidades masivas de datos y su aplicabilidad al desarrollo de aplicaciones del mundo real.

Palabras Clave: Argumentación Rebatible, Bases de Datos Federadas, Sistemas de Argumentación Masiva.

1. Contexto

Esta línea de investigación se llevará a cabo dentro del ámbito de colaboración entre el Laboratorio de Investigación y Desarrollo (LIDIA) del Dep. de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur; y el Laboratorio de Investigación y Desarrollo en Inteligencia Artificial Concordia (LIDIA Concordia) de la Facultad de Ciencias de la Administración, Universidad Nacional de Entre Ríos.

2. Introducción

El estudio de la Argumentación como forma de razonamiento en las Ciencias de La Computación ha cobrado cada vez mayor fuerza. La Argumentación representa un sofisticado mecanismo para la formalización del *razonamiento de sentido común*, que es de gran utilidad para la obtención de respuestas útiles en entornos de inconsistencia e incompletitud, de manera similar a la realizado por los humanos. Para lograr esto, en Argumentación una proposición es aceptada o no de acuerdo a un

análisis de las razones que se tiene para creer o no en la misma, donde estas razones o justificaciones son denominadas argumentos. De manera intuitiva, en un Sistema de Argumentación (SA), un argumento es un conjunto coherente de afirmaciones que soportan una conclusión. Además, la forma en que estos argumentos son considerados permite la automatización de este tipo de razonamiento. Es por esto que a mediados de los años '80 comenzaron los desarrollos del área de argumentación desde un punto de vista computacional, donde los argumentos son explícitamente construidos y comparados como medios para resolver problemas en una computadora. En la última década, la argumentación ha evolucionado como un atractivo paradigma para conceptualizar el razonamiento de sentido común, donde la información sobre la que se basa este razonamiento puede ser inconsistente e incompleta.

Los Sistemas de Argumentación son actualmente utilizados en el desarrollo de aplicaciones en sistemas legales, negociación multi-agente, soporte a la toma de decisiones entre otros, como puede verse en [PV00, CML00, GGTS00, CMG09, BBC⁺12], habiendo probado sobremanera su utilidad en esos campos. Dentro de estos Sistemas se encuentra un tipo particular denominado Sistemas de Argumentación Basados en Reglas (SABR) [PS97, DKT06, AK05, GS04]. En estos sistemas existe un conjunto de reglas de inferencia con las cuales, a partir de cierta información (antecedente) se puede inferir de manera tentativa nueva información (consecuente). En este tipo de sistemas, las reglas son almacenadas en una base de conocimiento, junto a otra información en forma de hechos o presuposiciones, que representa evidencia que el agente obtiene de su entorno. A partir de esta evidencia, el agente puede usar las reglas de inferencia para construir de manera automática argumentos a favor o en contra de una afirmación. Luego se evalúan todos los argumentos construidos, y se determina cuáles de ellos son aceptados, para poder en última instancia concluir si, de la base de conocimiento del agente, está afirmación puede asegurarse o no. Estos formalismos son no-monótonos, dado que la introducción de nueva información al sistema puede generar nuevos argumentos que contradicen a otros ya existentes, y por lo tanto, invalidar afirmaciones que antes estaban garantizadas. En general, en la mayoría de estos formalismos, argumentos y contra-argumentos son comparados utilizando un criterio de preferencia predeterminado.

Sin embargo, y a pesar de su utilidad en campos tan diversos como los mencionados, los Sistemas de Argumentación Basados en Reglas actuales tienen una particularidad. Al basar sus conclusiones en información local incluida en el mismo, las conclusiones pueden verse afectadas por la visión particular que el agente que utiliza ese programa tenga del dominio.

Respecto de la integración de Bases de Datos en una vista unificada, el hecho de que tanto la información como los esquemas de las Bases de Datos que la almacenan sean incompletos e inconsistentes dificulta la consecución de una vista única consolidada a través de las técnicas

utilizadas actualmente por los Sistemas de Gestión de Base de Datos (DBMS). Es en la resolución de estas dificultades donde es conveniente contar con un sistema que pueda resolver de manera escéptica problemas de inconsistencia e incompletitud, decidiendo por aquello que menos conflictos genere, especificando qué integrará la vista única y qué será descartado.

Dado esto último, y considerando además que en [FKIRS11] se han estudiado operadores para la Revisión de Creencias a través de la unión consistente de bases de creencias de agentes, sentando un precedente en la utilización Argumentación Rebatible a través de Programación en Lógica Rebatible en la unión de bases de información inconsistentes; en la presente investigación se propone el desarrollo de un sistema que utilice argumentación en la generación de una Vista Única de la Bases de Datos de los Agentes, en particular Argumentación Rebatible, para así poder manejar las posibles inconsistencias en las bases de datos accedidas.

Una vez alcanzado lo anterior se plantea otro problema: si bien en los últimos tiempos hemos visto desarrollarse técnicas eficientes para el almacenamiento y recuperación de grandes volúmenes de datos, el análisis y la interpretación de tal cantidad de información sigue siendo un problema sin solución.

Una primer aproximación a la solución de este problema fue propuesto por el área de Bases de Datos Deductivas. Estos sistemas combinan técnicas y herramientas de las bases de datos relacionales y los sistemas basados en reglas, y son en consecuencia capaces de manejar grandes cantidades de información extrapolando la información implícita en el conjunto. Sin embargo estos sistemas no son capaces de representar correctamente el conocimiento y el razonamiento de sentido común, en particular cuando la información es incompleta y potencialmente contradictoria [Sub92, LS94, LS01].

Es en este escenario en donde la Argumentación Rebatible se muestra como un buen punto de partida en la búsqueda de la solución integral al problema del razonamiento en base a información extensa pero contrapuesta.

3. Líneas de Investigación y Desarrollo

Esta línea de investigación se enfocará en la utilización de la Argumentación para el desarrollo de sistemas argumentativos capaces de soportar conclusiones en cantidades masivas de datos soportadas en Bases de Datos Federadas. Esto permitirá a una aplicación basada en estas tecnologías producir respuestas mas refinadas, ya que podrá crear una mayor cantidad de argumentos, al tener mas *casos base* donde apoyar sus conclusiones; y reducir la subjetividad al usar datos diversos provenientes de varias fuentes, aumentando de esta manera el consenso en las respuestas brindadas. Básicamente, podemos dividir a un programa DeLP integrado con Bases de Datos en:

- Datos de soporte para argumentos: Almacenados en varias bases de datos relacionadas, que son accedi-

das de manera federada, es decir, lógicamente se ven como una sola. De esta forma el conocimiento de soporte se actualiza a medida que las bases de datos son actualizadas. Sistemas externos pueden brindarnos sus conocimientos y experiencias a medida que actualizan las bases de datos (pensar por ejemplo en entornos multiagente).

- Reglas de dominio: Conocimiento que se tiene del dominio de aplicación particular. Se usan para relacionar datos de forma que podamos usarlos para establecer conclusiones.

Para el desarrollo del sistema expuesto, la presente línea de investigación se relaciona con uno en particular dentro de los Sistemas Argumentativos existentes: *De-feasible Logic Programming (DeLP)*. En DeLP el conocimiento es expresado a través de reglas fuertes (no rebatibles) o débiles (rebatibles). Este formalismo nos permite representar tanto información contradictoria a través de la negación clásica como información incompleta mediante el uso de la negación por falla, convirtiéndose en un medio apropiado para el desarrollo de sistemas con las características mencionadas

Para poder lograr la integración propuesta de DeLP con Bases de Datos Federadas, debemos encarar varios problemas. Por un lado, debemos disponer de un repositorio centralizado de información de soporte para argumentos, de manera que el acceso a las bases de datos que están detrás de este repositorio sea transparente para el sistema. Por otro lado, debemos definir estrategias eficientes de recuperación de información relevante dentro del universo de datos disponibles en esta vista consolidada; y métodos de resolución para los conflictos que puedan generarse por el hecho de usar información inconsistente o incompleta. A su vez, respecto de las reglas del dominio que efectivamente usan dicha información para construir argumentos, para poder obtener conclusiones refinadas y objetivas, las mismas deberían ser obtenidas de conocimiento experto y consensuado, y no ser la opinión subjetiva de un grupo cerrado de personas. Por último, la eficiencia en la realización de los procesos argumentativos es un tema de vital importancia si queremos poder aplicar estas tecnologías para desarrollar aplicaciones del mundo real.

3.1. Uso de Bases de Datos como Soporte de Procesos Argumentativos

Con respecto a la integración de Bases de Datos con Sistemas de Argumentación Rebatible, en particular DeLP, hay dos aspectos principales que tenemos que tener en cuenta:

- Federación de distintas bases de datos (creación de vista única integrada)
- Búsqueda de datos relevantes para argumentos dentro del universo de datos disponibles; y definición de métodos para resolución de conflictos entre datos contradictorios, entre bases de datos distintas o dentro de una misma.

Acerca del primer aspecto, la primer dificultad para la obtención de la vista unificada de las bases de datos consiste en si el esquema de la misma se puede obtener automáticamente a partir de las BDs que se quieren integrar, o si es necesario que sea definida previamente. Por ejemplo, para lograr la automaticidad analizaremos la factibilidad de utilizar definiciones semánticas de los campos de las BDs a integrar a través de ontología, y una posterior fusión de las mismas de manera escéptica [GCS10]. A su vez, sin importar como fue obtenida, una vez que lo hayamos logrado deberíamos identificar la correspondencia entre los esquemas de las BDs a integrar con el esquema de la vista, es decir qué campos de las BDs corresponden con qué campos de la vista. Finalmente, debemos proceder a llenar con los datos de las BDs la vista única, donde podemos encontrarnos con que una misma entidad puede tener datos diferentes en distintas BDs, de manera que para poder incluirla en la vista debemos de identificar esta correspondencia entre datos desiguales, es decir saber que se refieren a la misma entidad [ZR05], para poder decidir qué dato almacenaremos finalmente. Para resolver los dos problemas mencionados últimos se utilizarán enfoques basados en Revisión de Creencias, tanto para la identificación de correspondencia entre esquemas como entre datos.

Respecto del segundo aspecto, un vez armada la vista única debemos usar los datos allí almacenados como soporte de los argumentos, por lo tanto debemos poder identificar aquellos datos relevantes para el proceso argumentativo. Para la identificación de datos relevantes analizaremos tanto el uso de técnicas sintácticas (basadas en funtores y parámetros de un predicado) como semánticas (por ejemplo el uso de ontologías para la identificación). Una vez identificados los datos que ofrecen soporte para argumentos, debemos establecer como elegiremos a que dato damos preferencia, para poder resolver discrepancias cuando éstas aparezcan. Al respecto de esto último, investigaremos alternativas de criterios de preferencia entre datos provenientes de Bases de Datos. Esta preferencia puede estar basada tanto en factores externos al argumento como la confianza respecto de la fuente de dato (es decir, preferir una BD sobre otra), o ser intrínseca a los argumentos (por ejemplo, si preferimos aquél argumento que utiliza una mayor cantidad de datos).

3.2. Definición de Reglas de Dominio

Respecto de las reglas de dominio que serán utilizadas en la construcción de argumentos, hay distintas maneras de obtenerlas.

- Adquisición de Know-How por parte del equipo de desarrollo del sistema.
- Consulta a expertos del dominio.
- Obtención automática de reglas a partir del análisis de documentos y datos.

En la presente línea de investigación nos centraremos en la definición de estrategias que permitan el descubrimiento automático de reglas. Particularmente, estudiaremos dos alternativas:

- La posibilidad de obtener reglas de dominio mediante el procesamiento de lenguaje natural semi-estructurado como el encontrado en mapas conceptuales. El conocimiento en este tipo de documentos está estructurado de manera que facilita la interpretación del mismo, simplificando el procesamiento automático de su contenido.
- Minería de reglas a partir de los datos almacenados en bases de datos. Por ejemplo, a través de la búsqueda de contraejemplos para las reglas ya establecidas para el programa.

3.3. Eficiencia de Procesos Argumentativos

El proceso detrás de la construcción y comparación de argumentos en DeLP es un proceso muy costoso desde un punto de vista computacional [CF06]. Por lo tanto, es necesaria la definición e implementación de métodos más eficientes para la realización del proceso de dialéctica, de manera que pueda cumplir con los requerimientos de tiempo impuestos a la mayoría de las aplicaciones hoy en día.

Al respecto, avances se han hecho en [CCS07, CCS04]. En este línea, se progresará sobre lo conseguido en esos trabajos, para luego combinar el concepto de grafos de dialéctica expuesto allí con mejoras en el proceso de dialéctica llevado a cabo por DeLP, como por ejemplo la implementación de nuevas técnicas de poda basada en heurísticas [RGGS11] o alteraciones en el orden de prueba de construcción de argumentos, tratando de probar primero aquellos argumentos que se espera que sean las hojas del árbol de dialéctica.

4. Resultados y Objetivos

El objetivo general de este trabajo de investigación es el diseño y construcción de la infraestructura necesaria para la realización de procesos de argumentación sobre repositorios masivos de datos. Esto permitirá, por ejemplo, la definición de una nueva arquitectura para Sistemas de Soporte de Decisiones (DSS) que combine el acceso a múltiples bases de datos en busca de información con tecnología de argumentación masiva en base a la información conseguida, para la definición de consejos en el soporte a la toma de decisiones. Una extensión análoga podría definirse para los Sistemas de Recomendación (RS) [CMS07], en especial aquellos que basan su funcionamiento sobre una base de conocimiento y que utilizan tecnologías de Web Semántica, es decir, Sistemas de Recomendación Semánticos. Estos sistemas tienen su base de conocimiento definida a través de un esquema de conceptos, como una taxonomía o un tesoro, o una ontología; sin embargo, podría definirse a la misma mediante

una o varias Bases de Datos, y luego el Recomendador realizar el razonamiento de las recomendaciones con la información almacenada en ellos.

Respecto de los resultados y objetivos alcanzados, hemos diseñado y desarrollado un framework que permite la obtención de información relevante para procesos argumentativos de manera eficiente, integrando DeLP con tecnologías de Bases de Datos Relacionales, presentando los resultados en congresos nacionales e internacionales. Respecto del futuro, se ha comenzado a investigar el uso de formalismos de Revisión de Creencias en la integración de fuentes heterogéneas de información, con vistas a desarrollar un framework que permita la integración de varias Bases de Datos en una única vista unificada. Por último, respecto de la eficiencia en la realización de procesos argumentativos hemos comenzado el estudio y la implementación del trabajo presentado en [CCS07, CCS04] como forma de incrementar la velocidad de respuesta de un sistema basado en DeLP.

5. Formación de Recursos Humanos

En la presente línea de investigación se enmarca el desarrollo de una tesis de posgrado en el Doctorado en Ciencias de la Computación del Departamento de Ciencias e Ingenierías de la Computación de la Universidad Nacional del Sur.

Referencias

- [AK05] L. Amgoud and S. Kaci. An argumentation framework for merging conflicting knowledge bases: The prioritized case. In *Proc. of the ECSQARU-2005 Conf., LNAI 3571*, pages 527–538. Springer, 2005.
- [BBC⁺12] F. Bex, I. Bratko, C.I. Chesñevar, W. Dvorak, M.A. Falappa, S.A. Gaggl, A.J. García, M.P. González, T. Gordon, J. Leite, M. Možina, C. Reed, G.R. Simari, S. Szeider, P. Torroni, and S. Woltran. *Agreement Technology Handbook*, chapter Argumentation in Agreement Technologies, page (in preparation). Springer Verlag, Law, Governance and Technology Series, 2012.
- [CCS04] M. Capobianco, C. I. Chesñevar, and G. R. Simari. An argument-based framework to model an agent's beliefs in a dynamic environment. In Iyad Rahwan, Pavlos Moraitis, and Chris Reed, editors, *ArgMAS*, volume 3366 of *Lecture Notes in Computer Science*, pages 95–110. Springer, 2004.
- [CCS07] M. Capobianco, C. I. Chesñevar, and G. R. Simari. On the construction of dialectical databases. *Inteligencia Artificial, Revis-*

- ta *Iberoamericana de Inteligencia Artificial*, 11(35):89–100, 2007.
- [CF06] L.A. Cecchi and P. Fillottrani. On the complexity of delp through game semantics. In J. Dix and A. Hunter, editors, *Proc. 11th Intl. Workshop on Nonmonotonic Reasoning (NMR 2006)*, pages 386–394, 2006.
- [CMG09] C. Chesñevar, A. Maguitman, and M. González. *Argumentation in Artificial Intelligence*, chapter Empowering Recommendation Technologies Through Argumentation, pages 403–422. Springer Verlag, 2009.
- [CML00] C. I. Chesñevar, A. G. Maguitman, and R. P. Loui. Logical models of argument. *ACM Comput. Surv.*, 32(4):337–383, December 2000.
- [CMS07] C. I. Chesñevar, A. G. Maguitman, and G. R. Simari. Recommender systems based on argumentation. *Emerging Artificial Intelligence Applications in Computer Engineering, Series Frontiers in Artificial Intelligence and Applications*, 160:53–70, 2007.
- [DKT06] P. M. Dung, R. A. Kowalski, and F. Toni. Dialectic proof procedures for assumption-based, admissible argumentation. *J. of Artificial Intelligence*, 170(2):114–159, 2006.
- [FKIRS11] M. A. Falappa, G. Kern-Isberner, M. D. L. Reis, and G. R. Simari. Prioritized and non-prioritized multiple change on belief bases. *J. of Philosophical Logic*, 2011.
- [GCS10] S. A. Gómez, C. I. Chesñevar, and G. R. Simari. Reasoning with inconsistent ontologies through argumentation. *J. of Applied Artificial Intelligence*, 24(1&2):102–148, 2010.
- [GGTS00] A. J. Garcia, D. Gollapally, P. Tarau, and G. R. Simari. Deliberative stock market agents using jinni and defeasible logic programming, 2000.
- [GS04] A. J. García and G. R. Simari. Defeasible logic programming an argumentative approach. *Theory and Practice of Logic Programming*, pages 95–138, 2004.
- [LS94] L. V. S. Lakshmanan and F. Sadri. Probabilistic deductive databases. *Proc. of the International Conference logic Programming Symposium*, 1994.
- [LS01] L. V.S. Lakshmanan and N. Shiri. A parametric approach to deductive databases with uncertainty. *J. of Intelligent Information Systems*, 3(3), 2001.
- [PS97] H. Prakken and G. Sartor. Argument-based logic programming with defeasible priorities. *J. of Applied Non-classical Logics*, 7:27–75, 1997.
- [PV00] H. Prakken and G. Vreeswijk. *Logical systems for defeasible argumentation*. Kluwer Academic, 2000.
- [RGGSS11] N. Rotstein, S. Gottifredi, A. J. García, and G. R. Simari. A heuristics-based pruning technique for argumentation trees. In Salem Benferhat and John Grant, editors, *Scalable Uncertainty Management*, volume 6929 of *Lecture Notes in Computer Science*, pages 177–190. Springer Berlin / Heidelberg, 2011.
- [Sub92] V. S. Subrahmanian. Paraconsistent disjunctive deductive databases. *J. of Theoretical Computer Science*, 93, 1992.
- [ZR05] H. Zhao and S. Ram. Entity identification for heterogeneous database integration: a multiple classifier system approach and empirical evaluation. *J. of Information Systems*, 30(2):119 – 132, 2005.