4. Conclusiones y líneas futuras

El agente implementa un modelo muy sencillo y rápido de implementar, aunque su rendimiento varía fuertemente en función de los umbrales de decisión utilizados.

Una crítica razonable de este modelo es su carácter 'ad hoc' en función del dominio de aplicación; por ello, el agente funciona mejor en la Competición Nacional, dado que estaba pensado para actuar con un número reducido de agentes en el entorno y con pocas iteraciones, lo que contradice los principios generales que justifican el uso de información de reputación.

Si se revisa el resultado de la ejecución de las reglas de comportamiento definidas se apreciará como el agente *GIAA* tiene una marcada tendencia a "mentir" al resto de los participantes en la subasta, dado que la necesidad de mentir aumenta rápida y continuamente, lo que a lo largo de las iteraciones hace descender la reputación del mismo y en consecuencia, disminuyen las peticiones de tasación que son recibidas.

El modelo desarrollado está abierto a múltiples y diversas modificaciones que pueden mejorar su rendimiento. Por ejemplo, entre las tres versiones de la Competición Nacional, que varían únicamente el valor de los umbrales de decisión, se puede apreciar una mejora de más del 20%.

Como propuesta, se apuntan las siguientes líneas de actuación:

- Segmentar más los umbrales de toma de decisión, lo que permitiría una mejor adecuación de la acción a tomar según la situación evaluada.
- Evitar que el agente vaya perdiendo sinceridad de forma general.
- Aprovechar la información sobre reputación de terceros para actualizar la confianza en un (agente, era).

5. Agradecimientos

Investigación financiada con los proyectos CICYT TSI2005-07344-C02-02, MADRINET S-0505/TIC/0255, CAM CCG06-UC3M/TIC-0781 y AUTOPIA (IMSERSO 31/06).

6. Referencias

1. ARTTestbed: http://www.art-testbed.net
2. Fullam, K., T. Klos, G. Muller, J. Sabater, A. Schlosser, Z. Topol, K. S. Barber, J. Rosenschein, L. Vercouter, and M. Voss. (2005) "A Specification of the Agent Reputation and Trust (ART) Testbed: Experimentation and Competition for Trust in Agent Societies," The Fourth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2005), Utrecht, July 25-29, pp. 512-518.
3. II Competición Nacional ARTTestbed, IIIA (CSIC). Valencia, Marzo 2007. Web: http://megatron.iiia.csic.es/spanishART/
4. II Competición Internacional ARTTestbed. Hawai, Mayo 2007. Web: http://www.lips.utexas.edu/art-testbed/competiton_results_prelim2007.htm

Filosofía del diseño del agente *ConfianzaEras*, participante en la I competición nacional de ART Testbed.

Iván García-Magariño

Departamento de Ingeniería del Software e Inteligencia Artificial,
Facultad de Informática,
Universidad Complutense de Madrid.
email: ivan_gmg@fdi.ucm.es

Resumen Este artículo presenta la estrategia del agente llamado *ConfianzaEras*, que participó en la primera competición nacional ART Testbed. Experimentalmente, se observa que este agente colabora a la exclusión de los agentes no fiables.

Key words: trust, reputation, competition testbed, multi-agent systems

1. Introducción

En los sistemas multi-agente de *trust and reputation* [2,1], se asume que la verdad absoluta o global no está disponible a los agentes, y cada agente tiene una perspectiva local y subjetiva. Cada agente construye su perspectiva local según la calidad de la información que recibe. Modelar la calidad de la información es muy útil. Se denomina *confianza(trust)* a la confianza de un agente en los diferentes agentes según su interacción con ellos. La *reputación(reputation)* de un agente es la suma de *confianzas* del resto de los agentes que han obtenido del trato con él.

La iniciativa *Agent Reputation and Trust(ART) Testbed*[3] permite a los investigadores usar métricas objetivas sobre entornos de confianza y reputación y hacer experimentos fácilmente repetibles. En la competición ART Testbed, cada agente tiene conocimiento parcial sobre unos cuadros artísticos. Compra opiniones sobre los cuadros y opiniones sobre el resto de los agentes. Elige qué información compartir. A lo largo de las iteraciones, el que haga tasaciones más precisas tendrá más clientela. Ganará la competición el que más dinero tenga. Se han celebrado varias competiciones a nivel nacional e internacional.

En este artículo se presenta filosofía del diseño del agente llamado *ConfianzaEras*, que participó en la competición ART Tesbed, celebrada en la universidad Carlos III de Madrid (España), el 24-26 abril del 2006. También se menciona los resultados obtenidos para dicho agente.

Este artículo se organiza de la siguiente forma. A lo largo del siguiente apartado, se describe la estrategia del agente ConfianzaEras. En el apartado 3, se

estudia y valora cierto comportamiento emergente al usar dicho agente. En el apartado 4, se menciona el resultado de nuestro agente en la competición.

2. Descripción de la Estrategia del Agente *ConfianzaEras*

En este apartado, se describe la estrategia del agente *ConfianzaEras*. Para hacer más inteligible la explicación, se ha dividido la estrategia en diferentes aspectos, para así poder discutir cada uno de ellos por separados, sin perder de vista que la estrategia es un total de todos los aspectos.

2.1. Confianza en los agentes

Para realizar la estimación de qué *confianza* poner en la opinión de cada agente, esta estrategia se basa en lo fiable que fue la última vez. Para ello, se usa los errores cometidos para cada agente en la última interacción con él.

El error cometido en una tasación no sólo depende de la honestidad del agente, sino también del conocimiento que tenga dicho agente sobre el cuadro y más concretamente sobre la *era* a la que pertenece el cuadro.

Por ello, se tiene una confianza para cada par <agente,era>, calculada con la media de errores para ese par. Si no se tiene ningún dato para un par <agente,era> se usa la media de los errores de ese agente para todos los cuadros. En la primera iteración, que no se tienen datos, la estrategia asigna el máximo de confianza. Por tanto, ante un agente desconocido, nuestro agente es confiado y le supone buena voluntad.

La confianza es una estructura de datos en nuestro agente, que se usa en los siguientes pasos.

2.2. Realizar una Tasación

La tasación es una media ponderada de las opiniones que han dado otros tasadores y la del propio agente. Por tanto, lo importante es qué pesos damos a las opiniones de cada tasador. Pues bien, esta estrategia para cada par <agente,era> ha usado la *confianza*(apartado 2.1) para ese par. No se usa la reputación, por los motivos mencionados en el apartado 2.7.

2.3. Peticiones de Opiniones (selectivo) y Umbral de Confianza

Cada agente debe determinar en cada iteración a qué agentes les pide opinión. El precio de comprar una opinión es no negociable e igual para todos.

Los agentes a los que se le ha asignado una confianza muy baja probablemente envíen opiniones muy poco valiosas. Además el agente ConfianzaEras los tendrá muy poco en cuenta. Sin embargo, comprar estas opiniones cuesta el mismo precio que el resto. Por tanto, en la estrategia establecí un *umbral de confianza*. Sólo se pedirá opinión a los agentes en los que se tenga una confianza mayor al umbral de confianza. Éste umbral es un parámetro que se configura.

Se observa que, una vez que se clasifica a un agente como *engañoso* (por debajo del *umbral de confianza*), no se vuelve a solicitarle ninguna opinión en adelante. Esto se debe a que para volver a recalcular la confianza sobre un agente, sería necesario tener alguna opinión suya para evaluarle, lo que nunca llega a ocurrir. A este efecto, lo he denominado *efecto del prejuicio*.

Si se escoge el umbral de confianza suficientemente bajo, sólo se clasifica como engañosos a los agentes que hayan engañado de manera exagerada. Estos son muy susceptibles de engañar en el futuro. Por ello, no considero tan negativo el efecto del prejuicio.

2.4. Aceptar Peticiones de Opiniones (selectivo)

En esta estrategia, se ha incluido la posibilidad de vender las opiniones sobre cuadros de manera selectiva. Esto es, se ha definido dos parámetros:

- *aceptar_honest*: Determina si aceptar las peticiones de los agentes honestos, estos son los que superan el umbral de confianza.
- *aceptar_cheating*: Determina si aceptar las opiniones de los agentes engañosos, por debajo del umbral de confianza.

En un principio, se quiere vender las opiniones para sacar dinero. Por ello, a los agentes honestos decidí venderles las opiniones.

Sin embargo, los agentes engañosos representan una competencia mayor, ya que son agentes que están recibiendo o aprovechándose de la información de los demás, sin apenas compartir información valiosa, luego si no se les trata de forma diferente tienen más posibilidades de ganar.

Por ello, determiné tratarles de forma peor a los agentes clasificados como engañosos. Sin embargo, si les rechazaba venderle la opinión, puede que la información la obtuvieran de otros agentes y ganaran de igual forma. Además, mi agente no estaría ganando el dinero resultante de la venta de esa opinión. Por eso creí más conveniente aceptar la venta de opinión, y luego proporcionarle una información engañosa, esto es *pagarle con su misma moneda* . De esta forma obtendría el dinero resultante de la venta, no le proporcionaría información valiosa, y daría la posibilidad de que el agente engañoso fuere engañado con nuestra información, perjudicándole así, y tratando de colocarnos por encima de este rival engañoso.

2.5. Tiempo de Examinación o Sinceridad (selectivo)

Cuanto más *tiempo de examinación (CG)* se le dedique a examinar un cuadro más honesto se es en la opinión vendida.

La sinceridad de nuestro agente depende de cómo se haya clasificado anteriormente al agente que solicita la opinión. Se definen dos parámetros:

- *cg_honest*: Sinceridad que se tiene con los agentes honestos.

- *cg_cheating*: Sinceridad que se tiene con los agentes engañosos.

Por los motivos que se explicó en el apartado anterior, las opiniones ofrecidas a los agentes engañosos son engañosas.

A los agentes honestos, por un lado no interesa proporcionarle información muy honesta, ya que también son nuestros rivales. Por otro lado, tampoco interesa mentir en nuestras opiniones a todos los agentes ya que nuestro agente podría ganarse una mala reputación y que los agentes dejaran de solicitarle opiniones. Por tanto, se procura llegar a un equilibrio en el valor de *cg_honest*.

2.6. Distribuir la Reputación

En un principio, comunicar la reputación favorece a la comunidad de agentes o sistema multi-agente; ya que permite a la comunidad detectar a los agentes engañosos y excluirlos de la comunidad para que la comunidad funcione bien. En la misma línea, nos interesa desprestigiar a los agentes engañosos, ya que son más susceptibles de ganar, como se explicó anteriormente.

Además, interesa vender información sobre la reputación para obtener los beneficios de la venta.

Por tanto, nuestra estrategia acepta todas las solicitudes de reputación, y es sincera devolviendo el valor de confianza correspondiente.

2.7. Solicitar la Reputación

Nuestra estrategia no solicita ni compra información de reputación a otros agentes.

En primer lugar, el agente E que evaluamos (figura 1) puede tratar de manera diferente a nuestro agente *YO* que a otro agente *O*. Por ejemplo, podría ser engañoso con O y honesto con YO. Por tanto, la reputación no sería de gran ayuda a nuestro agente.

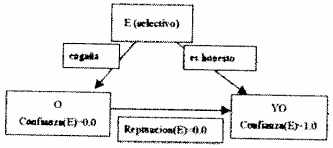

Figura 1. Ejemplo de Comunicación de Reputaciones

En segundo lugar, la reputación que llega a través de terceros puede ser engañosa, según las reglas del juego.

En último término, el no comprar información de reputación , hace ahorrar algo de dinero. La cantidad ahorrada es muy baja y por tanto es sólo un aspecto secundario.

Por estos motivos, se ha decidido no comprar reputaciones.

2.8. Fácil configuración del Agente

Cabe destacar que la implementación de nuestro agente permite diferentes configuraciones, según se escoja los valores de los parámetros (*umbral de confianza, aceptar_honest, aceptar_cheating, cg_honest y cg_cheating*).

3. Comportamiento Emergente de Exclusión de los Agentes No Fiables

Inicialmente se enfrentaron los agentes ejemplos de la plataforma y se observó que el agente *cheating* (engañoso) mantiene una cuenta bancaria superior al agente *honest* (honesto).

Sin embargo, en un sistema multi-agente de confianza y reputación es más deseable que los agentes engañosos sean excluidos para mejor funcionamiento del sistema multi-agente. Justamente se consigue el anterior objetivo, si se incluyen varios agentes con la estrategia *selectiva* descrita en el apartado 2. El motivo es que la estrategia trata favorablemente a los agentes honestos frente a los engañosos.

El comportamiento emergente de exclusión de los agentes engañosos se observó en la siguiente prueba experimental. En la prueba se usaron tres agentes *selectivos*, el agente Cheating y el agente Honest.

En las dos primeras iteraciones, se observó que el agente Cheating tiene más saldo que el resto de los agentes. Esto se debe a que, inicialmente, tanto los agentes selectivos como Honest se fían de Cheating, ya que no le conocen, y Cheating se aprovecha de ellos obteniendo mejor beneficios y engañando al resto de agentes.

Sin embargo, en la segunda iteración, los agentes selectivos se dan cuenta de que el agente Cheating es engañoso, y le tratan de manera especial en adelante. Los agentes selectivos dejan de comprarle sus opiniones, y le empiezan a enviar opiniones engañosas a Cheating. En la prueba, en las iteraciones tres y cuatro, se observó que los agentes selectivos alcanzaron más saldo que el agente Cheating. A partir de este instante, los tres agentes selectivos continuaron enviando opiniones valiosas al agente Honest y engañosas al agente Cheating.

El proceso en el que el saldo del agente Honest alcanza el saldo del agente Cheating fue más largo debido a que el agente Cheating seguía aprovechándose del agente Honest al recibir su información valiosa y enviarle información engañosa.

Finalmente, en la iteración doce el agente Honest tenía más saldo (13140$ frente a 12509$) y mayor mayor clientela (dos unidades frente a una), que el agente Cheating. Los agentes selectivos tuvieron resultados mucho más positivos (aproximadamente 22.000$ y tres unidades de clientela), gracias a tratar de manera eficaz la confianza.

En conclusión, los agentes *selectivos* con la estrategia descrita en el apartado 2 producen un comportamiento emergente de exclusión de los agentes engañosos. Rasmusson[4] señala lo deseable que es la exclusión de las fuentes de información no fiables. En esta exclusión, entre otras cosas, se basa el concepto de seguridad blanda (*soft security*) en entornos con algunas fuentes no fiables.

4. Conclusiones y Resultados en la Competición

Este artículo presenta la estrategia de un agente que participó en la competición nacional de *ART Testbed 2006*. Experimentalmente, se ha observado que dicha estrategia contribuye a la exclusión de los agentes poco fiables.

En la competición, el agente *ConfianzaEras* pasó a la final y *obtuvo el quinto puesto*. En la competición observé que, si los agentes gestionan bien la confianza, se produce la exclusión de los agentes *engañosos*, sin que se tenga que hacer ningún mecanismo explícito para ello. Esto se debe a que el resto de agentes dejan de comprarle opiniones al agente engañoso. Por otro lado, proporcionar opiniones engañosas a los agentes engañosos perjudica al propio agente, ya que dejará de vender esas opiniones. Por tanto, para los participantes de esta competición, probablemente hubiera sido más beneficioso hacer una venta de opiniones igual de honesta para todos los agentes.

5. Agradecimientos.

Este trabajo de investigación se ha realizado dentro del proyecto *Métodos y herramientas para modelado de sistemas multi-agente*, subvencionado por el Ministerio de Educación y Ciencia con referencia TIN2005-08501-C03-01 y ha contado con financiación de las ayudas de la Comunidad de Madrid y Universidad Complutense como *Grupo de investigación consolidado 910494*.

Referencias

1. K.S. Barber, K. Fullam, and J. Kim. Challenges for Trust, Fraud, and Deception Research in Multi-agent Systems. *Trust, Reputation, and Security: Theories and Practice*, 2631:8–14, 2003.
2. K.S. Barber and J. Kim. Belief Revision Process Based on Trust: Agents Evaluating Reputation of Information Sources. *Trust in Cyber-societies: Integrating the Human and Artificial Perspectives*, 2246:73–82, 2002.
3. K.K. Fullam, T.B. Klos, G. Muller, J. Sabater, A. Schlosser, Z. Topol, K.S. Barber, J.S. Rosenschein, L. Vercouter, and M. Voss. A specification of the Agent Reputation and Trust (ART) testbed: experimentation and competition for trust in agent societies. *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, pages 512–518, 2005.
4. L. Rasmusson and S. Jansson. Simulated social control for secure Internet commerce. *Proceedings of the 1996 workshop on New security paradigms*, pages 18–25, 1996.

Agent UNO: Winner in the 2007 Spanish ART Testbed competition

Javier Murillo and Víctor Muñoz

Universitat de Girona
Institut d'Informàtica i Aplicacions
Av. Lluís Santaló s/n 17071 Girona
{jmurillo, vmunozs}@eia.udg.es

Abstract. In multi-agent systems where agents compete among themselves, trust is an important aspect to have in mind. The ART Testbed Competition has been created with the aim of evaluating objectively different strategies that agents can use in this kind of environments. In this paper we present the winning strategy at the Spanish competition of 2007 with an analysis of the factors that have contributed to this success.

1 Introduction

In shared and competitive environments, agents interact with each other in order to achieve their goals. This interaction allows them to obtain better results than would get isolatedly. However, since agents are not interested in global outcome but only their own, maybe some of this interaction will be done with the intention of disserving them. In such situations agents need to use a trust and reputation mechanism, providing them with an uncertainty model allowing them to discern other agents' behaviors, by means of whom the agent could be able to select when and which agents to trust.

In recent years there has been a growing interest un trust mechanisms for multi-agent systems [9] and a good number of models and strategies have been proposed to deal with this [10, 3, 8]. Unfortunately, models have been tested in dissimilar problems. In consequence, with the goal of providing a "*common platform on which researchers can compare their technologies against objective metrics*" the Agent Reputation and Trust (ART) Testbed Competition was created in 2006, at both national and international level [7, 6, 4, 5]. This competition serves also as an impulse to promote research in this field and to design new strategies applicable in the real world. See [1] for more information about the ART Testbed and international competition.

In this paper we describe the winning strategy used in agent UNO for the second Spanish competition held in Valencia in March 2007. The paper is structured as follows. The next section describes the general functioning of the agent UNO. The next two sections show two fundamental components of the strategy, namely, the question and answer procedures. Then we present the results obtained with the agent in both national and international competitions of 2007.