

Jordi Sabater-Mir¹, Javier Carbó², Verónica Venturini², José Manuel Molina López²

¹Instituto de Investigación en Inteligencia Artificial, Consejo Superior de Investigaciones Científicas (CSIC), Campus UAB de Bellaterra;

²Universidad Carlos III de Madrid

<jsabater@iia.csic.es>,
<javier.carbo@uc3m.es>,
<veronica.venturini@uc3m.es>,
<josemanuel.molina@uc3m.es>

1. Introducción

La principal característica de los agentes autónomos es precisamente su autonomía. Como sucede en las sociedades humanas, donde cada individuo tiene sus propios intereses que le mueven a actuar de forma distinta, en un sistema multiagente abierto los agentes pueden tener comportamientos inesperados de acuerdo a sus objetivos individuales (que por otro lado no tienen por qué ser conocidos de antemano).

Por un lado la autonomía de los agentes es esencial pero por otro tiene que ser posible controlar, dentro de unos márgenes, el comportamiento de los individuos en cualquier sociedad. En sociedades artificiales este control se puede ejercer a distintos niveles:

■ **Control desde el punto de vista de la seguridad computacional:** A este nivel se trabaja para garantizar la capa más básica de control, como por ejemplo la identidad de los individuos o la integridad y autenticidad de los mensajes que se envían. Aquí se utilizan técnicas de criptografía y de seguridad computacional.

■ **Control desde el punto de vista institucional:** El siguiente nivel pretende dar garantías a nivel de protocolos de interacción. Una institución que actúa como marco de interacción entre agentes puede restringir las acciones de estos para que se cumplan ciertos requisitos. Por ejemplo, en una subasta, la institución puede impedir que un individuo pueda pujar si no hay una venta en curso o que un individuo pueda comprar algo si no ha demostrado antes su solvencia económica.

■ **Control desde el punto de vista social:** Aunque se pueda garantizar la identidad de los individuos y la integridad de los mensajes enviados utilizando técnicas de seguridad computacional y aunque tengamos una institución que regule los protocolos de interacción, existen aún aspectos del comportamiento que escapan a dicho control como por ejemplo la veracidad del contenido de los mensajes; ¿Cómo podemos garantizar que lo que dice un individuo es cierto? En este punto es donde los mecanismos de control social y en especial los mecanismos de confianza y reputación son útiles.

La confianza y la reputación en los sistemas multiagente

Resumen: Después de una introducción a los conceptos básicos, presentamos brevemente diversos modelos de confianza y reputación que han sido desarrollados por grupos de investigación en el ámbito de los países de habla hispana (en muchos casos en colaboración con grupos de otros países) y que son a su vez referentes a nivel internacional. Aunque lejos de ser exhaustiva, creemos que la selección de modelos ilustra adecuadamente tanto la problemática como las soluciones que se están aplicando actualmente en este campo.

Palabras clave: Confianza, reputación, sistemas multiagente.

Autores

Jordi Sabater-Mir es Científico Titular en el *Institut d'Investigació en Intel·ligència Artificial* (IIIA) del Consejo Superior de Investigaciones Científicas (CSIC) en Barcelona. Tiene un doctorado en Inteligencia Artificial por la UAB y fue becario Marie Curie en el Instituto de Ciencias Cognitivas del CNR italiano (ISTC-CNR) en Roma. Su investigación se ha desarrollado en el campo de los sistemas multiagente y en particular en el área de los modelos computacionales de credibilidad y reputación, simulación social basada en agentes y agentes cognitivos. Ha publicado más de 90 artículos en revistas científicas especializadas, congresos y *workshops* del tema. Es miembro del "EURAMAS board of directors" y fue el *PC chair* del congreso EUMAS 2009 en Chipre. Ha sido co-organizador de las últimas nueve ediciones del *workshop "Trust in Agent societies"*.

Javier Carbó es Profesor Titular de Universidad y miembro del Grupo de Inteligencia Artificial Aplicada de la Universidad Carlos III de Madrid. Es Licenciado en Informática por la Universidad Politécnica de Madrid y Doctor en Informática por la Universidad Carlos III de Madrid. Su tesis y su principal línea de investigación están centradas en la confianza y la reputación en sociedades de agentes. Hizo una estancia de investigación en la Universidad de Savoie (Francia), es autor de 2 capítulos de libro, 31 contribuciones a conferencias internacionales y 7 artículos en revista ISI-JCR dedicados a la confianza y reputación con agents. Asimismo participa en varios proyectos de investigación sobre aplicaciones de Inteligencia Ambiental.

Verónica Venturini es Ingeniera en Informática por la Universidad Católica de Salta (UCASAL, Argentina), y Máster en Ciencia y Tecnología Informática por la Universidad Carlos III de Madrid (UC3M), en el área de Inteligencia Artificial. Se encuentra desarrollando su tesis doctoral en el grupo de Investigación GIAA (Grupo de Inteligencia Artificial Aplicada) de la UC3M, relacionada a Sistemas Multi-Agentes para entornos de Inteligencia Ambiental. Ha publicado 4 artículos en congresos internacionales y 1 contribución a una revista incluida en el ISI-JCR. Actualmente, es docente de la cátedra de Inteligencia Computacional de la Universidad FASTA, sede Bariloche (Argentina).

José Manuel Molina López es doctor ingeniero de telecomunicación por la UPM, catedrático del área de Ciencia de la Computación e Inteligencia Artificial en la Universidad Carlos III de Madrid y director del departamento de Informática de dicha universidad, donde anteriormente ocupó los cargos de director del programa de doctorado en Ciencia y Tecnología Informática, Vicerrector de Estudiantes, Residencias y Actividades Culturales, adjunto al vicerrector del campus de Colmenarejo y subdirector de la titulación de Ingeniería Informática. Actualmente es el director del grupo de investigación de Inteligencia Artificial Aplicada <www.giaa.inf.uc3m.es> donde desarrolla diversos proyectos de investigación públicos y privados, centrados en la aplicación de técnicas de Inteligencia Artificial a diversos problemas de ingeniería. Es autor de más de 50 artículos en revistas internacionales con revisores y de más de 200 ponencias en congresos internacionales, de 16 libros y de diversas colaboraciones en obras colectivas.

Además de este control desde el punto de vista de la sociedad, en más de una ocasión los agentes necesitan cooperar con otros agentes o simplemente delegar tareas con lo cuál se hace necesario seleccionar entre posibles colaboradores. Aquí de nuevo la confianza

y la reputación pueden ser de gran ayuda al individuo para tomar decisiones.

Los mecanismos de confianza y reputación han estado presentes y funcionando en las sociedades humanas desde siempre. Dada la

“Por un lado la autonomía de los agentes es esencial pero por otro tiene que ser posible controlar, dentro de unos márgenes, el comportamiento de los individuos en cualquier sociedad”

naturaleza de los sistemas multiagente, no es de extrañar pues que casi desde el inicio de este paradigma se haya trabajado en adaptarlos para su uso en entornos virtuales y para ser usados por entidades artificiales.

En este artículo después de una introducción a los conceptos básicos se presentan brevemente diversos modelos de confianza y reputación que han sido desarrollados por grupos de investigación en el ámbito de los países de habla hispana (en muchos casos en colaboración con grupos de otros países) y que son a su vez referentes a nivel internacional. Aunque lejos de ser exhaustivos, la selección de modelos creemos ilustra tanto la problemática como las soluciones que se están aplicando actualmente en este campo.

2. Definiendo la confianza y la reputación

Como suele suceder, los conceptos aparentemente más sencillos son los que más problemas presentan a la hora de ser definidos. Los conceptos de confianza y reputación no son una excepción.

Como decía el sociólogo alemán Niklas Luhmann [1], “*la ausencia completa de confianza nos impediría incluso levantarnos por la mañana*”². La confianza es esencial en nuestra vida diaria ya que vivimos en un mundo incierto en el que tenemos que tomar decisiones constantemente basándonos en expectativas y manejando esa incertidumbre. Es la confianza (o desconfianza) en los otros individuos lo que nos permite tomar las precauciones necesarias para interactuar, delegar tareas y buscar colaboradores.

El mismo Luhman define la confianza³ de la siguiente forma: “*La confianza empieza donde el conocimiento [certeza] termina. La confianza proporciona las bases para tratar con imágenes del futuro inciertas, complejas y amenazadoras*”. Otra definición en la misma línea es la de Greg Eloffson que dice que “*la confianza es el resultado de un conjunto de observaciones que nos conducen a la creencia de que podemos depender de las acciones de un tercero, sin una garantía explícita, para conseguir un objetivo en una situación arriesgada*”. En todas estas definiciones vemos que el elemento clave es la incertidumbre del mundo y que la confianza se nos presenta como un mecanismo que nos permite lidiar con ella.

Hay múltiples elementos que contribuyen a la creación de la confianza. El más importante es

sin duda el conjunto de experiencias directas adquiridas en la interacción con los demás. Sin embargo estas experiencias no siempre están disponibles y puede ser muy costoso adquirirlas. Aquí es donde la reputación entra en escena como fuente alternativa de información para la evaluación de la confianza.

La reputación la podemos definir como “lo que una *entidad social* dice acerca de un sujeto relativo a su comportamiento”. Entendemos como *entidad social* un conjunto de individuos más un conjunto de relaciones entre estos individuos o propiedades que los caracterizan y que los identifican como un grupo delante de los otros miembros de la sociedad. Ejemplos de *entidades sociales* serían la familia, el vecindario, el club de golf, etc. Es lo que comúnmente se entiende como grupo social.

Cabe remarcar de la definición que lo que importa es lo que se “dice”. Aunque idealmente lo que se dice debería ser lo que se piensa no tiene por qué ser así en el caso de la reputación y de hecho en muchos casos no es así. Por otro lado esto nos indica otro aspecto importante de la reputación: *no puede existir reputación si no hay comunicación*. Podemos tener a los individuos de una entidad social que “piensan” lo mismo en referencia a un comportamiento dado, pero si no “comunican” (expresan) ese pensamiento no podemos hablar de reputación. Hay otros términos como el de “imagen compartida” que se deberían utilizar en ese caso.

La reputación es, además de una fuente para la confianza, un mecanismo de control social. Todas las sociedades humanas incorporan de una forma u otra el concepto de reputación y en muchas de ellas tiene una gran importancia. Aquellos que se comportan de acuerdo con lo esperado por la sociedad y adquieren una buena reputación tienen más posibilidades de interactuar con los demás mientras que los que van en contra de lo esperado se ven excluidos.

Tal como se ha dicho, prácticamente desde los inicios de la investigación en sistemas multiagente se detectó la necesidad de incorporar estos conceptos también en sociedades formadas por agentes autónomos, que probablemente, interactúan a su vez con humanos en sistemas híbridos. En los últimos años se han desarrollado multitud de modelos que hacen accesibles estos mecanismos a entidades artificiales [2]. Las siguientes secciones son una pequeña muestra de estos modelos.

3. ReGreT

ReGreT [3] es un modelo de confianza y reputación que tiene como principal característica la explotación de las relaciones sociales para mejorar el cálculo de los valores de confianza. Está orientado a sociedades de tamaño medio donde los individuos interactúan de forma más o menos regular y donde las relaciones sociales que se puedan establecer son duraderas.

ReGreT utiliza tres fuentes de información principal: las experiencias directas del individuo, la información proveniente de otros individuos y las ya comentadas relaciones entre los miembros de la sociedad. Cabe destacar que es un modelo no centralizado y por tanto la información de la que se nutre es aquella a la que tiene acceso el individuo, es decir, información más o menos parcial del mundo.

Un aspecto a destacar del modelo es que, además de proporcionar un valor para la confianza, calcula una medida de la credibilidad que tiene ese valor.

La **figura 1** muestra el esquema general del modelo. Las fuentes de información alimentan las dos principales fuentes para el cálculo de la confianza: por un lado la “*Direct trust*” (o confianza directa) que se basa en las experiencias directas del individuo (“*outcomes*”) y por otra la reputación.

La “*Direct Trust*” se calcula usando una agregación de los “*outcomes*” que el individuo ha obtenido hasta el momento. Un “*outcome*” no es más la diferencia entre lo esperado y lo sucedido en una determinada acción por parte de un tercero. Por ejemplo si el contrato de compra decía que el producto sería de calidad ‘A’ y llegaría en 10 días y al final la calidad se corresponde a lo esperado y el producto llega en 9 días, el “*outcome*” resultante será positivo (y por lo tanto influirá de forma positiva a la confianza en ese proveedor). Como medida de la credibilidad del valor de la “*Direct Trust*” se tiene en cuenta la antigüedad de las experiencias (no es lo mismo una experiencia reciente que una que se tuvo hace un año) y el número de éstas (no es lo mismo tener un par de experiencias que tener 100).

En el modelo ReGreT la reputación se divide en tres tipos: la “*witness reputation*” o reputación basada en la información proporcionada por otros individuos de la sociedad, la “*neighbourhood reputation*” que utiliza la reputación

“La reputación entra en escena como fuente alternativa de información para la evaluación de la confianza”

y el tipo de relación entre el individuo objetivo y otros miembros de la sociedad para extrapolar un valor de reputación (“dime con quien andas y te diré quien eres”). Y finalmente la “*System reputation*” que utiliza el rol que el individuo juega en la sociedad para asignar un valor de su reputación.

Cabe destacar en el cálculo de la “*witness reputation*” el uso de las relaciones sociales como elemento de ponderación a la hora de agregar la información que viene de otros individuos. Por ejemplo, el modelo tiene en cuenta que la información que proviene de un individuo que tiene una relación de cooperación con el individuo objetivo probablemente esté sesgada para favorecerlo. Aunque esto es difícil de saber con certeza, el modelo toma una posición conservadora y lo que hace es dar una relevancia menor a esa opinión.

En el cálculo de la reputación también se evalúa la credibilidad que tiene el valor calculado, para lo cuál se tienen de nuevo en cuenta las relaciones sociales.

Finalmente se agregan el valor de la “*Direct Trust*” con el de la reputación para obtener un valor de confianza, dando preponderancia a la “*Direct Trust*” si la medida de su credibilidad es alta (las experiencias directas, si son fiables, son la mejor fuente de información) o a la reputación en caso contrario.

4. Afras: Gestión de la reputación borrosa

Los agentes que componen el sistema denominado AFRAS (*A Fuzzy Reputation Agent System*) están dedicados a la gestión de la reputación borrosa. La reputación que un agente generará sobre cada uno de los demás agentes se basará en el comportamiento que éste ha mostrado en todas las comunicaciones cualquiera que fuera el papel que estaba desempeñando. Así la reputación que da nombre al sistema de agentes construido tiene carácter genérico.

Asumida esta interpretación genérica de la reputación, AFRAS representa este concepto mediante conjuntos borrosos ya que

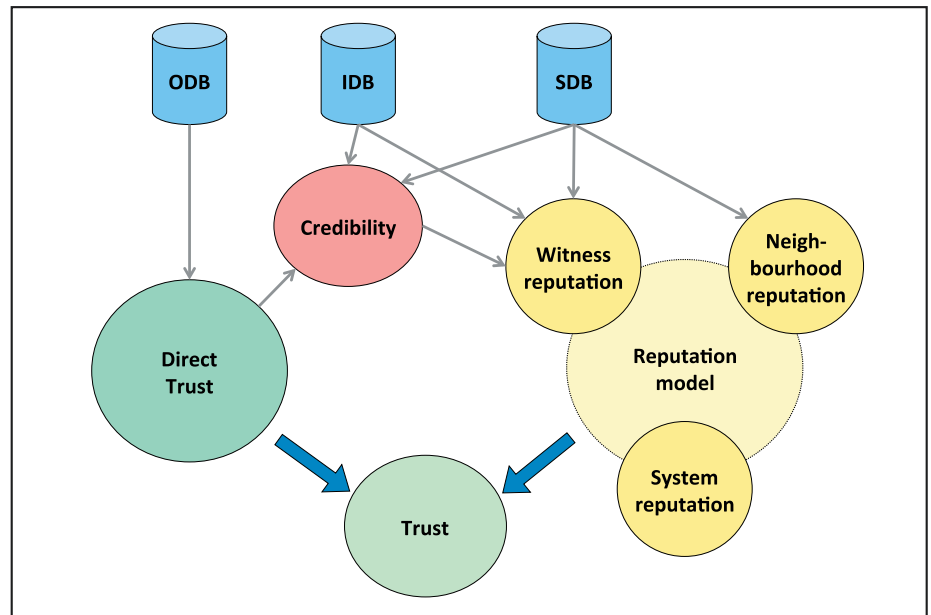


Figura 1. Esquema general de Regret.

la reputación basada en criterios subjetivos se expresa de una forma vaga e inconcreta, debido al carácter abierto y dinámico de los sistemas de agentes esta información puede ser incompleta e incierta.

Por lo tanto, AFRAS define cómo se va modificando el conjunto borroso de la reputación de un agente al producirse nuevos valores de satisfacción correspondientes a interacciones individuales definidas también como un conjunto borroso. Para ello AFRAS agrega estos valores mediante una media ponderada entre la reputación anterior y la nueva satisfacción producida, y determina usando un peso cuánto recordamos de la reputación previa del otro agente.

AFRAS interpreta este peso como el “rencor” de nuestro agente. Esta característica del agente va cambiando con el tiempo, modificándose en función del grado de decepción sufrida en una interacción. Donde el grado de similitud entre la satisfacción producida por la interacción y la reputación previa del vendedor, tomada como predicción de su comportamiento, determina el grado de decepción sufrida.

Adicionalmente al uso del “rencor” como peso en la media ponderada se aplica una función de olvido dependiente del instante del tiempo transcurrido entre la última actualización de la reputación y la valoración recibida recientemente.

Asimismo, el sistema de agentes AFRAS pretende emular el funcionamiento de la sociedad humana en la gestión y comunicación de esta reputación genérica subjetiva. En consecuencia, con este objetivo de reflejar esta relación social con un sistema de agentes, las decisiones se toman de acuerdo a atributos

que caracterizan el comportamiento social de un agente como el de un ser humano: el egoísmo, la susceptibilidad y la timidez de un agente determinan su comportamiento.

Consideraremos que un agente estará comportándose sin timidez si estuviera evitando el aislamiento, y en consecuencia, debería dar más opciones a los agentes desconocidos. El egoísmo decide si se ejecuta o no la intención de responder a una pregunta de otro agente en lugar de ejecutar la intención de preguntar a otro agente. La susceptibilidad representa cuantos celos tiene un agente a la hora de tomar decisiones con un coste económico.

El grado de solapamiento de estas características, también definidas como conjuntos borrosos, con la reputación del agente con el que se planea interactuar determina la decisión tomada. Al igual que ocurría con el rencor, estas características van adaptándose al grado de decepción que va sufriendo el agente en cuestión.

La arquitectura interna de los agentes AFRAS está inspirada en los principios del paradigma deliberativo ya que tiene tres distintos niveles de abstracción (ver figura 2). Estos se corresponden con el modelo del mundo, social, y mental que contiene cada agente. Estos modelos definen una arquitectura por capas, que se activa de forma ascendente y se ejecuta de forma descendente.

Este modelo ha sido probado en distintos escenarios ad hoc con diferente grado de éxito: evaluando la velocidad a la que converge al comportamiento real de agentes constantes sin recomendaciones [4], el grado de influencia positiva que tienen las recomendaciones [5], y la protección que proporciona ante confabulaciones [6].

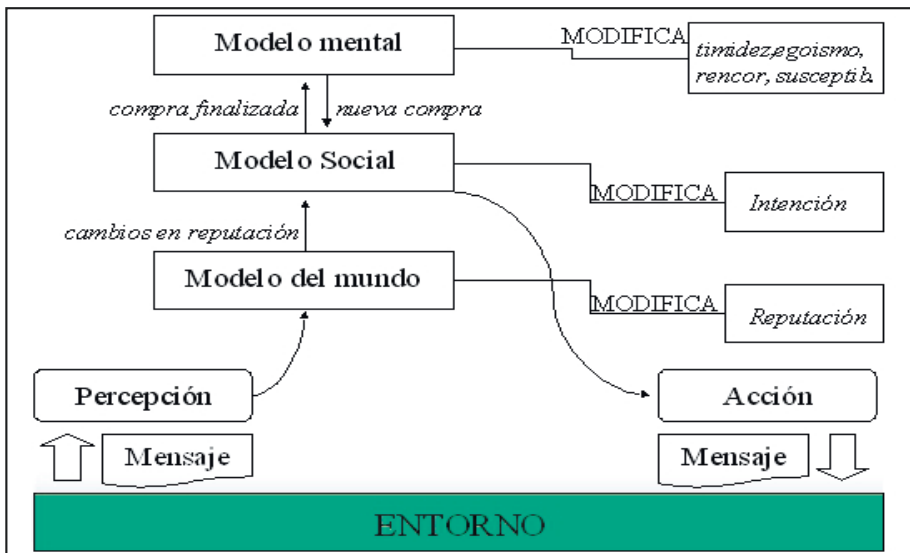


Figura 2. Arquitectura interna de un agente AFRAS.

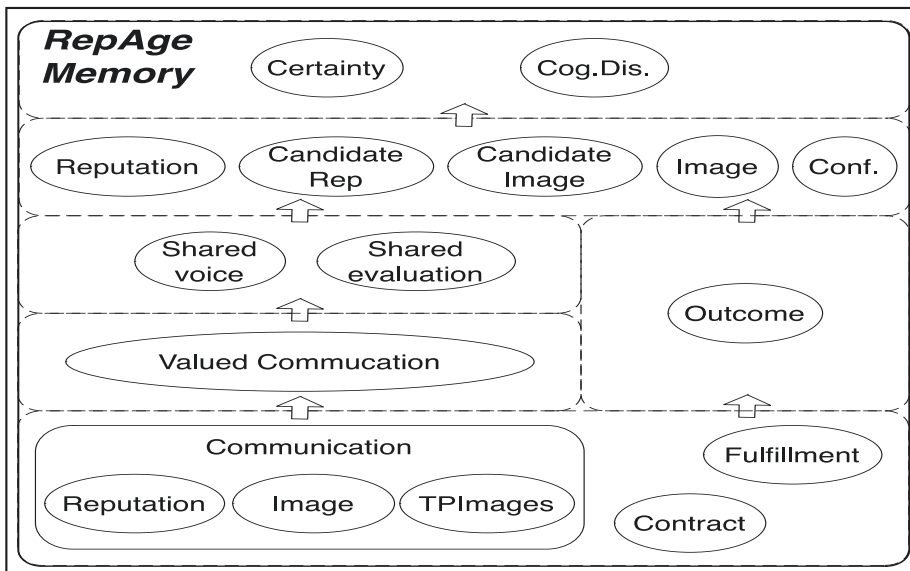


Figura 3. Arquitectura de Repage.

5. RepAge

El modelo RepAge [7] es un modelo de reputación que presenta una aproximación distinta a la de los demás modelos comentados hasta ahora. Tradicionalmente los modelos de confianza y reputación se han considerado como cajas negras a las que se les puede consultar para obtener valores lo más precisos posible sobre la confianza y la reputación de terceros. Como se realiza ese cálculo internamente es irrelevante desde la perspectiva de los demás componentes que integran el agente.

Este modelo, basado en una teoría cognitiva sobre la reputación, pone un énfasis especial en cómo se representa la información internamente en el modelo para precisamente explotar no sólo el valor final de la reputación, sino como se ha llegado a ese valor. Posteriores trabajos han mostrado como utilizar esa transparencia para, por ejemplo, dotar al agente de la capacidad de justificar

los valores de reputación. Esto permite a los agentes participar en procesos argumentativos para discutir sobre esos valores tal como hacemos los humanos. Al estar basado en una teoría cognitiva, es un modelo que facilita la interacción entre agente y humano ya que es capaz de presentar la información en una forma que el humano entiende fácilmente.

La figura 3 muestra los elementos que conforman la parte más importante de la arquitectura, la memoria del modelo RepAge. Esta memoria contiene predicados lógicos interrelacionados entre sí que representan distintos conceptos asociados a la construcción de la reputación. La idea es que la trama de predicados que se va generando describe como se ha llegado a ese valor.

Por ejemplo, el modelo guarda constancia de que la imagen del agente A asociada a la actividad “cocinar” es 0,8 (donde 0 es mala

reputación y 1 muy buena). A ese valor se ha llegado porque hay tres experiencias directas positivas (por ejemplo resultado de ir a cenar a casa de A) que lo avalan y una comunicación por parte de un informador (un amigo que también ha ido a comer a casa de A) que coincide con las experiencias directas. A su vez el informador es creíble porque en el pasado sus informaciones han sido exactas.

Este tipo de construcción en la memoria de RepAge permiten más tarde al agente argumentar con alguien que ponga en duda las capacidades de A como cocinero, intercambiando en lo que se conoce como “argumentation game” las bases de la valoración. El resultado puede ser la reafirmación en el valor ya calculado o su revisión a la luz de los argumentos dados por el otro agente.

6. Un modelo anticipatorio de Trust

En este modelo la confianza se centra en una única dimensión: la competencia, utilidad o satisfacción esperada de un agente respecto a un contexto particular. Más concretamente, la confianza como una creencia que estima la calidad del servicio esperado de un agente dado, basado en los servicios proporcionados directamente por ese agente y por información obtenida de otros agentes.

Típicamente un modelo de trust considera esas dos fuentes de información: experiencias directas y recomendaciones. En este modelo además se distinguen las recomendaciones proporcionadas por un agente sobre sí mismo. Así, este modelo distingue tres tipos de fuentes de confianza: Direct Trust (DT), Advertisements-based Trust (AT), and Recommendations-based Trust (RT).

Con objeto de adaptarse más rápidamente a la naturaleza dinámica e incierta de un entorno abierto, un agente puede anticiparse, es decir, tener creencias relativas a las expectativas sobre las posibles consecuencias de sus actos. Así en este modelo se distingue entre las creencias ‘históricas’, basadas en experiencias pasadas (DT), y creencias anticipatorias, que incluyen un componente de especulación, no necesariamente objetivo y racional (AT y RT).

Para representar la ignorancia y la incertidumbre, este modelo utiliza dos dimensiones relativas a la confianza (confidence) asociada a una creencia:

- Intimidad: es una medida relativa al número de interacciones que soportan una determinada creencia ($T \in \{DT, AT, RT\}$) sobre un agente j .
- Predictibilidad: es una medida basada en la dispersión o variabilidad de los datos hasta el instante t que soportan una determinada creencia ($T \in \{DT, AT, RT\}$) sobre un agente j .

“ Una de las propuestas más novedosas consiste en un modelo de reputación adaptado a las necesidades de los dominios de inteligencia ambiental ”

Estas dos dimensiones se combinan en un único valor de confianza de una creencia utilizando una t-norma (operador \otimes).

El *Direct Trust* de un agente representa el comportamiento pasado del agente de acuerdo a criterios objetivos, y se determina en este modelo mediante la suma ponderada de sus experiencias directas, donde el peso utilizado es una función de olvido, $\varphi(t, T)$ definida como:

$$\varphi(t, T) = \begin{cases} 1 & \text{si } T-t \geq \emptyset \\ \cos\left(\frac{\pi}{2\emptyset} \cdot (T-t)\right) & \text{si } 0 < T-t < \emptyset \end{cases}$$

Donde \emptyset es un parámetro que establece la máxima antigüedad de una experiencia para ser relevante (influir en las creencias). Así para un valor $\emptyset = 10$ tenemos la curva representada en la **figura 4**.

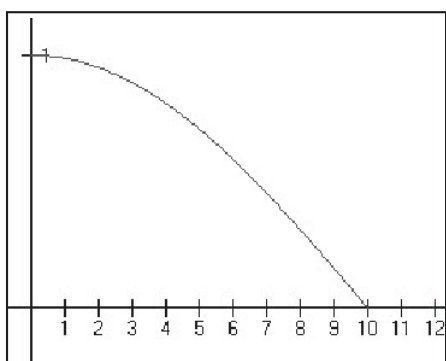


Figura 4. Representación de la curva de olvido de una experiencia.

Por otro lado, el *Advertisements-based Trust* (AT) de un agente a se refiere a la expectativa de comportamiento en el próximo instante $T+1$ en base al comportamiento anunciado por el propio agente y a las discrepancias pasadas entre sus anuncios (*advertisements*) y el comportamiento observado posteriormente. Este modelo da el mismo peso a ambos valores, calculándose el resultado como una resta del anuncio menos la discrepancia media observada en el pasado.

Los valores de *Recommendations-based Trust* (RT) de un agente se calculan de una forma análoga a los de AT mediante el comportamiento que otros agentes relataron sobre el agente objeto de estudio y la discrepancia de las opiniones que dieron en el pasado con los comportamientos observados a posteriori en el agente objeto de estudio. Este modelo calcula este valor de forma semejante al *Advertisements-based trust*, dando el mismo peso

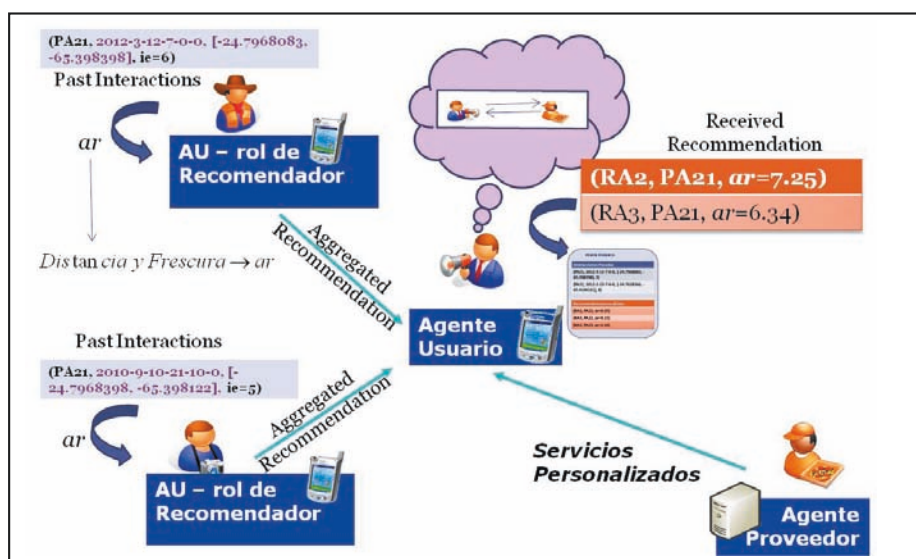


Figura 5. Esquema de las comunicaciones del modelo CALoR.

a ambos factores, calculándose el resultado como una resta de la recomendación menos la discrepancia media observada en el pasado.

Estos tres tipos de confianza deben integrarse en una única medida global de confianza (*Global Trust*) como una media ponderada de estas creencias (DT, AT y RT), utilizando como pesos sus valores de *confidence* correspondientes.

La experimentación de este modelo fue realizada con el ART *testbed* [8] y se encuentra detallada en [9].

7. CALoR: Un modelo de reputación para inteligencia ambiental

Una de las propuestas más novedosas consiste en un modelo de reputación adaptado a las necesidades de los dominios de inteligencia ambiental: CALoR (*CALoR: Context-Aware and Location Reputation Model*) de Venturini, Carbó y Molina. Este modelo está actualmente en desarrollo y se enmarca dentro del contexto de la arquitectura de agentes para aplicaciones descrito en [10].

La personalización que persigue todo sistema de inteligencia ambiental se alcanza mediante la localización. Un dato esencial entonces para el cálculo de la reputación en estos entornos es la distancia de interacción. La suposición que asume este modelo consiste en que a menor distancia una interacción se puede evaluar con mayor precisión/fiabilidad. Por ejemplo, al observar un cuadro en la visita a un museo (ver **figura 5**).

Sigue esta línea el modelo descrito en [11] para clasificar y filtrar información geográfica aportada colaborativamente. Sin embargo dicho modelo no tiene en cuenta la frescura de la información. En un modelo de reputación es necesario establecer una relación entre el tiempo en el que se asignó un valor de evaluación de la interacción y cuan reciente es este dato ya que suponemos dinámico el comportamiento de los agentes a lo largo del tiempo.

En este modelo propuesto por Venturini, Carbó y Molina, se propone una función que evalúa la experiencia directa que tuvo el recomendador, y que permitirá producir la recomendación pedida en relación a cuan reciente fue dicha interacción y su relación geoespacial con el lugar físico del proveedor de productos/servicios en el momento del intercambio.

A este valor lo denominaremos *Aggregated Recommendation* (ar), y tendrá la forma de una suma ponderada de las recomendaciones (interacciones del recomendador con el proveedor de interés), normalizada utilizando un ajuste necesario para considerar los valores de distancia dentro de un rango limitado.

Este valor de ajuste utilizado como peso de cada recomendación es:

- Directamente proporcional a un valor de frescura calculado con la fórmula propuesta en el modelo FIRE [12].
- Inversamente proporcional al logaritmo de la distancia desde la que interactuó el recomendador con el proveedor objeto de estudio.

El uso de la inversa del logaritmo de la distancia permite dar mayor relevancia a diferencias pequeñas entre dos interacciones cuando la distancia desde la que se interactúa es pequeña, y menor relevancia a diferencias grandes cuando la distancia desde la que se interactúa es grande (ver **figura 6**).

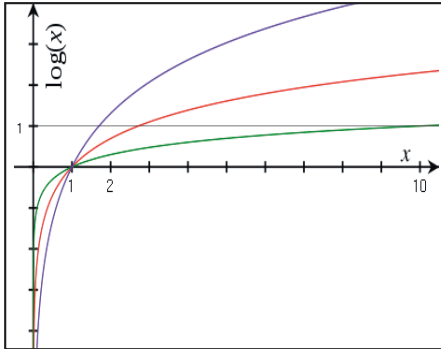


Figura 6. Gráfica de la función logarítmica.

9. Conclusiones

En este artículo hemos revisado brevemente algunos de los diversos modelos de confianza y reputación que han sido desarrollados por grupos de investigación en el ámbito de los países de habla hispana y que por sus correspondientes publicaciones pueden considerarse referentes a nivel internacional. Aunque esta revisión no es completa, la selección de modelos ilustra los diferentes enfoques, tanto de definición del problema como de representación de la información y de formulación matemática que se están aplicando actualmente en este campo.

Referencias

- [1] N. Luhmann. *Trust and Power*. Chichester: Wiley, 1979.
- [2] I. Pinyol, J. Sabater-Mir. Computational trust and reputation models for open multi-agent systems: a review. *Artificial Intelligence Review, Springer Netherlands, Online First*, pp. 1-25, 2011.
- [3] J. Sabater. Trust and Reputation for Agent Societies. *Monografies de l'institut d'investigació en intel.ligència artificial*, 20, IIIA-CSIC, 2003.
- [4] J. Carbo, J.M. Molina, J. Davila. Trust management through fuzzy reputation. *Int. Journal of Cooperative Information Systems, Vol. 12, Num. 1*, 2003.
- [5] J. Carbo, J.M. Molina, J. Davila. Fuzzy referral based cooperation in social networks of agents. *Artificial Intelligence Communications, Vol. 18, Num. 1*, 2005.
- [6] J. Carbo, J.M. Molina, J. Davila. Avoiding Malicious Agents in E-Commerce Using Fuzzy Recommendations. *Journal of Organizational Computing and Electronic Commerce, Vol. 17, Num. 1*, 2007.
- [7] J. Sabater-Mir, M-Paolucci, R. Conte. RePage: REPutation and imAGE among limited autonomous partners. *JASSS - Journal of Artificial Societies and Social Simulation*, 9 (2), 2006.
- [8] K. Fullam, T. Klos, G. Muller, J. Sabater, A. Schlosser, Z. Topol, K. S. Barber, J. Rosenschein, L. Vercouter, M. Voss. 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)*, pp. 512-518, 2005.
- [9] M. Gómez, Javier Carbo, Clara Benac. Honesty and trust revisited: the advantages of being neutral about other's cognitive models. *Autonomous Agents and Multi-Agent Systems, vol. 15, num 3*, pp 313-335. Springer (December 2007). ISSN: 1387-2532.

[10] V. Venturini, J. Carbo, J. M. Molina. Methodological design and comparative evaluation of a MAS providing Aml. *Expert Systems With Applications Journal*. marzo, 2012.

[11] M. Bishr, L. Mantelas. *A trust and reputation model for filtering and classifying knowledge about urban growth*. Publicado online: 29 de julio de 2008. Springer Science+ Business Media B.V., 2008. <http://www.iacm.forth.gr/_docs/pubs/4/mantelas-geojournal%202008.pdf>.

[12] T. D. Hunyh. *Trust and Reputation in Open Multi-Agent Systems (FIRE)*. Tesis doctoral, University of Southampton, 2006.

Notas

¹ "A complete absence of trust would prevent [one] even getting up in the morning".

² En la literatura sobre el tema se utiliza el término en inglés "trust" que aquí hemos traducido como "confianza". También se traduce en algunas ocasiones como "credibilidad".