

Agentes inteligentes aprenden a establecer relaciones entre socios en mercados bilaterales

Constanza Quaglia¹, Mercedes Canavesio¹, Ernesto Martinez^{1,2}

¹CIDISI Centro de Investigación y Desarrollo en Ingeniería en Sistemas de Información
Facultad Regional Santa Fe - Universidad Tecnológica Nacional (FRSF-UTN)
Lavaisse 610 - Santa Fe - SF - Argentina

²INGAR Instituto de Desarrollo y Diseño - CONICET-UTN - Santa Fe - SF - Argentina

Cotii.q@gmail.com, mcanaves@frsf.utn.edu.ar, ecmarti@conicet-santafe.gov.ar

Resumen. *En respuesta a presiones competitivas, pequeñas y medianas empresas, han comenzado a formar alianzas o redes entre ellas. Esto les permite a cada compañía incrementar sus habilidades para reaccionar y adaptarse a los cambios en su entorno de negocios, concentrarse en sus competencias claves, incrementar la disponibilidad de recursos y ganar en economía de escala. Así, el modelo de la compañía fractal basada en proyectos permite la integración virtual y temporal entre diferentes empresas para lograr objetivos particulares de negocios. El éxito de la aplicación del modelo radica en el establecimiento de relaciones cliente-servidor entre los gestores de los proyectos. Este trabajo analiza la incorporación de aprendizaje en los agentes de gestión, para seleccionar socios con quienes establecer relaciones cliente-servidor más beneficiosas. Para ello propone la aplicación del aprendizaje por refuerzo enfocado en el problema del bandido de 2-bazos.*

Abstract. *In response to global competition, small and medium companies have begun to form alliances and networks between them. This allows each company to increase their ability to react and adapt to changes in their business environment, focus on their core competencies, increase the availability of resources and gain economies of scale. Thus, the model of fractal-based company based-project enables virtual and temporal integration between different companies in order to achieve specific business objectives. The successful implementation of the model lies in the establishment of client-server relationships between project managers. This paper analyzes the integration of learning management agents to select partners with whom to establish the most beneficial client-server relationships. He advocates the application of reinforcement learning focused on the two-sided bandit problem.*

1. Introducción

La globalización de los mercados, los rápidos avances tecnológicos y la creciente personalización de los productos y servicios crean una fuerte competencia entre las empresas. Por ello, necesitan responder a demandas que son cada vez más personalizadas, incrementando permanentemente la variedad de sus productos y servicios, y al mismo tiempo reduciendo los costos e incrementando la calidad de los mismos [ElMaraghy, 2009].

Estas condiciones se acentúan en aquellas compañías dedicadas al diseño de productos y tecnologías innovadoras, como son las empresas que pertenecen al sector bio-industrial, donde cada nuevo proyecto involucra alta incertidumbre, costos y riesgos [Shah, 2004; Pisano, 2006]. Por ello, estas compañías deben concentrarse en la efectiva gestión de los procesos de desarrollo, tanto de sus nuevos productos como de sus métodos de manufactura [Pisano, 2006]. Para superar esta situación, fundamentalmente pequeñas y medianas compañías de este sector industrial, junto a organizaciones públicas y privadas, universidades y centros de investigación, están formando redes de asociación estratégicas.

Estas redes les permiten acelerar los procesos de diseño, controlar los costos de desarrollo, reducir el riesgo y explorar nuevos nichos de mercados [Shah, 2004]. Al mismo tiempo, los avances en las tecnologías de la información y las comunicaciones han facilitado la formación de estas alianzas entre empresas. Las redes o alianzas entre compañías ofrecen varios beneficios a cada uno de sus miembros, tales como:

- mejorar la asignación y apalancamiento de recursos internos,
- acceso a tecnologías, experiencias y habilidades externas las cuales no están presentes en la compañía o es más económico o rápido su contratación externa,
- reducción de tiempos de diseño y desarrollo,
- gestión distribuida del riesgo en cada proyecto.

Para capturar los beneficios que otorga una red de empresas, es necesario definir un modelo de empresa que influya por diseño el comportamiento de cada compañía involucrada y las relaciones que entre ellas se establecen. Atendiendo a esta problemática [Canavesio y Martinez, 2007; Canavesio y Martinez, 2014] han propuesto un modelo de compañía fractal basado en proyectos que permite la integración virtual y temporal entre empresas para el logro de una meta particular de negocio.

En este modelo, el proyecto es la unidad fractal de gestión, auto-gobernada, interdependiente y temporal que otorga a la red de empresas flexibilidad, rapidez de respuesta y agilidad para adaptarse a los cambios de su entorno de negocios. La unidad fractal de gestión o proyecto se concibe constituida por un agente gestor y un objeto gestionado. Este último puede ser un fin (meta) o un medio (recurso) necesario para el logro de aquel fin. De allí que los roles que los gestores de proyecto pueden asumir son el de gestor de fines o gestor de medios, respectivamente. Los actores que desempeñan tales roles se vinculan entre sí e interactúan a través de relaciones cliente-servidor. La eficiencia y la eficacia de los resultados obtenidos en el logro de sus metas, dependerá de la estrategia de selección de socios que cada gestor de proyecto implemente.

Enfoques tradicionales consideran sólo el precio como único criterio para llevar a cabo tal selección, pero es necesario utilizar técnicas de toma de decisiones multicriterio, que consideran calidad, despacho, costos y servicios además de incorporar factores sociales, políticos y de atención al cliente. Así, el proceso de selección de socios involucra información incompleta e incierta con respecto a los servidores candidatos, sus habilidades y experiencias [Zeydan, Colpan y Cobanoglu, 2011; Huang, Gao y Chen, 2011].

El concepto de relación cliente-servidor entre los gestores, es fundamentalmente importante para este modelo de empresa integrada, por lo que la selección, tanto sea del mejor servidor para cada cliente como el mejor cliente para cada servidor, es clave en la obtención del mayor beneficio en la integración entre las empresas. Por ello, es necesario dotar a los gestores de proyecto con capacidad de aprendizaje, que individualmente les permita definir y modificar su comportamiento sobre la base de sus preferencias, percepciones, habilidades e interacción con otros gestores de la compañía fractal.

En la literatura, autores como Hu y otros (2001), Valluri y otros (2005), Kim y otros (2008), Derhami y otros (2010), abordan esta problemática implementando diferentes problemas de aprendizaje por refuerzo [Sutton y Barto, 1998]. Los trabajos de estos autores, plantean el problema de toma de decisión sólo desde la perspectiva del agente que emite un requerimiento por recurso considerando a los proveedores como actores pasivos o al menos sin otorgarles la posibilidad de aprender sobre su entorno y decidir responder sólo aquellos requerimientos por recursos que más los satisfagan económicamente. Esta perspectiva también fue considerada en el trabajo de Canavesio y Martinez (2011), que aplica aprendizaje por refuerzo en los gestores de la compañía fractal basada en proyectos. Allí los autores proponen algoritmos de aprendizaje aplicados a los gestores de fines y a los gestores de medios, pero sus aprendizajes son independientes del logro de parejas cliente-servidor y del éxito de las mismas.

Por ello, este trabajo implementa algoritmos de aprendizaje por refuerzo [Sutton y Barto, 1998] que permiten analizar, en un mercado de coincidencias como el de la compañía fractal basada en proyectos, cómo los agentes clientes identifican servidores eficaces y de mínimo costo, mientras los servidores aprenden a detectar clientes, cuyos contratos les otorguen el mayor beneficio económico, conduciendo y premiando el logro de coincidencias entre clientes y servidores en cada instancia de negociación.

2. La compañía fractal basada en proyecto

La idea de la compañía fractal [Warnecke, 1993] es un modelo de empresa conceptual, que a través de unidades autónomas, descentralizadas e interdependientes, denominadas fractales, otorga a las empresas mayor flexibilidad y agilidad para adaptarse a los cambios en su entorno de negocios. Un fractal es definido como una estructura que describe un patrón idéntico, que se replica a sí mismo a distintos niveles de abstracción, de manera recursiva. En el modelo de empresa fractal propuesto por [Canavesio y Martinez, 2007, Canavesio y Martinez, 2014], la unidad fractal de gestión se concibe como un proyecto. Esta unidad fractal de gestión o proyecto combina distintas habilidades, destrezas y competencias necesarias para lograr un objetivo específico

(satisfacer un requerimiento por recursos, diseñar de un nuevo equipo de RX, desarrollar una prótesis, etc.).

Dentro de la red de empresas, cada proyecto posee las siguientes propiedades que están presentes en cada instancia, más allá de su nivel de abstracción:

- *Orientado a la meta*: cada proyecto produce una resultante específica para un cliente claramente definido.
- *Autónomo*: cada gestor de proyecto posee suficiente libertad para ejecutar actividades y gestionar los recursos involucrados para el logro de una dada meta.
- *Temporal*: cada proyecto debe lograr su meta en una cantidad limitada de tiempo.
- *Relación recursiva*: cada proyecto puede ser definido como parte de un super-proyecto o puede contener distintos niveles de sub-proyectos (sub-proyectos, sub-sub-proyectos, ..., etc.)
- *Ciclo de vida* que cubre cuatro etapas: definición, ejecución, control y cierre.
- La información y conocimiento relevante del proyecto es almacenada y compartida por todos los gestores intervinientes y les permiten mejorar sus futuras decisiones.

En el modelo de la compañía fractal basada en proyectos, la unidad fractal propuesta se compone de un gestor de proyecto que gestiona la misma y de un objeto que es gestionado por éste (Figura 1). El gestor de proyectos es un actor o agente inteligente, que posee suficiente libertad para tomar decisiones, ejecutar acciones, aprender y ajustar permanentemente su comportamiento a la dinámica del mercado. Dado que en el modelo, tanto los fines o metas como los medios o recursos son gestionados a través de proyectos, el gestor de un proyecto asumirá el rol de gestor de fines o gestor de medios, respectivamente. Ambos roles poseen funciones y responsabilidades claramente definidas. Así, un actor desempeñando del rol de gestor de fines será responsable de gestionar, coordinar e integrar las tareas multidisciplinarias e interdependientes necesarias para lograr la meta de su proyecto, haciendo un uso eficiente y eficaz de los mejores recursos que logre negociar para emplear en sus tareas. Por otro lado, el gestor de recursos será responsable por la provisión eficaz de los medios o recursos que hubiera comprometido para el logro de un dado fin.

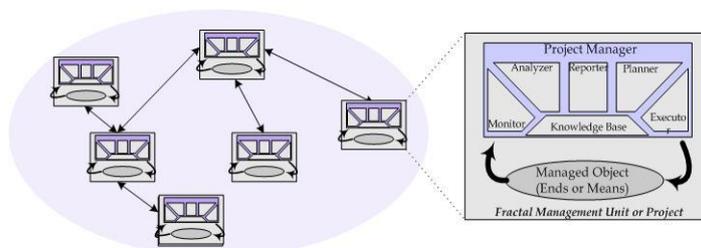


Figura 1. Estructura del proyecto como unidad fractal de gestión

En el modelo propuesto, las empresas se vinculan entre sí a través de relaciones cliente-servidor, entre gestores de proyectos (clientes) y gestores de recursos

(servidores). Estas relaciones cliente-servidor se establecen a través de algún mecanismo de negociación entre agentes interesados, en la asociación virtual y temporal para el logro de un objetivo de negocio particular. El modelo de la compañía fractal es un *mercado bilateral* (two-sided markets) [Sarne y Kraus, 2008; Rochet y Tirole, 2008; Kumar, Lifshits y Tomkins, 2010; Chen y Song, 2013], donde existen clientes que anuncian requerimientos por la provisión de recursos a diversos proveedores, y a la vez, existen proveedores de recursos que seleccionan a un conjunto de clientes a quienes desean proveerle sus recursos. Así se logran pares cliente con servidores de recursos que definen una relación cliente-servidor, por lo que el modelo se debe considerar como un *mercado de coincidencias* [Sotomayor, 2004].

Para ello, se propone la incorporación de algoritmos de aprendizaje por refuerzo en ambos roles de los gestores de proyecto, enfocándolo en el problema del bandido de 2-brazos (two-sided bandit problem) que les permitirá hallar coincidencias entre ellos.

3. La compañía fractal como un sistema multi-agente

La compañía fractal es un sistema social, en la que intervienen actores o agentes inteligentes, que desempeñan roles e interactúan, a través de relaciones cliente-servidor, para lograr objetivos de negocios específicos. Un agente es definido como una entidad física o virtual que en un entorno dado, es capaz de actuar de manera autónoma con el fin de alcanzar sus objetivos [Russell y Norving, 2004]. Además, un agente posee capacidad de aprendizaje, que le permite capitalizar experiencia, modificar su política de actuación y la representación interna del entorno para adaptar su comportamiento a los cambios del mismo.

La estructura organizacional y de gestión del modelo de la compañía fractal propuesto, se basa en el concepto de relaciones cliente-servidor, establecidas entre gestores de proyecto. En ellas, el gestor de proyecto cliente negocia la utilización de un recurso dado o la delegación del logro de una sub-meta a otro gestor de proyecto servidor. Por ello, elegir el mejor socio en cada oportunidad es clave para obtener el mayor beneficio de la integración entre las empresas. Este proceso de selección involucra información incompleta e incierta con respecto a los servidores candidatos, sus habilidades y experiencias. Además, algunas de las características consideradas (performance, calidad, confiabilidad, etc.) son subjetivas o cualitativas dificultando aún más la definición o valoración concreta de las mismas [Huang, Gao y Chen, 2011; Hu y Wellman, 2001; Ren y Anumba, 2002].

En toda organización, la selección de socios es una cuestión clave dado que por un lado buenos socios conducen a reducir costos operativos y mejorar la calidad de productos y/o servicios finales y por otro lado, la elección incorrecta puede causar problemas operativos y financieros [Zeydan, Colpan y Cobanoglu, 2011; Valluri y Croson, 2003; Kim, Bilsel y Kuman, 2008; Derhami, Saadatjoo y Saadatjoo, 2010]. Así, la selección de socios de negocios es un problema no determinístico, desafiante y complejo por lo que para resolverlo es necesario desarrollar un sistema inteligente que lo soporte [Derhami, Saadatjoo y Saadatjoo, 2010].

Por ello, es necesario dotar a los gestores de proyecto, en sus diferentes roles, con capacidad de aprendizaje, que individualmente les permita definir y modificar su comportamiento en base de sus preferencias, percepciones, habilidades, nivel de

conocimiento, comunicación e interacción con otros gestores en la compañía fractal. Este proceso de aprendizaje les permitirá mejorar constantemente sus desempeños y reducir la incertidumbre al aprender sobre las preferencias, creencias y estrategias de los otros agentes, para así competir eficazmente con ellos.

3.1. Aprendizaje por refuerzo en agentes que seleccionan socios en mercados bilaterales

El aprendizaje por refuerzo [Sutton y Barto, 1998] es un enfoque computacional para entender y automatizar el aprendizaje orientado al logro de metas y toma de decisiones en una secuencia. Mientras un agente interactúa con su entorno, aprende por prueba y error cual acción ejecutar. En cada episodio, el agente selecciona una acción posible en el actual estado y la ejecuta, causando que el entorno se mueva al siguiente estado. El agente recibe una recompensa que refleja el valor de la acción tomada (Figura 2).

El objetivo del agente es maximizar la suma de las recompensas acumuladas desde un estado inicial hasta que alcanza el estado final. Inicialmente, el agente desconoce el curso de acción a tomar en función del contexto. A través de la interacción, el agente descubre qué acciones tienen mayor recompensa tras un análisis retrospectivo de los resultados (aciertos y errores) que ha obtenido. La implementación de agentes que aprenden por refuerzo se lleva a cabo utilizando una estructura compuesta por los siguientes elementos:

- *Política*, define el objeto de optimización y mejora el conocimiento disponible por el agente.
- *Función recompensa*, define el objetivo que se espera satisfacer al final de cada episodio.
- *Función valor o utilidad* proporciona una medida de la efectividad de una política dada.
- El *modelo del entorno* imita el comportamiento del mismo.

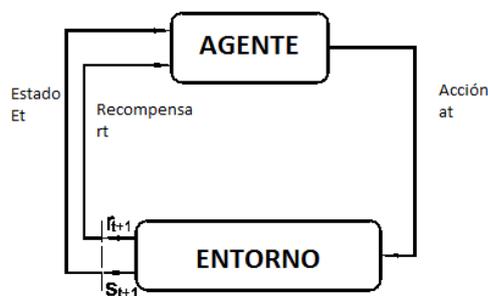


Figura 2. Aprendizaje por refuerzo (Adaptado Sutton y Barto, 1998)

En el caso particular del problema de bandido de n -brazos, un agente debe elegir cuál de los n brazos jalar en cada período de tiempo para maximizar la recompensa recibida, mientras simultáneamente trata de estimar la distribución de recompensas de cada uno de los brazos. El agente debe decidir entre jalar el brazo con el valor más alto esperado y jalar el brazo que le permita aprender más sobre su distribución de recompensas. Se propone especializar esta situación al problema del bandido de 2-brazos, donde un brazo obtiene una recompensa basado en quien lo jaló y que él puede rechazar a quien lo hizo.

El problema de aprendizaje enfocado como el problema del bandido de 2-brazos es una formulación natural para mercados en los cuales existen dos tipos diferentes de agentes que deben lograr coincidencias entre ellos, repetidas veces. Ejemplos de estos mercados son el de citas, en el cual hombres y mujeres van en varias ocasiones a citas mientras aprenden sobre sus candidatos; el mercado laboral, en donde empleadores y potenciales empleados aprenden el uno del otro durante las entrevistas [Rochet y Tirole, 2006; Das y Kamenica, 2005; Das, 2006] como así también el mercado de la compañía fractal basada en proyectos, cuando los gestores de proyecto deben decidir con quién asociarse para establecer relaciones cliente-servidor. Si bien un simple problema del bandido de 2-brazos no podrá capturar todos los aspectos de estos mercados si puede proveer un útil punto de partida para estudiarlos.

3.1.1. El modelo de aprendizaje

Un plan de proyecto está integrado por cientos o miles de tareas que requieren de diferentes recursos y habilidades para su ejecución. El gestor de proyecto dispone de un centro de potenciales proveedores, algunos calificados como confiables y capaces de proveer recursos con el nivel de calidad requerido y otros que no reúnen tales condiciones. A priori el gestor de proyecto desconoce esta calificación de los potenciales proveedores por lo que, deberá aprender a distinguirlos.

Para ello, el gestor de proyecto utiliza un parámetro denominado *factor de credibilidad*, que considera la capacidad de un proveedor de cumplir en tiempo y forma sus contratos. Inicialmente, todos los gestores de recursos son considerados confiables para proveer de recursos para la tarea ta . Mientras el gestor de proyecto tenga poca experiencia sobre la credibilidad de los proveedores, incurrirá en elevados costos, por elegir servidores económicos e ineficaces o eficaces pero onerosos. Cuando el agente ha adquirido conocimiento respecto a quienes son los servidores confiables y además capaces de proveerle los mejores recursos para una dada tarea, se reduce la búsqueda, la incertidumbre y los costos asociados debido a que las negociaciones se circunscriben sólo a este grupo de proveedores.

Por otro lado, los proveedores de recursos deben aprender qué contratos de provisión aceptar para incrementar beneficios y credibilidad ante sus clientes. Durante su aprendizaje, un gestor de recursos novato desconoce su potencial y por lo tanto aceptará todos los contratos adjudicados para la provisión de recursos sin discriminar las tareas, incurriendo en elevados costos adicionales que reducirán considerablemente beneficios y credibilidad, debido a los incumplimientos. Esta situación se irá revirtiendo a través de sucesivas interacciones que le permitirán al servidor reconocer a que tareas es capaz de proveerle recursos.

En la compañía fractal existen F gestores de fines y M gestores de medios, que interactúan durante T períodos de tiempo o episodios intentando aprender y conocer la confiabilidad y capacidad de los socios entre quienes establecer relaciones cliente-servidor exitosas.

El mecanismo de emparejamiento está basado en el algoritmo de Gale-Shapley [Das y Kamenica, 2005], complementado con el algoritmo de aprendizaje por refuerzo denominado Q-learning [Sutton y Barto, 1998]. En el algoritmo de Gale-Shapley, cada agente conoce a priori sus preferencias con respecto a los demás agentes para aceptar o rechazar la asociación. El modelo supone que los agentes desconocen su entorno y les asignan igual valor de confiabilidad y capacidad a sus potenciales socios. Por ello, luego de cada iteración, los agentes irán aprendiendo a conocer su entorno y definiendo de este modo sus preferencias. Para ello, se emplea el Q-Learning que se observa en la Figura 3.

```
Inicializar  $Q(a_i, t_a, p)$  arbitrariamente
Repetir  $N$  episodios
  Elegir  $a_i, t_a, p$  utilizando la política  $\epsilon$ -greedy
  Aplicar la acción  $a_i, t_a, p$ , observar  $r, B$  para la acción seleccionada
  Actualizar  $Q_{t+1}(a_i, t_a, p) = Q_t(a_i, t_a, p) + \alpha[r_{t+1} - Q_t(a_i, t_a, p) + B_{t+1}(a_i, t_a, p)]$ 
FinRepetir
```

Figura 3. Algoritmo Q-learning para implementar el aprendizaje

Cada gestor de fines i tiene una lista de valores Q_{ij}^f , donde j es un gestor de medios. En cada episodio, el gestor de fines envía un requerimiento por recursos a un gestor de medios, según una política ϵ -greedy [Sutton y Barto, 1998] de selección de proveedor. Es decir que el agente puede tomar dos alternativas posibles: explorar, con una probabilidad asociada ϵ o explotar, con una probabilidad de $1-\epsilon$. Si decide explorar, entonces elegirá un gestor de medios al azar. En el caso que decida explotar, elegirá aquel que tenga el máximo valor de Q . Finalmente, el agente le envía una solicitud de asociación al gestor de medios seleccionado.

Del mismo modo, cada gestor de medios i tiene una lista de valores Q_{ij}^m , donde j representa un gestor de fines. En cada iteración, la decisión del gestor de medios está limitada a la elección entre aquellos gestores de fines que le hayan enviado requerimiento por recursos. Cuando un gestor de fines i elige a un gestor de medios j , pueden darse varias situaciones:

1. El gestor de medios no está asociado con otro gestor de fines.
2. El gestor de medios está asociado a otro gestor de fines k tal que $Q_{ji}^m > Q_{jk}^m$, es decir que el valor de Q para el que le propone supera al de su socio actual.
3. El gestor de medios está asociado a otro gestor de fines k tal que $Q_{ji}^m \leq Q_{jk}^m$, es decir que el valor de Q del que propone no supera al de su socio actual.
4. El gestor ya estaba asociado con i .

A estas alternativas se le suma la política de ϵ -greedy, con la cual el agente puede explorar o explotar, aplicando también el algoritmo de la figura 3. Finalmente, la decisión se toma de la siguiente manera:

1. Si el agente está solo, acepta.
2. Si no, aplica ϵ -greedy:
3. Si decide explorar, acepta.
4. Si decide explotar, acepta si la nueva propuesta tiene mejor valor de Q que su socio actual (caso 2) o si ya estaba asociado con ese mismo gestor (caso 4), y rechaza si la nueva propuesta tiene valor de Q inferior que su actual socio (caso 3).

Al comenzar la simulación todos los valores de Q se inicializan en 0. Al final de cada episodio los gestores actualizan sus variables internas. El gestor de fines actualiza su valor de Q correspondiente al gestor de medios que eligió, según la ecuación (1)

$$Q_{t+1} = Q_t + \alpha[r_{t+1} - Q_t + B_{t+1}] \quad (1)$$

donde r_{t+1} es la recompensa obtenida de dicha asociación y B_{t+1} es el factor de credibilidad que tiene el gestor de medios con quien estableció la relación cliente-servidor. Este factor evalúa cuán bien el servidor se desempeñó.

El factor de credibilidad se establece inicialmente en 100 para todos los gestores servidores, dado que los considera a todos igualmente confiables y capaces de satisfacer sus requerimientos. Luego, de cada episodio se calcula el valor del mismo de acuerdo con la ecuación (2).

$$B_{t+1} = \begin{cases} B_t, & \text{proveedor capaz} \\ B_t - 25, & \text{proveedor incapaz} \end{cases} \quad (2)$$

De esta manera se busca que los gestores aprendan a elegir entre aquellos que son capaces de proveer los recursos requeridos y a descartar aquellos que no lo son, penalizándolo al asignarles un valor de Q más bajo.

Por otro lado, cada gestor de medios actualiza el valor de Q del gestor de fines que lo eligió y con quien estableció la relación cliente-servidor, según las ecuaciones (3), (4) y (5). En la ecuación (4) B representa la recompensa obtenida y P es una penalidad que se define en la ecuación (5). Esta penalidad permite que el gestor de medios aprenda a aceptar aquellos contratos para el cual es capaz de proveer los recursos.

$$Q_{t+1} = Q_t + \alpha[r_{t+1} - Q_t] \quad (3)$$

$$r_t = B_t - P_t \quad (4)$$

$$P_{t+1} = \begin{cases} P_t, & \text{si es capaz} \\ P_t + 25, & \text{si no es capaz} \end{cases} \quad (5)$$

3.1.2. Resultados

Se realizaron simulaciones con diferentes escenarios, para obtener información acerca del comportamiento de los agentes en diferentes situaciones y comprobar la validez del modelo propuesto.

Las simulaciones realizadas fueron de 1000 iteraciones, utilizando un valor de ϵ de 0,1 y α de 0,1. Los valores de recompensas son obtenidos de una distribución normal con una media de 60 y desviación estándar 20.

Escenario 1: igual cantidad de gestores de medios y de fines.

Se realizaron corridas con cinco gestores de cada tipo. En el primer caso se observó que el gestor de fines 1 obtiene su mayor recompensa al elegir y ser aceptado por el gestor de medios 4, para establecer relaciones cliente-servidor, por lo que con el transcurso de los episodios aprende a seleccionar a dicho gestor como proveedor confiable (Figura 4).

Por otro lado, la curva de aprendizaje del gestor de medios 1 muestra un comportamiento similar, aprende a aceptar al gestor de fines 2, ya que es el que le da mayor recompensa (Figura 5) al ser elegido por este gestor de fines y aceptar prestarle sus recursos. La recompensa promedio obtenida por ambos gestores tiende a aumentar con el tiempo y a largo plazo se estabiliza, luego que han aprendido sobre sus preferencias en relación a sus potenciales socios para una dada tarea y recurso.

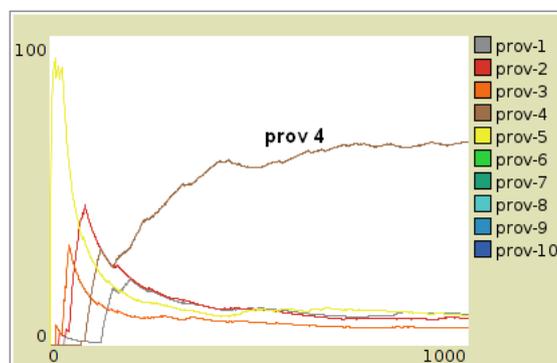


Figura 4. Curva de aprendizaje del gestor de fines 1.

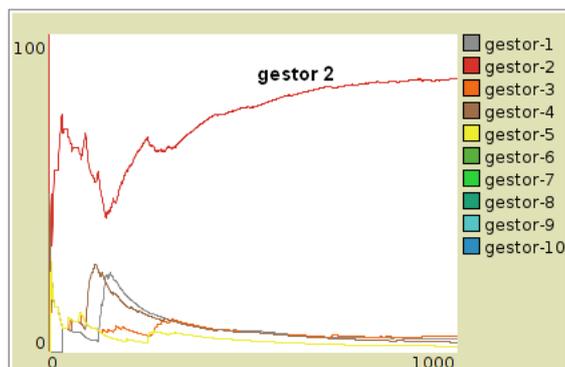


Figura 5. Curva de aprendizaje del gestor de medios 1.

Por otro lado se observe que el único gestor capaz de proveer al gestor de fines 1, era el gestor de medios 2. Sin embargo, éste último obtiene sus mejores recompensas con otros gestores de fines que le envían requerimientos por recursos, por lo que no elige al gestor de fines 1 como socio en ninguna iteración. Así, la curva de aprendizaje del gestor de fines 1 (Figura 6) muestra que éste no desarrolla una preferencia por ningún gestor de medios en particular, porque ninguno ellos han aceptado proveerle sus recursos.

En todos los casos la recompensa promedio de los gestores de cada tipo tiene una curva similar, aumenta las primeras iteraciones y tiende a estabilizarse en valores similares, como muestra la Figura 7. Esto se debe a que al principio, los gestores no tienen información de los demás, con lo cual deben explorar para ir conociéndolos y las recompensas obtenidas en promedio son bajas. A medida que transcurre el tiempo, aprenden a seleccionar el más conveniente, por lo tanto las recompensas se elevan y luego se estabilizan, ya que los agentes comienzan a explotar el conocimiento que ya tienen.

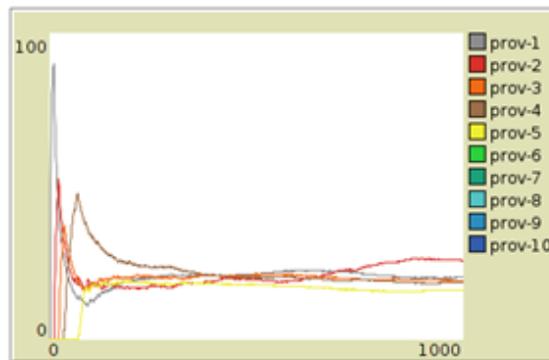


Figura 6. Curva de aprendizaje para el gestor de fines 1.

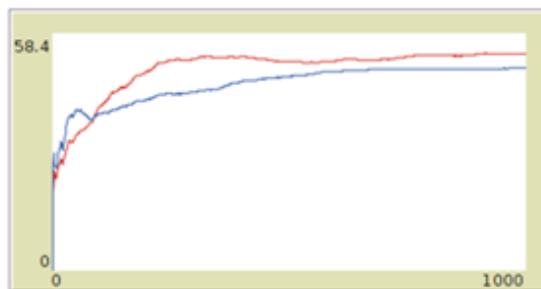


Figura 7. Recompensa promedio de cada tipo de gestor.

Escenario 2: más gestores de medios que gestores de fines

En este caso, se simuló con cinco gestores de fines y diez gestores de medios. Las diferentes simulaciones muestran que los gestores de fines en general no tienen una clara preferencia por un gestor de medios en particular, como se puede apreciar en la figura 8, dado que cuentan con un vasto grupo de potenciales proveedores de recursos entre los cuales seleccionar su mejor socio.

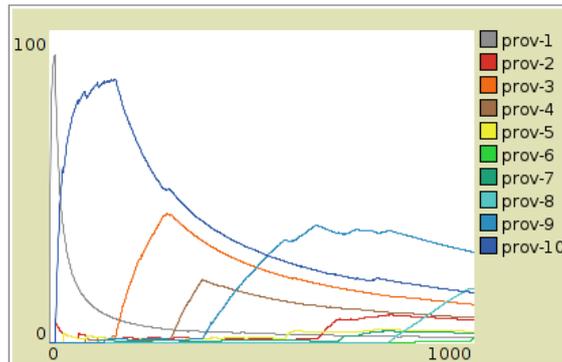


Figura 8. Curva de aprendizaje del gestor de medios 1.

En la Figura 9 se observa que la recompensa promedio obtenida por cada tipo de gestor es más alta en el caso de los gestores de fines. Esto se debe a que al haber más gestores de medios, los gestores de fines tienen mayores posibilidades de elecciones entre potenciales proveedores, por lo tanto es probable que un gestor de medios quede sin poder asociarse o se vea obligado a aceptar contratos con una baja recompensa. La elección de socio para los gestores de medios, está fuertemente ligada a la elección o no por parte de un gestor de fines del envío de requerimientos por recursos. Al ser mayor la oferta de servidores, éstos deben competir por lograr mayor cantidad de contratos sin que supere su capacidad y no disminuya su credibilidad.

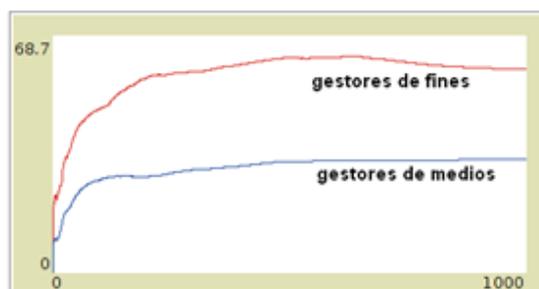


Figura 9. Recompensa promedio cuando hay más gestores de medios que de fines.

Escenario 3: más gestores de fines que gestores de medios

En este caso, se simuló con diez gestores de fines y cinco gestores de medios. Los gestores de medios en algunas ocasiones tenían preferencia por algún gestor determinado, pero sin hacer una diferencia marcada con el resto, nuevamente este comportamiento se atribuye a que por ser un número menor de servidores que de clientes, éstos últimos se ven obligados a aceptar contratos por recursos con gestores con bajo nivel de credibilidad con el fin de alcanzar los objetivos de sus respectivos proyectos. Esto se evidencia también en la Figura 10, donde se observa que los gestores de medios obtuvieron una mayor recompensa que los gestores de fines. Bajo estas condiciones, para los gestores de fines aumenta la competencia por el logro de contrato con los gestores de medios más confiables.



Figura 10. Recompensa promedio cuando hay más gestores de fines que de medios.

5. Conclusiones

El trabajo presenta brevemente el modelo de la compañía fractal basada en proyectos para la integración virtual y temporal entre diferentes compañías para el logro de un dado objetivo de negocios. Este modelo de empresa aplica a compañías donde permanentemente encaran proyectos de innovación con alta incertidumbre, costos y riesgos, como por ejemplo empresas del sector bio-ingenieril o farmacéutico.

La estructura organizacional, temporal y virtual, entre las empresas en el modelo de empresa integrada, se sustenta en el concepto de relaciones cliente-servidor, establecidas entre gestores de proyectos y a través de las cuales los mismos interactúan. La eficiencia y la eficacia de los resultados obtenidos en el logro de sus metas, dependerá de la estrategia de selección de socios que cada gestor de proyecto implemente. Por ello, se plantea la incorporación de aprendizaje en los agentes que se desempeñen como gestores de proyecto, para que los mismos descubran sus preferencias y su entorno para sí lograr coincidencias cliente-servidor provechosas para sus negocios.

La compañía fractal como se ha la propuesto es un mercado de coincidencias donde los gestores clientes requieren recurso para las tareas de sus proyectos y los servidores de recursos seleccionan el conjunto de clientes a quienes desean proveerles sus recursos.

Las simulaciones realizadas sobre la incorporación de aprendizaje en los agentes de gestión para establecer relaciones cliente-servidor, mostró cómo los clientes aprenden a requerir recursos sólo a proveedores confiables, es decir que poseen la capacidad de proveer de recursos con la calidad requerida por la tarea. Al mismo tiempo, los agentes que se desempeñan como gestores de recursos aprenden a identificar a qué tareas son capaces de proveer un dado recurso para maximizar sus beneficios y credibilidad ante sus clientes. De esta manera, aquellos proveedores a quienes no se les adjudican contratos de provision de recursos, deberán revisar su situación y puntos débiles para lograr que su permanencia en la compañía fractal sea beneficiosa tanto individualmente como colectivamente. Además, se observa de los escenarios planteados que cuanto más equilibrada es la cantidad entre gestores de fines y gestores de medios, se logra mayor competencia y con ello aprendizaje de los agentes intervinientes.

Si bien este trabajo en particular se basa en un modelo sencillo de la compañía fractal basada en proyectos, esta simplificación nos provee un útil punto de partida para estudiar la incidencia del aprendizaje en los agentes gestores de proyecto en mercados

bilaterales y cómo los mismos logran conocer sus preferencias y entorno para lograr coincidencias entre gestores clientes y servidores. La actual utilización de variables de recompensa y capacidad podrían extenderse para abarcar aspectos como beneficio económico de las interacciones, restricciones de tiempo de entregas, restricciones de calidad, etc. Además, como trabajo futuro se diseñará una arquitectura cognitiva para implementar los agentes gestores de proyecto de la compañía fractal.

Referencias

- Canavesio, M, Martinez, E (2007) Enterprise modeling of a Project-oriented fractal company for SMEs networking. *Computer in Industry* Nro. 58 Pp 794-813.
- Canavesio, M, Martinez, E (2014) Formal specification and modeling of a project-oriented fractal company using situation calculus. *Revista Electronica Argentina-Brasil de Tecnología de la Información y las Comunicaciones*. Vol 1. Nro 1.
- Chen,J., Song, K., (2013) Two-sided matching in the loan market. *International Journal of Industrial Organization* Nro. 33 Pp. 145-152.
- Das, S., Kamenica, E., (2005) Two-sided bandits and the dating market. *Proceeding 19th Intenational Joint Conference on Artificial Intelligence IJCAI'05*. Pp. 947-952.
- Das, S., (2006). *Dealers, Insiders, and Bandits: learning and its effects on market outcomes*. PhD Thesis. Massachusetts Institute of Technology.
- Derhami, V, Saadatjoo, M, Saadatjoo, F. (2010) *A reinforcement learning approach for price offer in supplier selection process*. Srpinger-Verlog. Pp. 326-329
- ElMaraghy, H. (2009) *The cognitive factory. In changeable and reconfigurable manufacturing systems* Pp. 1015–1024. Springer. London
- Hu, J.; Wellman, M. P.(2001). Learning about other agents in a dynamic multiagent system. *Journal of Cognitive Systems Research*. Vol. 2, Pp. 67-79.
- Huang, B.; Gao, Ch.; Chen, L.(2011) Partner selection in a virtual enterprise under unceatain information about condidates. *Expert Systems with Applications*. Nro. 38, Pp. 11305-11310.
- Kim, T, Bilsel, R., Kuman, S. (2008) Supplier selection in dynamic competitive environments. *International journal of services operations and iformatics* Vol. 3 Nro. 3-4 Pp. 283-293.
- Kumar,R., Lifshits,Y., Tomkins,A. (2010) Evolution of two-sided markets. *ACM Proceeding of the third ACM international Conference on web search and dataming*. ISBN 978-1-60558889-6 Pp. 311-320.
- Pisano, G. (2006). *Science Business: the promise, the reality, and the future of Biotech*. Harvard Business School Press.
- Ren, Z.; Anumba,C.J.(2002). Learning in multi-agent systems: a case study of construction claims negotiation. *Advanced Engineering Informatics*. Vol. 16, Pp. 265-275.

- Rochet, J, Tirole, J. (2006). Two-sided markets: a progress report. *The RAND Journal of Economics* Vol. 37 (3) Pp. 645-667.
- Rochet,J, Tirole, J. (2008) Tying in two-sided markets and the honor all cards rule. *International journal of industrial organization* Vol. 26 Nro. 6 Pp. 1333-1347.
- Russell, A.; Norving,P.(2004). *Inteligencia artificial. Un enfoque moderno*. Pearson Educacion SA.
- Sarne,D., Kraus, S, (2008) Managing parallel inquiries in agents` two-sided search. *Artificial intelligent* Vol. 172 (4-5) Pp 541-569
- Shah, N. (2004). Pharmaceutical supplier chains: Key issues and strategies for optimization. *Computers and Chemical Engineering*. Vol. 28, Pp. 929-941.
- Sotomayor, M. (2004) Implementation in the many-to-many matching market. *Games and economic behaviour*. Nro. 46, Pp. 199-212.
- Sutton, R. Barto,A (1998) *Reinforcement learning. An introduction*. MIT Press.
- Valluri, A.; Croson, D.C. (2003). Agent learning in supplier selection models. *Decision Support Systems*. Vol. 39. Pp. 219-240.
- Warnecke, H.J. (1993) *The fractal campany. A revolution incorporate culture*. Springer-Varlag. Berlin.
- Zeydan, M.; Colpan,C.; Cobanoglu,C.(2011) A combined methodology for supplier selection and performance evaluation. *Expert Systems with Applications*. Nro. 38, Pp. 2741-2751.