

# Revisión de Metodologías para la Evaluación de Rendimiento en Sistemas Multiagente

1<sup>st</sup> Bryan Mendoza

Facultad de Ingeniería

Universidad de Cuenca

Cuenca, Ecuador

bsteven.mendoza7@ucuenca.edu.ec

**Abstract**—La evaluación del rendimiento en sistemas multiagente (MAS) es un área clave pero aún poco desarrollada. A pesar de su creciente uso en entornos complejos, existen pocos métodos estandarizados para medir su eficiencia. Este artículo analiza cuatro enfoques que abordan la evaluación desde distintas perspectivas, incluyendo modelos basados en teoría de grafos, estructuras GQM y lógica difusa. A partir del estudio de sus aplicaciones, se identifican sus ventajas, limitaciones y la necesidad de enfoques más integradores. Los resultados evidencian que una evaluación completa del rendimiento en MAS requiere considerar múltiples propiedades de forma simultánea.

**Keywords**—Multi-Agent Systems, MAS evaluation, Communication, Performance Evaluation, Measurement, Goal-Question-Metric, Fuzzy Model.

## I. INTRODUCCIÓN

Los sistemas multiagente (MAS) han emergido como una solución eficaz para abordar problemas complejos en entornos distribuidos, dinámicos y abiertos [1]. Estos sistemas están compuestos por múltiples agentes autónomos que interactúan entre sí y con su entorno, cooperando para alcanzar objetivos individuales o colectivos [2]. Gracias a su flexibilidad, escalabilidad y capacidad de adaptación, los MAS se han aplicado con éxito en dominios tan diversos como la robótica, la logística, los sistemas inteligentes de transporte, el comercio electrónico, la industria del software, la gestión del conocimiento, y la salud [3].

A medida que estas aplicaciones se vuelven más críticas y complejas, la necesidad de evaluar rigurosamente el rendimiento de los MAS se vuelve indispensable [4]. No basta con asegurar que un sistema funciona, es necesario saber qué tan bien lo hace, bajo qué condiciones, y cuáles de sus componentes o comportamientos contribuyen o limitan su desempeño. Evaluar el rendimiento permite no solo validar la eficiencia del sistema, sino también identificar cuellos de botella, oportunidades de mejora y comparaciones objetivas entre distintas arquitecturas o implementaciones [5].

Sin embargo, a pesar de la importancia de este tema, actualmente existe una carencia de metodologías estandarizadas para la evaluación de rendimiento en MAS [5]. La diversidad de arquitecturas, plataformas de desarrollo, objetivos funcionales y propiedades emergentes de los MAS hace que la medición de su desempeño sea un desafío. La mayoría de los

enfoques existentes se enfocan en aspectos parciales como la eficiencia computacional, el tráfico de mensajes o el tiempo de respuesta, dejando de lado dimensiones fundamentales como la inteligencia, la autonomía, la capacidad de aprendizaje o la adaptabilidad [5]. Esto ha llevado a que muchas evaluaciones sean poco representativas, subjetivas o difíciles de replicar.

Es por ello que, diversos investigadores han propuesto enfoques novedosos que buscan abordar el rendimiento de los MAS desde múltiples perspectivas [3]. Este artículo tiene como objetivo analizar y sintetizar cuatro propuestas metodológicas recientes, cada una con un enfoque distinto, pero complementario, sobre la medición del rendimiento en MAS. A través de modelos basados en teoría de grafos, estructuras GQM, lógica difusa y métricas organizacionales, estos trabajos exploran dimensiones como la comunicación, racionalidad, autonomía, reactividad, adaptabilidad, inteligencia, seguridad, escalabilidad, y otras propiedades clave que influyen en el comportamiento global de un sistema multiagente.

## II. PERFORMANCE EVALUATION OF MULTIAGENT SYSTEMS: COMMUNICATION CRITERION

El artículo [6] aborda un problema fundamental en el campo de los MAS, la ausencia de una metodología estándar para evaluar su rendimiento. A pesar del crecimiento del interés en los MAS y su aplicación en áreas como redes, automatización, diagnóstico industrial, tráfico aéreo o robótica, no existe una herramienta de medición que permita evaluar su eficiencia ni compararlos objetivamente. Frente a esta necesidad, los autores proponen un modelo de evaluación centrado específicamente en uno de los atributos clave de los MAS, la comunicación.

El objetivo principal es establecer una metodología que permita analizar y medir características organizacionales de un MAS, comenzando por la comunicación. Esta característica es esencial, ya que constituye el soporte de la interacción entre agentes y es fundamental para el funcionamiento colectivo del sistema.

### A. Metodología de Evaluación

Para desarrollar el modelo, los autores adaptan la conocida metodología de evaluación de rendimiento de Jain, reorganizando los pasos en dos fases diferenciadas: una fase de

elección y una fase de implementación. En la fase de elección se definen los objetivos, se identifican las características del sistema, se selecciona la característica a evaluar (en este caso, la comunicación) y se elige la técnica de evaluación.

Por otro lado, la fase de implementación contempla el levantamiento de servicios del sistema, la fijación de la carga de trabajo, la ejecución de la evaluación, el análisis de los resultados y la presentación de conclusiones al usuario.

### *B. Sistema de Evaluación de Comunicación*

El sistema de evaluación propuesto se basa en la teoría de grafos y está compuesto por dos módulos principales.

- **Módulo de observación:** Este módulo actúa como un espía del sistema, recolectando información sin intervenir directamente en su funcionamiento, lo cual evita distorsiones en el rendimiento. Esta observación se realiza de forma pasiva mediante el uso del agente Sniffer de la plataforma JADE, que almacena los mensajes en un archivo de trazas.
- **Módulo de medición:** Este módulo analiza la comunicación desde tres perspectivas: estructural, sintáctica y estadística. Recibiendo mayor atención al aspecto estructural debido a la naturaleza del modelo. El sistema está modelado por un grafo dirigido donde los nodos son agentes y los arcos indican enlaces de comunicación ponderados por el número de mensajes intercambiados. Posteriormente, analizan sus propiedades como el grado de participación del agente y el grado de solicitud del mismo en el acto de comunicación. También se usaron índices útiles para describir la estructura del grafo como la complejidad de la red de comunicación, conectividad, carga de la red y puntos de articulación.

En cuanto al aspecto sintáctico, se evalúan dos elementos: la tipología de los mensajes, mediante el análisis de los performativos utilizados y la complejidad del contenido de cada mensaje, clasificándolo como simple, medio o complejo.

### *C. Experimentación*

Para validar el sistema de evaluación, se aplicó en una aplicación MAS diseñada para detectar y localizar fallos en un sistema industrial. Esta aplicación fue implementada en JADE e incluye cinco agentes de detección (D1 a D5), un agente de localización (LOC) y un agente de interfaz (INT). Los mensajes intercambiados fueron capturados por el Sniffer, procesados por un analizador ACL y usados para construir el grafo de comunicación y calcular las métricas.

Los resultados mostraron que la red de comunicación es poco compleja y tiene un grado de comunicación reducido. Sin embargo, todos los agentes participan en la comunicación. El análisis también evidenció una distribución desigual de la carga comunicativa, siendo los agentes LOC e INT los más activos. La tipología de los mensajes fue limitada, utilizando únicamente el acto comunicativo INFORM, lo cual concuerda con la naturaleza de la aplicación. Además, todos los mensajes eran de complejidad media.

En resumen, el artículo presenta una herramienta eficaz para evaluar la comunicación en MAS, que permite analizar desde una perspectiva organizacional sin interferir con el comportamiento del sistema. Los autores señalan como trabajo futuro el perfeccionamiento del módulo de observación y la extensión de la evaluación a otras características relevantes de los MAS.

## III. AN EVALUATION METHOD FOR MULTI-AGENT SYSTEMS

El artículo [7] propone un método sistemático para evaluar MAS, abordando la falta de métricas adecuadas que consideren propiedades esenciales como la racionalidad, autonomía, reactividad y capacidad de adaptación al entorno. La mayoría de trabajos previos se han centrado en métricas individuales de agentes o en aspectos estructurales como comunicación o arquitectura, los autores sostienen que la evaluación debe considerar también la complejidad del entorno y el comportamiento colectivo del sistema.

### *A. Definición del plan de métricas*

El método propuesto está diseñado para MAS estáticos, es decir con agentes fijos durante la ejecución y se fundamenta en el enfoque Goal-Question-Metric (GQM), el cual estructura el análisis en función de metas, preguntas específicas y métricas cuantificables. Además, la propuesta combina dos perspectivas: intra-agente, la cual hace referencia a la evaluación interna de cada agente e inter-agente que hace referencia a la interacción entre agentes y con el entorno. El plan de evaluación define cinco metas principales:

- 1) **Complejidad del entorno:** Se evalúa mediante la inobservabilidad, inaccesibilidad de recursos, inestabilidad, grado de competencia entre agentes y uso de modelos de confianza/reputación. La combinación de estas variables se traduce en un puntaje total que clasifica el entorno como de baja, media o alta complejidad.
- 2) **Racionalidad:** Se mide en función de la capacidad del agente para seleccionar acciones que maximicen su éxito, considerando factores como tipo de agente, habilidad para planificar, aprender, y construir modelos internos del entorno y de otros agentes. Se calcula un valor numérico que representa el grado de racionalidad tanto individual como del sistema completo.
- 3) **Autonomía:** Se evalúa a través de la proactividad (capacidad de actuar sin intervención externa), diagnóstico de errores, aprendizaje, y papel organizacional dentro del MAS.
- 4) **Reactividad:** Se refiere a la habilidad del agente para percibir y responder a cambios del entorno. Incluye la eficiencia perceptiva, la velocidad para alcanzar objetivos y minimizar comunicación innecesaria durante la ejecución.
- 5) **Adaptabilidad al entorno:** Evalúa la capacidad de cada agente para ajustar su comportamiento frente a estímulos externos y gestionar diferentes situaciones. Se consideran aspectos como habilidad para elegir acciones en

función de objetivos, asumir múltiples roles, compartir tareas, aprender, encontrar heurísticas útiles y manejar excepciones.

Para cada meta, las métricas definidas permiten generar valores normalizados que luego se interpretan de acuerdo con la complejidad del entorno. Así, se determina si las capacidades internas del MAS son adecuadas o no para operar eficazmente en dicho contexto.

#### *B. Aplicación del método de evaluación*

El método fue aplicado al sistema GeCo\_Automotive, una plataforma de gestión del conocimiento en empresas del sector automotriz. El MAS evaluado está compuesto por dos agentes estáticos, uno clasificador de recursos y otro de búsqueda. Ambos colaboran para sugerir recursos formativos apropiados para cada usuario. Se evaluó el entorno como de baja complejidad, con acceso parcial al conocimiento de los usuarios pero alta disponibilidad de recursos y se obtuvieron los siguientes resultados:

Baja racionalidad: los agentes tienen comportamientos reactivos simples, sin planificación ni aprendizaje.

Baja autonomía: los agentes no diagnostican errores, no negocian, ni realizan comunicaciones autónomas.

Alta reactividad: los agentes ejecutan acciones rápidamente sin procesamiento interno, adecuado para entornos simples.

Baja adaptabilidad: los agentes solo gestionan excepciones, pero no aprenden ni modifican su comportamiento ante nuevas situaciones.

Los resultados obtenidos indican que, si bien el MAS es adecuado para su entorno poco complejo, existen debilidades en su autonomía y adaptabilidad que podrían afectar su funcionamiento si el contexto se vuelve más dinámico o incierto.

En resumen, el artículo presenta un enfoque novedoso y estructurado para evaluar MAS que va más allá de las métricas tradicionales. Su principal aporte radica en el uso de métricas de alto nivel que permiten valorar aspectos clave de la inteligencia de los agentes, así como su adecuación al entorno. Además, el enfoque GQM permite adaptar la evaluación a distintos contextos y aplicaciones.

#### **IV. METHODOLOGY FOR PERFORMANCE EVALUATION OF DISTRIBUTED MULTI AGENT SYSTEM**

El artículo [8] propone una metodología para evaluar el rendimiento de sistemas multiagente distribuidos (DMAS), abordando la falta de enfoques integrales que consideren todas sus características genéricas. Tradicionalmente, los métodos de evaluación se han centrado en propiedades organizacionales o en aspectos limitados como el tiempo de respuesta, la tasa de mensajes o el modelado de la comunicación. Sin embargo, estas aproximaciones no capturan adecuadamente características fundamentales de los MAS como la autonomía, el aprendizaje, la flexibilidad o la capacidad de adaptación, elementos esenciales que influyen directamente en la eficiencia y comportamiento general del sistema.

#### *A. Modelo de evaluación del desempeño MAS: Modelo MPE*

Los autores, proponen el modelo MPE (Multi-Agent Performance Evaluation), que combina tres técnicas fundamentales: modelado analítico, medición directa y simulación. Este modelo se estructura en tres fases: observación, medición y verificación de conformidad. En la fase de observación, se recolectan datos en tiempo real sobre la actividad de los agentes, con el fin de identificar las propiedades más relevantes que afectan su desempeño. Estas propiedades genéricas incluyen la comunicación, la autonomía, el aprendizaje y la flexibilidad. En la fase de medición, se aplican funciones de mapeo extraídas de la teoría de grafos para traducir estas propiedades en métricas cuantificables. Por último, en la etapa de verificación, se comparan los resultados con escenarios de referencia (mejor y peor caso) para validar el modelo.

Las métricas se derivan de funciones topológicas propias de grafos, y permiten representar con precisión varios aspectos del comportamiento del sistema. Por ejemplo, para evaluar la comunicación se utilizan métricas como el número de conexiones, la carga de red, la complejidad de la red y la conectividad. Para la autonomía, se mide la independencia social de los agentes mediante el análisis del número de conexiones entrantes y salientes. El aprendizaje se refleja a través de la existencia de bucles en el grafo, lo que indica la capacidad de los agentes para memorizar o autoevaluarse. La flexibilidad se estima utilizando el índice de alcanzabilidad y el coeficiente de agrupamiento, que indican la persistencia del agente y su capacidad para integrarse en comunidades dentro del sistema.

#### *B. Aplicación y Resultados*

Estas métricas se aplicaron a dos sistemas reales desarrollados con la plataforma JADE: un Sistema de Gestión de Cursos (CMS) y un Sistema de Monitoreo Remoto de Pacientes (RPMS). En el CMS, participaron agentes de profesores, estudiantes e intermediarios. Donde se observó una alta autonomía en los agentes intermediarios y una red compleja con múltiples conexiones, lo que sugiere una organización bien distribuida y automatizada. También se identificaron nodos con bucles, indicando una capacidad de aprendizaje sólida. Por otro lado, el RPMS involucraba agentes médicos, pacientes y de monitoreo. Aunque mostró una alta reactividad con agentes capaces de responder rápidamente a eventos clínicos, su nivel de autonomía y aprendizaje era menor. La complejidad de red también resultó significativamente más baja que en el CMS.

Los resultados indican que los sistemas altamente reactivos como RPMS tienden a depender más de recursos externos y requieren mayor supervisión, lo que reduce su eficiencia general. En cambio, los sistemas más automatizados, como CMS, muestran un mejor balance entre autonomía, flexibilidad y rendimiento, siendo menos dependientes de la intervención humana y más sostenibles en escenarios complejos.

El artículo concluye que el rendimiento de un MAS no puede ser evaluado únicamente desde una perspectiva estructural o basada en tasas de comunicación. Es fundamental

considerar la naturaleza organizativa del sistema, su capacidad de adaptación y el grado de autonomía de sus agentes. El modelo MPE ofrece una visión integral que permite cuantificar estos factores de manera coherente y aplicable en contextos reales.

## V. PERFORMANCE EVALUATION OF A MULTI-AGENT SYSTEM USING FUZZY MODEL

El artículo [9] propone un modelo general para evaluar el rendimiento de MAS utilizando la lógica difusa en combinación con el modelo Goal/Question/Metrics (GQM). La evaluación del rendimiento se basa en tres criterios fundamentales: inteligencia, seguridad y escalabilidad. Cada uno de estos criterios se define mediante un modelo GQM, en el que se establece un nivel de objetivos, que se descomponen en preguntas y, finalmente, en métricas cuantificables. Los valores medidos pueden ser numéricos simples o resultados de un proceso de cuantificación difusa, permitiendo así obtener un porcentaje de rendimiento para cada criterio y, por ende, un rendimiento global del sistema.

El modelo GQM utilizado permite estructurar la evaluación, ya que cada objetivo se asocia a varias preguntas, y cada pregunta a varias métricas. Por ejemplo, en el caso del criterio de inteligencia se consideran cuatro objetivos: adaptación, racionalidad, autonomía y reactividad. Para este criterio se definen diecisiete métricas específicas, y el valor final de la inteligencia de un agente se obtiene mediante la integración, a través de controladores de lógica difusa, de la información proveniente de cada uno de los objetivos. De forma similar, el criterio de seguridad se fundamenta en conceptos como la confidencialidad, autenticación, integridad, control de acceso, no repudio y disponibilidad. Cada uno de estos aspectos se evalúa mediante métricas que toman valores binarios según se cumpla o no el requisito evaluado. Asimismo, el criterio de escalabilidad se estructura considerando dos objetivos: uno relacionado con la capacidad de adaptación y otro que valora específicamente la capacidad de escalado (scale-up/scale-down) de los agentes. Este último se mide a través de diversas preguntas que exploran si los agentes, sus conocimientos, tipos, tareas o acciones pueden escalar de manera eficaz sin requerir recursos adicionales.

### A. Metodología de Evaluación

La propuesta metodológica es implementada en un sistema denominado PEMASFM (Performance Evaluation of Multi-Agent System Fuzzy Model). La arquitectura del PEMASFM integra una base de datos que almacena todo el modelo GQM de los criterios de rendimiento, así como los tipos de funciones a utilizar para la agregación de las métricas, que pueden ser de lógica difusa, de promedio aritmético o funciones identidad. El usuario ingresa los valores medidos para cada métrica, y el sistema, apoyado en herramientas como MATLAB, Microsoft SQL Server y Microsoft Visual Studio, procesa estos datos para calcular, mediante controladores de lógica difusa, los valores de cada nivel jerárquico, desde las preguntas hasta los objetivos y, finalmente, cada criterio de evaluación. Los

resultados se muestran en forma de porcentajes y gráficos de barras, facilitando así la visualización del rendimiento de cada agente y del sistema en conjunto.

En cuanto a la implementación, el artículo detalla varios controladores difusos específicos. Se presenta un controlador general de evaluación que integra los resultados de los criterios de inteligencia, seguridad y escalabilidad para emitir un resultado global de rendimiento. Los controladores específicos, por ejemplo, para el criterio de inteligencia, cuentan con conjuntos de reglas difusas que permiten evaluar cada uno de los objetivos y combinarlos para obtener una valoración final. De forma análoga, se desarrollan controladores para los criterios de seguridad que evalúan aspectos esenciales como la confidencialidad, autenticación, integridad y control de acceso repartidos en dos grandes objetivos y para la escalabilidad, que analiza la capacidad de los agentes para adaptarse y escalar tanto en entornos individuales como en múltiples plataformas.

### B. Aplicación del método de evaluación

El modelo se valida mediante un estudio de caso aplicado al "Code Advisory System", una aplicación MAS destinada a mejorar la eficiencia de programadores en C#. Este sistema integra agentes con roles dinámicos, tales como un agente de interfaz que gestiona la comunicación del usuario, un agente "How-To-Do" que suministra fragmentos de código, y un agente "What-Is-Wrong" encargado de detectar errores en el código. La aplicación del PEMASFM a este sistema produjo resultados cuantitativos: el rendimiento general fue del 75%, con el criterio de inteligencia alcanzando un 81%, el de escalabilidad un 83% y el de seguridad apenas un 12%. Estos valores muestran que, mientras las capacidades adaptativas y de escalabilidad son elevadas, la seguridad del sistema es deficiente.

Se destaca que el PEMASFM ofrece un marco robusto y versátil para la evaluación integral del rendimiento en MAS. El modelo no solo permite cuantificar la performance de los agentes a través de métricas estructuradas y reglas difusas, sino que también ofrece perspectivas claras para identificar áreas de mejora, como se evidenció en el caso de la baja seguridad en el estudio de caso. Este modelo puede convertirse en un estándar para la medición de rendimiento en MAS, dado que su estructura dinámica permite incorporar nuevos criterios y adaptarse a diversas configuraciones y requerimientos operativos.

Este enfoque integral y riguroso brinda una herramienta útil tanto para desarrolladores como para gestores y propietarios de MAS, facilitando la toma de decisiones basada en datos y ofreciendo un camino para la mejora continua del rendimiento en entornos complejos.

## VI. CONCLUSIÓN

La evaluación del rendimiento en MAS representa un desafío crucial y aún insuficientemente resuelto dentro del campo de los sistemas inteligentes distribuidos. A través del análisis

de los cuatro enfoques revisados en este artículo, se evidencia que existen múltiples formas de abordar esta problemática, cada una enfocada en diferentes dimensiones del rendimiento, desde la evaluación de la comunicación entre agentes mediante teoría de grafos, hasta modelos estructurados como GQM, técnicas de lógica difusa y análisis de propiedades organizacionales.

Entre los modelos estudiados, aquellos basados en teoría de grafos permiten obtener una visión estructural precisa de la red de agentes y sus interacciones, mientras que las metodologías GQM y las reglas difusas ofrecen una capacidad mayor de adaptación a distintos contextos y niveles de abstracción. Asimismo, la aplicación práctica de estos modelos en plataformas reales como JADE y casos concretos como sistemas de diagnóstico, monitoreo remoto o asistencia a programadores, demuestra su aplicabilidad y utilidad para identificar tanto fortalezas como debilidades en sistemas MAS.

Uno de los aspectos encontrados tras esta revisión, es que no existe un modelo universal que capture integralmente todas las características que definen el rendimiento de un MAS. Por el contrario, los modelos actuales tienden a enfocarse en ciertos aspectos particulares como la eficiencia comunicativa, la inteligencia de los agentes, su autonomía, la seguridad o la escalabilidad. Lo que conlleva, la necesidad de enfoques híbridos o integradores, capaces de combinar varias métricas de distinta naturaleza.

En conjunto, los enfoques revisados aportan herramientas valiosas para avanzar hacia una evaluación más sistemática, objetiva y completa del rendimiento en MAS, lo que resulta fundamental en escenarios donde la confiabilidad, eficiencia y adaptabilidad del sistema son determinantes. No obstante, el campo aún requiere marcos metodológicos más robustos y generalizables, que integren múltiples criterios sin sacrificar simplicidad ni claridad operativa.

## REFERENCES

- [1] M. Wooldridge, *An Introduction to MultiAgent Systems*. John Wiley & Sons, 2009.
- [2] J. Ferber, *Multi-Agent Systems: An Introduction to Distributed Artificial Intelligence*. Addison-Wesley, 1999.
- [3] A. Omicini, "Sma: Sistemas multi-agentes," *Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial*, vol. 8, no. 21, pp. 11–27, 2004.
- [4] T. J. Bench-Capon, "Evaluating multi-agent systems," *AI Communications*, vol. 15, no. 4, pp. 245–247, 2002.
- [5] S. Gómez, L. Montoya, and J. A. Botía, "Performance evaluation in multi-agent systems: A survey," *The Knowledge Engineering Review*, vol. 28, no. 2, pp. 151–176, 2013.
- [6] F. Ben Hmida, W. Lejouad Chaari, and M. Tagina, "Performance evaluation of multiagent systems: communication criterion," in *Agent and Multi-Agent Systems: Technologies and Applications: Second KES International Symposium, KES-AMSTA 2008, Incheon, Korea, March 26-28, 2008. Proceedings 2*, pp. 773–782, Springer, 2008.
- [7] P. Di Bitonto, M. Laterza, T. Roselli, and V. Rossano, "An evaluation method for multi-agent systems," in *Agent and Multi-Agent Systems: Technologies and Applications: 4th KES International Symposium, KES-AMSTA 2010, Gdynia, Poland, June 23-25, 2010, Proceedings. Part I 4*, pp. 32–41, Springer, 2010.
- [8] A. Ali, M. Aslam, J. Janjua, and M. Chaudhry, "Methodology for performance evaluation of distributed multi agent system," *The Nucleus*, vol. 54, no. 2, pp. 75–82, 2017.

- [9] S. A. D. Aly and H. M. S. Badoor, "Performance evaluation of a multi-agent system using fuzzy model," in *2018 First International Workshop on Deep and Representation Learning (IWDRL)*, pp. 7–12, IEEE, 2018.