Instituto Tecnológico de Costa Rica

Escuela de Ingeniería Mecatrónica

Resumen del estado del arte sobre los sistemas motivacionales en arquitecturas cognitivas

Diana Cerdas Vargas

Coruña, 23 de julio de 2025

Sistemas motivacionales en arquitecturas cognitivas

En esta sección se incluye un resumen del estado del arte de los sistemas motivacionales en las arquitecturas cognitivas aplicadas en la robótica autónoma. En primer lugar se hace una breve introducción a la definición formal de arquitectura cognitiva, los tipos de arquitecturas cognitivas que existen. Finalmente se incluyen los ejemplos más destacados de arquitecturas cognitivas aplicados en la robótica autónoma y sus sistemas motivacionales.

0.1 Arquitecturas cognitivas en robótica

Para lograr alcanzar estos niveles más altos de autonomía en la robótica, se debe dar a los robots una capacidad de percepción activa, toma de decisiones, adaptación continua, y en los casos más avanzados, aprendizaje abierto y a lo largo del tiempo (*lifelong open-ended learning*)[1]. Es acá donde surgen las arquitecturas cognitivas. En esta sección se va a profundizar sobre las arquitecturas cognitivas en la robótica, algunos tipos de arquitecturas que existen y sus enfoques para el descubrimiento y aprendizaje de los objetivos.

0.1.1 ¿Qué es una arquitectura cognitiva?

De acuerdo con [2][3], una arquitectura cognitiva se considera un marco teórico y computacional que modela mecanismos de procesamiento cognitivo que permiten crear comportamientos inteligentes. Estos mecanismos, por lo general, están inspirados en el funcionamiento del cerebro humano. Tienen como objetivo integrar componentes como la percepción, memoria, razonamiento, aprendizaje, acción y motivación.

- **Percepción**: Es un proceso que transforma la información cruda que recibe el sistema del entorno por medio de sensores, en una representación interna del sistema para poder realizar tareas cognitivas [2]. La percepción es lo que le permite al robot generar su "visión del mundo", conocer qué objetos tiene a su alrededor, saber qué acciones puede ejecutar en cada momento, entre otras. Por lo general, las percepciones más comunes se relacionan a la visión y el tacto.

- **Memoria**: Esta es una parte fundamental en las arquitecturas cognitivas. Acá se incluye la memoria de corto y largo plazo. La memoria de corto plazo se relaciona con puntos intermedios que facilitan el proceso de aprendizaje. Mientras que la memoria de largo plazo hace referencia al conocimiento base y algunas acciones o problemas que el robot ya logró aprender por completo [2].

- **Toma de decisiones**: Se refiere a la capacidad de la arquitectura para seleccionar las acciones que debe realizar en cada momento. Por lo general se tienen dos tipos: los deliberativos y los reactivos. La reactividad son todas aquellas acciones que realiza el robot de manera inmediata, sin "pensar", similar a los reflejos humanos [2]. Por otro

lado, los sistemas deliberativos de toma decisiones son todos aquellos procesos cognitivos que pueden hacer que un robot seleccione una acción u otra. Cada arquitectura cognitiva pude tener un sistema deliberativo diferente, algunos ejemplos podrían ser: motivaciones, estados afectivos, emociones, estados de ánimo, impulsos, entre otros. Es importante destacar que en toda arquitectura cognitiva se debe encontrar un balance entre las acciones reactivas y deliberativas para garantizar que el robot logre su propósito y que no se quede atascado haciendo acciones reactivas o aprendiendo tareas que no son necesarias [3]

- **Aprendizaje**: Es la capacidad que tienen estos sistemas para mejorar su rendimiento con el tiempo. Al igual que en los seres humanos, el aprendizaje en las arquitecturas cognitivas se basa en la experiencia del robot. Existe una gran variedad de tipos de aprendizaje, dependiendo del enfoque que le den los diseñadores a la arquitectura cognitiva, en sus casos de aplicación, método de entrenamiento deseado y los algoritmos [2].

- **Metacognición**: De acuerdo con Flavell, citado por [2], la metacognición se puede ver como "pensar sobre el pensamiento". Son todas aquellas habilidades que permiten realizar una introspección sobre los procesos internos de razonamiento [2]. Se podría decir que es todo lo relacionado con la explicabilidad del sistema y la capacidad de identificar, explicar y corregir las decisiones que toma el sistema.

0.1.2 Tipos de arquitecturas cognitivas

Además de su aplicación en la robótica autónoma, las arquitecturas cognitivas se pueden aplicar a una gran variedad de áreas diferentes, algunas de las más comunes de acuerdo con [2] son: para experimentos psicológicos, modelado del desempeño humano en un entorno de tareas específico, interacción humano-robot y humano-computadora (HRI/HCI), procesamiento de lenguaje natural (NLP), análisis de grandes cantidades de datos, resolver problemas complejos de visión por computadora, juegos, entre otros.

Es por esta razón que existen diversos tipos de arquitecturas cognitivas, dependiendo de su objetivo final. En el caso de este trabajo, se va a limitar el estudio de arquitecturas cognitivas en el área de robótica autónoma. Dentro de esta área, se pueden encontrar tres tipos concretos: simbólicas, emergentes e híbridas.

Arquitecturas cognitivas simbólicas

También conocidas como arquitecturas cognitivistas [2]. Estas arquitecturas se destacan por representar conceptos por medio de símbolos que pueden ser manipulados por medio de un conjunto de reglas predefinidas. Además, este tipo de arquitecturas se caracterizan por representar de manera intuitiva el conocimiento, lo que las hace altamente usadas. Sin embargo, cabe mencionar que también tienen una menor capacidad de adaptarse a cambios en el dominio donde operan, lo que las hace menos flexibles y robustas [2].

Arquitecturas cognitivas emergentes

Se les conoce como arquitecturas conexionistas. Este tipo de arquitecturas se caracterizan por tener un aprendizaje altamente paralelo, incluso se podría decir que es similar al de las redes neuronales. Esto debido a que en ambos casos el flujo de la información se va propagando en diferentes nodos desde la entrada hasta la salida. Las arquitecturas emergentes resuelven las limitaciones de adaptabilidad y flexibilidad de las arquitecturas simbólicas, pero al ser estructuras más complejas y que ejecutan una serie de acciones en paralelo pierden cierto grado de explicabilidad [2]. Esto no significa que sean una caja negra, como el caso de las redes neuronales, pero resulta más complicado de comprender y explicar la forma en la que aprenden.

Arquitecturas cognitivas híbridas

Este tipo de arquitecturas trata de combinar lo mejor de los dos mundos: la robustez de las arquitecturas emergentes con la simplicidad y transparencia de las arquitecturas simbólicas. No existe ningún tipo de restricciones sobre cómo se debe hacer la combinación de ambos tipos, por lo que han surgido una gran variedad de combinaciones [2]. No obstante, no todas las arquitecturas híbridas desarrolladas analizan o describen detalladamente los elementos simbólicos que utilizan y por qué lo hacen.

0.1.3 Diferentes enfoques motivacionales para el descubrimiento y aprendizaje de objetivos

Cuando se habla de los enfoques motivacionales en el contexto de arquitecturas cognitivas, se hace referencia a los mecanismos intrínsecos que impulsan al robot para explorar su entorno, que le permita adquirir conocimiento y descubrir nuevos objetivo de manera autónoma. Por lo general estos mecanismos son basados en los procesos biológicos que se dan en el cerebro del ser humano y/o en los animales. Algunos ejemplos son: impulsos internos, la curiosidad, las emociones, estados de ánimo o los estados afectivos, entre otros [2].

Actualmente, existe una amplia cantidad de arquitecturas cognitivas que se han desarrollado a lo largo de los años para tratar de dotar autonomía a los robots. En esta sección se van a destacar algunas de las arquitecturas cognitivas desarrolladas y cuál es su enfoque motivacional para el descubrimiento y aprendizaje de objetivos.

Arquitectura DNC

La arquitectura DNC (*Dynamic Neural Curiosity*) está inspirada en el funcionamiento del cerebro humano, especialmente en cómo se regula la atención y la curiosidad. Esta arquitectura permite que un robot descubra y aprenda nuevos objetivos por sí mismo, sin necesidad de instrucciones externas. El robot alterna entre dos modos: uno de exploración, donde

busca cosas nuevas en su entorno, y otro de aprendizaje, donde intenta alcanzar los objetivos que ha descubierto. Para decidir cuándo cambiar entre estos modos, el sistema utiliza señales como la novedad de lo que percibe o qué tanto está mejorando al intentar una tarea. Los objetivos se descubren observando cambios en el entorno provocados por las propias acciones del robot, y se priorizan aquellos que representan un reto o que aún no han sido dominados. Gracias a este enfoque, el robot puede desarrollar habilidades cada vez más complejas de forma autónoma, guiado únicamente por su curiosidad y progreso interno [4].

Arquitectura IMGEP

La arquitectura IMGEP (*Intrinsically Motivated Goal Exploration Processes*) permite que un robot explore y aprenda por sí mismo a través de metas que él mismo genera. En lugar de recibir instrucciones externas, el robot crea sus propios objetivos y decide cuáles intentar alcanzar según qué tanto está aprendiendo. Esta decisión se basa en su progreso de aprendizaje: si una meta le resulta cada vez más fácil, entonces ha aprendido algo útil. Así, el robot se enfoca en las metas donde nota que está mejorando, y deja de lado aquellas donde ya no progresa. Las metas se definen como funciones que evalúan resultados, por ejemplo, mover un objeto a cierta posición o lograr cierto efecto en el entorno. IMGEP también organiza el aprendizaje de forma gradual, comenzando con tareas más simples y avanzando hacia otras más complejas. Gracias a este enfoque, el robot desarrolla nuevas habilidades de forma autónoma, motivado internamente por su propia mejora y curiosidad [5].

Arquitectura GRAIL

La arquitectura cognitiva GRAIL (*Goal-discovering Robotic Architecture for Intrinsically-motivated Learning*) utiliza un enfoque motivacional basado en la competencia. En este, el robot es capaz de descubrir y seleccionar sus propias metas a partir de sus señales sensoriales (percepciones), por lo que no depende de recompensas externas. En esta arquitectura se le da recompensa al robot en el progreso que hace al aprender a mejorar una habilidad, y esto impulsa al robot a aprender [6].

Para descubrir los objetivos, esta arquitectura hace que el robot explore su entorno y detecte cambios o efectos generados por sus acciones. Por ejemplo, tomar un objeto, mover un objeto, encender una luz, entre otros. Estos cambios luego se transforman en un nuevo objetivo. La arquitectura prioriza los objetivos en los que el robot tenga una mayor mejora en su desempeño, que lleva al final a que el robot ejecute primero los objetivos más simples y luego avanza a tareas más complejas [6].

Arquitectura H-GRAIL

La arquitectura H-GRAIL (*Hierarchical Goal-discovering Robotic Architecture for Intrinsically-motivated Learning*) es una versión extendida de GRAIL que permite a un robot aprender de manera más organizada y adaptarse a tareas más complejas. H-GRAIL funciona a partir de

dos tipos de motivación interna: una para descubrir nuevas metas (cuando el robot nota algo interesante o nuevo en el entorno), y otra para mejorar en alcanzar metas ya conocidas. El sistema decide en cada momento si explorar (buscar nuevas metas) o explotar (practicar una meta para mejorar su habilidad). Además, organiza las metas en pasos más pequeños llamados sub-metas, lo que permite al robot aprender tareas con varios pasos interdependientes. Esta arquitectura también está diseñada para adaptarse cuando cambian las condiciones del entorno, reorganizando sus planes y prioridades. Gracias a esta combinación de curiosidad, progreso y organización jerárquica, H-GRAIL permite un aprendizaje autónomo más robusto y flexible [7].

Arquitectura ASMO

Existen varias arquitecturas cognitivas que utilizan como motivaciones intrínsecas los impulsos, que provienen del término en inglés *drives*. Estos impulsos representan necesidades fisiológicas básicas como la alimentación, la seguridad, entre otros. Por lo general, estos impulsos se van "desvaneciendo" conforme se va satisfaciendo la necesidad. Tanto en el caso de los seres vivos como en los robots, a medida que estos trabajan para alcanzar los objetivos, pueden existir varios impulsos que estén afectando su comportamiento al mismo tiempo [2].

Un ejemplo de aplicación de este enfoque motivacional es la arquitectura ASMO. (*Attentive Self-Modifying Architecture*) está inspirada en cómo los humanos prestan atención y cambian su comportamiento según lo que les interesa o necesitan en un momento dado. En este modelo, el robot tiene varios procesos mentales que "compiten" por su atención, y aquellos que son más urgentes, relevantes o motivadores son los que se activan primero. ASMO incorpora un sistema motivacional simple, compuesto por tres impulsos o *drives*: seguir objetos de color rojo, recibir elogios y alcanzar un estado interno de bienestar o "felicidad". Estos impulsos afectan cómo se distribuye la atención del robot y qué acciones decide realizar. Además, ASMO es capaz de modificar su propio comportamiento con el tiempo, cambiando las prioridades entre procesos en función de sus experiencias. Aunque no genera metas de forma autónoma como otras arquitecturas, su diseño le permite adaptarse a diferentes situaciones guiado por sus impulsos internos y su atención dinámica [8].

Lista de necesidades identificadas

En esta sección se incluyen la lista de necesidades identificadas para poder desarrollar este proyecto. Estos son todos aquellos criterios que se consideraron de vital importancia para el cliente, el Grupo Integrado de Ingeniería (GII).

1. El sistema permite que el usuario exprese sus propósitos en lenguaje de alto nivel y es capaz de interpretarlo.
2. El sistema convierte los propósitos del humano en misiones válidas.
3. El sistema genera funciones matemáticas de impulso (funciones de drive) que permiten evaluar el nivel de satisfacción de una misión.
4. El sistema se incorpora con el código de la arquitectura cognitiva existente en Python y ROS2 y esta se comunica con el robot.
5. El sistema realiza acciones de alineación para asegurar que el propósito del humano esté completamente definido.
6. El sistema prioriza las misiones de acuerdo con la importancia que tenga para cumplir con el propósito asignado y considerando las necesidades (needs) ya definidas.
7. El sistema es fiel a los conceptos de la arquitectura e-MDB.
8. El sistema proporciona retroalimentación clara al usuario sobre cuáles son las misiones (missions) y sus funciones de impulso (funciones de drive), además de cómo y cuándo se están satisfaciendo.
9. El sistema se prueba en un robot real en un ambiente de simulación.

Bibliografía

[1] A. Romero, F. Bellas, and R. J. Duro, "A perspective on lifelong open-ended learning autonomy for robotics through cognitive architectures," *Sensors*, vol. 23, DOI 10.3390/s23031611, no. 3, p. 1611, 2023. [En línea]. Disponible: https://www.mdpi.com/1424-8220/23/3/1611

[2] I. Kotseruba and J. K. Tsotsos, "40 years of cognitive architectures: core cognitive abilities and practical applications," *Artificial Intelligence Review*, vol. 53, DOI 10.1007/s10462-018-9646-y, no. 1, pp. 17–94, 2020.

[3] P. Langley, J. E. Laird, and S. Rogers, "Cognitive architectures: Research issues and challenges," *Cognitive Systems Research*, vol. 10, DOI 10.1016/j.cogsys.2006.07.004, no. 2, pp. 141–160, 2009.

[4] Q. Houbre and R. Pieters, "Dynamic neural curiosity enhances learning flexibility for autonomous goal discovery," *arXiv preprint arXiv:2412.00152*, 2024.

[5] S. Forestier, V. G. Santucci, A. Baranes, and P.-Y. Oudeyer, "Intrinsically motivated goal exploration processes with automatic curriculum learning," *Journal of Machine Learning Research*, vol. 23, no. 139, pp. 1–45, 2022.

[6] V. G. Santucci, G. Baldassarre, and M. Mirolli, "Grail: a goal-discovering robotic architecture for intrinsically-motivated learning," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 8, no. 3, pp. 214–231, 2016.

[7] A. Romero, G. Baldassarre, R. J. Duro, and V. G. Santucci, "A motivational architecture for open-ended learning challenges in robots," *arXiv preprint arXiv:2506.18454*, 2025.

[8] R. Novianto, B. Johnston, and M.-A. Williams, "Attention in the asmo cognitive architecture," in *2010 10th International Conference on Intelligent Systems Design and Applications*, pp. 1343–1348. IEEE, 2010.