

Aprendizado procedimental e sensório-motor em robôs cognitivos

Leonardo L. Rossi^{1,3}, Leticia M. Berto², Eric Rohmer¹,
Paula P. Costa¹, Ricardo R. Gudwin¹, Esther L. Colombini², Alexandre S. Simões³

¹Faculdade de Engenharia Elétrica e Computação – Universidade Estadual de Campinas (Unicamp) – Av. Albert Einstein, 400 – Campinas, SP – CEP 13083-852 – Brasil.

²Instituto de Computação
Universidade Estadual de Campinas (Unicamp) – Campinas, SP – Brasil.

³Departamento de Engenharia de Controle e Automação
Universidade Estadual Paulista (Unesp) – Sorocaba, SP – Brasil.

Abstract. *This work presents a cognitive agent based on the CONAIM model (Conscious Attention-Based Integrated Model) that can incrementally learn procedures for carrying out experiments inspired by the first three sensory-motor substages of Jean Piaget's Theory. The agent was implemented using the Cognitive Systems Toolkit (CST) and validated with mobile robots in simulated environments. Results shown that this approach can solve complex object tracking experiments incrementally. **Keywords:** Cognitive Robotics, Reinforcement Learning, Incremental Learning.*

Resumo. *Este trabalho apresenta um agente cognitivo baseado no modelo CONAIM (Conscious Attention-Based Integrated Model) que pode aprender incrementalmente procedimentos para a realização de experimentos inspirados nos três primeiros subestágios sensório-motores da Teoria de Jean Piaget. O agente foi implementado usando o Cognitive Systems Toolkit (CST) e validado com robôs móveis em ambientes simulados. Os resultados demonstram que a abordagem pode resolver experimentos complexos de rastreamento de objetos através do aprendizado incremental. **Palavras-chave:** Robótica Cognitiva, Aprendizado por Reforço, Aprendizado Incremental.*

Dissertação de mestrado defendida em 5/7/2021. **Orientador:** Prof. Dr. Alexandre da Silva Simões. **Co-orientadora:** Profa. Dra. Esther Luna Colombini¹

1. Introdução

Inspirados em como os humanos constroem o conhecimento por meio da interação com o mundo, pesquisadores de arquiteturas cognitivas buscam reproduzir este comportamento com criaturas artificiais [Paraense et al. 2016]. Porém, o desenvolvimento de habilidades cognitivas em máquinas requer a coordenação de mecanismos complexos que dependem uns dos outros. O processo de desenvolvimento dessas habilidades é incremental e evolutivo, conforme apresentado por Piaget [Piaget 1952].

¹Este trabalho foi realizado no âmbito do PPI-Softex com apoio do MCTI por meio do Termo de Cooperação Técnica [01245.013778/2020-21].

Neste trabalho, um agente cognitivo baseado no modelo CONAIM (*Conscious Attention-Based Integrated Model*) foi proposto [Colombini et al. 2017] e implementado com o *Cognitive Systems Toolkit* (CST) [Paraense et al. 2016] para execução em uma robô humanoide. O agente foi projetado para aprender incrementalmente procedimentos para realizar experimentos inspirados nos três primeiros subestágios sensório-motores da Teoria de Jean Piaget [Piaget 1952]. Ao longo do trabalho, são demonstradas as funções cognitivas necessárias em cada subestágio para formar reações circulares e como a adição de novas funções ajuda o agente a resolver tarefas complexas, anteriormente não resolvidas. O agente é treinado seguindo o paradigma do Aprendizado por Reforço [Sutton and Barto 2015], demonstrando que o agente é capaz de aprender a aprender novos comportamentos baseados na reutilização de conhecimento anterior.

Como principais contribuições deste trabalho, podemos listar:

1. O desenho e implementação do fluxo top-down do CONAIM no CST que pode ser incorporado em qualquer agente que seja implementado com o CST;
2. O projeto e implementação de um único mecanismo de aprendizado procedural em CST que pode aprender e reutilizar esquemas incrementalmente para os três primeiros subestágios sensório-motores da Teoria de Piaget;
3. A modelagem de um conjunto de ambientes para experimentos sensório-motores com robôs humanoides;
4. A implementação e avaliação computacional de experimentos sensório-motores para **Object Tracking** nas três primeiras subestágios sensório-motores da Teoria de Piaget executados com o humanoide Marta, conforme proposto por [Berto 2020].

Também investigamos a relação entre atenção, aprendizagem e posição a cada instante de atuadores físicos e virtuais, analisando a relação do desenvolvimento cognitivo com memórias e aprendizagem no modelo CONAIM. Por fim contribuímos com a área disponibilizando nosso código em: <https://github.com/CST-Group/cst>.

2. CONAIM e CST

O modelo CONAIM (*Conscious Attention-Based Integrated Model*) [Colombini et al. 2017] é um modelo formal baseado em atenção para consciência da máquina. O CONAIM incorpora diversos aspectos relevantes para um agente cognitivo (memórias, esquema corporal, motivação, atenção, entre outros) e é capaz de lidar com múltiplos sistemas sensoriais, múltiplos processos de extração de características, tomada de decisão e aprendizado [Colombini et al. 2017]. Em seu ciclo cognitivo, são utilizados mecanismos **top-down** e **bottom-up**. Durante o ciclo **bottom-up**, os sensores fornecem dados externos, armazenados em uma janela de tempo, na memória sensorial para formar mapas de características, mapa atencional e mapa de saliência [Colombini et al. 2017]. A modelagem completa está detalhada em [Colombini et al. 2017]. No ciclo **top-down**, a modulação atencional do sistema dependerá do estado de atenção global (por exemplo, o nível da bateria da robô). Também dependerá do objetivo do agente e da dinâmica atencional do estado atual [Colombini 2014].

O CST (*Cognitive Systems Toolkit*) [Paraense et al. 2016] é um kit de ferramentas aberto baseado em Java para construir arquiteturas cognitivas. A arquitetura CST é orientada a *codelets*, que são pequenos pedaços de código, implementados

como funções assíncronas, e que rodam em paralelo com tarefas simples e definidas [Paraense et al. 2016]. Um **objeto de memória** (MO) é um sinal ou representação utilizados, com outros MOs, por *codelets* para armazenar e acessar dados [Paraense et al. 2016].

3. Robótica de Desenvolvimento e a Teoria de Piaget

O desenvolvimento de agentes artificiais com comportamentos autônomos e adaptativos e aprendizado incremental são objetivos da pesquisa em robótica cognitiva e robótica de desenvolvimento (DevRobotics) [Cangelosi and Schlesinger 2014]. A área surgiu devido à necessidade de robôs executarem tarefas que exigem níveis comparáveis de inteligência humana em ambientes complexos e imprevisíveis, envolvendo adaptação e evolução [Cangelosi and Schlesinger 2014]. Os modelos e experimentos da área são inspirados nos princípios e mecanismos de desenvolvimento observados no início da vida, envolvendo robôs desempenhando as mesmas habilidades cognitivas das crianças, como nos experimentos propostos por Jean Piaget. Um conceito relevante na teoria de Piaget [Piaget 1952] são os **esquemas**, que representam redes de estruturas mentais que ajudam a lembrar conceitos específicos e entender o ambiente. Quando processos mentais simples se tornam mais sofisticados, novos esquemas são desenvolvidos e o comportamento também se torna mais complexo e adequado ao ambiente.

No desenvolvimento mental, segundo Piaget, **adaptação** ou **aprendizagem** é a tendência de ajustar os processos mentais ao ambiente alterando as estruturas cognitivas [Piaget 1952]. O processo de adaptação envolve equilibrar os processos de assimilação e acomodação. **Assimilação** e **Acomodação** são processos inseparáveis, complementares e simultâneos [Piaget 1952]. **Assimilação** é a criação de novos esquemas seguindo o mesmo ciclo ou sequência de esquemas existentes para interpretar experiências e tomar decisões. **Acomodação** é o processo complementar que envolve a alteração de esquemas existentes como resultado de novas informações adquiridas através da assimilação. A assimilação pode originar **reações circulares**, repetições de ciclos adquiridos ou em processo de aquisição [Piaget 1952]. A reação circular resulta da assimilação de um resultado interessante, até então desconhecido para o sujeito, que foi produzido pela redescoberta ou repetição da ação. As reações circulares podem ser primárias, secundárias ou terciárias. **Reações Circulares Primárias** são comportamentos derivados de reflexos, atividades do próprio corpo que formam novos esquemas através da coordenação dos sentidos; **Reações Circulares Secundárias** são derivadas de comportamentos intencionais que direcionam o interesse para resultados externos ao invés do corpo do bebê. **Reações Circulares Terciárias** são o esforço do sujeito em buscar novas experiências.

4. Experimentos de aprendizado incremental

Experimentos com arquiteturas cognitivas no campo da DevRobotics são fortemente baseados em teorias de desenvolvimento infantil. No entanto, um ponto comum nesses experimentos é a falta de padronização para conduzir e avaliar o desenvolvimento do agente. Para ajudar a resolver essas questões, [Berto 2020] propôs um conjunto de experimentos incrementais para o cenário robótico de acordo com os estágios de desenvolvimento sensorio-motor de Piaget juntamente com os resultados esperados. Ela também definiu a relação entre *cognits* [Fuster 2006] e os esquemas propostos por Piaget no desenvolvimento do estágio sensorio-motor [Piaget 1952].

O conceito de *cognit* [Fuster 2006] é usado por [Berto 2020] para representar elementos de processamento de memória com um nível de avaliação. Os *cognits* foram inseridos no contexto do estágio sensório-motor de Piaget, na seguinte caracterização: **1º Subestágio. Uso de Reflexos:** criação de *cognits* inatos (ações, estados e mecanismo de aprendizagem); **2º Subestágio. Reações Circulares Primárias:** criação e adaptação de *cognits* baseados em funções, mas não em intenções; **3º Subestágio. Reações Circulares Secundárias:** criação e adaptação de *cognits* para formar planos, objetivos e intenções. Neste trabalho construímos um agente para realizar os experimentos propostos por [Berto 2020] relacionados ao **rastreamento de objetos**, onde o robô deve aprender de forma incremental as habilidades de rastreamento de objetos através de sensores RGB-D. [Berto 2020] também disponibiliza uma lista completa dos experimentos cognitivos propostos e sua metodologia de análise.

5. Materiais e métodos

Durante o desenvolvimento do mestrado, foi projetado um modelo de robô cognitivo para investigar quais módulos cognitivos devem estar presentes em um agente que possa aprender incrementalmente a realizar experimentos com níveis de dificuldade crescentes. Primeiro, o agente foi modelado usando módulos CONAIM implementados com o CST. Foram adicionados módulos à medida que o agente evolui do 1º para o 3º subestágio sensório-motor, de forma que este responda a tarefas mais complexas. Os módulos ativos para cada subestágio são demonstrados na Figura 2. Foi empregado aprendizado por reforço para o agente aprender esquemas para memória procedimental. Para corresponder aos experimentos de rastreamento de objetos propostos por [Berto 2020], a acuidade visual do robô foi incrementada a cada subestágio alterando a resolução da câmera. Também foi incrementada a lista de comportamentos disponíveis, refinando as ações à medida que o agente cresce.

Robôs e ambiente: A robô humanoide Marta, um robô feminino de 1,1 m de altura, foi projetada e construída fisicamente por nosso grupo de trabalho. Marta tem 25 graus de liberdade, e sua cabeça pode realizar movimentos de *pitch* e *yaw*. O software de controle da robô foi desenvolvido de acordo com o modelo cognitivo detalhado nas próximas seções, totalmente desenvolvido em *Java* usando o *Cognitive Systems Toolkit* (CST) [Paraense et al. 2016]. Também foram criadas várias cenas simuladas no CoppeliaSim para os experimentos. Nas cenas, Marta está sentada em um pequeno espaço delimitado e uma visão ampla de seu entorno. Uma arena foi delimitada fora desse primeiro espaço, e blocos coloridos (azuis e verdes) foram distribuídos aleatoriamente. Um segundo robô, um Pioneer P3DX vermelho, navega aleatoriamente na arena como um distrator móvel. Este robô foi modelado com comportamentos reativos usando o Algoritmo Braitenberg [Braitenberg 1984]. Marta foi equipada com uma câmera RGB-D na cabeça, inspecionando o mundo em quatro canais distintos de cores (R, G e B) e informações de distância (D). Ambos os robôs e o ambiente são mostrados na Figura 1.

Modelo cognitivo: O sistema cognitivo da humanoide Marta foi modelado de acordo com o CONAIM [Colombini et al. 2017]. O primeiro nível da arquitetura é o **sistema atencional** [Colombini 2014], responsável por coletar os dados do ambiente, compactar essas informações e selecionar os pontos mais relevantes da cena. As entradas do sistema recebem múltiplas informações sensoriais, dos quatro canais da câmera (R, G, B, D), com alguma resolução previamente configurada. Essas observações geram quatro mapas

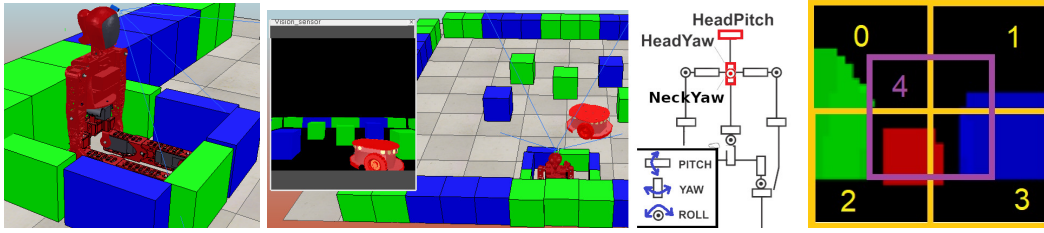


Figura 1. O ambiente de simulação. Da Esq. para Dir. a) Robô Marta equipada com uma câmera RGB-D; b) Ambiente com blocos coloridos distribuídos e um robô Pioneer P3DX atuando como distrator. A visão de Marta da cena é mostrada à esq. (c) Graus de liberdade da agente. Motores utilizados (em vermelho). (d) Divisão do espaço visual da agente para o mapa de características *top-down* de região.

de características *bottom-up* distintos, que carregam as informações sobre os sinais mais discrepantes em cada canal. Também foram adotados três mapas de características *top-down*, que podem enfatizar aspectos particulares de cor, distância e região dos dados de entrada de acordo com o objetivo da agente. Todos os mapas de características são ponderados e combinados com o mapa atencional, modulado por efeitos de inibição de retorno (IOR), para geração do mapa de saliência.

O segundo nível da arquitetura é o **Sistema Cognitivo**, responsável por modular a relação entre a robô e o ambiente, bem como pela evolução cognitiva da agente. Nos presentes experimentos, apenas alguns módulos do sistema cognitivo CONAIM [Simões et al. 2017] foram considerados. Esses módulos foram ativados incrementalmente durante os experimentos sucessivos. Cada procedimento/esquema $m_p \in M_P$ representa um conhecimento aprendido armazenado na memória procedimental M_P . Inicialmente, a memória de trabalho (M_W) recebe o mapa de saliência (\mathcal{L}) – empregado como estado no aprendizado por reforço – emergindo do sistema atencional. Como a agente não tem conhecimento no início, um novo procedimento m_p é criado em M_P , e a agente cognitiva pode gradualmente aprender algo sobre ele. Se a agente tiver algum conhecimento anterior armazenado em M_P que se ajuste ao estado atual, ocorre um *recall* de procedimentos (R_P). Este conhecimento será considerado pelo *Decision Maker* (D). Em alguns experimentos, também foi modelado um conjunto de motivações $mv_i \in M_V$ para explorar o uso de novas ações em alguns estados. A volição VO é a função responsável por transformar as motivações da agente em tarefas que o processo de decisão também irá considerar. Uma função de aprendizado procedimental (L_P) é responsável por criar ou atualizar o conteúdo de $m_p \in M_P$, neste caso atuando respectivamente de forma análoga à assimilação e acomodação na teoria de Piaget. Um esquema do modelo cognitivo adotado é mostrado na Figura 2 (a). A Figura 2 (b) mostra o esquema de implementação da arquitetura CONAIM+CST para o aprendizado incremental proposto.

5.1. Formulação do Aprendizado do Agente

Estados: Foram adotadas diferentes resoluções de imagem entre os experimentos: 64x64 pixels (EXP-01), 128x128 pixels (EXP-02) e 256x256 pixels (EXP-03). Os mapas \mathcal{F}_R , \mathcal{F}_G , \mathcal{F}_B e \mathcal{F}_D *bottom-up* foram calculados com um *average pool* sobre a observação de cada mapa de cores no tempo t , e então a diferença entre a média de cada região e a média da imagem. Ou seja, os elementos mais discrepantes em cada mapa a cada vez que

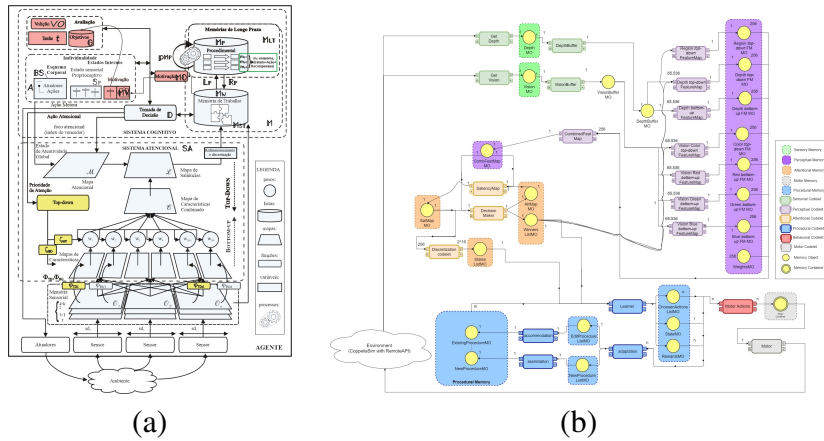


Figura 2. (a) Modelo interno da agente humanoide Marta. O modelo é iniciado no EXP 1 somente com os elementos em azul. No EXP 2 são adicionados os elementos em vermelho. No EXP 3 são adicionados os elementos em amarelo. (b) Esquema de implementação da arquitetura CONAIM+CST com a robô Marta e que recebe estímulos atencionais *bottom-up* e *top-down*.

t são destacados. Mapas de características *top-down* (\mathcal{F}_{color} , \mathcal{F}_{dist} e \mathcal{F}_{reg}) permitem que a agente direcione deliberadamente sua atenção para os elementos desejados. Cada valor de pixel foi comparado com um objetivo específico (cor, distância ou região espacial) para construção desses mapas. Particularmente no mapa de características *top-down* (\mathcal{F}_{reg}), o espaço visual foi dividido em 5 regiões distintas, conforme mostrado na Figura 1 (c). O mapa de características combinado (\mathcal{C}) calcula a média ponderada, pixel por pixel, dos i mapas habilitados. O mapa atencional (\mathcal{M}), que contém o mecanismo de Inibição de Retorno (IoR) [Colombini 2014] é aplicado sobre (\mathcal{C}) para gerar o mapa de saliência (\mathcal{L}). Por fim, o vetor de estados (S) é calculado na Memória de Trabalho (M_W) usando um operador *MaxPool* com um kernel 4x4 e 4 passos sobre o mapa de saliência (\mathcal{L}), gerando um estado final 4x4, com uma discretização de 2 níveis obtida com um limite. Dessa forma, o sistema tem um total de 65.536 (2^{16}) estados.

Ações: A robô foi autorizada a realizar 17 ações possíveis (A), divididas em três grupos: ações motoras, virtuais e atencionais. As Ações Motoras (A_m) são as ações sobre os atuadores físicos do pescoço da robô, capazes de girar os motores da cabeça *pitch* e *yaw*. As Ações Virtuais (A_v) são internas à agente e simulam o movimento dos olhos. O atuador virtual seleciona um ponto no espaço visual onde a agente foca (fóvea). As Ações de Atenção (A_a) são divididas em dois subgrupos. O primeiro grupo de ações envolve direcionar a cabeça da robô na direção do ponto mais saliente da imagem (vencedor). O segundo grupo refere-se às ações *top-down*. Ele pode enfatizar cores, distâncias ou regiões específicas nos mapas de recursos. Algumas dessas ações possíveis foram habilitadas ou desabilitadas para cada um dos três conjuntos de experimentos.

Aprendizado: Foi utilizado o algoritmo de aprendizado por reforço Q-learning [Sutton and Barto 2015]. Os elementos de memória $m_p \in M_P$ foram modelados como *QTables* capazes de armazenar pares Estado-Ação ($S \rightarrow A$) para procedimentos particulares. Os estados (S) foram modelados a partir dos mapas de saliência (\mathcal{L}) que representam o ambiente. O reforço recompensa positivamente a robô se houver um sincronismo espaço-temporal entre o estímulo visual (ponto mais saliente da imagem) e o foco atual

da robô (motor ou virtual). Não há recompensa se não houver tal sincronismo, e a recompensa é fortemente negativa se a robô perder o equilíbrio. O mecanismo de aprendizado permanece inalterado durante todos os experimentos.

5.2. Experimentos propostos

Os experimentos foram propostos com base nos cenários para avaliar a capacidade da agente em aprender a rastrear outros agentes ou objetos, propostos por [Berto 2020]. Durante o treinamento, a cada episódio a agente é inicializada em uma posição fixa no ambiente e um ruído aleatório é adicionado aos seus atuadores. O episódio é encerrado quando uma das seguintes condições é atingida: **i)** a agente atingiu o número máximo de passos/ações; **ii)** a robô cai ou excede os limites de seus atuadores motorizados, ou **iii)** a robô não tem saliências por várias iterações. **Parâmetros de simulação.** Número máximo de episódios: 200; Número máximo de passos: 500; Número máximo de iterações sem saliências: 5. **Parâmetros de aprendizado.** Uma política ϵ -greedy foi empregada com taxa de exploração começando em 0,95 e decaindo linearmente até 0 no último episódio. Taxa de aprendizado α : 0,9; Taxa de desconto temporal γ : 0,99. **Acuidade visual / Resoluções do sensor de visão.** 64x64 para EXP-01; 128x128 para EXP-02; 256x256 para EXP-03. **Recompensas.** +1 de recompensa para cada novo dado inserido na memória procedural; +1 para manter ou direcionar um atuador (motor ou virtual) ao vencedor da atenção; -10 se a agente cair ou exceder os limites dos atuadores físicos, ou se a agente não tiver saliências em seu ciclo atencional por várias iterações; e, apenas para o EXP-03, +1 quando a agente identifica regiões nas quais determinada característica desejada é destacada de acordo com o processo *top-down*.

EXP-01. 1º Subestágio: Uso de Reflexos. Este conjunto de experimentos investiga um processo computacional proposto para modelar as **reações reflexas**. Neste conjunto, nenhuma intencionalidade ou motivação foi adotada.

EXP-02. 2º Subestágio: Reações Circulares Primárias. Neste conjunto de experimentos, foi investigado se as reações reflexas na agente, iniciadas no 1º subestágio, podem evoluir para comportamentos semelhantes às reações circulares primárias propostas por Piaget. Assim como no experimento anterior, apenas os estímulos *bottom-up* foram considerados. **Motivação.** É aplicada uma função de motivação (MO), relacionada à curiosidade da agente sobre os efeitos de suas ações, para incentivar a agente à explorar esquemas que ainda não foram explorados no episódio atual. Todo o conteúdo de M_P aprendido anteriormente é preservado.

EXP-03. 3º Subestágio: Reações Circulares Secundárias. Neste último conjunto de experimentos, foi investigado um processo computacional baseado em quais comportamentos associados às reações circulares secundárias propostas por Piaget podem ser observados. Ou seja, foi verificado se a agente poderia selecionar intencionalmente uma ação que lhe permitisse atingir um objetivo. O mecanismo de atenção *top-down* é usado nesta fase. Todo o conteúdo de M_P aprendido anteriormente é preservado.

6. Resultados e Discussão

Para cada um dos três conjuntos de experimentos propostos para treinar e validar a arquitetura integrada, foram executados 200 episódios por experimento. Ao final de cada episódio, a recompensa obtida e o número de ações realizadas são redefinidos e os atuadores da robô retornam para a posição inicial. O robô Pioneer 3-DX é posicionado

aleatoriamente na cena. A Figura 3 demonstra a recompensa resultante e o número de ações executadas por episódio para cada experimento de aprendizado executado. Pode-se notar que, como a agente reutiliza o conhecimento dos subestágios anteriores, tanto a recompensa quanto o número de ações são maiores para um agente mais desenvolvido ($3^o > 2^o > 1^o$).

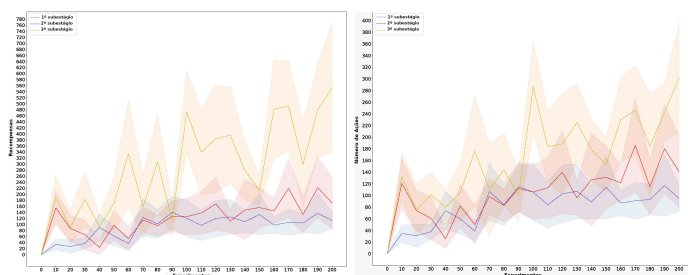


Figura 3. Recompensa resultante (esquerda) por episódio e número de ações (direita) para cada experimento de aprendizado para todos os subestágios.

6.1. EXP-01 - 1º Subestágio: Uso de Reflexos

O experimento demonstra o curso atencional de seleção para percepção para a agente executando apenas o curso de atenção *bottom-up*. Os resultados obtidos para este experimento são demonstrados na Figura 4. Nos primeiros episódios, a agente explora os limites de seus atuadores devido à alta taxa de exploração, enquanto promove o refinamento de pares estado-ação para o atuador virtual de seleção de fóvea. A agente estabeleceu seu foco atencional no robô Pioneer P3DX, enquanto este se movia em regiões mais próximas à humanoide. Com o distanciamento do robô Pioneer P3DX, a agente orientou sua atenção para outros objetos próximos e seu próprio corpo. Os estímulos obtidos durante a exploração do corpo do agente reforçaram os reflexos utilizados. A ausência de um sistema de motivação fez com que a agente realizasse as ações reforçadas em maior quantidade, mesmo quando os estímulos que promoviam esse reforço não estavam mais presentes, e em menor quantidade as ações que não participavam dessas interações.

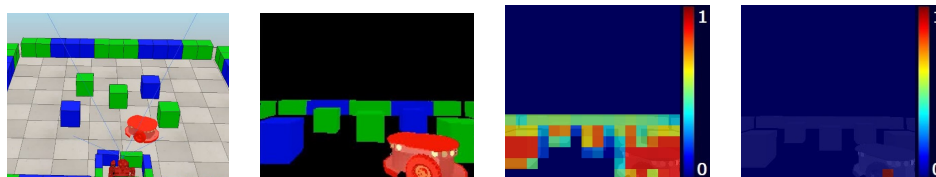


Figura 4. EXP-01: 1º Subestágio. Dados sensoriais obtidos no 1º episódio do Aprendizado Procedimental. Esq. para dir. (a) Visão geral da cena no *CoppeliaSim* ($t = 1s$); (b) visão da câmera de Marta ($t = 1s$); (c) Mapa de Saliência ($t = 3s$); (d) Vencedor do ciclo atencional ($t = 3s$).

6.1.1. Validação do Aprendizado - *Tracking* de Objetos

Neste experimento utilizou-se a QTable resultante do final do último episódio do Aprendizado Procedimental. Como forma de avaliação da política aprendida, foram executados 100 episódios de teste, sem atualização dos parâmetros de aprendizado, para cada experimento com um máximo de 500 ações em cada.

Experimento A - 1º Subestágio - Objeto em posição fixa e com cor primária. A robô Marta foi posicionada a 80cm do robô Pioneer P3DX. Os resultados obtidos para este experimento são demonstrados na Figura 5. A agente inicialmente direcionou sua atenção para o robô Pioneer P3DX, que permaneceu estacionário durante este experimento, como sugerido por Berto (2020). [Berto 2020]. No entanto, a ação dos ciclos excitatórios e inibitórios promovidos pelo Sistema Atencional direcionou o foco atencional para regiões mais próximas à humanoide. A atuação dos reflexos reforçados durante a exploração do corpo do agente na Aprendizagem Procedimental resultou na alternância dos atuadores do agente entre as regiões mais próximas da humanoide e do robô Pioneer P3DX. Como esperado, a robô aprendeu a responder a estímulos salientes usando reflexos.

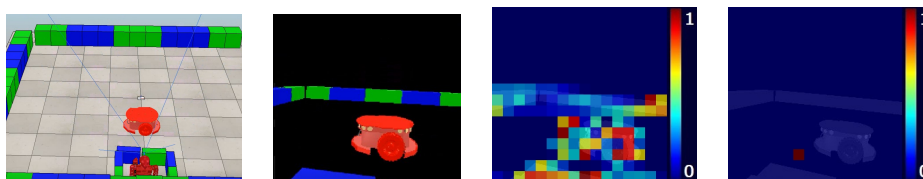


Figura 5. EXP-01: 1º Subestágio. Dados sensoriais obtidos no Experimento A. Esq. para dir. (a) Visão geral da cena no *CoppeliaSim* ($t = 40s$); (b) visão da câmera de Marta ($t = 40s$); (c) Mapa de Saliência ($t = 43s$); (d) Vencedor do ciclo atencional ($t = 43s$).

Experimento B - 1º Subestágio - Objeto movendo-se lentamente e de cor primária. Marta foi posicionada 80cm à frente do robô Pioneer P3DX, o qual possui um algoritmo Braitenberg e se move com velocidade constante de 0,1m/s. Os resultados obtidos para este experimento são demonstrados na Figura 6. A agente direcionou sua atenção para o robô Pioneer P3DX enquanto este permanecia nas regiões mais próximas ao agente (região frontal). No entanto, à medida que o Pioneer P3DX se desloca para as áreas laterais, a atuação dos reflexos reforçados durante o Aprendizado Procedimental resultou novamente no direcionamento do foco atencional para as regiões mais próximas à humanoide. Assim, Marta não conseguiu rastrear o objeto em movimento fora de seu campo visual, conforme esperado para esta fase de desenvolvimento.

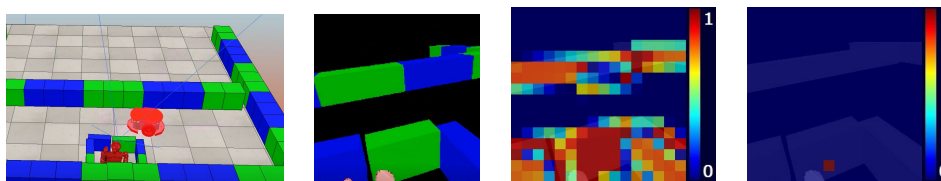


Figura 6. EXP-01: 1º Subestágio. Dados sensoriais obtidos no Experimento B. Esq. para dir. (a) Visão geral da cena no simulador *CoppeliaSim* ($t = 40s$); (b) visão da câmera de Marta ($t = 40s$); (c) Mapa de Saliência ($t = 43s$); (d) Vencedor do ciclo atencional ($t = 43s$).

6.2. EXP-02 - 2º Subestágio: Reações Circulares Primárias

Neste experimento, Marta continua com apenas elementos de percepção *bottom-up*. Com a implementação de um modelo de motivação, a agente passa a explorar possíveis ações

que não possuem esquemas definidos na Memória Procedimental M_p . As reações reflexas desenvolvidas no 1º subestágio podem gerar reações circulares primárias, estabilizando o aprendizado de determinados conjuntos de ações. Os resultados obtidos para este experimento são demonstrados na Figura 7. O uso da QTable do subestágio anterior permite que a agente tenha um melhor desempenho na tarefa de rastreamento de objetos. O Sistema Atencional do 2º Subestágio, durante o Aprendizado Procedimental, permitiu que a agente estabelecesse seu foco atencional no robô Pioneer P3DX enquanto este se movia em regiões mais próximas à humanoide, como no 1º subestágio. Com a retirada do robô Pioneer P3DX, a agente voltou a orientar sua atenção para objetos próximos e seu próprio corpo. No entanto, com a ação do sistema de motivação, a agente foi motivada a explorar todas as ações possíveis para cada novo esquema não encontrado na Memória Procedimental. Esse comportamento minimizou o desempenho das ações reforçadas desenvolvidas no 1º subestágio, permitindo à agente adquirir maiores recompensas e promovendo a formação de reações circulares primárias.

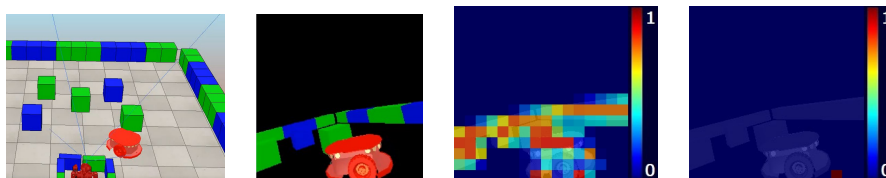


Figura 7. EXP-02: 2º Subestágio. Dados sensoriais obtidos no 1º episódio do Aprendizado Procedimental. Esq. para dir. (a) Visão geral da cena no simulador *CoppeliaSim* ($t = 1s$); (b) visão da câmera de Marta ($t = 1s$); (c) Mapa de Saliência ($t = 3s$); (d) Vencedor do ciclo atencional ($t = 3s$).

6.2.1. Validação do Aprendizado - *Tracking* de Objetos

Experimento A - 2º Subestágio - Objeto em movimento e de cor primária. Utilizou-se a QTable resultante do final do último episódio de Aprendizado Procedimental. Marta foi posicionada com o robô Pioneer P3DX fora de seu campo de visão, conforme ilustrado na cena da Figura 8 (a). O robô Pioneer P3DX possui um algoritmo Braitenberg e se move com velocidade constante de 0,1m/s. A agente direcionou sua atenção para as regiões mais próximas de seu corpo quando o robô Pioneer P3DX saiu de seu campo de visão. A utilização das reações circulares primárias promoveu um desempenho com maior número de ações realizadas em relação ao subestágio anterior. O aumento da acuidade visual e o refinamento do movimento dos atuadores neste subestágio também permitiram que o agente tenha maior controle sobre seus atuadores.

6.3. EXP-03 - 3º Subestágio: Reações Circulares Secundárias.

Neste subestágio, a humanoide Marta possui um algoritmo cognitivo-atencional que possui todos os elementos demonstrados na Figura 2, com elementos de percepção *bottom-up* e *top-down*, e modelos de intencionalidade e motivação. A agente é capaz de explorar as reações circulares primárias desenvolvidas nas etapas anteriores e desenvolver reações circulares secundárias. Os resultados obtidos para este experimento são demonstrados na Figura 9. O Sistema de Atenção do 3º Subestágio, durante o Aprendizado Procedimental, permitiu que a agente estabelecesse seu foco de atenção no robô P3DX. Ao utilizar

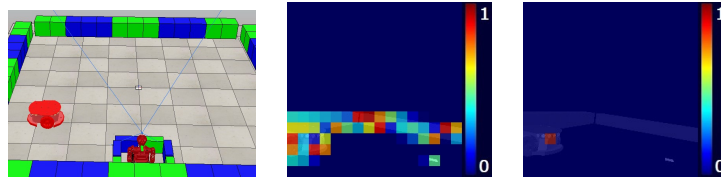


Figura 8. EXP-02: 2º Subestágio. Dados sensoriais obtidos no Experimento A. Esq. para dir. (a) Visão geral da cena no simulador *CoppeliaSim* (t = 1s); (b) Mapa de Saliência (t = 3s); (c) Vencedor do ciclo atencional (t = 3s).

as ações atencionais, a agente foi capaz de acompanhar o movimento do P3DX mesmo quando este estava nas regiões mais distantes da humanoide e fora de seu campo de visão.

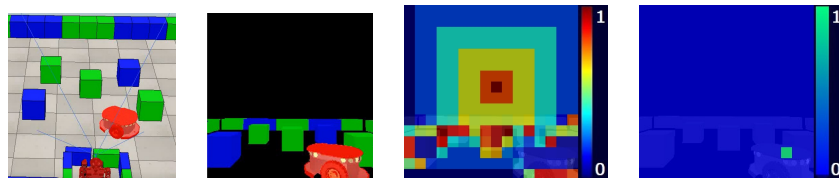


Figura 9. EXP-03: 3º Subestágio. Dados sensoriais obtidos no 1º episódio do Aprendizado Procedimental. Esq. para dir. (a) Visão geral da cena no simulador *CoppeliaSim* (t = 1s); (b) visão da câmera de Marta (t = 1s); (c) Mapa de Saliência (t = 3s); (d) Vencedor do ciclo atencional (t = 3s).

6.3.1. Validação do Aprendizado - *Tracking* de Objetos

Experimento A - 3ª Subestágio - Objeto em movimento e cor primária. Utilizou-se a QTable resultante do último episódio de Aprendizado Procedimental. Marta foi posicionada com o robô Pioneer P3DX fora de seu campo de visão. O robô Pioneer P3DX possui algoritmo Braitenberg e se move com velocidade constante de 0,1m/s. Nesse experimento, a agente manteve seu foco de atenção no robô Pioneer P3DX mesmo quando ele estava distante. Esse comportamento permitiu que o agente realizasse diferentes ações de forma mais uniforme para receber mais recompensas com as diferentes ações realizadas. O aumento da acuidade visual neste subestágio e o refinamento da movimentação dos atuadores permitiram que o agente tivesse maior controle sobre seus atuadores do que nos subestágios anteriores. A utilização de reações circulares secundárias desenvolvidas durante este subestágio promoveu um desempenho com maior número de ações realizadas em relação às subestágios anteriores.

7. Conclusão

Neste trabalho, foi proposto um mecanismo incremental de aprendizado procedimental que pode criar esquemas e reutilizar esquemas previamente aprendidos, inspirado nos subestágios de desenvolvimento sensório-motor de Piaget. Foram investigados quais módulos em uma arquitetura cognitiva são necessários para controlar um robô que interage com seu ambiente enquanto realiza um conjunto de experimentos sensório-motores com dificuldade crescente. Foi demonstrada a importância da motivação e da atenção para formar reações circulares primárias e secundárias a partir de reflexos. Essa abordagem

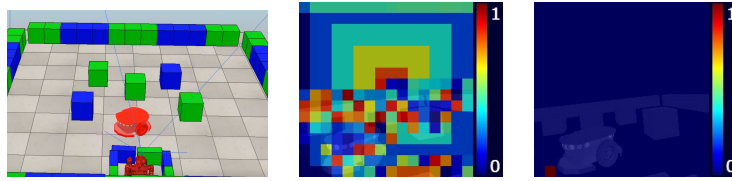


Figura 10. EXP-03: 3º Subestágio. Dados sensoriais obtidos no Experimento A. Esq. para dir. (a) Visão geral da cena no simulador *Coppeliasim* (t = 1s); (b) Mapa de Saliência (t = 3s); (c) Vencedor do ciclo atencional (t = 3s).

permitiu empregar um único mecanismo incremental que evolui ao longo do tempo. Os experimentos demonstraram a viabilidade do uso de uma arquitetura cognitivo-atencional baseada no CONAIM e implementada com CST. Foram implementados com sucesso os experimentos correspondentes aos três primeiros subestágios propostos por [Berto 2020] para rastreamento de objetos. Além deste trabalho, o seguinte artigo foi desenvolvido durante o mestrado e aceito para publicação em congresso internacional: BERTO, L. M., ROSSI, L. L., ROHMER, E., COSTA, P. D. P., SIMOES, A. S., GUDWIN, R. R. and COLOMBINI, E. L. **Learning over the Attentional Space with Mobile Robots**. IEEE International Conference on Development and Learning (ICDL). 2020.

Referências

- Berto, L. M. (2020). Exploring cognitive functions in robotics. *Institute of Computing of the University of Campinas (Master thesis)*.
- Braitenberg, V. (1984). *Vehicles: Experiments in Synthetic Psychology*. Cambridge MA: MIT Press.
- Cangelosi, A. and Schlesinger, M. (2014). *Developmental Robotics: From Babies to Robots*. The MIT Press.
- Colombini, E. L. (2014). An attentional model for intelligent robotics agents. *Instituto Tecnológico de Aeronáutica, São José dos Campos (PhD Thesis)*.
- Colombini, E. L., da Silva Simões, A., and Ribeiro, C. H. C. (2017). An attentional model for autonomous mobile robots. *IEEE Systems Journal*, 11(3):1308–1319.
- Fuster, J. M. (2006). The cognit: A network model of cortical representation. *International Journal of Psychophysiology*, 60:125–132.
- Paraense, A. L. O., Raizer, K., de Paula, S. M., Gudwin, R. R., and Rohmer, E. (2016). The cognitive systems toolkit and the cst reference cognitive architecture. *Biologically Inspired Cognitive Architectures*, 17:32–48.
- Piaget, J. (1952). *The origins of intelligence in children*. The origins of intelligence in children. W Norton & Co, New York, NY, US.
- Simões, A. d. S., Colombini, E. L., and Ribeiro, C. H. C. (2017). CONAIM: A Conscious Attention-Based Integrated Model for Human-Like Robots. *IEEE Systems Journal*, 11(3):1296–1307.
- Sutton, R. and Barto, A. (2015). *Reinforcement Learning: An Introduction*. MIT Press. Cambridge, Massachusetts.