EVOLUÇÃO DE REDES NEURAIS ARTIFICIAIS EM UMA ARQUITETURA COGNITIVA BIOLOGICAMENTE INSPIRADA PARA NAVEGAÇÃO AUTÔNOMA EM LABIRINTOS

Luciene de O. Marin*, Mauro Roisenberg†, Edson R. De Pieri*

*Laboratório de Controle e Automação (LCA), Departamento de Automação e Sistemas (DAS), Universidade Federal de Santa Catarina, Caixa Postal: 476

†Laboratório de Conexionismo e Ciências Cognitivas (L3C), Departamento de Informática e Estatística (INE), Universidade Federal de Santa Catarina, Caixa Postal: 476

Emails: luciene@das.ufsc.br, mauro@inf.ufsc.br, edson@das.ufsc.br

Abstract— This work proposes an evolution of ANN arrangements into a neural architecture biologically inspired, that allows autonomous navigation of mazes for a robotic agent. During the evolution is described the online learning techniques, as well as the performances, advantages and disadvantages of each proposed arrangement. The base construction of the cognitive architecture is based in the theories of cognitive maps and latent learning. Without any priori knowledge, the simulated robot explores completely the maze and then navigates from a start place to a destination place. The cognitive architecture provides that complex behaviors happen concomitantly with the reflexive and reactive behaviors, implemented by AANs. From the results of simulations, it is verified that the best arrangement is composed by a recurrent ART, switching MLP nets with reinforcement learning. This way are obtained desirable performances, supporting the stability-and-plasticity dilemma related to the learning of state-action mappings.

Keywords— Mobile Robot Navigation, ART - Adaptive Resonance Theory, Multi-Layer Perceptron, Cognitive Maps, Reinforcement Learning.

Resumo— Este trabalho propõe uma evolução de arranjos de RNAs dentro de uma arquitetura neural biologicamente inspirada, que permite a navegação autônoma de labirintos por um agente robótico. Durante a evolução são descritas as técnicas de aprendizado em tempo de operação, bem como os desempenhos, vantagens e desvantagens de cada arranjo proposto. A construção base da arquitetura cognitiva se fundamenta nas teorias de mapas cognitivos e aprendizado latente. Sem nenhum conhecimento a priori, o robô simulado explora por completo o labirinto e em seguida navega de um local de origem a um local objetivo. A arquitetura cognitiva provê que comportamentos complexos ocorram concomitantemente com os comportamentos reativos e reflexivos, implementados pelas RNAs. A partir dos resultados das simulações, constata-se que o melhor arranjo é o composto por uma rede ART re-alimentada, comutando redes MLPs com aprendizagem por reforço. Desta forma são obtidos desempenhos desejáveis, atendendo ao dilema da estabilidade e plasticidade relacionado ao aprendizado de mapeamentos de estado-ação.

Palavras-chave— Navegação de Robôs Móveis, rede ART - *Adaptive Resonance Theory*, redes *Multi-Layer Perceptrons*, Mapas Cognitivos, Aprendizagem por Reforço.

1 Introdução

A incerteza dos sensores unida à imprecisão dos atuadores e à dinâmica de ambientes reais fazem do projeto de controladores de robôs móveis um problema difícil. Assim torna-se desejável favorecer os robôs com capacidades de aprendizado onde o mesmo adquira de maneira autônoma seu sistema de controle e adapte seu comportamento a situações nunca experimentadas. Por isto Redes Neurais Artificiais (RNAs) são ferramentas bastante utilizadas na implementação da percepção, navegação e controle de sistemas robóticos, devido às suas capacidades intrínsecas de generalização, tolerância a falhas, paralelismo e seus algoritmos de aprendizado.

Historicamente o grande sucesso das RNAs se deve ao algoritmo de aprendizagem supervisionada, *back-propagation* (Rumelhart et al. 1986). Porém, RNAs com este paradigma de aprendizado são de uso limitado no desenvolvimento de robôs móveis por muitas razões, dentre elas, devido à necessidade de uma fase de treinamento prévia onde, após a aprendizagem, a rede não pode facilmente ser modificada de forma incremental. Por outro lado, o que se espera de um robô móvel autônomo é que ele não necessite ser pré-treinado e sim que descubra seu mundo e se adapte em tempo real (Damper et al. 2000).

Muitos mecanismos neurais são capazes de obter um aprendizado permanente e em tempo de operação onde a maioria das aplicações são voltadas a ambientes dinâmicos e imprevisíveis, assim como os tratados em (Fierro & Lewis 1998, Skoundrianos & Tzafestas 2004, Arleo et al. 2004). Quanto a tarefas de navegação, uma técnica comumente empregada é a Aprendizagem por Reforço (AR) (Sutton & Barto 1998). Ela é uma alternativa promissora, com respaldo em teorias psicológicas de aprendizado (Pearl 2000). Sua função é garantir que um agente aprenda determinado comportamento mediante interações de tentativa-e-erro em ambientes dinâmicos. Exemplos de aplicações mostrando a capacidade de adaptação de técnicas de AR na aprendizagem de caminhos, via interações dinâmicas com ambientes físicos são

vistos em (Tan et al. 2002, Xu et al. 2003, Arleo et al. 2004).

Este trabalho tem por objetivo unir os conceitos clássicos da aprendizagem por reforço com o paradigma conexionista, para implementação de comportamentos reflexivos. Utilizar a importante capacidade de plasticidade e estabilidade de redes ART para o implementação dos comportamentos reativos. Mostrar o passo a passo da evolução dos arranjos de RNAs propostos para este fim. E por fim construir uma arquitetura cognitiva capaz de aprender comportamentos complexos em tempo de operação, permitindo a navegação autônoma e eficiente de robôs móveis em labirintos. Sendo sua construção inspirada no comportamento de ratos que constroem mapas cognitivos biológicos de seus ambientes, com propriedades de aprendizado latente(Tolman 1932).

O restante deste artigo está organizado como segue: a seção 2 apresenta aspectos teóricos gerais dos paradigmas de RNAs utilizados e da técnica de aprendizagem por reforço. A seção 3 mostra o passo a passo da evolução realizada na arquitetura cognitiva de base, apresentando três propostas de arranjos para o controle reflexivo e reativo, seus respectivos desempenhos e fraquezas. A seção 4 descreve as simulações realizadas e os resultados obtidos. Na seção 5 são apresentas as conclusões e direções futuras desta pesquisa.

2 Fundamentação teórica

RNAs são ferramentas promissoras e muito utilizadas na implementação de robôs baseados em comportamentos (Fierro & Lewis 1998, Arleo et al. 1999, Roisenberg et al. 2004, Vieira et al. 2004). Porém, em muitos casos a aprendizagem é feita a priori e posteriormente o agente é colocado no ambiente, como em (Nehmzow & McGonigle 1994). Devido às dinâmicas de ambientes reais, mecanismos capazes de fazer a aprendizagem permanente e em tempo de operação são mais desejáveis. Tanto aqueles controlados por RNAs quanto outros controlados por diferentes abordagens da Inteligência Artificial, como por exemplo, lógica nebulosa (Abreu & Correia 2001).

Um modelo proposto por Carpenter & Grossberg (1988), conhecido como ART (do acrônimo em inglês *Adaptive Resonance Theory*) atende ao requisito da aprendizagem em tempo de operação e principalmente resolve o dilema da *plasticidade* × *estabilidade*. Uma rede ART é apta a manter o equilíbrio entre as propriedades de plasticidade (discriminação) e de estabilidade (generalização). Ou seja, ela é capaz de criar uma nova categoria de padrões, quando estimulada por padrões não reconhecidos e ainda agrupar padrões similares na mesma categoria, quando reconhecidos. Uma regra de similaridade que define onde agrupar os padrões é determinada por um grau de semelhança entre um padrão dado e padrões previamente armazenados. E assim, a rede responde rapidamente a dados aprendidos previamente e ainda é capaz de aprender quando novos dados são apresentados.

São vários os tipos de redes ART que podem ser treinadas segundo o paradigma supervisionado (Carpenter et al. 1991, Carpenter et al. 1992), quando utilizam um agente externo que indica a resposta desejada para o padrão de entrada, ou não-supervisionado (auto-organizável) que é o caso da rede ART1 utilizada como componente da arquitetura cognitiva implementada neste trabalho.

Já a aprendizagem por reforço (AR) é um paradigma de aprendizagem baseado em comportamento, onde as interações entre o aprendiz e seu ambiente procuram alcançar um objetivo específico, apesar da presença de incertezas(Sutton & Barto 1998). O fato de que esta interação é realizada de maneira não supervisionada torna a AR particularmente atrativa para situações dinâmicas. Há duas abordagens de AR: (i) a *abordagem clássica*, onde a aprendizagem ocorre através de um processo de punição e recompensa com o objetivo de alcançar um comportamento global altamente qualificado e a (ii) *abordagem moderna*, que se fundamenta na técnica matemática conhecida como programação dinâmica (Haykin 2001). Neste trabalho, a abordagem clássica de AR é aplicada ao aprendizado on-line das redes MLPs, através das políticas de aprendizados que serão descritas na seção seguinte.

3 A evolução da arquitetura cognitiva

Esta seção apresenta o passo a passo da evolução da arquitetura cognitiva proposta. Como pode ser visto na Fig. 1, as três arquiteturas distintas possuem seus respectivos conjuntos de RNAs que compreendem a porção responsável pela implementação dos comportamentos reflexivos[1] e reativos[2] que equivalem a "evitar obstáculos" e "seguir corredores" respectivamente.

Os módulos fixos. As três arquiteturas apresentadas possuem os seguintes módulos em comum: Percepção, Políticas de Aprendizado, Mapa Cognitivo e Ação. A seguir cada um deles será descrito com maior detalhe.

Percepção. Este módulo recebe os dados dos sensores de proximidade e do ângulo orientado do robô e fornece a informação do estado do ambiente, ou seja, se existem obstáculos nas direções Oeste, Norte, Leste e Sul. Ele também identifica bifurcações e becos sem saídas, acionando o módulo de Mapa Cognitivo.

[1] Comportamentos cuja a ação emergente é função unicamente das entradas sensórias.

[2] Comportamentos cuja a ação emergente depende não apenas das entradas sensórias, mas também do estado interno do agente.

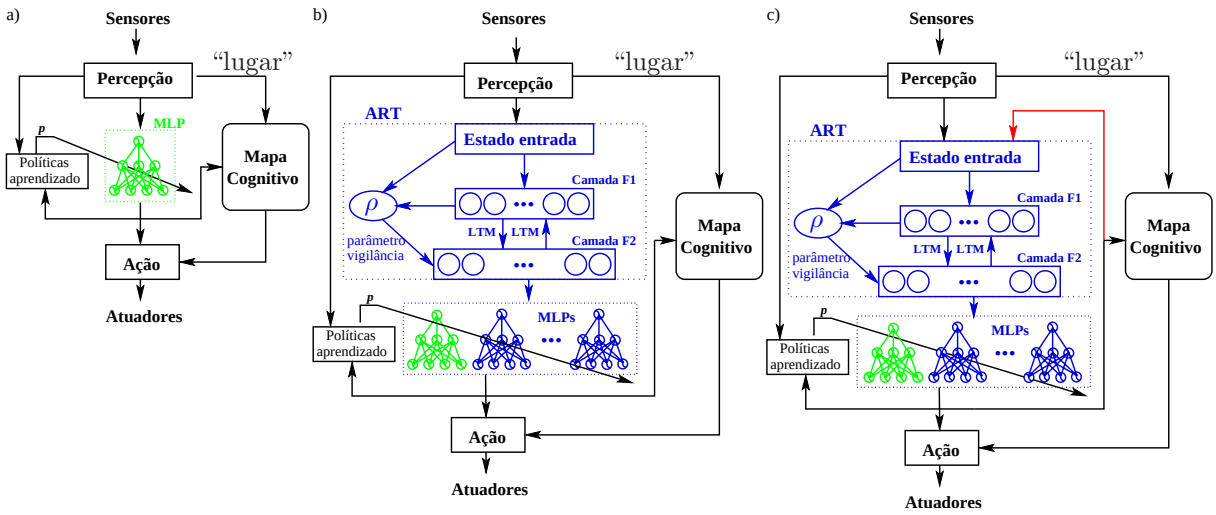

Figura 1: As arquiteturas cognitivas propostas.

Mapa cognitivo. O sistema de Mapa Cognitivo é acionado toda vez que o robô se encontra em um beco sem saída ou em uma bifurcação. Nestes locais o robô associa informação êxtero-cêntrica (do termo em inglês *allothetic*), vinda dos sensores, com informação egocêntrica (do termo em inglês *idiothetic*) que corresponde a sua informação interna de hodometria. Portanto, becos sem saída e bifurcações servem de marcos para a construção do mapa cognitivo do ambiente.

Políticas de aprendizado. A construção do módulo de Políticas de aprendizado é inspirado no modelo clássico da AR, no qual a aprendizagem das redes MLPs acontece através de um processo de punição e recompensa, com o objetivo de alcançar os comportamentos reflexivo e reativo. A seguir um exemplo de uma das políticas de aprendizado:

1. `SE` há obstáculos a Oeste e a Leste `ENTÃO`
   (a) `SE` a saída da rede MLP é igual a Oeste `OU` Leste `ENTÃO`
   /**Comportamento de evitar obstáculos**/
   Punir a rede MLP, alterando um de seus pesos
   (b) `SENÃO SE` a direção da última ação está livre `E` esta ação não foi repetida `ENTÃO`
   /**Comportamento de seguir corredor**/
   Punir a rede MLP, alterando um de seus pesos
   (c) `SENÃO` Recompensar a rede MLP (preservando seus pesos atuais).

Ação. O módulo de ação é responsável por posicionar a frente do robô para a direção (oeste, norte, leste ou sul) correspondente à ação ditada pela rede MLP. Em seguida ele fornece comandos aos atuadores para que o robô execute um passo de controle naquela direção.

A evolução do controle reflexivo/reativo. A seguir será mostrado o passo a passo da evolução dos arranjos de RNAs dentro das porções de controle reflexivo/reativo de cada uma das arquiteturas cognitivas propostas.

1^a *proposta: uma rede MLP.* A primeira arquitetura proposta contém uma única rede MLP com treinamento on-line em sua porção reflexiva/reativa, como ilustrado na Fig. 1 a). A desvantagem deste esquema é que o aprendizado de um estado é totalmente "esquecido" assim que a rede se depara com um novo estado do ambiente, ou seja, ele não satisfaz ao dilema *plasticidade* × *estabilidade*.

2^a *proposta: uma rede ART1 comutadora de redes MLPs.* Neste esquema, como mostrado na Fig. 1 b), uma rede do tipo ART1 desempenha a função de um comutador de redes MLPs, através de seu neurônio vencedor da camada F2. Seu estado de entrada, neste caso, consiste apenas do estado do ambiente. Neste esquema, o dilema da *plasticidade* × *estabilidade* ainda não é atendido sob o seguinte aspecto: sempre que o robô percorrer corredores de mesma direção, o estado do ambiente será igual e assim a mesma MLP será acionada. Porém, como aqui se desconsidera o sentido da direção em que o robô percorre o corredor, a rede MLP sempre sofrerá punições, mesmo tendo aprendido uma ação correta em um instante anterior. De um modo geral, isto torna o aprendizado on-line das redes MLPs custoso, de forma que a porcentagem de punições não decresce.

3^a *proposta: uma rede ART1 recorrente comutadora de redes MLPs.* Com o resultado da 2^a proposta, conclui-se que somente a informação do estado do ambiente é insuficiente para que a rede ART1 satisfaça o dilema da *plasticidade* × *estabilidade* relacionado ao aprendizado on-line das redes MLPs, que implica no decréscimo de punições aplicadas. Para isto se faz necessária a informação adicional da última ação executada pelo robô. Este laço de realimentação, mostrado na Fig. 1 c), permite à rede ART1 uma maior capacidade de discriminação e armazenamento de padrões que se referem a estados do ambiente junto

com o estado interno do robô. A confirmação da vantagem desta proposta sobre as anteriores comprova-se após a estabilidade do aprendizado auto-organizável da rede ART1 ao discriminar um eficiente mapeamento de estado-ação, de maneira que o aprendizado on-line das redes MLPs execute somente "recompensas". Com este arranjo torna-se possível a concepção de dois estados (anterior e atual) que são captados pela rede ART1, permitindo que possam ser tratadas situações aparentemente idênticas (com relação ao ambiente) de maneiras diferentes, apoiando-se na ação tomada no passo anterior.

4 Simulações e Resultados

As implementações das arquiteturas foram feitas no software *WSU Khepera Robot.*[3] Este programa simula um robô Khepera®(Mondada et al. 1993) que conta somente com as informações de seus sensores de proximidade, ângulo orientado e hodometria.

A primeira etapa de testes consistiu da implementação do controle do robô através das três arquiteturas cognitivas propostas (Fig. 1) para então computar a quantidade de punicoes/recompensas aplicadas às redes MLPs, bem como a quantidade de classes criadas pelas redes ART1. Assim verifica-se os desempenhos dos aprendizados por reforço e auto-organizável das respectivas redes. A Tab. 1 mostra os resultados obtidos para cada arquitetura. Na simulação das três propostas o robô primeiramente explora todo o labirinto, realizando a construção de seu mapa cognitivo e então navega entre o lugar Início e o lugar Objetivo. Para esta tarefa são necessários em média 1400 passos (Tab. 1). Constata-se a eficiência da 3^a proposta com relação à quantidade de punições (901) aplicadas às redes MLPs e ao número de classes (29) criadas pela rede ART1, que permitem uma melhor discriminação dos padrões de entrada. Já a Fig. 2 mostra o desempenho do aprendizado por reforço das redes MLPs nos três testes, computando a cada passo a quantidade de punições executadas. Também se comprova a eficiência da 3^a proposta ao verificar-se que ao término da fase de exploração e início da navegação as redes MLPs não mais recebem punições (3^o gráfico da Fig. 2). Nos gráficos da Fig. 3 são mostrados as ativações dos neurônios da camada F2 durante as trajetórias do robô ao realizar a exploração e em seguida a navegação no labirinto (3^a arquitetura proposta). São estes neurônios que acionam as redes MLPs.

Tabela 1: Medidas de desempenho dos aprendizados on-line das redes MLPs e ART1 em cada uma das arquiteturas propostas.

Arquitetura (Fig.1)	Punições	Classes (n^o de MLPs)	Passos (Exploração + Navegação)
1^a proposta	1674	1	1429
2^a proposta	3454	14	1405
3^a proposta	901	29	1386

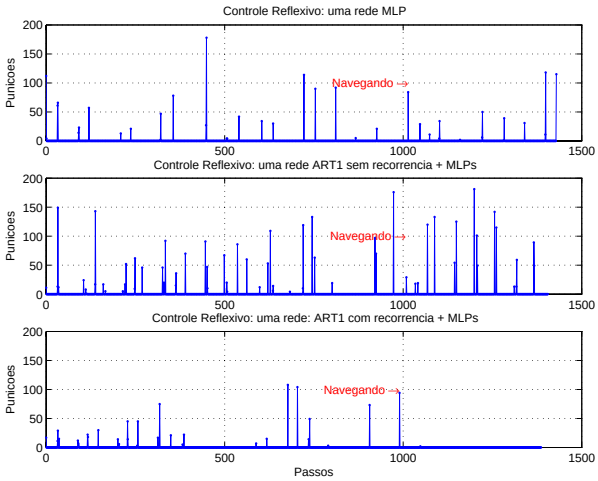

Figura 2: Desempenhos do aprendizado on-line das redes MLPs no controle reflexivo/reativo das arquiteturas propostas.

Na segunda etapa de testes, considerou-se a 3^a arquitetura proposta com modificações no valor do parâmetro de vigilância, ρ, das respectivas redes ART1. A partir daí foram testadas 5 arquiteturas, considerando-se os valores de ρ iguais a 0,6, 0,7, 0,8, 0,9 e 1,0 respectivamente. Para cada arquitetura foram computados a quantidade de classes criadas e a porcentagem de ações corretas executadas pelo controle reflexivo/reativo (MLPs), durante 8000 passos. Portanto, para um dado conjunto de padrões a serem classificados, um valor elevado de ρ irá resultar em uma discriminação mais refinada entre classes do que se tivesse um valor mais baixo. Isto se comprova no primeiro gráfico da Fig. 4, onde o menor número de classes (20) é obtido com $\rho = 0,6$ e o maior (31) com $\rho = 0,9$ e $\rho = 1,0$. Conseqüentemente quanto menor o parâmetro de vigilância da rede ART1, menos eficiente será a aprendizagem por reforço das redes MLPs. Dado que baixos ρ's ocasionam maior ocorrência de punições, isto faz com que as curvas da porcentagem de acertos tornem-se mais acidentadas, conforme pode ser visualizado no segundo gráfico da Fig. 4.

[3]O simulador *WSU Khepera Robot* foi desenvolvido pela *Wright State University* e *Ohio Board of Regents*. Seu uso é dirigido pela Licença Pública da KSIM versão 1.0. O código fonte, documentação, e o texto da licença encontram-se juntos com esta versão do programa. Na falta desta, uma distribuição completa pode ser obtida em: http://gozer.cs.wright.edu/ksim/ksim/html. A versão 7.2 foi liberada em 6 de julho de 2004

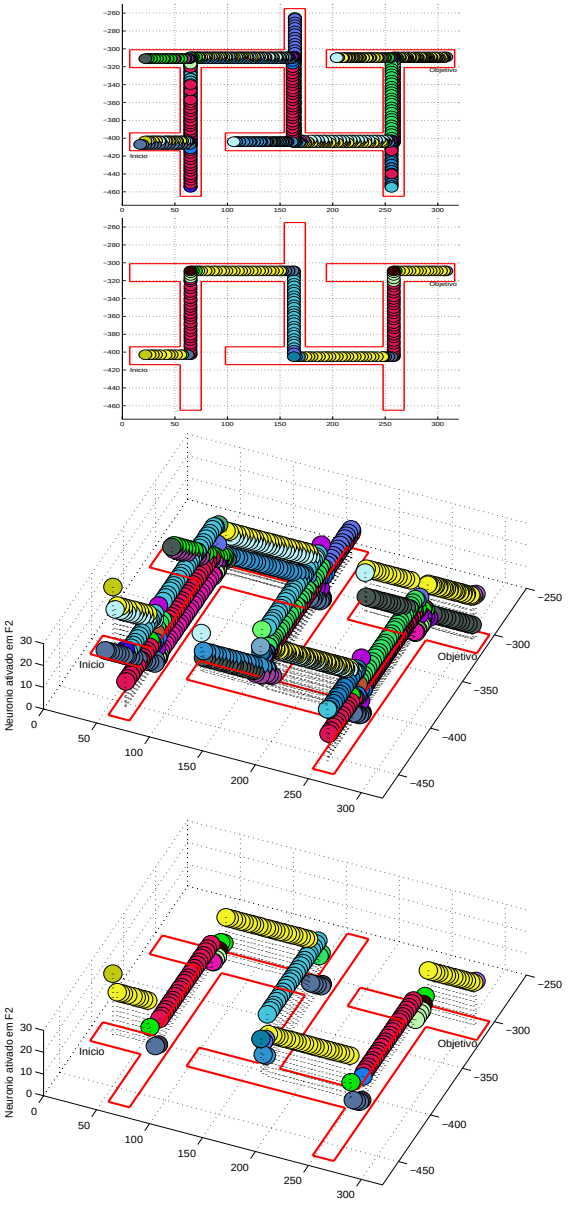

Figura 3: Neurônios ativados na camada F2(rede ART1) da 3^a arquitetura proposta, visualizados no plano e no espaço.

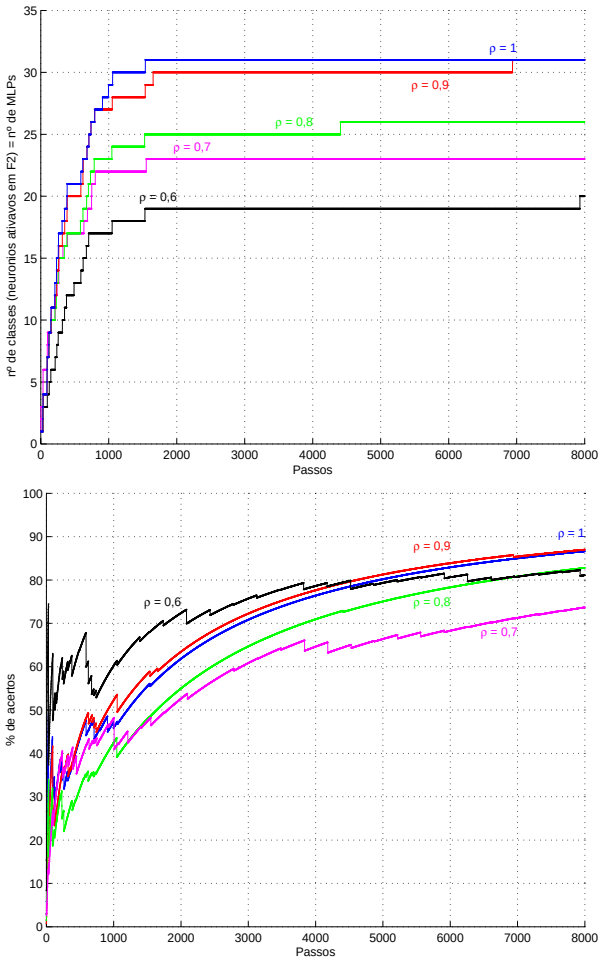

Figura 4: N^o de classes criadas pelas redes ART1 (1^o gráfico) e curvas da porcentagem de acertos das redes MLPs (2^o gráfico), variando-se os parâmetros de vigilância, ρ.

5 Conclusões e Trabalhos Futuros

Este trabalho apresentou uma evolução de arranjos de redes neurais do tipo ART1 e MLPs inseridas na porção de controle reflexivo/reativo de uma arquitetura cognitiva biologicamente inspirada, a fim de trazer conceitos clássicos da aprendizagem por reforço à luz do paradigma conexionista. Foram descritas as técnicas de aprendizado em tempo de operação e seus respectivos desempenhos com relação ao número de punições aplicadas às redes MLPs. Com os resultados obtidos observou-se que a primeira arquitetura proposta (uma única rede MLP utilizada no controle reflexivo/reativo) é incapaz de produzir eficientes mapeamentos de estado-ação devido às suas características intrínsecas de plasticidade e estabilidade. Isto determinou a criação da segunda proposta (uma rede ART1 comutando redes MLPs), a qual possui a capacidade de armazenar novos padrões sem "esquecer" os já aprendidos. Porém, apenas com a informação do estado do ambiente, a rede ART1 tornou-se incapaz de realizar uma correta discriminação dos padrões de entrada, ocasionando sempre punições ao relacionar pares de estado-ação. Portanto o melhor arranjo arquitetural correspondeu àquele composto por uma rede ART1 re-alimentada comutando redes MLPs com treinamento on-line por reforço. Desta forma, a arquitetura cognitiva como um todo é mais eficiente na classificação de padrões que reúnem o estado do ambiente e o estado interno do robô. Isto faz com que as redes MLPs sofram o menor número de punições possíveis, de forma a acelerar o processo do aprendizado em tempo de operação dos comportamentos reflexivos e reativos.

Esta arquitetura também abre perspectivas interessantes no desenvolvimento de sistemas computacionais de navegação biologicamente inspirados e trabalhos futuros serão dirigidos na imple-

mentação desta arquitetura cognitiva para a navegação autônoma de robôs móveis reais em ambientes dinâmicos.

Referências

Abreu, A. & Correia, L. (2001). A fuzzy behavior-based architecture for decision control in autonomous vehicles, *Proceedings of the 2001 IEEE International Symposium on Intelligent Control - México* pp. 370–375.

Arleo, A., del R. Millán, J. & Floreano, D. (1999). Efficient learning of variable-resolution cognitive maps for autonomous indoor navigation, *IEEE Transactions on Robotics & Automation* **15**(6): 990–1000.

Arleo, A., Smeraldi, F. & Gerstner, W. (2004). Cognitive navigation based on nonuniform gabor space sampling, unsupervised growing networks, and reinforcement learning, *IEEE Transactions On Neural Networks* **15**(3): 639–652.

Carpenter, G. A., Grossber, S. & Reynolds, J. (1991). ARTMAP: A self-organizing neural network architecture for fast supervised learning and pattern recognition, *IEEE Conference - Neural Networks for Ocean Engineering,* pp. 863–868.

Carpenter, G. A. & Grossberg, S. (1988). The ART of adaptive pattern recognition self-organizing by a neural network, *IEEE Computer* **21**(3): 77–88.

Carpenter, G. A., Grossberg, S., Markuzon, N., Reynolds, J. H. & Rosen, D. B. (1992). Fuzzy artmap: A neural network architecture for incremental supervised learning of analog multidimensional maps, *IEE Transactions on Neural Networks* **3**(5): 698–713.

Damper, R. I., French, R. L. B. & Scutt, T. W. (2000). Arbib: An autonomous robot based on inspirations from biology, *Robotics and Automonous Systems* **1**(31): 247–274.

Fierro, R. & Lewis, F. L. (1998). Control of a nonholonomic mobile robot using neural networks, *IEEE Transactions on Neural Networks* **9**(4): 589–600.

Haykin, S. (2001). *Redes neurais: princípios e práticas*, 2 edn, Bookman®.

Mondada, F., E.Franzi & Lenne, P. (1993). Mobile robot miniaturization: a tool for investigation in control algorithms, *Int. Symposium on Experimental Robotics. Kyoto, Japan* .

Nehmzow, U. & McGonigle, B. (1994). Achieving rapid adaptations in robots by means of external tuition, *in: D.T. Cliff, P. Husbands, J.-A. Meyer, S. W. Wilson (Eds.), From Animals to Animats 3: Proceedings of the Adaptive Behavior, MIT Press, Cambridge, MA* pp. 301–308.

Pearl, J. (2000). *Causality*, Cambridge University Press, New York.

Roisenberg, M., Barreto, J. M., de Almeida Silva, F., Vieira, R. C. & Coelho, D. K. (2004). Pyramidnet: A modular and hierarchical neural network architecture for behavior based robotics, *ISRA - International Symposium on Robotics and Automation*, Querétaro, México, pp. 32–37.

Rumelhart, D. E., Hinton, G. E. & Williams, R. (1986). *Learning representations by back-propagation errors*, Nature 323.

Skoundrianos, E. N. & Tzafestas, S. G. (2004). Fault diagnosis on the wheels of a mobile robot using local model neural networks, *IEEE Robotics & Automation Magazine* pp. 83–90.

Sutton, R. S. & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*, MIT Press, Cambridge, MA.

Tan, K. C., Tan, K. K., Lee, T. H., Zhao, S. & Chen, Y. J. (2002). Autonomous robot navigation based on fuzzy sensor fusion and reinforcement learning, *Proceedings of the 2002 IEEE International Symposium on Intelligent Control. Vancouver, Canada* .

Tolman, E. C. (1932). Cognitive maps in rats and men, *Psychol. Rev.* **55**: 189–208.

Vieira, R. C., Roisenberg, M. & Furtado, O. J. (2004). Formal languages aspects as a tool for representation and implementation of behavior-based robotics., *In: IEEE Conference on Robotics, Automation and Mechatronics* pp. 959–963.

Xu, X., Wang, X.-N. & He, H.-G. (2003). A self-learning reactive navigation, *Proceedings of the Second International Conference on Machine Learning and Cybernetics* pp. 2384–2388.