Tese apresentada à Pró-Reitoria de Pós-Graduação e Pesquisa do Instituto Tecnológico de Aeronáutica, como parte dos requisitos para obtenção do título de Mestre em Ciências no Programa de Pós-Graduação em Engenharia Eletrônica e Computação, Área de Informática.

Fernanda Monteiro Eliott

APRENDIZADO AUTÔNOMO PARA ROBÔS MÓVEIS BASEADO EM EMOÇÕES ARTIFICIAIS

Tese aprovada em sua versão final pelos abaixo assinados:

Prof. Dr. Carlos Henrique Costa Ribeiro
Orientador

Prof. Dr. Celso Massaki Hirata
Pró-Reitor de Pós-Graduação e Pesquisa

Campo Montenegro
São José dos Campos, SP – Brasil
2010

Dados Internacionais de Catalogação-na-Publicação (CIP)
Divisão de Informação e Documentação

Eliott, Fernanda Monteiro
Aprendizado Autônomo para Robôs Móveis Baseado em Emoções Artificiais / Fernanda Monteiro Eliott. São José dos Campos, 2010.
Número de folhas no formato 126f.

Tese de mestrado – Curso de Engenharia Eletrônica e Computação, Área de Informática – Instituto Tecnológico de Aeronáutica, 2010. Orientador: Prof. Dr. Carlos Henrique Costa Ribeiro.

1. Modelos computacionais bioinspirados. 2. Robôs autônomos. 3. Aprendizado por reforço. I. Comando-Geral de Tecnologia Aeroespacial. Instituto Tecnológico de Aeronáutica. Divisão de Ciência da Computação. II.Título

REFERÊNCIA BIBLIOGRÁFICA

ELIOTT, Fernanda Monteiro. **Aprendizado Autônomo para Robôs Móveis Baseado em Emoções Artificiais**. 2010. 126f. Tese de mestrado em Informática – Instituto Tecnológico de Aeronáutica, São José dos Campos.

CESSÃO DE DIREITOS

NOME DO AUTOR: Fernanda Monteiro Eliott
TÍTULO DO TRABALHO: Aprendizado Autônomo para Robôs Móveis Baseado em Emoções Artificiais
TIPO DO TRABALHO/ANO: Tese de Mestrado / 2010

_____

Fernanda Monteiro Eliott
Rua Pirassununga, 141 – Jd. das Indústrias
CEP 12240-160 – São José dos Campos – SP

APRENDIZADO AUTÔNOMO PARA ROBÔS MÓVEIS BASEADO EM EMOÇÕES ARTIFICIAIS

Fernanda Monteiro Eliott

Composição da Banca Examinadora:

Prof. Dr.	Nei Yoshihiro Soma	Presidente - ITA
Prof. Dr.	Carlos Henrique Costa Ribeiro	Orientador - ITA
Prof. Dr.	Takashi Yoneyama	Membro Interno- ITA
Prof. Dr.	Aluizio Fausto Ribeiro Araújo	Membro Externo- UFPE

ITA

Agradecimentos

A ***todos*** os queridos professores do depto de Filosofia da USP, dentre os quais destaco: Pablo Rubén Mariconda, Osvaldo Frota Pessoa Junior e Caetano Ernesto. Aos Professores Drs. Franklin Leopoldo e Silva e Marco Antonio Zingano, cujos ensinamentos deram suporte à seção 2.1 desta dissertação. Aos Professores Drs. João Vergílio Gallerani Cuter e Luiz Henrique Lopes dos Santos, que ajudaram a direcionar o foco da minha pesquisa, além de me apresentarem Wittgenstein.

Ao professor Dr. do IME – USP Ernesto G. Birgin, que leciona admiravelmente.

Em memória do gênio, o professor Dr. Henrique Schützer Del Nero.

Ao depto de Informática do ITA como um todo, ao professor Dr. Luiz Alberto Vieira Dias, cujas aulas de Teste de *Software* influenciaram os testes sobre o código desenvolvido e, também, o texto do Apêndice. Em especial, ao professor Dr. Nei Soma por todo o conhecimento transmitido quando de seu curso CT234 que, para mim, funcionou como um "divisor de águas".

Ao mestre Dr. Carlos Henrique Ribeiro toda a gratidão e carinho pela orientação e admiráveis interesse e competência para lidar com pessoas e objetos multidisciplinares.

À CAPES pelo apoio financeiro.

À minha indescritível e preciosa família.

"Se antes era o Sol que devia mudar, a Terra que devia converter-se em apenas mais um planeta e não centro de um universo de dimensão reduzida, agora é a mente que não mais faz o mundo girar em torno dela, mas gira em torno do cérebro".

Henrique Schützer Del Nero

Resumo

Especialistas da área de neurofisiologia têm proposto a consideração dos sentimentos como parte dos processos cognitivos, e não de uma alma imaterial: tem sido defendido que as emoções não devem mais ser entendidas como opostas às decisões inteligentes, mas sim como parte e elemento decisivo para estas. Consequentemente se tornaram defensáveis a introdução de emoções artificiais no aprendizado de agentes artificiais, bem como a construção de modelos homeostáticos computacionais para estes. Nesta tese são relatados experimentos sobre uma arquitetura de controle baseado em comportamento e fundamentada sobre a simulação de processos hormonais e emocionais. São apresentadas e discutidas a arquitetura e modificações sobre esta, ou seja, a separação, da estrutura de aprendizado baseado em emoções, em diferentes redes neurais artificiais, uma rede para cada emoção. Os resultados mostraram que é razoável considerar modelos computacionais para processos emocionais que possam sustentar seleção de comportamento autônomo inteligente.

Abstract

Consideration of feelings as cognitive processes rather than as an immaterial part of a soul has been proposed by leading experts in neurophysiology. It is advocated that emotions should not be seen anymore as some opposite of intelligent decisions, but rather as part and decisive element for it. It thus became defendable to introduce artificial emotions for behavioural learning in artificial agents and to build computational homeostatic models for these agents. This work reports experiments on a behaviour-based control architecture based on simulated emotions and hormonal processes and presents and discusses a modification to it, namely the separation of the learning structure based on emotions into different artificial neural networks, one for each emotion. Results show that it is feasible to consider computational models of emotional processes that can support intelligent autonomous behaviour selection.

Lista de Ilustrações

Sumário

1 Introdução

O senso comum teve por costume considerar as decisões inteligentes como opostas às emoções e sentimentos. Estes fariam parte de uma alma ou um espírito, enquanto as decisões inteligentes ocorreriam através da razão, física. Porém, o estudo das consequências em pessoas que sofreram lesões no cérebro e que, após tais lesões, passaram a apresentar comportamento diferente e dificuldade para priorizar ações diárias e simples, assim como tomar decisões, resultou em uma defesa completamente diferente por parte de especialistas como (SCHACHTER 1964), (SACKS 1985) e (DAMÁSIO 1999). Tal defesa consiste em categorizar as emoções e sentimentos como cognitivos. Se as emoções e os sentimentos forem concebidos como físicos, sua reprodução (ao contrário do que seria se cridas como espirituais) passa a ser pensável.

De acordo com (PICARD 1995), as emoções, seriam necessárias para o nosso comportamento criativo, mas, principalmente, estudos neurológicos indicariam que a tomada de decisão desprovida de emoção poderia ser tão inadequada quanto a tomada de decisão com emoção exagerada. Dessa forma, desenvolver computadores que tomem decisões inteligentes poderia requerer computadores que tivessem algum tipo de "emoção".

Como poderiam ser usadas emoções artificiais no controle de um agente autônomo? Especificando: como poderiam ser usadas emoções artificiais no controle de um agente autônomo que se adapte ao seu ambiente usando técnicas de aprendizado por reforço? (GADANHO 1999) busca fornecer respostas práticas a tal questão. Tendo por base a objetividade científica, pretende-se, nesta tese, testar e analisar a arquitetura de controle baseado em comportamento de (GADANHO 1999) e propor e analisar uma modificação, aqui indicada, desta mesma arquitetura. Para tanto, a aplicação da arquitetura se dará no contexto de um agente artificial solitário autônomo, cujas motivações se originam em um Sistema

Homeostático. Emoções artificiais foram usadas no contexto deste agente solitário, muito embora, como (GADANHO 1999) ressalta, as conotações sociais associadas às emoções se enderecem também a robôs sociais. Para o teste e a modificação, objetivos principais desta tese, será seguida uma visão realista, ou seja, serão assumidas as premissas de que a exterioridade existe e que é possível percebê-la, que as sensações transmitem algo dessa exterioridade, e que as emoções são processos cognitivos (DAMÁSIO 1994) e não parte de uma alma imaterial e, por isso, passíveis de reprodução e modelamento artificial.

1.1 Autonomia e Robótica

Segundo (RUEBENSTRUNK 1998), é grande a influência dos princípios de (SIMON 1967; SLOMAN 1987; TODA 1993) sobre o desenvolvimento de agentes artificiais autônomos e, a despeito da multiplicidade das pesquisas neste campo, seria comum o conceito da autoridade das emoções sobre o controle, ou seja, sobre o como estas poderiam, em relação a agentes (robôs móveis autônomos) que têm uma tarefa a cumprir em um ambiente incerto, aperfeiçoar decisões autônomas, mecanismos de atenção, e estabelecer ações prioritárias. Para (GADANHO 1999), se se considerarem as emoções como essenciais para a razão humana, já haverá a alusão à importância daquelas para o atingir a automotivação necessária para sustentar a autonomia.

O conceito de autonomia no contexto humano, é definido por (CORREA, SOUSA 2002) como relacionado à idéia de autogoverno, tornando possível aos indivíduos regularem suas condutas a partir de regras próprias; além disso, (MACEDO 1991) ressalta que o conceito de autonomia pressupõe auto-organização e identidade própria; por fim, para (HEGEL 1807), a autonomia de um sujeito depende também das atitudes que os outros tomam em relação a esse sujeito. Segundo (GADANHO 1999), no campo da Robótica, autonomia

assume vários significados: desde automaticidade, o funcionar sem intervenção (geralmente o sistema é dito autônomo quando é capaz de concluir sua tarefa sem intervenção, seja esta humana ou de qualquer outro sistema (YAVNAI 1989)), a autosuficiência (o robô é capaz de se recarregar sem qualquer assistência) e a, inclusive, automotivação (definir as regras que governam seu próprio comportamento). A definição de autonomia adotada em (GADANHO 1999), e de interesse especial para esta tese, é a de autogoverno, ou seja, fazer e seguir as próprias regras.

A autonomia tem aplicações interessantes em diversas áreas, como em (McFARLAND 1994) que, após observar estados cognitivos de animais sob condições ambientais normais, modelou o comportamento animal utilizando robôs, e aplicou tais observações na definição de um conjunto de ações combinadas para garantir uma ação intencional; ou, também, em sistemas híbridos, no campo de Teoria de Agentes, que combinam abordagens deliberativas e reativas (FERGUSON 1992).

Ao buscar fundamentar um comportamento autônomo para a arquitetura e a tarefa do agente, (GADANHO 1999) considerou algumas capacidades comportamentais extraídas do comportamento animal por (HALLAM; HAYES 1992). São elas:

- A primeira diz respeito à Homeostase, termo criado por Walter Cannon (BAYLISS, 1966) e que versa sobre o equilíbrio dinâmico de um sistema aberto - de acordo com (PUTTINI; JÚNIOR 2007), (BERNARD 1865) teria influenciado a criação do conceito Homeostase através da sua pesquisa sobre os mecanismos de regulação orgânica, em que defenderia a dependência da vida em relação à constância do ambiente interno. No contexto de um agente artificial, este teria objetivos homeostáticos – o robô deve ter algumas variáveis internas e mantê-las dentro de um limite. Por exemplo: o nível de energia (Seção 2.3.1) pode ser utilizado como base para motivação interna; assim, o equilíbrio dinâmico do agente em relação ao

ambiente se daria através da manutenção, pelo agente, do nível de energia dentro de um intervalo pré-estabelecido.

- Percepção – o robô deve possuir sensores variados e capacidade de extrair a informação subjacente aos dados captados por cada sensor, sendo particularmente um desafio para o robô autônomo o lidar com uma percepção fecunda, o que deve ocorrer através de mecanismos rápidos e eficientes de seleção e fusão sensorial.
- Movimento – o robô deve ser capaz de se mover competentemente em seu ambiente e apresentar ações elaboradas, tais como mover objetos, além de ter um repertório de movimentos que lhe possibilitem flexibilidade de escolha.
- Reação e aprendizado – o robô deve ser capaz de exibir reações rápidas a alguns dos estímulos providos por seu ambiente e, associadamente, aprender a relevância de tais estímulos. O agente deve ter capacidades adaptativas sem, no entanto, prejudicar seu desempenho.
- Navegação – embora não seja essencial, é adequado que o robô tenha uma referência, um ponto de retorno, pois a habilidade de retornar aos pontos referenciais pode permitir comportamentos mais complexos.

Em (GADANHO 1999), devido à dificuldade em se preverem todos os cenários com os quais o robô irá se defrontar, a autonomia poderia auxiliá-lo a lidar com situações inesperadas; assim, o modo usado para atingir autonomia em um agente se deu através de um módulo de aprendizado que aperfeiçoa o seu desempenho por aprendizado por reforço ao interagir com o seu ambiente. Em (GADANHO 1999) não seria suficiente como uma capacidade de aprendizado para um agente autônomo (no sentido de autogoverno), o detectar regularidades em seu ambiente através de auto-organização, sendo também necessários um tipo de motivação interna, a fim de decidir o que fazer (o Sistema Homeostático se direciona a

uma tal função), e o possuir algum mecanismo para estabelecer objetivos sem qualquer ajuda. Tais necessidades, por sua vez, conduzem a uma outra: à da possessão de mecanismos internos que permitirão a determinação das características cruciais à sua interação com o ambiente e, também, as conotações positivas e/ou negativas associadas.

O Sistema Homeostático projetado e desenvolvido em (GADANHO 1999) apresenta redes neurais recorrentes (com realimentação), em que emoção e percepção influenciam um ao outro. Através desta influência é alcançada a persistência de estados emocionais. O modelo supriria o agente com estados emocionais coerentes com sua interação contextual com o ambiente, ao atribuir valores às características relevantes de sua interação. Os estados emocionais do robô considerados neste modelo são: Felicidade, Medo, Tristeza e Raiva. Com o objetivo de mostrar se há vantagem no modelo desenvolvido, ou seja, em ter emoções representando o papel de automotivação em um robô autônomo, (GADANHO 1999) realizou os seus experimentos em um robô móvel Khepera (MONDADA 1994), e os resultados teriam mostrado que emoções artificiais podem ser usadas com sucesso como fonte de reforço se a arquitetura de controle for selecionada adequadamente.

1.2 Objetivos

Desde que o foco geral da pesquisa em (GADANHO 1999) foi, através de sua arquitetura, investigar como emoções artificiais podem auxiliar no controle de um agente solitário autônomo que se adapta ao seu ambiente usando técnicas de Aprendizado por Reforço (reforços estes fornecidos através de seu próprio Sistema Homeostático) e que, como resultado dos reforços e em resposta ao seu ambiente, decide qual comportamento executar, o objetivo deste trabalho será reproduzir e propor modificações (o aprendizado ocorrerá de forma especializada para cada possível emoção artificial) sobre tal modelo computacional e realizar as análises experimentais pertinentes. Por fim, tem-se também por objetivo adquirir o conhecimento necessário para, em trabalhos futuros (Seção 5.1.2), desenvolver e testar uma arquitetura direcionada à simulação social.

1.3 Estrutura Geral da Tese

Esta tese está organizada da seguinte forma: o Capítulo 2 introduz brevemente as sensações segundo uma concepção filosófica, e as emoções segundo (DAMÁSIO 1994). São apresentados alguns modelos artificiais da literatura dotados de sistema homeostático, detalhando-se, especificamente, o de (GADANHO 1999). O Capítulo 3 percorre passo a passo a Arquitetura de Controle Baseado em Comportamento de (GADANHO 1999), que apresenta como um de seus componentes principais o Sistema Homeostático explicado no Capítulo 2; ao final, são propostas modificações sobre esta arquitetura. O Capítulo 4 apresenta os resultados e comparações obtidos após a sujeição destas arquiteturas, tanto a original quanto a modificada, em uma tarefa de navegação robótica. Finalmente, o Capítulo 5 apresenta as conclusões finais e propostas para trabalhos futuros.

2 Modelos Artificiais para Sistemas Baseados em Emoção

2.1 Introdução

O que os sentidos nos fornecem seria uma representação da realidade, de objetos reais, ou elaboração epistemológica? Segundo (HAACK 1978), o senso comum, se valendo do bom senso, crê na existência e independência do mundo exterior; o bom senso pode coincidir com as conclusões científicas a respeito de determinado objeto de investigação, como no caso da independência do mundo exterior, para uma visão realista. Assim, o realismo científico assenta na tese ontológica de que o mundo exterior existe, é independente de nós, e podemos conhecê-lo epistemicamente; esta possibilidade de conhecimento do mundo permite que se veja a ciência como uma busca por informações verdadeiras acerca da estrutura do mundo - como as teorias científicas se aproximariam da verdade, o valor de verdade das proposições dependeria de sua correspondência com o mundo.

Um pressuposto geralmente embutido nos modelos computacionais que simulam processos homeostáticos, e na ciência em geral, é o da existência do mundo exterior e a crença do especulador como descobridor e não como criador (COSTA, N. 1997). As sensações são fundamentais nesse sentido, e têm sido objeto em diversas correntes filosóficas, seja tendo no especulador um descobridor, seja um criador.

Uma das visões mais inquietantes sobre a autoridade das sensações talvez seja a de (BERKELEY 1710), em que *realidade* é realidade percebida, inexistindo realidade externa. Ainda segundo Berkeley, o dado sensível seria uma realidade dada ao sujeito que percebe as coisas, pois o ato sensível seria uma ideia na mente do sujeito, mas, ao mesmo tempo, ideia seria também o objeto sensível. Sendo assim, as percepções seriam imanentes, pois, se não as concebêssemos como pertencentes ao espírito, não poderíamos concebê-las de modo algum, porque a ideia não existiria fora do espírito, e a existência de uma ideia consistiria em ser

percebida. Ter ideia sem ter percepção seria incongruente, visto que uma é a outra e as ideias, sendo todas definidas pela percepção, estariam separadas do sujeito mesmo estando dentro dele. Essa separação seria tão manifesta, que a interpretaríamos como exterioridade.

Desde os filósofos mais antigos já havia o interesse pelas sensações e a reflexão acerca da sua relação com a realidade. Aristóteles se direcionou a tal questão sob uma perspectiva realista: em sua obra *De Anima,* as sensações ligam cada um ao que é externo ao seu corpo; assim, quando sentimos algo, este algo foi causado por um sensível (aquele percebido pelos nossos sentidos) específico, significando que as sensações nos transmitem apenas os particulares. Mas seria a partir destes que daríamos conteúdo às nossas memórias, estas à experiência e, com esta, estaríamos aptos a adquirir o conhecimento universal, que é o conhecimento científico. Portanto, segundo a argumentação no *De Anima*, as sensações estão extremamente vinculadas à intelecção.

Além de refletir sobre o vínculo entre sensação e intelecção, também é interessante considerar o ente (aquilo o que é) como matéria e a sua forma específica e, além disso, as categorias em que se podem colocar os sensíveis. Aristóteles, no *De Anima*, coloca o sensível como um ente, ou seja, matéria e forma, que tanto pode ser natural: árvore, criança..., como não-natural: cadeira, mesa, etc. Além disso, categoriza o sensível em três modos: o comum, percebido por qualquer dos órgãos sensoriais por não ser próprio a nenhum órgão sensorial específico (movimento, tamanho, unidade, número, repouso, figura); o próprio (aqueles que, exceto por acidente, têm um dos órgãos sensoriais próprio para processar sua percepção, e não há engano quanto à sensação propriamente dita, mas, por exemplo, “quanto ao que é ou onde está o colorido, ou quanto a o que é ou onde está aquilo que produziu o ruído”); e os sensíveis segundo acidente, que ocorrem quando percebo uma característica própria não pelo seu órgão sensorial próprio, ou seja, quando correlaciono sensações para um mesmo ente, por exemplo:

“a sensação de que a bile é amarga e amarela”. Mas os próprios são os preponderantemente sensíveis, “bem como aquilo em relação a que naturalmente é a essência de cada sensação”.

Ao simular sensações em um agente artificial reproduzindo a autoridade que as sensações parecem ter sobre nós, os resultados que se obtêm estão sob a premissa de tal reprodução. Na teoria aristotélica o especulador também é descobridor, tomando como prova as próprias sensações para se basear na convicção de que o que apreendemos através dos sentidos realmente existe. Os sentidos nos transmitem a realidade, por isso o mundo é real, e as sensações são a evidência. Já sobre as emoções, o enfoque de Aristóteles é voltado para a ética; na obra *Ética a Nicômaco,* defende que as emoções são o que altera os nossos juízos: na alma existem as paixões (apetites, audácia, inveja...), as faculdades (coisas em virtude das quais se diz que somos capazes de sentir as paixões) e as disposições de caráter (nossa posição boa ou má com relação às paixões). A virtude seria uma disposição de caráter relacionada com a escolha e consistente em uma mediania relativa a nós e, dentre todas as coisas, deveríamos nos prevenir contra o agradável e o prazer, pois não poderíamos julgá-los com imparcialidade.

2.1.1 Mente Material

Se todas as partes do cérebro forem perfeitamente substituídas por réplicas artificiais, ainda assim existirá a mente, a subjetividade? Se se acreditar no Materialismo, na mente como um processo físico, cognitivo, pode-se responder que sim. Já se houver uma premissa dualista, não. Segundo uma visão dualista, mente e cérebro são diferentes e a mente não é material. De acordo com muitos autores, como (DENNETT 1991), por exemplo, um dos legados negativos do dualismo cartesiano seria a crença em uma mente material, mas centralizada, tal qual um palco em que todas as experiências conscientes seriam apresentadas, e na sequência em que teriam ocorrido no indivíduo. Esse resquício cartesiano, ou seja, a centralidade da consciência, sobre a visão materialista é definido, em (DENNETT 1991), como o “teatro

cartesiano": uma metáfora sobre como as experiências conscientes podem se localizar no cérebro.

Segundo (DENNETT 1991) envolvemos a consciência em um mistério desnecessário, por parecer que nossas experiências e pensamentos conscientes não podem ser ocorrências cerebrais, mas algo feito de material diferente, e localizado em um espaço também diferente. A mente consciente pareceria não poder ser o cérebro, porque nada neste poderia ser o recheio em que estariam: as experiências e pensamentos conscientes; o "eu"; a fonte do que faz com que eu me importe com certas coisas; exercer o papel de responsabilidade moral. (DENNETT 1991) explicita uma objeção comum ao dualismo: mente e corpo devem interagir, mesmo sendo de natureza distinta: os órgãos sensoriais, através do cérebro, informam a mente que, por sua vez, age sobre o corpo.

2.1.2 Emoção: Elemento da Cognição

Se se pensa em algo como imaterial, a reprodução física *desse* algo se torna inconcebível. Por exemplo, se se está sob a premissa de uma consciência imaterial, seria razoável tentar modelá-la e testá-la? Porém, se se considera algo como físico, uma possível reprodução desse algo pode ser complexa, mas teoricamente factível. Houve o costume de se pensarem as emoções como algo alheio, separado ou até mesmo inverso às decisões racionais; entretanto, pesquisas em diversas áreas do conhecimento, como de (SCHACHTER, 1964) e as do neurofisiologista António Damásio (DAMÁSIO 1994), permitiram ponderações acerca das emoções como parte dos processos racionais de tomada de decisão. Ao se considerarem as emoções como elementos da cognição e não mais como de uma alma imaterial, tornaram-se defensáveis a inclusão de sistemas homeostáticos como mecanismos de atenção em agentes artificiais.

A consciência no contexto da Robótica, assim como as emoções, desperta polêmicas, recebe diferentes definições e é objeto de estudo em diversas áreas. Sob uma visão

simplificada de consciência (consistindo na percepção do que ocorre no mundo), o advento desta em um agente artificial se daria a partir de suas capacidades de possuir um mapa interno do mundo e, através de seus sensores e atuadores, construir modelos de objetos do mundo e posicioná-los no mapa, sendo o agente apto a se reconhecer como entidade do mundo e posicionar-se nele (GUDWIN 2005).

2.1.3 Hipótese do Marcador Somático de (DAMÁSIO 1994)

As emoções auxiliariam nossos processos de tomada de decisão diminuindo gasto computacional e de tempo: através delas filtramos a abundância de dados dos nossos ambientes e focamos naqueles mais relevantes; assim não gastamos tempo analisando cada um dos aspectos do nosso ambiente, mas sim considerando apenas aqueles em que os nossos mecanismos de atenção estão focados. Através das emoções também conseguimos tomar decisões e organizar, hierarquizar as nossas ações como, por exemplo, amarrar o cadarço do sapato, pegar as chaves do carro e sair para trabalhar – vide (DAMÁSIO 1994) para detalhes sobre dificuldades concernentes ao planejamento e a associação entre deficiência na tomada de decisão e danos sobre emoções e sentimentos em pessoas que sofreram lesões nas áreas pré-frontais do cérebro.

As arquiteturas baseadas em processos homeostáticos têm sido influenciadas por (DAMÁSIO 1994). Segundo este, todas as emoções gerariam sentimentos, mas nem todo sentimento geraria emoção. Emoção seria a combinação de um processo avaliatório mental, com respostas a esse processo, em sua maioria dirigidas ao corpo, resultando em um estado emocional do corpo, mas também dirigidas ao cérebro, resultando em outras alterações mentais. Os sentimentos seriam o processo de monitoramento do corpo e nos ofereceriam a cognição dos nossos estados viscerais e músculo-esqueléticos: a cognição seria a experiência dessas mudanças, a percepção de todas as mudanças que constituiriam a resposta emocional. A nossa atenção convergiria substancialmente para os sinais do corpo quando houvesse

sentimento associado a emoções (havendo partes dele passando de segundo para primeiro plano de nossa atenção).

Segundo a hipótese do Marcador Somático de (DAMÁSIO 1994), os Marcadores Somáticos seriam adquiridos por meio da experiência e tanto sob o controle de um sistema interno de preferências quanto sob influências de um conjunto externo de circunstâncias - entidades e fenômenos com os quais o organismo tem de interagir, e convenções sociais e regras éticas. Os Marcadores Somáticos não precisariam ser pensados como se fossem sentimentos, embora ressaltassem de modo camuflado e através de um mecanismo de atenção, determinados componentes em detrimento de outros e controlassem sinais essenciais a alguns aspectos do processo de decisão e planejamento.

Os sentimentos de fundo (DAMÁSIO 1994) corresponderiam aos estados do corpo intermediários às emoções e seriam provavelmente os de maior ocorrência ao longo da vida. (DAMÁSIO 1994) defende que os sentimentos de fundo seriam o âmago de nossa representação do 'eu', teriam origem em estados corporais de fundo e não em estados emocionais, e, mesmo podendo ser agradáveis ou desagradáveis, não seriam demasiado positivos ou negativos.

2.2 Modelos Homeostáticos

O foco desta tese será o modelo de (GADANHO 1999); porém, é relevante conhecer, mesmo que simplificadamente, outros modelos da literatura que também façam uso de sistemas homeostáticos. As emoções artificiais são usadas para direcionar objetivos, estabelecer prioridades e dirigir a atenção de um sistema para um aspecto significativo (interno ou externo). (VELÁSQUEZ 1997) desenvolveu o modelo chamado Cathexis (concentração de energia emocional em um objeto ou ideia), baseando-se na teoria de (MINSKY 1985), em que as emoções são diversos subsistemas; (VELÁSQUEZ 1998)

descreve uma arquitetura de agente que integra emoções, comportamentos e mecanismos, criando um processo tomador de decisão que tem como componente principal o modelamento de alguns dos aspectos das emoções. (CAÑAMERO 1997), também sob influência de (MINSKY 1985), investigou como as emoções podem afetar o comportamento de uma criatura artificial integrada por diversos agentes modelados de acordo com parâmetros psicológicos. Mais recentemente, (MOIOLI et al. 2008) desenvolveu um modelo estendido de um sistema endócrino e uma estrutura de rede neural artificial do tipo GasNet (HUSBANDS et al. 1998), e sob controle evolutivo.

Para uma explanação mais aprofundada sobre os modelos emocionais, vide (RUEBENSTRUNK 1998). A seguir será feita uma breve exposição de alguns dos modelos homeostáticos mais conhecidos na literatura, inclusive o de (GADANHO 1999) – o qual será retomado em detalhes na seção subsequente, visto que a arquitetura completa (detalhada no capítulo 3), da qual o Sistema Homeostático faz parte, será testada e analisada no Capítulo 4.

2.2.1 Modelo de (FOLIOT; MICHEL 1998)

Em (FOLIOT; MICHEL 1998) é assumido que as emoções são a base da cognição porque provêem um modelo funcional; assim, o objetivo é mostrar como estruturas baseadas em emoção podem contribuir para a emergência da cognição ao criar condições de aprendizado razoáveis. A emoção é representada em dois níveis: no do processo, que pode elucidar e diferenciar emoções (avaliação de estímulo); e no do estado, que pode fornecer informações sobre o sistema. Estes dois níveis possuem duas implicações distintas: a avaliação servirá para extrair as informações relevantes que serão consideradas; e, para que ocorra aprendizado da mudança de objetivo, o estado será diretamente vinculado ao processo.

Para explorar essa estrutura baseada em emoção, dois experimentos são descritos: no mais simples, a avaliação é um simples sinal ruim/bom que cria uma associação. Já o segundo experimento, mais complexo, usa sinais de reforço, sensores de proximidade e simulação de

uma câmera colorida para que haja o aprendizado da associação entre os estímulos visuais e os comportamentos adequados. O sinal deste experimento é baseado na teoria de avaliação (*appraisal*) que cria um esquema. O conceito "*appraisal*" é fundamental a algumas teorias sobre as emoções (FRIJDA 1993; LEDOUX 1996; ROSEMAN, ANTONIOU, JOSE 1996; SCHERER 1997; LEÓN, HERNÁNDEZ 1998). Em (FOLIOT; MICHEL 1998) "*appraisal*" é definido como a capacidade de avaliar a relevância de um estímulo ao usar critérios baseados na novidade, no desprazer, na importância de um objetivo, no potencial do controle e na taxa de normalidade - a fim de manter o sentido original dos trabalhos, nos próximos modelos, a palavra "*appraisal"* não será traduzida.

Ainda no segundo experimento, o aprendizado se dá quando ocorre um valor alto de desprazer, produzindo um novo esquema que contém os estímulos mais recentes como entrada dos sensores - o esquema associa entradas sensoriais a respostas motoras. É feita então uma análise para saber qual objetivo está associado ao estímulo, e checar se esse mesmo objetivo permite que haja retorno a um estado normal. Se esse estado normal for alcançável dentro de um pequeno limite de tempo, a representação é associada ao esquema, caso contrário, o esquema é destruído.

Através deste experimento, é pretendido mostrar como as emoções podem ser vistas como um sistema avaliatório operando automaticamente tanto no nível perceptual quanto no cognitivo, mensurando eficiência e relevância, sendo pretendido que as informações introduzam mudanças nos mecanismos de atenção e representação. Para (FOLIOT; MICHEL 1998) pareceria que o encapsulamento de ambos estaria na origem das emoções, e que os processos correspondentes estariam continuamente ativos na cognição, ao menos significando que está tudo bem para o agente. E os vários sentimentos emocionais resultariam da intensidade da avaliação do estímulo, ranqueando desde 'sentimento vago' a 'choques emocionais'.

2.2.2 Modelo de Agente Adaptativo de (BOTELHO; COELHO 1998)

(BOTELHO; COELHO 1998) descreve um modelo em que o comportamento adaptativo é alcançado através de aprendizado emocional, sendo este tido como válido para o aprimoramento do comportamento de agentes artificiais autônomos. Como a emoção é um processo que opera no controle da arquitetura, o comportamento se aperfeiçoa à medida que o processo emocional do agente se aperfeiçoa também. A avaliação afetiva é feita por um mecanismo ativo, enquanto a avaliação cognitiva, por um mecanismo cognitivo; esta distinção na arquitetura teria, como pontuada pelos autores, semelhanças com (DAMÁSIO 1994), visto que a amídala tem sido tida como responsável por muitos fenômenos afetivos e, o córtex visual, por manter representações cognitivas do ambiente. O trabalho em (BOTELHO; COELHO 1998) pretende mostrar que novas motivações podem ser criadas como resultado de processos emocionais, e revelar que pelo menos parte da memória episódica de um agente deve ser acessível à emoção e ao aprendizado emocional.

2.2.3 TABASCO de (PETTA; STALLER 1998)

"*Appraisal*" é definida em (PETTA; STALLER 1998) como avaliação contínua das hipóteses ambientais para o agente, sendo elemento constituinte da geração de emoção, fazendo mediação entre eventos e emoções, explicando o porquê de um mesmo evento poder suscitar diferentes emoções, assim como para o entendimento do que diferencia uma emoção de outra. Haveria muitas teorias para especificar quais e quantos seriam os critérios avaliatórios (*appraisal*) minimamente necessários para diferenciação de emoção, mas haveria um grande consenso em relação a esse critério incluindo: "The perception of a change in the environment that captures the subject's attention (novelty and expectancy), the perceived pleasantness or unpleasantness of the stimulus or event (valence), the importance of the stimulus or event to one's goal or concerns (relevance and goal conduciveness or motive consistency), the notion of who or what caused the event (agency or responsibility), the estimated ability to deal with the event and its consequences (perceived control,

power or coping potential), and the evaluation of one's own actions in relation to moral standards or social norms (legitimacy), and one's self-ideal" (REEKUM, SCHERER 1997), citado em (PETTA; STALLER 1998).

TABASCO é a arquitetura de (PETTA; STALLER 1998) para agentes em que as emoções são modeladas como um processo adaptativo relacionado à interação agente – ambiente. Em TABASCO é modelado um processo de *appraisal*, geração de tendências de ação e de seguir padrões. Os dois maiores componentes desta arquitetura de agente são: Percepção e *Appraisal* (*Perception and Appraisal*); e Ação. Na Figura 2.1 encontra-se um esquema geral da arquitetura TABASCO. O módulo chamado *Perception and Appraisal* processa os estímulos ambientais e modela o processo avaliatório (*appraisal)*. O módulo *Appraisal Register* faz uma intermediação entre o *Perception and Appraisal* e os componentes de *Action,* detectando e combinando as saídas avaliatórias das três camadas (*sensory, schematic, conceptual*) do primeiro e influenciando o segundo, baseando-se no estado avaliativo do mundo. O *Action Monitoring* monitora o planejamento e processos de execução no módulo *Action* enviando os resultados para o módulo *Perception and Appraisal*, onde são integrados ao processo *appraisal*. O módulo *Action* contém os comandos dos motores para habilidades reativas e as tendências de ação, enquanto as entradas sensoriais são processadas no componente *Perception and Appraisal*.

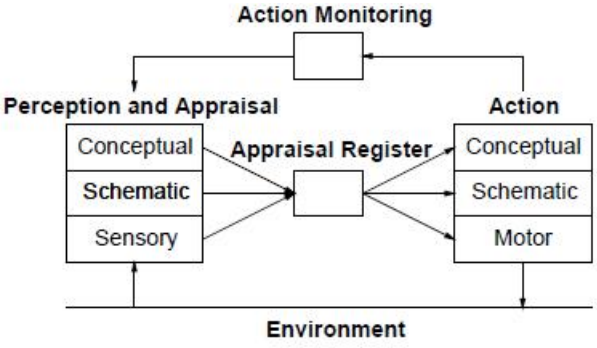


FIGURA 2.1 Arquitetura TABASCO, retirada de (PETTA; STALLER 1998).

2.2.4 Modelo de (GADANHO 1999) e Desdobramentos

Na arquitetura apresentada em (GADANHO 1999) foram adotadas as concepções de (DAMÁSIO 1994) acerca de sentimento e emoção, assim como a da hipótese do Marcador Somático. Os Marcadores Somáticos (DAMÁSIO 1994) nos ajudariam a tomar decisões rápidas sem que gastássemos tempo e grande capacidade de processamento. Segundo (GADANHO 1999), uma dificuldade crucial para o aprendizado de agentes artificiais seria a determinação de quando o controle deve reavaliar sua decisão relacionada ao comportamento que deve ativar; assim, em (GADANHO 1999) emoções artificiais foram usadas para tal determinação e consideradas úteis neste aspecto, pois teriam possibilitado um melhor desempenho com menos esforço computacional do que a melhor combinação de mecanismos de interrupção que utilizava intervalos de tempo.

Três comportamentos foram selecionados em (GADANHO 1999), sendo que o agente deve aprender a coordená-los de modo que os processos homeostáticos sejam úteis quando da determinação da função de recompensa; do tempo gasto para aprendizado da tarefa; e de quando reavaliar o comportamento escolhido em resposta ao estado do mundo. Os comportamentos selecionados em (GADANHO 1999), e também utilizados nos experimentos, são: Evite Obstáculos; Busque por Luz; Siga Paredes.

2.2.4.1 Arquitetura ALEC

Um dos desdobramentos da Arquitetura Baseada em Emoção de [(GADANHO 1999), (GADANHO; HALLAM 2001 b), (GADANHO; CUSTÓDIO 2002 a)] foi a adição de um Sistema Cognitivo que complementaria as capacidades de adaptação baseadas em emoção através de regras de conhecimento explícitas e extraídas da interação do agente com o seu ambiente. Segundo (GADANHO; CUSTÓDIO 2002 b), estes dois mecanismos de aprendizado, um correlacionado ao sistema cognitivo e o outro às emoções, teriam

contribuído positivamente para o desempenho de aprendizado do agente. Tal arquitetura, representada na Figura 2.2, foi denominada ALEC (*Asyncronous Learning by Emotion and Cognition*) e é baseada no modelo Clarion (SUN; PETERSON 1998).

A adição do Sistema Cognitivo teria permitido um aumento significativo na velocidade de aprendizado do agente, porém apenas uma pequena melhora em termos de desempenho. A arquitetura ALEC implicaria em que, enquanto as associações emocionais seriam mais poderosas quanto à capacidade de cobrir estados, seriam pobres em poder explanatório e poderiam introduzir erros de supergeneralização. Já o conhecimento cognitivo, seria restrito ao aprendizado sobre relações de causalidade simples e de curto prazo: as informações cognitivas seriam as mais precisas, mas devido à impossibilidade de consultar todos os eventos que o agente experiencia, seriam selecionadas apenas as que parecem mais importantes - o que, segundo (GADANHO 2003) seria consistente com a ideia defendida por (CYTOWIC 1993), de que as emoções dariam um sentido intuitivo do que é correto enquanto a cognição construiria um modelo da realidade que permitiria ao indivíduo analisar decisões.

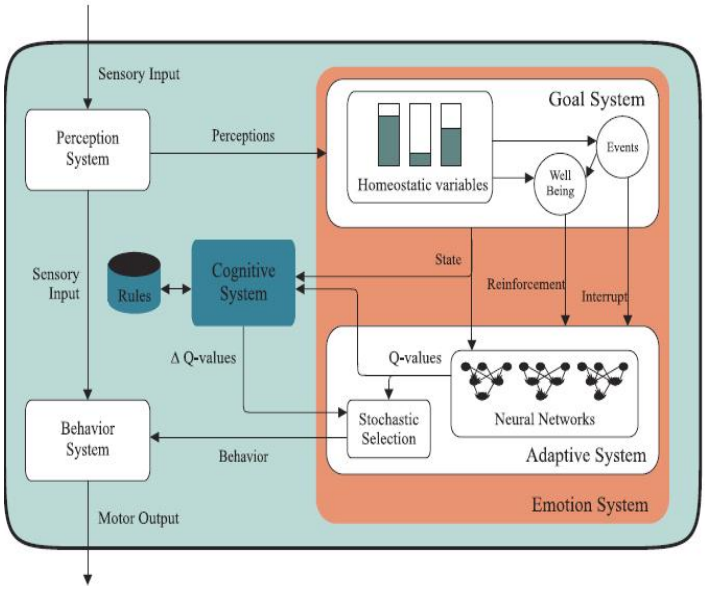


FIGURA 2.2 Arquitetura ALEC, retirada de (GADANHO 2003)

A influência de (DAMÁSIO 1994) e a hipótese do Marcador Somático continuam presentes: na arquitetura ALEC, a forma que o Sistema Emocional interage com o Sistema

Cognitivo se assemelha à hipótese do Marcador Somático.

> "In his hypothesis, Damásio suggested that humans associate high-level cognitive decisions with special feelings that have good or bad connotations depending on whether choices have been emotionally associated with positive or negative long-term outcomes. If these feelings are strong enough, a choice may be immediately followed or discarded. Interestingly, these markers do not have explanatory power and the reason for the selection may not be clear. In fact, although a decision may be reached easily and immediately, the person may subsequently feel the need to use high-level reasoning capabilities to find a reason for the choice. Meanwhile, a fast emotion-based decision could be reached which, depending of the urgency of the situation, may be vital. ALEC shows similar properties when it uses emotional associations to guide the agent. These associations (or more specifically, the utility values) are akin to the somatic-markers suggested by Damásio (1994) in that both provide a long-term indication of the "goodness" of the options available to the agent. Furthermore, the cognitive system can correct the emotion system when it reaches incorrect conclusions. By knowing the exceptions from previous experiences, the cognitive system may choose to override the emotion reactions, which— though powerful — can be more unreliable" (GADANHO 2003).

Resultados experimentais sugeririam que o uso do Sistema Cognitivo, juntamente com o Emocional, poderiam beneficiar sistemas robóticos. Outra vantagem proporcionada por ALEC seria o permitir conhecimento inato sobre o mundo em duas formas distintas, como preferências e não-preferências no Sistema Emocional, ou regras de ação simples no Sistema Cognitivo. Segundo (GADANHO 1999), abriria muitas possibilidades de pesquisa o fato de ALEC obter conhecimento explícito sobre o mundo; além disso, extensões a esta arquitetura poderiam consistir em adicionar conhecimento mais específico ao Sistema Cognitivo, o qual poderia ser usado para o planejamento de tarefas mais complexas.

2.2.4.1.1 Modelo Clarion

O Modelo Clarion (SUN et al 1998, 2001; SUN 2002) é uma arquitetura para modelar processos cognitivos em um sentido psicológico. Um conceito considerado chave nesta arquitetura seria a premissa da dicotomia entre os conhecimentos implícito e explícito; outras

características importantes seriam a de o agente poder aprender por si próprio sem considerar se o conhecimento do domínio foi provido *a priori* ou externamente; e a de os subsistemas interagirem entre si constantemente: precisariam trabalhar juntos para atingir processamento cognitivo. (SUN 2006) resalva que a interação social somente se faz possível devido às habilidades inatas (mesmo que parcialmente) de agentes cognitivos para refletir sobre seus próprios comportamentos e modificá-los dinamicamente. Aprendizados tipo *top-down* e *botton-up* foram adicionados à Arquitetura Clarion, ou seja, por aprendizado *top-down* entende-se (SUN, ZHANG 2004) aquele que partiria do conhecimento explícito para o implícito; e, por sua vez, o *bottom-up* seria o aprendizado que partiria do conhecimento implícito para o explícito.

Os subsistemas de Clarion são: o ACS, que denota o subsistema centrado em ação; o NACS, subsistema não centrado em ação; o MS, subsistema motivacional e, por fim, o MCS, subsistema metacognitivo. O papel do ACS seria controlar as ações sem considerar se estas se direcionam a movimentos físicos externos ou a operações mentais internas. Já o do NACS, o manter o conhecimento geral, tanto implícito quanto explícito. O do MS, prover motivações fundamentais para a percepção, ação e cognição, em termos de prover impulso e *feedback,* isto é, indicando se os resultados são satisfatórios ou não. O do MCS seria o de monitorar, dirigir e modificar as operações do ACS dinamicamente, assim como as operações de todos os outros subsistemas. Os mecanismos de Clarion, especialmente o Metacognitivo e o Motivacional, seriam uma contribuição singular ao modelamento de cognição e simulação social.

2.3 Sistema Homeostático de (Gadanho 1999)

Para desenvolver a Arquitetura de Controle Baseado em Comportamento, (GADANHO 1999) teve por inspiração (DAMÁSIO 1994); portanto, os termos *Emoção e Sentimento* serão adotados de acordo com (DAMÁSIO 1994). A Tabela 2.1 define a utilização destes termos.

Emoção	Combinação do processo mental avaliativo com suas respostas para este, na maior parte em direção ao próprio corpo, mas também ao cérebro. Conjunto de mudanças que ocorrem no corpo ou no cérebro e que geralmente é originado por um conteúdo mental.
Sentimento	Processo de monitoramento do corpo e que nos oferece a cognição dos nossos estados viscerais e músculos-esqueléticos. Percepção das mudanças provocadas por emoções.

TABELA 2.1. Definições para sentimentos e emoção segundo (DAMÁSIO 1994).

De modo metafórico, o conteúdo mental que origina a emoção seria: “Concebo a essência do viver mental das emoções como algo que podemos ver através de uma janela que abre directamente sobre uma imagem continuamente actualizada da estrutura e do estado do nosso corpo” (DAMÁSIO 1994).

O termo *sensação* é definido de acordo com a Tabela 2.2.

Sensações	Medições diretamente advindas dos sensores do agente: 1 – sensores de proximidade de obstáculo; 2 - sensores de intensidade de luz.

TABELA 2.2. Definições para os termos *sensação* e *agente*.

Para efeitos de simulação, a partir desta seção considera-se o agente artificial como sendo um robô Khepera (MONDADA 1994), que possui oito sensores, os quais medem proximidade de obstáculo e intensidade de luz. As medições diretamente advindas dos sensores são transformadas em sensações. Segundo (GADANHO 1999), os sentimentos foram escolhidos de acordo com sua relevância para a navegação animal autônoma em ambientes desconhecidos; já as emoções, foram selecionadas tendo por base aquelas que seriam adequadas e úteis para a interação do robô e seu ambiente e, ainda, por serem as emoções mais universalmente expressas (EKMAN 1992), juntamente com o nojo, que não foi utilizado em (GADANHO 1999).

Em (GADANHO 1999) um Sistema Homeostático é usado para resolver o que implicitamente requeria algum tipo de mecanismo de atenção. A Figura 2.3 ilustra o esquema geral do Sistema Homeostático, que é uma rede neural artificial recorrente com quatro saídas e sete entradas. A oitava entrada não alimenta a rede - como em (DAMÁSIO 1994), nem todo sentimento geraria emoção. Além disso, na Figura 2.3 estão destacadas, em azul, as localizações, na tese, das descrições dos cálculos de cada parte do Sistema Homeostático.As partes do Sistema Homeostático serão detalhadas a partir da Seção 2.3.1.

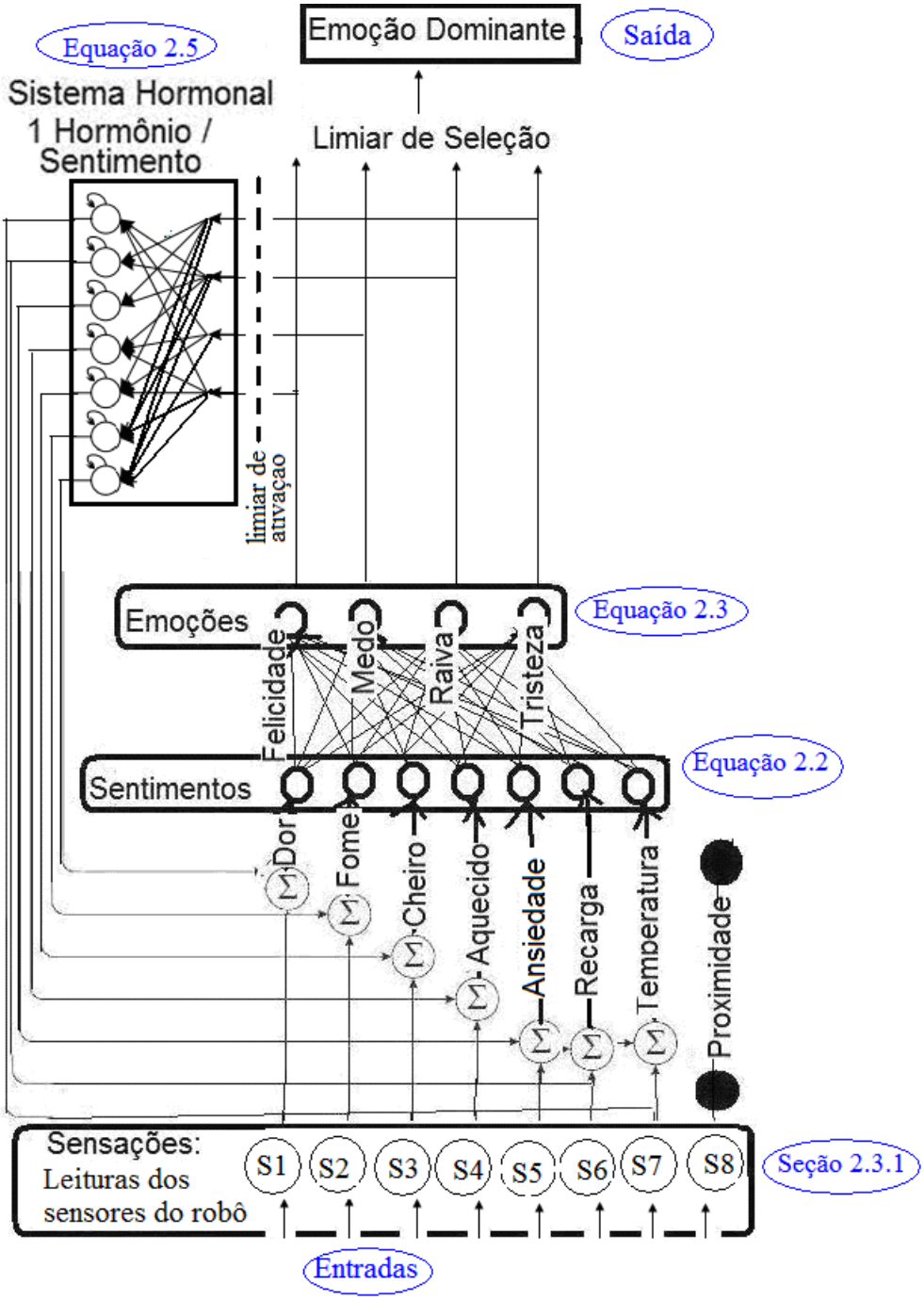


FIGURA 2.3 Sistema Homeostático adaptado de (GADANHO 1999).

2.3.1. Especificações para as Sensações

Segue uma explicação de como as sensações e os sentimentos foram modelados segundo (GADANHO 1999). Os sentimentos são: Sto = {Fome, Dor, Ansiedade, Temperatura, Recarregando, Cheiro, Aquecido, Proximidade}. A seguir são listados os sentimentos e as suas respectivas sensações associadas: {S1, S2, S3, S4, S5, S6, S7, S8}.

- Dor - recebe valores da sensação S1: existente apenas se o agente se chocar com algum objeto. Neste caso, a intensidade da dor será proporcional ao número de sensores de distância com valores altos, caso contrário, será zero.
- Fome- recebe valores da sensação S2: significa o déficit de energia do agente. A sensação correspondente, nomeada s2, recebe valores de acordo com a Equação (2.1):

$$S2_t = 1 - nv_t \qquad (2.1)$$

Onde: $S2$ = sensação que alimenta o sentimento Fome;
nv_t = nível de energia do agente na iteração atual

$$nv_t = \begin{cases} 1 \text{ se o tanque do agente estiver completo} \\ 0 \text{ se estiver vazio} \cdot \end{cases}$$

- Cheiro- recebe valores da sensação S3: ativo apenas quando o agente aciona/colide em uma fonte de energia (toda fonte de luz do ambiente é uma fonte de energia). Sua intensidade é diretamente proporcional ao número de iterações em que a energia continuará disponível (= 200 iterações).
- Aquecido- recebe valores da sensação S4: diretamente dependente da intensidade de energia percebida pelos sensores de luminosidade do agente, é o valor normalizado do sensor de luz que está detectando maior luminosidade no momento; quanto maior a intensidade de luz recebida por aquele sensor, maior o valor desta sensação.

  O nível de energia diminui proporcionalmente à atividade dos motores. Se o agente estiver completamente parado, levará 100 mil iterações para gastar toda a sua energia; já se estiver correndo com velocidade máxima, levará 20 mil iterações. Demorará 200

iterações para recuperar toda a sua energia (partindo do nível zero de energia) se receber intensidade máxima de luz nos seus sensores traseiros.

- Ansiedade - recebe valores da sensação S5: reiniciada ao valor zero, junto com o hormônio associado ao sentimento Ansiedade, sempre que há uma avaliação sobre qual dos comportamentos deverá ser executado. Se o agente andar uma boa distância, sua Ansiedade diminuirá; caso contrário, aumentará. Para o cálculo da Ansiedade, há dois parâmetros diferentes: um que regula o seu aumento (*BoredomRaiseSteps* = 1000 iterações), e outro que regula a sua diminuição (*BoredomLowerSteps* = 200 iterações). O parâmetro *BoredomRaiseSteps* é o número de iterações em que o agente deve estar completamente parado para alcançar o valor máximo da sensação que alimenta o sentimento Ansiedade.
- Recarga- recebe valores da sensação S6: existente apenas se o agente estiver adquirindo energia. Neste caso, é diferente de zero e diretamente proporcional à intensidade de luz recebida pelos sensores *traseiros* do agente – mas apenas se a intensidade de luz for alta o suficiente para aumentar o nível de energia do robô, ou seja, maior do que 60% da luminosidade máxima possível, a fim de evitar que o robô reabasteça quando está se afastando da fonte de energia (luz). Se o nível de energia for alto (maior do que 0,95), esta sensação é multiplicada por $(1\text{-}nv_t)$ isto é, pela sensação de fome, definida na Equação (2.1).
- Temperatura- recebe valores da sensação S7: existente apenas se os motores do agente estiverem se movendo com valores altos. Se a atividade total do motor for acima do limiar *TempRaiseTh* (vide o apêndice), então a temperatura irá se elevar. Se o valor for baixo, ou seja, atividade do motor menor do que *TempLowerTh*, então a temperatura irá diminuir. Quando os motores estão com força total (atividade dos motores = 20), o agente usa *TempRaiseSteps* (= 200 iterações) para ir de Temperatura zero para

Temperatura máxima. Levará *TempLowerSteps* (= 1000 iterações) para baixar sua temperatura de volta para zero com os motores desligados. O aumento da Temperatura é diretamente proporcional à atividade total do motor e a diminuição é diretamente proporcional a um menos a atividade total do motor.

- Proximidade- recebe valores da sensação S8: reflete a proximidade do obstáculo mais próximo percebido pelos sensores de distância: é o valor normalizado do sensor de distância de maior valor.

2.3.2 Especificações para as Emoções / Relações com os Sentimentos

A conotação das emoções é tal que a única positiva é a Felicidade; as negativas são o Medo, a Raiva e a Tristeza. As emoções dependem dos sentimentos com valores mais altos que o agente tem no momento.

Para que assim aconteça, os pesos, as dependências dos sentimentos sobre as emoções são definidas de modo que a Emoção Dominante (a emoção que tem o valor mais alto na iteração considerada) sinaliza para o agente a sua situação atual de acordo com a Tabela 2.3:

Descrição	Estado	Alterador	Conotação
Felicidade	•Sem colisões e razoável nível de energia; •Se movendo, ou recarregando, ou próximo a uma fonte de energia.	Intensificação: Combinação dos fatores.	Positiva
Medo	Se houver colisão.	Intensificação: Não captação de fonte de luz.	Negativa
Tristeza	Pouca energia e sem recarregar.	Redução: valores altos de Fome ou Ansiedade.	Negativa
Raiva	•Se permanecer em uma mesma região. •Valores altos do sentimento de Fome.	Intensificação: Combinação dos fatores.	Negativa

TABELA 2.3. Descrição das dependências entre sentimentos e emoções.

As dependências acima mencionadas, dos sentimentos sobre as emoções, são definidas de acordo com a Tabela de pesos 2.4, que se encontra disponível em (GADANHO 1999) - as dependências foram estabelecidas por (GADANHO 1999) de forma experimental.

Utilizou-se exatamente esta Tabela 2.4 em todos os experimentos descritos nesta tese. Entretanto, esta tabela, que não deixa de ser uma simplificação, pode ser alterada de acordo com as ligações que se queira estabelecer entre os sentimentos e as emoções; por exemplo: se se julgar a tristeza como complementar à felicidade, a Tabela 2.4 se mostrará insatisfatória, visto não contemplar tais emoções desta maneira. Na seção 5.1.1 serão consideradas possíveis modificações a esta tabela de pesos.

	Fome	Dor	Ansiedade	Temperatura	Recarregando	Cheiro	Aquecido	Bias
Felicidade	-0,2	-0,3	-0,2	0,2	0,4	0,3	0	0,1
Tristeza	0,7	0	0,1	-0,2	-0,4	0	-0,2	-0,1
Medo	-0,2	0,7	-0,2	0,1	-0,2	-0,2	0	0
Raiva	0,2	0,1	0,7	-0,2	-0,2	0	0	0

TABELA 2.4. Dependências entre sentimentos e emoções, traduzida de (GADANHO 1999).

As características das emoções são:

1. Têm valor positivo ou negativo;
2. Persistem no tempo: alterações repentinas entre diferentes emoções são inibidas através de ação hormonal;
3. Sua ocorrência depende das entradas diretas dos sensores e do histórico emocional recente do agente;
4. O estado emocional pode ser neutro ou dominado por uma emoção – o que implica na existência de um mecanismo para decidir qual emoção, se alguma, é dominante em qualquer momento.

Se o valor da Emoção, seja qual for e quantas forem, estiver acima do limiar de ativação (LA = 0,2), influenciará o agente através do Sistema Hormonal, ou seja, as emoções que estiverem acima do limiar de ativação influenciarão para que sejam produzidos os hormônios coerentes com aquelas emoções em particular. As emoções que tiverem valores inferiores ou iguais ao limiar de ativação não influenciarão o Sistema Hormonal.

Para se poder dizer que o agente está sob influência de uma emoção dominante, é necessário que a emoção que tiver o maior valor dentre as quatro apresente valor superior ao limiar de seleção (LS = 0,2). Se nenhuma das quatro emoções apresentar valor superior ao limiar de seleção, não haverá uma emoção dominante e então o estado emocional do agente será neutro – neste caso diz-se que há ausência de emoção dominante, ou que esta é Neutra.

Devido à circularidade do sistema, ou seja, ao fato de a rede neural que constitui o Sistema Homeostático ser de realimentação, o limiar de seleção acaba sendo dependente do limiar de ativação: as emoções acima do limiar de ativação influenciam os hormônios, automaticamente influenciando para que a emoção que ativou o hormônio seja intensificada posteriormente através dos sentimentos - ao mencionar a dependência do limiar de seleção em relação ao limiar de ativação, (GADANHO 1999) observa que mesmo assim provavelmente o limiar de seleção não deveria ser menor para assegurar que as emoções dominantes estivessem sempre ativas.

2.3.3 Sistema Hormonal

As reações do corpo que elevam a intensidade das emoções são também aquelas estimuladas pela emoção: cada emoção tenta influenciar o estado do agente de tal forma que seu estado resultante combine com o estado que suscitou aquela emoção em particular. Os valores hormonais são responsáveis pela memória do sistema emocional e dependem de seus próprios valores prévios (exceto na primeira iteração, quando tanto os valores dos hormônios

atuais quanto os dos hormônios da iteração anterior valem zero) e das influências emocionais (IE). As características dos hormônios são (GADANHO 1999):

- Há um hormônio associado a cada sentimento (exceto o sentimento Proximidade);
- Seus valores podem ser positiva ou negativamente altos o suficiente para esconder totalmente as reais sensações da percepção do agente sobre si próprio;
- O mecanismo hormonal introduz uma competição entre as emoções a fim de obter o controle sobre o agente, o qual, por fim, é o que seleciona qual emoção será dominante. Por outro lado, os sentimentos do agente não são unicamente dependentes de suas sensações, mas também dependem do seu estado emocional, isto é, da intensidade de suas emoções.

As influências emocionais são calculadas através da mesma Tabela 2.4 usada para calcular as emoções. Para o cálculo do valor dos hormônios, é utilizado um parâmetro que assume dois diferentes valores (para uso do parâmetro, Seção 2.3.4), ou seja:

- $\bar{p}$ usado quando as emoções e suas influências (as influências emocionais, IE) se elevam.
- $\underline{p}$ quando as intensidades das emoções estão diminuindo.

A queda das emoções é lenta, enquanto a emergência de novas emoções é mais rápida.

2.3.4 Equações do Sistema Homeostático

Cada sentimento possui uma sensação e um hormônio que lhe são associados (exceto o sentimento de Proximidade, que não possui um hormônio que lhe seja associado). Há oito sensações e, também, oito sentimentos, todos recebendo valores dentro do intervalo [0,1]. O oitavo sentimento, o de Proximidade de obstáculo, não possui um hormônio associado; assim,

neste único caso, sensação e sentimento recebem o mesmo valor, advindo dos sensores do agente. Para o cálculo de todos os outros sentimentos, faz-se de acordo com a Equação (2.2):

$$\mathrm{Sto(i)_n} = \left(\mathrm{CF} \times \mathrm{H(i)}_{n-1}\right) + \mathrm{Sens(i)_n} \qquad (2.2)$$

$$\mathrm{Sto(i)_n} = \begin{cases} 1 \ \text{Se} \ \mathrm{Sto(i)_n} > 1 \\ 0 \ \text{Se} \ \mathrm{Sto(i)_n} < 0 \\ \mathrm{Sto(i)_n} \ \text{caso contrário} \end{cases}$$

Onde $\mathrm{Sto(i)_n}$, $\mathrm{Sens(i)_n}$ e $\mathrm{H(i)_{n-1}}$, $\mathrm{i} = 1, \ldots, 7$ são, respectivamente, os sentimentos e sensações correspondentes na iteração *n,* e hormônios na iteração *n-1*; e CF (CF = 0,9) é um coeficiente hormonal (constante). As emoções são baseadas nos sentimentos (Sto) internos e atuais do agente e são calculadas a partir do peso linear e simples das suas dependências específicas (que são retiradas da Tabela 2.4) em relação aos sentimentos.

Há quatro emoções $\mathrm{E(i)}$, $\mathrm{i} = 1, \ldots, 4$, as quais são inicializadas com os valores de bias e somadas ao produto entre os sentimentos e suas dependências específicas sobre as emoções, de acordo com a Equação (2.3):

$$\mathrm{E(i)_n} = \mathrm{Bias\,(i)} + \sum_{\mathrm{j}=1}^{7} \mathrm{Sto\left(j\right)_n} \cdot \mathrm{DS\left(i,j\right)} \qquad (2.3)$$

$$\mathrm{E(i)_n} = \begin{cases} 1 \ \text{Se} \ \mathrm{E(i)_n} > 1 \\ 0 \ \text{Se} \ \mathrm{E(i)_n} < 0 \\ \mathrm{E(i)_n} \ \text{caso contrário} \end{cases}$$

onde DS(i,j) vem da Tabela 2.4, e $\mathrm{E(i)_n}$ i = 1,..,4, j = 1,..,7, são, respectivamente, as dependências dos sentimentos sobre *aquela* emoção específica, na iteração n. As dependências dos sentimentos sobre emoções, DS, são os pesos que conectam sentimentos a emoções (Tabela 2.4). As emoções ficam dentro do intervalo [0, 1].

A Figura 2.4 exemplifica, na iteração *n*, o cálculo feito através da Equação (2.3) acima.

	BIAS	FOME: Sto(1)	DOR: Sto(2)	TÉDIO: Sto(3)	TEMPERATURA:Sto(4)	RECARREGANDO:Sto(5)	CHEIRO: Sto(6)	AQUECIDO: Sto(7)
E(1): Felicidade	Bias(1) +	DS(1,1)*Sto(1)+	DS(1,2)*Sto(2)+	DS(1,3)*Sto(3)+	DS(1,4)*Sto(4)+	DS(1,5)*Sto(5)+	DS(1,6)*Sto(6)+	DS(1,7)*Sto(7)
E(2): Tristeza	⇨	⇨	⇨	⇨	⇨	⇨	⇨	⇨
E(3): Medo	⇨	⇨	⇨	⇨	⇨	⇨	⇨	⇨
E(4): Raiva	⇨	⇨	⇨	⇨	⇨	⇨	⇨	⇨

FIGURA 2.4 Calculando os valores das emoções.

As influências emocionais, IE (i), i = 1..., 7, usadas para definir os valores hormonais, são calculadas de acordo com a Equação (2.4). O limiar de ativação define a ativação de uma emoção: se a emoção tiver valor acima deste limiar, então está ativa.

$$IE(i)_n = \sum_{j=1}^{4} Aux(j)_n \qquad (2.4)$$

Onde: $i = 1, \ldots, 7; \quad j = 1, \ldots 4;$

$IE(i)_n = \text{Influências Emocionais no passo } n;$

$$Aux(j)_n = \begin{cases} 0 \text{ se } E(j)_n < \text{Limiar de Ativação} \\ DS(j,i) \cdot E(j)_n \qquad \text{caso contrário} \end{cases}$$

Para se ter o resultado final, tanto para o cálculo das emoções quanto para o dos sentimentos, ambos devem passar por uma equação de limiar, ou seja, devem se enquadrar ao intervalo [0,1]. Já as influências emocionais e os hormônios, não. Ao se considerarem a Tabela 2.4 de dependências emoção / sentimento e a Equação (2.4) das influências emocionais, será percebido que o próprio modelo restringe os valores das influências emocionais sobre os sentimentos. A Tabela 2.4 mostra como são calculados os valores mínimos e máximos de cada uma das sete influências emocionais sobre os sentimentos.

Emoção \ Sentimento	Fome		Dor		Ansiedade	
Felicidade	-0,2 * 1 +	-0,2 * 0 +	-0,3 * 1+	-0,3 * 0+	-0,2*1 +	-0,2*0 +
Tristeza	0,7 * 0 +	0,7 * 1 +	0 * 0 +	0 * 0 +	0,1*0 +	0,1*1 +
Medo	-0,2 * 1 +	-0,2 * 0 +	0,7 * 0 +	0,7 * 1 +	-0,2*1 +	-0,2*0 +
Raiva	0,2 * 0	0,2 * 1	0,1 * 0	0,1 * 1	0,7*0	0,7*1
	Mínimo	Máximo	Mínimo	Máximo	Mínimo	Máximo
Influências Emocionais	= -0,4	= 0,9	= -0,3	= 0,8	= -0,4	= 0,8

Emoção \ Sentimento	Temperatura		Recarregando		Cheiro		Aquecido	
Felicidade	0,2 * 0 +	0,2 * 1 +	0,4 * 0 +	0,4*1 +	0,3 * 0 +	0,3 * 1+	0 * 0 +	0 * 0 +
Tristeza	-0,2 * 1 +	-0,2 * 0 +	-0,4 * 1+	-0,4 * 0 +	0 * 0 +	0 * 0 +	0 * 0 +	0 * 0 +
Medo	0,1 * 0 +	0,1 * 1 +	-0,2 *1 +	-0,2 * 0 +	-0,2 *1 +	-0,2*0+	-0,2* 1 +	-0,2 * 0 +
Raiva	-0,2 * 1	-0,2 * 0	-0,2 * 1	-0,2 * 0	0*0	0 * 0	0*0	0 * 0
	Mínimo	Máximo	Mínimo	Máximo	Mínimo	Máximo	Mínimo	Máximo
Influências Emocionais	= -0,4	= 0,3	= -0,8	= 0,4	= -0,2	= 0,3	= -0,2	= 0

TABELA 2.4. Cálculo das influências emocionais máximas e mínimas do sistema.

Por exemplo, seguindo a Tabela 2.4, o menor valor das influências emocionais sobre o sentimento de Fome é -0,4. Todas as emoções que têm peso negativo em relação ao sentimento de Fome são as que influenciam seu decréscimo e, as que têm peso positivo, seu incremento. Assim, se forem consideradas com valor máximo (*um*) as primeiras e, com valor mínimo (*zero*), as segundas, tem-se:

$$\left(0{,}2\times 0\right)+\left(-0{,}2\times 1\right)+\left(0{,}1\times 0\right)+\left(-0{,}2\times 1\right)=-0{,}4.$$

A emergência de novas emoções é mais rápida do que a diminuição das já existentes, pois há dois parâmetros diferentes usados no cálculo dos hormônios. O parâmetro α usado em cada iteração *n* recebe valor de acordo com o que está ocorrendo com o agente: se a situação deve estimular (quando as emoções e suas influências estão aumentando: p↑) ou desestimular (quando as intensidades das emoções estão diminuindo: p↓) a produção de hormônios. Finalmente, os valores dos hormônios são calculados de acordo com a Equação (2.5):

$$H(i)_n = \begin{cases} 0 & \text{se } n = 0 \\ \alpha_n \cdot H(i)_{n-1} + (1 - \alpha_n) \cdot IE(i)_{n-1} & \text{caso contrário} \end{cases} \quad (2.5)$$

Onde: $i = 1, \ldots, 7$;

$IE(i)_{n-1}$ = Influência Emocional (i) da iteração anterior;

$$\alpha = \begin{cases} 0{,}996 \quad \underline{p} & \text{se } |H(i)_{n-1}| \geq |IE(i)_n| \\ 0{,}98 \quad \bar{p} & \text{caso contrário} \end{cases}$$

Na Figura 2.3, no Sistema Hormonal, pode-se perceber que há um loop em cada hormônio: tal ocorre devido à atualização do hormônio, Equação (2.5), ser feita considerando-se o histórico do valor hormonal (valor do hormônio na iteração anterior).

2.3.5 Comportamento do Sistema Homeostático

Os valores dos hormônios podem aumentar rapidamente, permitindo uma ágil construção de um novo estado emocional, e diminuir vagarosamente, consentindo na persistência de um estado emocional mesmo quando a causa que o produziu cessou - outro dos traços característicos das emoções (GADANHO 1999).

Objetivando mostrar a dinâmica do modelo em termos da interação ambiente/robô, (GADANHO 1999) gerou o exemplo (Figura 2.6 a) em que o agente colide com um obstáculo. A colisão produz uma sensação que será capturada pelo sentimento Dor. Assumindo que a emoção Medo tem forte dependência em relação ao sentimento Dor, então a intensidade do Medo deverá aumentar se a intensidade de Dor for alta o suficiente para ativar a emoção Medo, e a emoção Medo produzirá hormônios. Em particular, o hormônio associado ao sentimento Dor irá aumentar velozmente durante a colisão; o que, por sua vez, fará com que a emoção Medo aumente e possivelmente influencie as outras emoções existentes. Quando o robô finalmente consegue se livrar da colisão, ainda terá Dor, mas não mais devido à sensação da colisão, e sim porque o hormônio associado à Dor tem um valor alto. Assim a emoção Medo persistirá enquanto o hormônio gradualmente diminui o seu

valor. Isso significa que, enquanto o robô se distancia do obstáculo, ainda haverá Dor. Todavia, esta irá desaparecer assim que diminuir o risco de outra colisão.

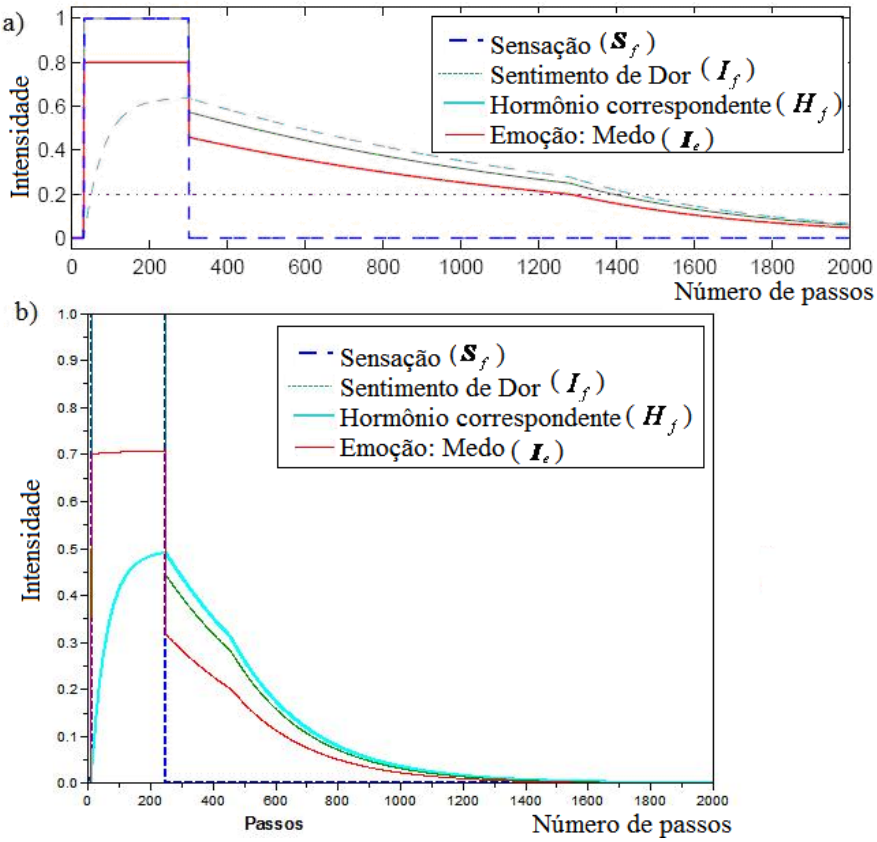


FIGURA 2.5 Gráficos da resposta emocional à colisão. a) adaptada de (GADANHO 1999) e b) gerada no contexto desta tese.

Os gráficos da Figura 2.5 mostram, durante uma situação de colisão, a resposta da emoção Medo para uma sensação sobre a qual tem a dependência indireta (visto haver o sentimento Dor fazendo a mediação emoção-sensação) de acordo com o peso da Tabela de Dependências Emoção/Sentimento (GADANHO 1999). É importante observar que a tabela de pesos para a geração dos gráficos, a) e b) foram diferentes, pois, no primeiro a tabela usada por (GADANHO 1999) não foi a correspondente à Arquitetura Baseada em Comportamento (a ser descrita no Cap. 4), mas Baseada em Ação. Já na segunda, a tabela utilizada foi a da Arquitetura Baseada em Comportamento, Tabela 2.4.

A análise de (GADANHO 1999) para a Figura 2.5 a), sendo a mesma para a Figura 2.5 b): assumindo *feedback* do hormônio correspondente ao sentimento Dor inicialmente zero

e o valor da sensação respectiva, $S_f = 1$, a intensidade da emoção Medo, I_e é alta: 0,8 em a) e 0,7 em b) sendo o maior valor possível para este exemplo de emoção (os valores dos item *a* e *b* são diferentes porque as tabelas de dependências utilizadas foram diferentes, mas o comportamento do Sistema Homeostático é o mesmo). A influência do hormônio H_f só é notória depois de a sensação retornar para o valor zero. Antes disso, a intensidade do sentimento I_f está saturada. Quando o estímulo desaparece, a intensidade da emoção sofre uma queda repentina no seu valor porque se torna dependente unicamente do valor total do hormônio H_f acumulado enquanto a sensação estava ativa. Os valores do hormônio e da emoção gradualmente decaem para zero sem a presença da sensação. Quando a intensidade da emoção decai para valores abaixo do limite de ativação (0,2), a influência da emoção sobre o hormônio cessa e os valores do Parâmetro $\underline{p}$ atuam.

(GADANHO 1999) observa que deve ser percebido que as escalas de tempo envolvidas na persistência de uma emoção, depois que o estímulo tiver cessado, são muito pequenas, particularmente quando em presença de um novo estímulo que favoreça uma outra emoção. E que poder-se-ia somente especular sobre humores, dentro do contexto do seu Sistema Homeostático, em relação aos valores residuais de hormônios que devem existir no sistema e que não são fortes o suficiente para estimular a existência de uma emoção dominante – o que poderia ser consistente com a teoria de que os humores seriam diferenciáveis das emoções em termos do nível de excitação (PANKSEPP, 1995). Segundo (GADANHO 1999) os valores residuais de hormônios poderiam agir como humores no sentido de que poderiam favorecer o aparecimento de certas emoções, mas o fato de as escalas de tempo envolvidas na persistência dos valores residuais serem curtíssimas, tal interpretação acerca dos humores se tornaria inconsistente. Temperamentos diferentes poderiam ser alcançados por diferentes dependências emocionais sobre os sentimentos, ou mudando outros parâmetros do sistema.

3 Arquitetura de Controle e Modificações

Neste capítulo serão descritos a tarefa do agente e os comportamentos disponíveis, que o agente deverá aprender a coordenar, para concretizá-la. Em seguida será apresentada a arquitetura descrita em (GADANHO 1999) e que tem como parte integrante o Sistema Homeostático, descrito no Capítulo 2. Em especial, pontos importantes desta mesma arquitetura serão detalhados do modo que foram desenvolvidos para a obtenção dos resultados. Ao final deste capítulo serão apresentadas as modificações feitas à arquitetura original e que se caracterizam, basicamente, pelo aumento no número de redes neurais para que o aprendizado ocorra de acordo com a emoção dominante na situação que ativou o aprendizado.

3.1 Tarefa do agente

A tarefa do agente considerada nesta tese foi a mesma que a descrita em (GADANHO 1999), ou seja, cada fonte de luz simulada no ambiente do robô é uma fonte de energia que disponibiliza alguns itens de energia, sendo MaxFoodItems a quantidade máxima de itens. Havendo uma quantia limitada de itens de energia em cada fonte, após liberá-los todos, um a um, a fonte será exaurida. Após algumas iterações (um número aleatório entre zero e vinte mil), a fonte com menor quantidade de itens (desde que esta seja menor do que MaxFoodItems) irá recuperar um item de energia.

Para que a fonte de energia disponibilize um item para o agente, este deve colidir naquela, fazendo com que haja energia disponível por um período de tempo. Havendo energia disponível, o agente deve rapidamente se posicionar de ré para a fonte de energia, visto que sua obtenção somente ocorre através de valores altos de leitura de luminosidade nos sensores traseiros do robô. O item liberado fica disponível por MaxFoodAvaibleSteps (= 200 iterações)

e, portanto, somente durante este intervalo de tempo o agente pode adquirir energia. Os sensores de luminosidade do agente são sensíveis à fonte de energia quando esta tem itens de energia disponíveis e, também, enquanto o robô está recarregando. Porém, quando a fonte de energia não tem mais nenhum item disponível, se torna um obstáculo qualquer para o agente.

O nível de energia do agente, bem como as intensidades das sensações, estão entre zero e um. A diminuição no nível de energia é proporcional à atividade total do agente, isto é, à soma dos valores absolutos da velocidade dos motores Esquerdo e Direito: para gastar toda a sua energia, o agente leva EnergyAutoStopSteps (= 100 mil iterações) se estiver parado e, se estiver se movendo com velocidade total, EnergyAutoRunSteps (= 20 mil iterações). O nível de energia do robô irá aumentar somente se os valores de intensidade de luz recebidos nos sensores traseiros forem altos o suficiente – 60% do valor máximo. Se esta condição for satisfeita, o aumento de energia será diretamente proporcional à luz recebida pelos sensores. Se o agente receber luz com intensidade máxima sobre os seus sensores traseiros, EnergyRechargSteps (= 100 iterações) será o número de iterações necessários para recuperar toda a sua energia.

3.2 Comportamentos do Agente

Segundo (GADANHO 1999), se a tarefa do robô autônomo for complexa, usualmente será necessário introduzir uma hierarquia dos comportamentos, que pode ser apenas uma decomposição da tarefa em um conjunto de tarefas simples, associadas a comportamentos. O que seria um comportamento nesse caso? “Um comportamento relaciona estímulos sensoriais a ações produzidas sobre os atuadores do robô” (RIBEIRO et al. 2001). Neste contexto de aplicação, os atuadores são os motores Esquerdo e Direito de um robô Khepera (MONDADA 1994) e os comportamentos são sequências pré-estabelecidas das ações: “vá em frente”, “pare”, “vire à direita” e “vire à esquerda”.

Os três comportamentos que o agente deverá aprender a coordenar (GADANHO 1999) são:

Evite obstáculos: se os sensores de proximidade não detectarem obstáculo por perto, permaneça parado; caso contrário, desvie.

Busque por Luz: vá em direção à fonte de energia (luz) mais próxima; se nenhuma fonte for detectada pelos sensores de luminosidade, fique parado.

Siga paredes: se não houver paredes por perto, siga em frente com velocidade máxima. Uma vez que uma parede é detectada pelos sensores de proximidade, siga-a.

Segundo (GADANHO 1999), a decomposição da tarefa do agente em comportamentos teria introduzido a necessidade de determinar quando ativar o controle (quando reavaliar o comportamento previamente selecionado e selecionar um novo, Seção 3.3.5), e teria sido descoberto empiricamente que uma seleção correta dos mecanismos de ativação de controle seria um passo crucial para a concretização da tarefa de aprendizado.

3.3 Arquitetura de Controle Baseada em Comportamento

Como brevemente mencionado ao longo desta tese, (GADANHO 1999) buscou investigar, através de sua arquitetura, como emoções artificiais podem auxiliar no controle de um agente que se adapta ao seu ambiente usando técnicas de Aprendizado por Reforço (a função de reforço utilizada será explicada na Seção 3.3.1) e que, como resultado dos reforços e em resposta ao seu ambiente, decide qual comportamento executar.

3.3.1 Aprendizado por Reforço (AR)

O Aprendizado por Reforço (AR) é uma técnica de aprendizado que consiste em como mapear situações em ações de modo a maximizar um sinal numérico de recompensa, tendo a própria experiência funcionando como instrutor. Suas características mais marcantes são a busca por tentativa e erro e a recompensa recebida de modo intermitente e *a posteriori*. Todos os agentes de AR têm objetivos, podem perceber aspectos do seu ambiente e escolher ações que modificam o estado deste seu ambiente (SUTTON, BARTO 1998).

Para realizar a sua tarefa, o agente precisa saber o que fazer diante de qualquer estado do mundo com o qual possa se deparar: esta solução é chamada política (NORVIG, RUSSELL 1995). Em (GADANHO 1999), o aprendizado da política se dá através da utilização do algoritmo *Q - Learning* (WATKINS, 1989; WATKINS, DAYAN, 1992) sobre uma arquitetura neural, sendo que a correção dos pesos neurais se dá através do algoritmo de Retropropagação - *Back-Propagation* (WERBOS 1994).

O *Q-learning* é um algoritmo de Diferenças Temporais, uma combinação entre os conceitos do algoritmo Monte Carlo e Programação Dinâmica: como o primeiro, nos métodos de Diferenças Temporais pode haver aprendizado diretamente de experiências cruas, sem um modelo da dinâmica do ambiente; e, como o segundo, os métodos de Diferenças Temporais têm suas estimativas atualizadas baseando-se, em parte, sobre outras estimativas aprendidas, sem que haja espera por um resultado final (SUTTON, BARTO 1998).

O retorno predito através do *Q – Learning* é uma função da ação e do estado. Neste algoritmo, a política pode ser em função das predições de retorno sem que haja uma estrutura de dados separada (como, por exemplo, escolher a ação com maior predição de retorno), ou como sugerido na Figura 3.1, retirada de (SUTTON 1992), estrutura também utilizada em (GADANHO 1999), vide a seção 3.3.5.2.

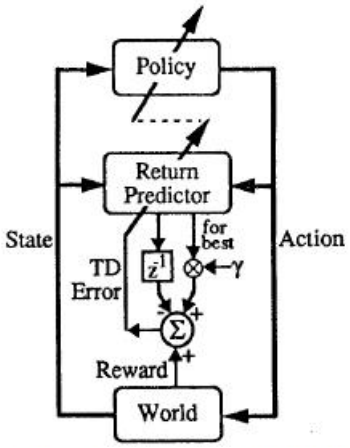


FIGURA 3.1 Arquitetura Q – Learning. A linha rotulada "for best" é a predição do retorno para a melhor ação; a outra saída de "Return Predictor" é o retorno predito para a ação efetivamente escolhida. Retirada de (SUTTON 1992).

3.3.1.1 Sinal de Reforço e Sistema Homeostático

As emoções artificiais, descritas no Capítulo 2 e obtidas através do Sistema Homeostático, foram utilizadas nas técnicas de AR sob vários aspectos em (GADANHO 1999). Serão apresentados nesta tese os usos que foram reproduzidos, os quais são:

- Medida de aprendizado;
- Detecção dos eventos significativos – Seção 3.3.4;
- Especificações de reforço.

As especificações de reforço providas pelas emoções se dão da seguinte maneira: as emoções dominantes que têm uma conotação negativa provêm reforço negativo; já a emoção que tem conotação positiva (para este caso sendo Felicidade a única), reforço positivo. Se a emoção dominante for Neutra, isto é, se nenhuma emoção tiver valor alto o suficiente para superar o limiar de ativação, o reforço recebido é zero. O sinal do reforço é estabelecido segundo a conotação da emoção dominante, de acordo com a Equação (3.1):

(3.1)

$$R = \begin{cases} 0 \text{ se } \forall \text{ e } \in \text{ Emoções, } Ie_n < \text{Limiar de Ativação} \\ ED \text{ se a Emoção Dominante for Felicidade} \\ -ED \text{ caso contrário} \end{cases}$$

Onde: R = Reforço; ED = Emoção Dominante;
Ie_n = Itensidade da Emoção Dominante no passo atual;
Limiar de Ativação = 0,2 ·

3.3.2 Arquitetura de (LIN 1993)

A Arquitetura de Controle Baseado em Comportamento (GADANHO 1999) foi inspirada na Arquitetura de Aprendizado proposta por (LIN 1993). A característica principal desta arquitetura (Figura 3.2) é a utilização de redes neurais artificiais, as quais aprendem a função de utilidade, uma rede por ação – sendo que a aquisição da política de ações do agente se baseia no algoritmo *Q-Learning* (WATKINS 1989). A utilidade de um par estado-ação é a recompensa imediata correspondente a esse estado e à ação (indexada pela rede neural correspondente), somada à utilidade descontada esperada do próximo estado, supondo-se que o agente escolhe ações ótimas a partir de então (NORVIG; RUSSELL 1995).

As redes neurais (*Utility Networks* na Figura 3.2) são sem realimentação *(feed-forward)* com uma única camada escondida, uma para cada ação, e são treinadas através do algoritmo de Retropropagação com *momentum* igual a 0,9 (vide Seção 3.3.5.1).

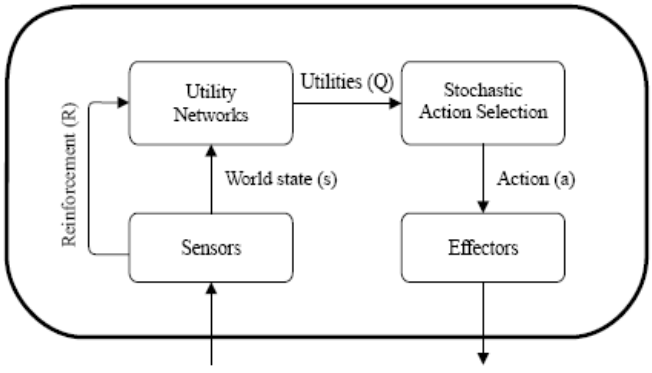


FIGURA 3.2 Arquitetura de aprendizado de (LIN 1993), retirada de (GADANHO 1999).

Segundo (GADANHO 1999), embora este não seja um dos melhores métodos para se treinar uma rede, teria sido escolhido por permitir o aprendizado incremental requerido para o aprendizado em tempo real. A entrada das redes neurais consiste no estado do mundo S_n; a única saída de cada rede neural modela a função de utilidade para cada uma das ações a segundo a Equação (3.2):

$$Q_n\left(S_n, a\right) = R_{n+1} + \gamma\, eval\left(S_{n+1}\right) \qquad (3.2)$$

Esta função representa o reforço cumulativo descontado esperado que o agente receberá após executar a ação a em resposta ao estado do mundo S_n na iteração n. O reforço recebido já no próximo estado S_{n+1} é R_{n+1}. A utilidade do estado S_{n+1}, ou eval (S_{n+1}), é o reforço cumulativo descontado esperado se a política ótima for escolhida pelo agente a partir do instante $n+1$. O valor γ é o fator de desconto, fixado em 0,9. Em cada iteração de aprendizado do algoritmo, os pesos da rede neural associada à última ação a tomada são atualizados para o estado anterior S_{n-1}. A iteração do algoritmo de Retropropagação é feita usando o valor alvo $Alvo_n\left(S_{n-1}, a\right)$, Equação (3.3). O cálculo do valor alvo depende da melhor estimativa atual $Q_n\left(S_n, k\right)$ dentre cada ação k determinada pelas saídas das redes quando usado como entrada o estado atual S_n:

$$Alvo_n\left(S_{n-1}, a\right) = R_n + \gamma \max\left\{Q_n\left(S_n, k\right) \mid k \in \text{ações}\right\} \quad (3.3)$$

Quando as piores ações são realmente muito ruins para o agente, uma seleção "gulosa" (ou seja, pela ação que tiver maior predição de retorno), pode ser insatisfatória; assim, uma alternativa seria o uso das regras de seleção de ação chamadas *Softmax*. Um exemplo deste método é aquele que usa a distribuição Boltzmann-Gibbs: temperaturas altas fazem com que as ações fiquem equiprováveis; e baixas temperaturas causam uma diferença maior, quando

da seleção de probabilidade, entre as ações com valores de predição diferentes. Se a temperatura tender a zero, a seleção de ação será pela que tiver maior valor de predição, ou seja, seleção gulosa (SUTTON, BARTO 1998).

Sendo assim, para a seleção de ação, (LIN 1993) usa a seleção de ação probabilística baseada na distribuição Boltzmann-Gibbs. A probabilidade de selecionar a ação a quando a temperatura é T, é dada pela Equação (3.4):

$$P_n\left(S_n, a\right) = \frac{e^{\frac{Q_n\left(S_n, a\right)}{T}}}{\sum_{K\varepsilon ações} e^{\frac{Q_n\left(S_n, k\right)}{T}}} \qquad (3.4)$$

Para a arquitetura de (LIN 1993), o valor da temperatura é aumentado se o agente permanecer em uma mesma pequena área por longo período, e é reduzida a zero nas fases de teste, isto é, durante os testes a ação com o valor mais alto de utilidade é sempre a selecionada.

3.3.3 Dinâmica da Arquitetura de Controle Baseada em Comportamento

Antes de detalhar cada parte da arquitetura de (Gadanho 1999), será apresentado um panorama geral acerca da mesma. Em toda iteração, são feitos:

- Sistema Homeostático: cálculo das oito sensações e sentimentos, das quatro emoções, das sete influências emocionais e hormônios (Capítulo 2);
- Comportamentos: cálculo das variáveis usadas pelos três comportamentos (Seção 3.2);
- Detector de Eventos: determinação da detecção de um evento (Seção 3.3.4);
- Submódulo Adaptativo: cálculo das funções de utilidade de cada um dos comportamentos em resposta ao mundo atual (Seção 3.3.5);

- Submódulo Adaptativo: são guardados os valores dos sentimentos da iteração atual, assim como os da rede neural correspondente ao comportamento escolhido para execução na iteração atual – se, na iteração seguinte, um evento for detectado, estes valores serão utilizados para correção dos pesos da rede neural (Seção 3.3.5.1);
- Os motores direito e esquerdo do agente recebem valores de velocidade de acordo com o comportamento executado.

Se o Detector de Eventos (a ser explicado na Seção 3.3.4), que funciona como um mecanismo de atenção, detectar que algo significativo ocorre (um evento), o comportamento executado na iteração imediatamente anterior será avaliado, podendo ser novamente escolhido, ou não. Apenas neste momento há aprendizado, ou seja, quando um evento é detectado, os valores de utilidade (para o estado do mundo S_n) dos três comportamentos são calculados para que aquele que tiver maior valor seja usado na construção do valor alvo (Seção 3.3.5.1.1). Além disso, o valor de utilidade do comportamento executado na iteração anterior para a situação do mundo S_{n-1} é aprendido. Observe que apenas são corrigidos os pesos da rede neural artificial do comportamento específico executado por último. Enquanto o Detector de Eventos não detecta um evento, um mesmo comportamento é continuamente executado, e não há aprendizado. As redes neurais artificiais devem ser capazes de generalização, ou seja, de produzir valores de saída (valores de utilidade) coerentes para a execução de comportamentos em situações similares às dos comportamentos aprendidos.

A Figura 3.3 ilustra a arquitetura como um todo, que será explicada em detalhes no presente Capítulo, exceto pelo Sistema Homeostático, explicado no Capítulo 2.

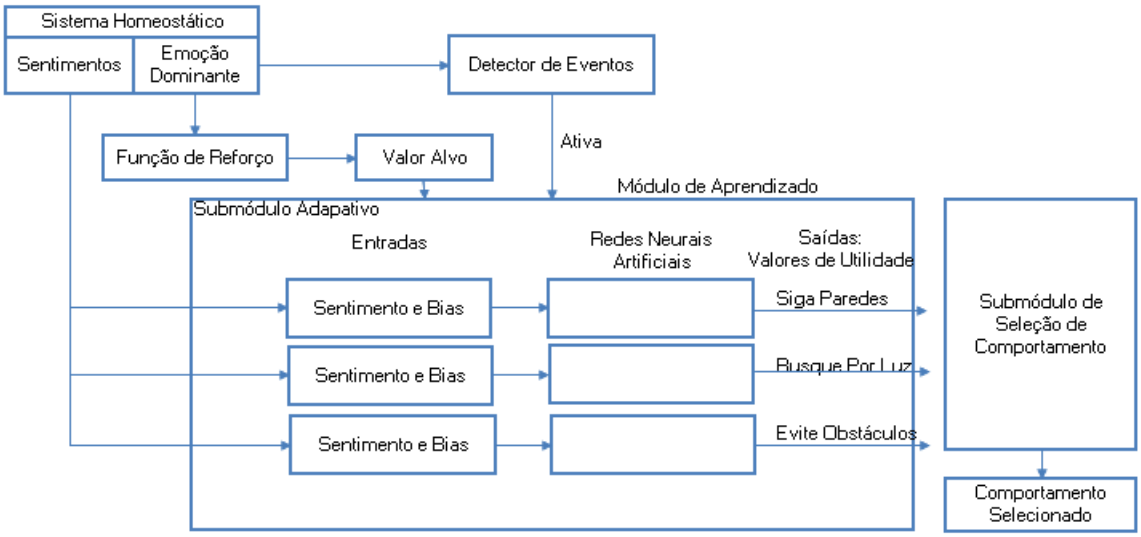


FIGURA 3.3 Esquema completo da Arquitetura de Controle Baseado em Comportamento

3.3.4 Detector de Eventos

Como a tarefa do agente deve ser realizada a partir da utilização de três comportamentos, é necessária a determinação do quando ativar o controle: do quando reavaliar o comportamento que tem sido executado até o estado presente e, se for o caso, selecionar um diferente. Para a concretização da tarefa de aprendizado com sucesso, (GADANHO 1999) observou a crucialidade de um mecanismo que ativasse o controle; além disso, considerou que, se as emoções pudessem prover uma estimação do estado do ambiente, seria razoável utilizá-las para ativar o controle. Dessa forma, baseando - se na suposição de que o estado emocional do agente refletiria a ocorrência de eventos significativos para este, (GADANHO 1999) delineou um mecanismo de detecção de eventos. Tal mecanismo consiste em ativar o controle (ou seja, ocorrer aprendizado) sempre que for encontrada alguma mudança expressiva no estado emocional do agente. O estado emocional do agente funciona, portanto, como um mecanismo de chamada de atenção a respeito de alterações relevantes do estado do ambiente.

Após a construção do mecanismo de Detecção de Eventos, (GADANHO 1999) o comparou experimentalmente a um outro mecanismo, que ativava o controle em intervalos

regulares (i.e. permitia o aprendizado em intervalos regulares). O primeiro mecanismo teria se mostrado melhor para o aprendizado do agente. Poder-se-ia perguntar: mas por que não ativar o controle em toda e qualquer iteração? Avaliar e selecionar um comportamento a cada iteração não é adequado, porque:

- Geralmente o comportamento precisa ser executado por iterações consecutivas;
- As diretrizes sequenciais dos comportamentos perderão sentido (já que não haverá uma sequência de execução), fazendo com que os comportamentos se pareçam mais com ações;
- Os comportamentos não serão avaliados corretamente porque será difícil atingir seu potencial de sucesso: o aprendizado do agente será limitado;
- Haverá gasto computacional desnecessário e inadequado.

Os comportamentos não precisam ser avaliados e selecionados a cada iteração, mas precisam ser avaliados após intervalos de tempo. Como estabelecer tais intervalos? Estes devem ser longos o suficiente para que o comportamento faça sentido, mas curtos o suficiente para que um comportamento inadequado em resposta ao estado do mundo não perdure – o que faria com que o nível de energia do agente diminuísse drasticamente.

Em (GADANHO 1999) várias tentativas foram feitas para que um intervalo ideal fosse encontrado, um que compatibilizasse a duração do comportamento e a dinâmica da interação do agente com seu ambiente. Dentre os intervalos testados (a cada 10, 35, 60 e 110 iterações), teria se adequado melhor o que detectava um evento a cada 35 iterações, pois levaria à obtenção de maiores valores de reforço e menor número de colisões, ao mesmo tempo em que permitiria manutenção de energia. Após definir um intervalo de detecção adequado, teriam sido comparados os controladores:

1. Ativa o aprendizado após intervalos fixos de iterações;
2. Ativa o controle em todas as iterações da simulação;
3. Ativa o controle apenas quando há uma mudança significativa no estado emocional.

O controlador do item 3 apresentaria níveis de emoção e de energia semelhantes aos do item 1, porém, seria melhor em evitar colisões, e aprenderia com um número menor de eventos. (GADANHO 1999) ressalta que o desempenho do controlador que ativa o aprendizado após intervalos fixos de tempo pode ser, em parte, dependente da tarefa, visto o agente dever persistir em seus comportamentos para se locomover de uma fonte de energia para outra.

Em suma, o mecanismo de Detecção de Eventos é o responsável por definir quando uma situação deve ser aprendida pelo agente (Figura 3.4).

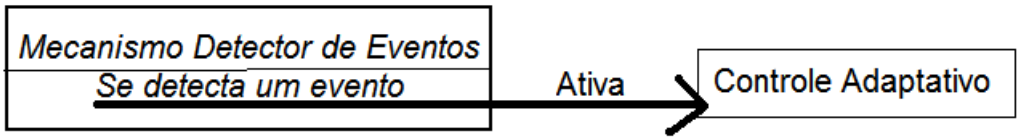


FIGURA 3.4 Detector de Eventos ativando o Módulo de Aprendizado.

Por exemplo, se o agente colidir em algum obstáculo, deverá aprender, através de um reforço negativo, que o comportamento que fez com que colidisse não deveria ter sido usado naquela situação; assim como, se por um acaso recarregar energia, deverá aprender, através de um reforço positivo, que o comportamento que fez com que recarregasse foi bem empregado. Não haverá aprendizado apenas quando emoções com valores altos surgirem, mas também quando houver uma variação repentina porque, se a emoção que tinha valor alto está diminuindo sua intensidade, a situação do mundo que causou tal emoção provavelmente já se dissolveu, sendo necessária a reavaliação do comportamento empregado.

3.3.4.1 Especificações para o Detector de Eventos e o papel do Sistema Homeostático

O mecanismo Detector de Eventos detecta um evento e ativa o Módulo de Aprendizado quando ocorrer pelo menos uma das seguintes instâncias:

1. A emoção dominante da iteração presente é diferente daquela da iteração imediatamente anterior (Equação 3.5).

$$ED_n \neq ED_{n-1} \qquad (3.5)$$

Onde: ED = Emoção Dominante;

$ED \in \{\text{Neutra, Felicidade, Medo, Tristeza, ou Raiva}\}$;

$n = \text{iteração atual};\ n-1 = \text{iteração anterior}\cdot$

2. É alcançado o valor máximo de dez mil iterações sem que qualquer evento seja detectado;
3. Sendo a emoção dominante da iteração presente a mesma que aquela da iteração imediatamente anterior, se o valor presente de emoção, após subtraído da média (HME) de intensidades de emoção do último período entre eventos, for maior do que um limiar permitido (LP = 0,02) e ξ vezes maior do que o desvio-padrão das intensidades de emoção do último período entre eventos (HDP). Este limiar permitido (LP) teria sido necessário para que fossem desprezadas variações de intensidade insignificantes; caso contrário, em situações em que o desvio padrão fosse bem pequeno, variações imperceptíveis seriam captadas pelo mecanismo de Detecção de Eventos.

A constante ξ em (GADANHO 1999) é considerada importante para o mecanismo de Detecção de Eventos: para valores de ξ menores ou iguais a *dois*, não faria muita diferença estabelecer um valor máximo de iterações sem aprendizado (item 2 acima); porém, para valores de ξ maiores, o limite de iterações sem aprendizado seria essencial porque, quanto maior o valor de ξ, menor a quantidade de eventos detectados, pois se torna mais difícil para

o mecanismo Detector de Eventos discriminar entre diferentes intensidades de uma mesma emoção.

Logo na primeira iteração, o valor da emoção dominante já é registrado e assim sucessivamente, até que um evento seja detectado; quando isso ocorrer, se o número de iterações sem detecção de eventos (i.e., número de iterações consecutivas sem aprendizado) for maior ou igual a dois, a média e o desvio-padrão desses valores gravados são calculados (Equação 3.8); e, novamente, são recalculadas, do zero, o desvio padrão e as médias. Em todas as iterações ocorre o cálculo descrito na Equação (3.6).

Cada período entre eventos pode conter qualquer número de iterações, porém sempre um número menor ou igual a dez mil (de acordo com o item 2 acima). Os eventos são contados de modo incremental: evento 1, evento 2, ..., evento n, evento n+1...

A Equação (3.7) serve para calcular, de acordo com o descrito no item 3 acima, se um evento será detectado, ou seja, se a situação requer aprendizado.

$$SED = \sum_{i=evento_n}^{evento_{n+1}} ED_n \tag{3.6}$$

$$DPaux = \sum_{i=evento_n}^{evento_{n+1}} \left(ED_n - HME\right)^2$$

Onde: SED = soma das Emoções Dominantes;

$\left(evento_n : evento_{n+1}\right)$ período entre eventos;

HME = *histórico da* média de intensidades de *ED* do último período entre eventos: $\left(evento_{n-1} : evento_n\right)$

DPaux = variável auxiliar para o cálculo do desvio − padrão, *vide E*quação 3.8

$$DMH_n = |ED_n - HME_{evento_{n-1}:evento_n}| \tag{3.7}$$

Considera − se que um evento é detectado se:

$$\left\{\text{período entre eventos} \geq 2 \wedge DMH > LP \quad \wedge \quad DMH > \left(\xi \times HDP\right)\right\}$$

Onde: LP = limiar permitido, $LP = 0{,}02$;

HDP = Histórico do desvio padrão do último período anterior a um evento;

DMH = módulo da diferença entre a ED atual e a média de EDs guardadas no último período entre eventos: $\left(evento_{n-1} : evento_n\right)$.

Se um evento tiver sido detectado, seja pelos itens 1, 2 ou 3 acima, é feito o cálculo para guardar os históricos emocionais (Equação 3.8).

(3.8)

$$\text{Se período, iterações entre eventos} \geq 2 \begin{cases} \text{HDP} = \sqrt{\dfrac{\text{DPaux}}{\left(\text{evento}_{n+1} - \text{evento}_n\right) - 1}} \\ \text{HME} = \dfrac{\text{SED}}{\text{evento}_{n+1} - \text{evento}_n} \end{cases}$$

$$\text{Se } n = 0: \quad \begin{cases} \text{HDP} = 0 \\ \text{HME} = 0 \cdot \end{cases}$$

Quando um evento é detectado, o sentimento Ansiedade é zerado e o estado emocional do agente é novamente calculado. A emoção dominante resultante deste cálculo constará como a primeira emoção a fazer parte do período entre eventos que se inicia.

O expediente de reiniciar o sentimento Ansiedade pode parecer estranho, mas, segundo (GADANHO 1999), haveria coerência porque seria comum o desaparecimento deste sentimento quando há a reavaliação dos comportamentos tomados até então e é feita uma nova decisão - mesmo quando esta se resume a reiterar o comportamento anteriormente tomado.

3.3.5 Módulo de Aprendizado

O Módulo de Aprendizado é formado por dois submódulos:

- Submódulo Adaptativo (de Memória Associativa): utiliza o algoritmo *Q-learning* para aprender uma política de ações sobre uma arquitetura baseada em redes neurais artificiais que estima os valores de utilidade. Os pesos sinápticos das redes são atualizados através do algoritmo de Retropropagação.
- Submódulo de Seleção de Comportamento: levando em consideração os sentimentos do agente e as avaliações previamente recebidas, seleciona um dentre três possíveis comportamentos: C= {Evite Obstáculos; Busque por Luz; Siga Paredes}.

Na Figura 3.5, adaptada de (GADANHO 1999), é apresentado um esquema geral do Módulo de Aprendizado.

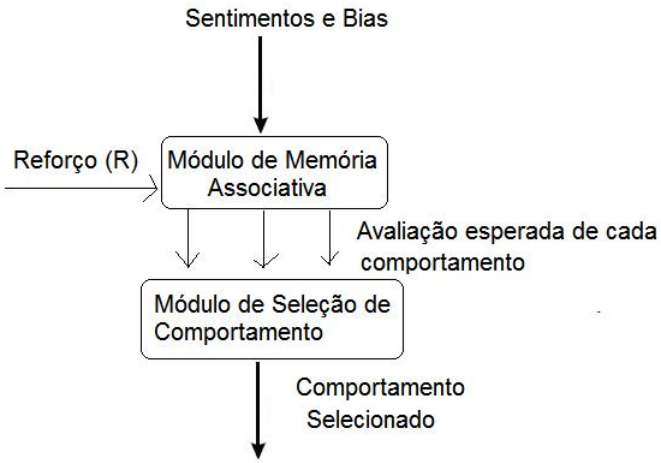


FIGURA 3.5 Esquema geral do Módulo de Aprendizado - figura adaptada de (GADANHO 1999)

3.3.5.1 Submódulo Adaptativo (Memória Associativa)

Este submódulo é constituído por três redes neurais artificiais, cada uma correspondendo a um comportamento, ou seja, uma rede para Busque por Luz, outra para Evite Obstáculos e, finalmente, outra para Siga Paredes. As redes são usadas para associar os sentimentos do agente aos valores esperados quando da execução dos comportamentos; por exemplo, a rede responsável pelo comportamento Evite Obstáculos fornecerá um valor do quão bom ou ruim será o estado resultante S_{n+1} após a execução deste comportamento em resposta ao estado do mundo atual S_n, ou seja, após a transição $T\left(S_n, c, S_{n+1}\right)$.

As redes neurais são sem realimentação, com três camadas, e apresentam as características:

9 unidades de entrada, uma para cada sentimento, e uma para bias;

1 camada escondida com 15 unidades. Em (GADANHO 1999) relata-se que teria sido experienciado que as redes neurais com 10 ou 15 unidades na camada escondida teriam

apresentado os melhores resultados sem aumentar muito o tempo de processamento para cada iteração de aprendizado – teriam sido testadas também redes com 5, 6, 8 e 25 unidades. A Tabela 3.1 sumariza a estrutura das redes neurais artificiais.

Entrada:	Camada escondida:	Saída:
9 unidades: Bias e 8 Sentimentos	1 camada com 15 unidades	1 unidade

TABELA 3.1. Estrutura das redes neurais artificiais.

Os pesos entre a camada escondida e a saída são inicializados com valores aleatórios; já os pesos entre a camada de entrada e a escondida são iniciados com zero, para que as redes possam prover uma avaliação inicial neutra.

A unidade de saída produz a utilidade prevista pela rede neural se o comportamento por ela representado for executado em resposta ao estado do mundo atual. Ou seja, são três redes, portanto três saídas, cada uma modelando a função de utilidade correspondente ao comportamento representado pela rede, de acordo com a Equação (3.9).

$$Q_n\left(S_n, c\right) = R_{n+1} + \gamma\, eval\left(S_{n+1}\right) \tag{3.9}$$

Onde: $c \in C$; $C = \{\text{Busque por Luz, Evite Obstáculos, Siga Paredes}\}$

$\gamma = 0{,}9$, o fator de desconto ;

R_{n+1} = reforço imediato recebido na iteração $n + 1$;

$eval\left(S_{n+1}\right)$ = o reforço cumulativo descontado a partir do estado S_{n+1} após a execução do comportamento c no estado S_n e supondo que o agente, a partir de então, segue uma política ótima ·

A função de ativação utilizada, tanto na camada escondida quanto na de saída, foi a tangente hiperbólica, Equação (3.10).

$$\tanh\left(\beta x_j\right) = \frac{1 - e^{-2\beta \cdot x_j}}{1 + e^{-2\beta \cdot x_j}} \qquad (3.10)$$

Onde: $\beta = 0{,}25$

onde x_j é o campo local induzido, ou seja, a soma ponderada das entradas e bias, do neurônio j, e $\tanh\left(\beta x_j\right)$ é a saída do neurônio.

As entradas são os oito sentimentos $Sto1_n$, $Sto2_n$, …, $Sto8_n$ e bias e, a saída, o valor de utilidade ($Q_n\left(S_n, c\right)$, $c \in Comportamentos$) se o comportamento específico c representado pela rede neural for aplicado em resposta aquele estado do mundo $S_n = \left[Sto1_n, Sto2_n, \ldots, Sto8_n \right]$. Sendo três redes/comportamentos, têm-se três valores de utilidade.

3.3.5.1.1 Valor Alvo

Ao mapearem-se as emoções dominantes como o valor do reforço, significa-se que o agente avalia a eficiência da sua política através do seu próprio Sistema Homeostático: sua emoção dominante lhe fornece uma noção acerca da eficiência de sua política, ou seja, sua emoção dominante ministra uma medida de aprendizado. A conotação negativa atual (iteração n) de uma emoção dominante indica para o agente que a decisão passada, a execução do comportamento x para o estado de mundo da iteração anterior n-1, foi errônea. É válido o mesmo raciocínio para o caso de se terem emoções dominantes Neutra ou Positiva.

O valor alvo fornecido à rede neural é dado pela Equação (3.11).

$$\text{Alvo}\left(\mathrm{S_{n-1}}, \mathrm{c}\right) = \mathrm{R_n} + \gamma \max\left\{\mathrm{Q_n}\left(\mathrm{S_n}, \mathrm{k}\right) \mid \mathrm{k} \in \text{comportamentos}\right\} \tag{3.11}$$

Onde: $\gamma = 0{,}9$, o fator de desconto;

$\max \mathrm{Q_n}\left(\mathrm{S_n}, \mathrm{k}\right) =$ o comportamento que apresenta o valor de utilidade mais alto para a iteração atual n;

$\mathrm{R_n} =$ Reforço na iteração n, sendo que o reforço é a ED, Emoção Dominante.

$Alvo\left(s_{n-1}, c\right)$ é o valor de utilidade que a rede neural que teve seu comportamento associado executado na iteração anterior n-1 deveria ter predito como saída dado o vetor de entrada. Por exemplo: independentemente de ter sido detectado um evento ou não na iteração n-1, o comportamento executado foi Evite Obstáculos e, nesta mesma iteração n-1, foram guardados a rede, o vetor de entrada da rede, assim como o valor de utilidade predito deste comportamento em resposta ao cenário $S_{n-1} : Q_{n-1}\left(S_{n-1}, \textit{Evite Obstáculos}\right)$. Na iteração n, os valores de utilidade dos três comportamentos são calculados: $\left\{Q_n\left(S_n, \textit{Evite Obstáculos}\right), Q_n\left(S_n, \textit{Busque por Luz}\right), Q_n\left(S_n, \textit{Siga Paredes}\right)\right\}$, um evento é detectado (deverá haver aprendizado) e, por isso, a rede responsável pelo comportamento Evite Obstáculos deverá ter seus pesos corrigidos. É feita uma comparação entre os três

valores de utilidade porque, o maior valor, será o $\max\left\{Q_n\left(S_n, k\right) \middle| k \in \textit{comportamentos}\right\}$ usado no cálculo do valor alvo para a iteração anterior n-1. Ou seja, a correção é "atrasada": na iteração atual se calcula o valor alvo da iteração anterior n-1 e, na iteração atual n, os pesos da rede associada ao comportamento executado são corrigidos.

Ao receber o valor alvo, a rede neural deverá ter seus pesos corrigidos para que a sua saída se aproxime do valor alvo. Essa correção é feita através do algoritmo de Retropropagação com taxa de aprendizado α de 0,3 e *momentum* de 0,9.

Ao analisar a Equação (3.11) do valor alvo, nota-se que este variará entre -1,9 e 1,9: as emoções dominantes, que são o reforço, variam entre -1 e 1 e os valores de utilidade dos comportamentos, fornecidos pelas redes, estão no mesmo intervalo, logo, os valores máximo e mínimo são como descritos na Equação (3.12).

$$\begin{aligned} &Alvo_n(s_{n-1}, c) = -1 + 0{,}9 \times (-1), \quad Alvo_n(s_{n-1}, c) = -1{,}9 \\ &\qquad\qquad ou \\ &Alvo_n(s_{n-1}, c) = 1 + 0{,}9 \times 1, \qquad\quad Alvo_n(s_{n-1}, c) = 1{,}9 \end{aligned} \tag{3.12}$$

Porém, ao utilizar a função de ativação tangente hiperbólica sobre a unidade de saída, os valores de saída aprendidos pelas redes neurais artificiais estarão entre -1 e 1. Para se adequar a isso, os valores alvo fornecidos à rede quando de seu aprendizado precisam ser normalizados para este intervalo.

O algoritmo de Retropropagação consiste em se apresentar à rede o valor alvo $Alvo(s_{n-1}, c)$, Equação (3.11), o qual é comparado à saída $Q_{n-1}\left(S_{n-1}, c\right)$ produzida, gerando o erro. Cada peso sináptico é então corrigido proporcionalmente ao negativo da derivada parcial do erro de sua unidade em relação ao peso, sendo este processo retropropagado até os pesos entre a camada escondida e as entradas.

Na iteração n é possível calcular o erro da saída da rede da iteração anterior, n-1. O erro da unidade de saída (ES) se dá pela subtração do valor alvo pelo valor de utilidade

predito pela rede, multiplicado pela derivada da função de ativação da saída, tal como descrita na Equação (3.13).

$$ES_n = \left(\text{Alvo}\left(S_{n-1}, c\right) - Q_{n-1}\left(S_{n-1}, c\right)\right) \times \tanh'\left(\beta x_j\right) \tag{3.13}$$

$$\text{Onde: } \tanh'\left(\beta x_j\right) = \beta \times \left(1 - Q_{n-1}(s_{n-1}, c)\right)^2$$

ES = erro da unidade de saida

Lembrando que $\beta = 0{,}25$

Após calcular o erro da unidade de saída, os pesos entre a camada escondida e a saída devem ser corrigidos e o erro deve ser retropropagado para que sejam corrigidos os pesos entre as entradas e a camada escondida. A Equação (3.14) mostra como o erro é calculado e como são corrigidos os pesos (uso da taxa de aprendizado e do termo de momentum) em todas as camadas e unidades da rede.

(3.14)

$$\text{Para: } i = 1{:}15 \left\{ \begin{array}{l} \Delta wb = \alpha \times CA(i)_{n-1} \times ES + \left(\mu \times MS(i)_{n-1}\right) \\ MS(i)_n = \Delta wb \\ DV = \beta \times \left(1 - \left(CA(i)_{n-1}\right)^2\right) \\ EU = ES \times PS(i)_{n-1} \times DV \\ PS(i)_n = PS(i)_{n-1} + \Delta wb \\ \text{Para } j = 1{:}9 \left\{ \begin{array}{l} \Delta wa = \left(\alpha \times EU \times Sto\left(j\right)_{n-1}\right) + \left(\mu \times ME\left(i,j\right)_{n-1}\right) \\ ME\left(i,j\right)_n = \Delta wa \\ PE\left(i,j\right)_n = PE\left(i,j\right)_{n-1} + \Delta wa\left(i,j\right) \end{array} \right. \end{array} \right.$$

Onde: $\alpha = 0{,}3$: taxa de aprendizado; $\mu = 0{,}9$: termo de momentum;

$i = 1, \ldots, 15$; pois são 15 unidades na camada escondida;

$j = 1, \ldots, 9$; pois são 8 sentimentos e bias;

Δwa e Δwb: gradiente descendente incremental;

$ME\left(i,j\right)$ o *momentum* do peso de entrada i,j;

$MS\,(i)$ o *momentum* do peso de saída i;

$CA_z\,(x)$ = valor de saída do $x-$ézimo neurônio da camada escondida no instante de tempo z;

EU = erro da unidade da camada escondida;

n = iteração atual; $n-1$ = iteração anterior.

Este processo de aprendizado ocorrerá sempre que um evento for detectado, e apenas em uma rede neural por evento, a que estiver associada ao comportamento executado na iteração imediatamente anterior. O critério de parada é o número total de iterações para um experimento completo, ou seja, 1.200.000 (para maiores esclarecimentos sobre os experimentos, vide Capítulo 4).

Sumarizando, o agente, através de seus atuadores (motores esquerdo e direito) interage com seu ambiente e o percebe através de suas sensações, que alimentam seu Sistema Homeostático. Este provê uma emoção dominante que, dependendo de sua intensidade, fará com que um evento seja detectado - além de prover o reforço.

Em todas as iterações o vetor de entradas é guardado, assim como o valor de utilidade calculado (iteração n) pela rede neural artificial associada ao comportamento executado. São também armazenados os valores de saída da camada escondida, pois, se um evento for detectado na próxima iteração n+1, a rede que terá seus pesos atualizados necessitará desses valores.

Se um evento não for detectado não há aprendizado, e o comportamento executado na iteração atual n será o mesmo que o da iteração anterior n-1.

Se um evento for detectado, o Mecanismo Detector de Eventos ativará o Módulo de Aprendizado, o sentimento Ansiedade e seu Hormônio associado serão ambos reiniciados e o Sistema Homeostático será reavaliado. Através do algoritmo *Q-learning* a rede neural responsável pelo comportamento mais recentemente executado receberá o valor que deveria ter produzido na saída e, através do algoritmo de Retropropagação, corrigirá os seus pesos para se adequar ao valor alvo. Após a correção, é calculado novamente o valor de utilidade do comportamento associado à rede neural que foi corrigida. Por fim, os valores de utilidade fornecidos pelas três redes neurais artificiais para os três comportamentos são passados para o Submódulo de Seleção de Comportamento.

3.3.5.2 Submódulo de Seleção de Comportamento

Quando um evento é detectado, o Submódulo de Seleção de Comportamento recebe os valores de utilidade para os três comportamentos em resposta ao estado do mundo atual $S_n : \left\{ Q_n\left(S_n, Evite\ Obstáculos\right), Q_n\left(S_n, Busque\ por\ Luz\right), Q_n\left(S_n, Siga\ Paredes\right) \right\}$.

Este submódulo apenas é ativado, isto é, seleciona um comportamento, se um evento tiver sido detectado; caso contrário, o comportamento selecionado quando do último evento, será continuamente executado. Se ativado, é feita uma seleção estocástica do comportamento a ser executado baseando-se na distribuição Boltzmann-Gibbs. Para a temperatura T, a probabilidade de selecionar o comportamento $k \in Comportamentos; k = \left\{ Busque\ por\ Luz,\ Evite\ Obstáculos,\ Siga\ Paredes \right\}$ é dada pela Equação (3.15).

$$P_n\left(S_n, comportamento\right) = \frac{e^{\frac{Q_n\left(S_n,\ comportamento\right)}{T}}}{\sum_{K \varepsilon comportamentos} e^{\frac{Q_n\left(S_n,\ k\right)}{T}}} \qquad (3.15)$$

Quando se diminui a temperatura, a probabilidade se concentra em um subconjunto menor dos estados de menor energia; assim, se o agente permanecer em uma mesma pequena área por longo período, o valor da temperatura é aumentado – para que explore seus comportamentos e saia daquela pequena região a fim de encontrar outras fontes de energia pelo seu ambiente. A forma de medir, nesta tese, se o agente está em uma mesma pequena região há muito tempo, se deu através da sensação correspondente ao sentimento Ansiedade, ou seja, se, após cem iterações consecutivas, a média de tal sensação se apresentar alta.

Em (GADANHO 1999), determinou-se como sendo cem o número de iterações para certificação da distância percorrida pelo agente. Este número de cem iterações teria sido escolhido para que fosse pequeno o suficiente para medir a maior parte do movimento do

agente e, ao mesmo tempo, grande para evitar que fossem equiparadas situações em que o agente está se movendo velozmente em uma pequena região à quando está realmente cobrindo uma região mais extensa.

3.4 Modificações sobre a Arquitetura

3.4.1 Modificação no Submódulo de Seleção de Ação

Dentre os métodos de seleção de ação, como alternativa à política Boltzmann-Gibbs, um dos mais simples seria o que tem por regra o selecionar uma das ações com maior predição de retorno (escolha gulosa). Este método reitera ações com valores de predição mais altos, perscruta (tradução para *exploitation* (BARRETO 2008)) o conhecimento atual visando maximizar recompensas imediatas, sem haver gasto de tempo ao experimentar ações aparentemente inferiores para o estado do mundo. Uma alternativa diferente seria o $p - greedy$: se portar de modo "guloso", ou seja, buscando sempre as ações com maiores predições de sucesso e, à vezes, com probabilidade P, selecionar uma ação aleatória, explorando o ambiente (se a ação aleatória for diferente da que tem maior avaliação). No limite, com número considerável de execuções sob este método de seleção de ação, todas as ações serão experimentadas um número infinito de vezes, assegurando que a escolha de ação convirja para a escolha da ação ótima [Watkins, 1986]. Esta alternativa foi testada, ou seja, o uso da arquitetura de (GADANHO 1999) com modificações no Submódulo de Seleção de Comportamento para uma seleção gulosa e, com probabilidade P, seleção de ação aleatória. Os resultados serão mostrados no Capítulo 4.

3.4.2 Modificação no Submódulo Adaptativo (Memória Associativa)

Em (GADANHO 1999), na seção Trabalhos Futuros, há a ponderação de se aperfeiçoar o Modelo Emocional, pois a influência das emoções sobre a memória não teria sido explorada. A sugestão seria a de se associar cada uma das quatro Emoções a diferentes mecanismos de atenção, mais especificamente, a redes neurais para se computarem os valores de utilidade de cada um dos três comportamentos. O agente então seria projetado para se

lembrar das experiências associadas aos estados emocionais, o que produziria uma categorização dos eventos memorizados de acordo com o tipo de problema.

A partir desta sugestão, fizeram-se alterações no Submódulo Adaptativo da arquitetura. Ao invés de uma rede neural para cada um dos três comportamentos, desenvolveram-se quatro categorias (uma para cada possível emoção) para cada um dos comportamentos, totalizando doze redes neurais ao total.

Cada rede corresponde a uma emoção quando dominante: se o agente apresenta maior valor para a emoção Felicidade, qual o melhor comportamento a ser ativado para que assim se mantenha ou potencialize tal emoção? E se apresenta maior valor para a emoção Tristeza (sinalizando estar quase sem energia), qual comportamento ativar para não gastar muito da energia que resta, mas, ainda assim, se locomover a fim de obter mais energia? Sendo assim, a cada iteração, de acordo com a Emoção Dominante, três redes neurais artificiais calculam os valores de utilidade dos comportamentos em resposta ao estado do mundo atual.

São quatro as emoções, mais a ausência de emoção dominante, isto é, a emoção Neutra. A emoção Tristeza, ocorrente apenas quando o agente está com pouca energia, quase nunca ocorre para os agentes que aprendem a sua tarefa, pois, uma vez exaurida a energia do agente, diz-se que este experimento em particular falhou. Mesmo assim o agente está susceptível a momentos com pouca energia e, quando ocorre, deve saber rapidamente o que fazer porque, ao acabar sua energia, acaba também a simulação. O problema em se ter uma rede para a emoção Tristeza é que esta rede pode ficar subtreinada. Assim, por simplicidade, a emoção Tristeza passou a ser tratada juntamente com a Neutralidade, ou, ausência de emoção dominante.

Para comparação, a Figura 3.6 ilustra um esquema representativo desta arquitetura Modificada e a Figura 3.3 da arquitetura proposta em (GADANHO 1999) é novamente apresentada. Resultados comparativos são apresentados no Capítulo 4 a seguir.

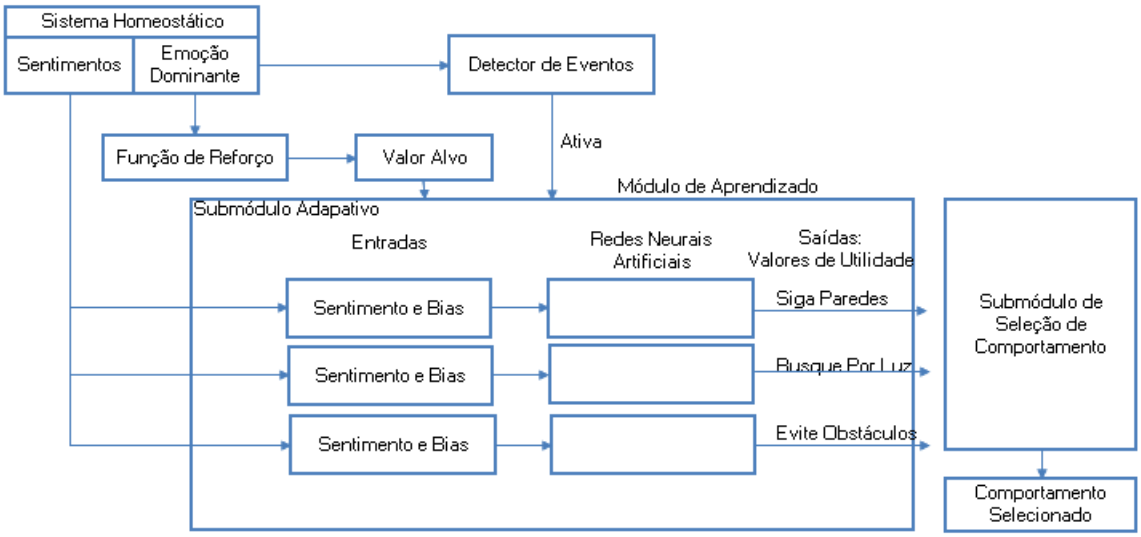


FIGURA 3.3 Esquema completo da Arquitetura de Controle Baseado em Comportamento.

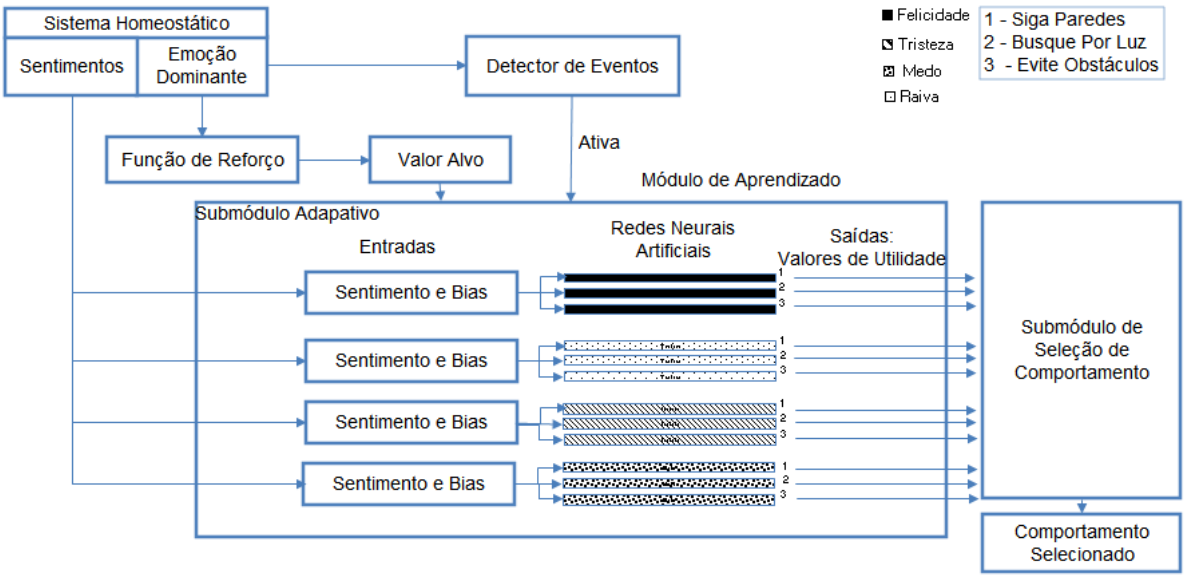


FIGURA 3.6 Esboço da Arquitetura Modificada.

4 Resultados Experimentais

A Arquitetura de Controle Baseado em Comportamento descrita em (GADANHO 1999) foi pesquisada e reproduzida. Doravante ambas, a arquitetura original, cujos gráficos podem ser vistos em (GADANHO 1999) e a aqui reproduzida (gráficos no capítulo 4) serão respectivamente denominadas: controlador Original e controlador Base. As modificações sobre a arquitetura do controlador Base (descritas na Seção 3.4) geraram um terceiro controlador, o controlador Modificado. Para que houvesse um parâmetro de comparação de desempenho entre os controladores, foi construído um controlador para referência, o controlador Referência**.** Na Seção 4.1 serão detalhados os controladores e, também, apresentadas duas tabelas (Tabelas 4.1 e 4.2) que sumarizam o conjunto de controladores considerados.

4.1 Controladores Usados e Comparados nos Experimentos

4.1.1 Controlador Base

Trata-se da implementação da arquitetura de Controle Baseado em Comportamento de (GADANHO 1999). Esta arquitetura está explicada ao longo desta tese da seguinte maneira: no Capítulo 2, apresenta-se o Sistema Homeostático, parte integrante da arquitetura e, no Capítulo 3, a arquitetura como um todo. Esta arquitetura de Módulo de Aprendizado, que motivou e que foi reproduzida na presente tese, também foi denominada, em (GADANHO 1999), “Emocional” porque a função de reforço utilizada advém de uma rede neural de realimentação fortemente motivada fortemente influenciada pela hipótese do Marcador Somático e pelas considerações acerca de emoções e sentimentos descritos em (DAMÁSIO 1994), e que busca simular processos homeostáticos naturais. A partir da ligação entre sensação, sentimento, emoção e hormônio artificiais, se obtém uma resposta emocional, a

qual, por fim, é utilizada para saber o quão bem sucedido foi o comportamento aplicado pelo agente.

4.1.2 Controlador Referência

Um segundo tipo de controlador, em que não há aprendizado, foi projetado para coordenar adequadamente os seus três comportamentos em resposta à sua interação com o ambiente, recebendo, assim, os maiores reforços possíveis – porém, deve ser notado que este controlador servirá apenas como referência, dado que seu desempenho não seja garantidamente ótimo. Este controlador, apesar de não aprender, possui o mesmo Sistema Homeostático (Capítulo 2) que os outros controladores. Assim, seu desempenho pode ser comparado ao dos que aprendem e, também, pode ser conhecida uma medida de maximização de recompensas (na verdade, emoção dominante com conotação positiva, ou seja, Felicidade) para avaliar o agente de aprendizado – admitindo que este controlador pré-projetado receba os melhores reforços possíveis. Este controlador foi denominado "Referência" e o correspondente de (GADANHO 1999), *Hand-crafted.* A intenção para a construção deste controlador "desenhado à mão" é a comparação, o saber o quão bem sucedido poderia ser o controlador Base ao final do seu aprendizado. Detalhes a respeito da implementação do controlador Referência podem ser encontrados no Apêndice.

4.1.3 Controlador Modificado

Este controlador tem arquitetura semelhante a do controlador Base, exceto pelo Submódulo Adaptativo, equipado com nove redes neurais artificiais a mais do que o primeiro (cada associando uma emoção a um comportamento) – controlador explanado na Seção 3.4.2.

4.1.4 Controladores *p-greedy*

Na Seção 3.4.1, foi mencionada a estratégia *p-greedy*, que consiste na escolha do comportamento com maior predição de sucesso e, com probabilidade p, de um comportamento de forma aleatória. Regras de seleção *Softmax* (Seção 3.3.2) são mais adequadas neste contexto, pois um comportamento erroneamente selecionado traz consequências realmente ruins, como o agente não conseguir recarregar sua energia ou colidir definitivamente em um obstáculo, produzindo influências negativas sobre o Sistema Homeostático através dos hormônios. Porém, para certificação da real vantagem e influência deste tipo de seleção de comportamento sobre o aprendizado do agente, foram construídos controladores sob uma modificação da estratégia *p-greedy*, descrita a seguir. A comparação entre estratégias diferentes permite que se relacionem suas correspondentes táticas de exploração e perscrutação. Para isso, os Submódulos de Seleção de Comportamento dos controladores Modificado e Base sofreram uma alteração, gerando outros dois controladores: P – Base e P – Modificado.

A modificação em ambos consiste em optar sempre pelo comportamento guloso (ou seja, o que apresentar maior predição de utilidade em resposta ao estado do mundo), perscrutando ao máximo o conhecimento adquirido; porém, se o agente permanecer em uma mesma pequena região por muito tempo (100 iterações consecutivas, como mencionado na Seção 3.3.5.2), a escolha será aleatória, dando a possibilidade de exploração.

A Tabela 4.1 sumariza o conjunto de controladores desenvolvidos e, a Tabela 4.2, por sua vez, o conjunto de controladores apresentados em (GADANHO 1999).

Controlador:	Aprendizado?	Estratégia de Exploração:
Referência (= Hand-crafted, GADANHO 1999)	Não	-
Base (= Original, GADANHO 1999)	Sim	Boltzmann-Gibbs
P – Base	Sim	P-greedy
Modificado	Sim	Boltzmann-Gibbs
P – Modificado	Sim	P-greedy

TABELA 4.1 - Conjunto de controladores desenvolvidoss.

Controlador:	Aprendizado?	Estratégia de Exploração:
Original (GADANHO 1999)	Sim	Boltzmann-Gibbs
Hand-crafted (GADANHO 1999)	Não	-

TABELA 4.2 - Conjunto de controladores considerados.

4.2 Resultados

Nos experimentos de (GADANHO 1999), trinta diferentes robôs de uma mesma arquitetura eram colocados em trinta posições iniciais aleatórias, estando cada robô com carga completa de energia e todos os outros valores iniciais zerados. Nos experimentos aqui reportados, as posições iniciais de *todos* os robôs foram sempre as mesmas e todas as variáveis dependentes de valores aleatórios foram iniciadas com sementes aleatórias.

Na Figura 4.1 estão identificados o mundo utilizado em todos os experimentos, bem como as fontes de energia e a posição inicial dos robôs. Os experimentos foram feitos em um simulador de robôs Khepera, o WSU (PERRETTA, GALLAGHER, 2003) e em Linguagem Java (para detalhes sobre o simulador, vide o Apêndice).

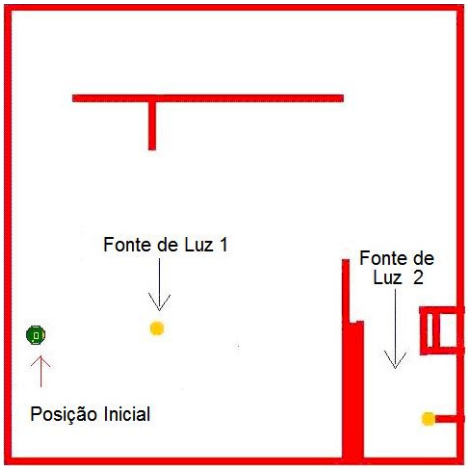


FIGURA 4.1 Mundo do agente no simulador WSU, com indicação da posição inicial do robô e das fontes de energia.

Os dados coletados e graficamente apresentados na Seção 4.2 foram assim considerados:

- Controlador: constarão nos gráficos, sob o nome de cada controlador específico, as médias, por período, do conjunto de trinta simulações diferentes e bem sucedidas de uma mesma arquitetura. Ou seja, cada controlador específico opera sobre trinta diferentes robôs que nunca atingiram nível zero de energia e não colidiram em obstáculo de forma definitiva (simulações que obtiveram sucesso).
- Período: quarenta mil iterações consecutivas.
- Simulação: cada simulação compreende trinta períodos de quarenta mil iterações cada, totalizando um conjunto de 1.200.000 iterações consecutivas.
- Média por período: a média (de alguma grandeza) sobre todas as trinta simulações bem sucedidas, em uma mesma arquitetura, no período considerado.
- Simulação mal sucedida ou falha: simulação em que houve colisão definitiva ou gasto completo de energia.
- Simulação bem sucedida: aquela em que não acabou a energia do robô e, também, em que não houve colisão tal que o robô ficasse terminantemente impedido de se locomover.

No caso do controlador Referência, quase todas as simulações são bem sucedidas: como já é desenhado para tirar o maior proveito possível de um ambiente que tenha obstáculos e fontes de energia, apenas falha quando colide permanentemente em obstáculo, o que ocorreu em cinco dos experimentos. Como o conjunto de execuções bem sucedidas por controlador terá trinta simulações diferentes, foram feitas trinta e cinco simulações: cinco falharam e trinta foram bem sucedidas.

Já no caso dos outros controladores, bem sucedido significa aprender a adquirir energia. Como as fontes de energia têm uma quantidade limitada de itens, o agente também precisa aprender a procurar por diferentes fontes de energia. A simulação fracassa quando um agente fica completamente sem energia, pois não conseguiu aprender a adquiri-la em tempo, ou se colidir permanentemente em um obstáculo. Para o índice de simulações bem sucedidas dos controladores Base e Modificado, Seção 4.2.2.

O eixo x, dos gráficos deste capítulo, em geral enumerará cada um dos trinta períodos das simulações. Já o eixo y, a média, por controlador (sendo trinta simulações por controlador), em cada um dos trinta períodos. Além disso, inexiste distinção entre as fases de aprendizado e a de desempenho, visto que um robô autônomo deve se adaptar continuamente ao seu ambiente (GADANHO 1999). No decorrer deste capítulo serão apresentados, para comparação qualitativa, alguns gráficos traduzidos e adaptados de (GADANHO 1999); porém, precisa ser notado que, nestes gráficos, uma simulação completa compreende sessenta períodos de cinquenta mil iterações cada.

4.2.1 Aspectos Gerais do Aprendizado da Seleção de Comportamento

Se se observarem os comportamentos disponíveis (Seção 3.2), perceber-se-á que apenas um, o Seguir Paredes, permite que o agente percorra o seu ambiente. Se os outros dois comportamentos forem selecionados erroneamente e repetidas vezes, o Sistema Homeostático ficará saturado pelo sentimento Ansiedade, que sinaliza que o agente não tem percorrido

grandes distâncias. Provavelmente esta seria uma razão para que o sentimento de Ansiedade e seu hormônio associado sejam zerados sempre que o Módulo de Aprendizado é ativado – caso contrário, os comportamentos herdariam o reforço negativo saturado pela Ansiedade e, assim, ocorreria aprendizado errôneo. Observe ainda que, se o agente estiver em uma mesma pequena área por muito tempo, aumenta a probabilidade de que qualquer um dos comportamentos seja escolhido.

Se o agente se mantiver em situação de colisão, o sentimento Dor apresentará valores altos e, dependendo do tempo em que ficar sem se mover ao não ativar o comportamento correto, que seria o Evite Obstáculos, o sentimento Ansiedade também aumentará. O agente deve aprender rapidamente que situação de colisão somente é boa se houver detecção de energia (ou seja, sentimento Aquecido com valores altos) e, uma vez que tal colisão tenha ocorrido, acionar logo em seguida o comportamento Evite Obstáculos para que seus sensores de luminosidade traseiros recebam valores altos de energia. Deve, então, permanecer parado enquanto houver energia disponível, ou seja, não ativar o comportamento Seguir Paredes neste momento, pois este faria com que se distanciasse da fonte de energia. Quando não houver mais energia disponível, deve aprender a se movimentar para que encontre outras fontes de energia. Porém, como o gasto de energia é proporcional à velocidade dos motores, os agentes de aprendizado tendem a evitar velocidades altas para guardar energia. Se o agente demorar muito para explorar seu ambiente, sua energia irá diminuir e os reforços negativos, suas emoções dominantes, alimentadas pelos sentimentos Ansiedade, Fome e Dor saturarão o Sistema Homeostático e farão com que as redes neurais artificiais aprendam padrões errados e, talvez, não consigam corrigir os seus pesos antes que o agente atinja nível zero de energia – se isso acontecer, a simulação falha.

4.2.2 Avaliação do Desempenho: Sucesso

As arquiteturas modificadas gastam quase 15% a mais de tempo para concluir o experimento completo do que as outras arquiteturas, pois são baseadas na adaptação de pesos de um conjunto maior de redes neurais artificiais. Para os experimentos feitos, a vantagem da arquitetura Modificada se dá em relação à taxa de sucesso quando do uso da estratégia *p-greedy* para seleção de comportamentos. Dos controladores P-Base, aproximadamente 30% não conseguiram concluir a tarefa. Já em relação ao P-Modificado, todos concluíram com sucesso, com exceção de um, que colidiu de forma definitiva em um obstáculo. Ou seja, a estratégia *p-greedy* para a arquitetura original não é uma boa opção; porém, o uso de mecanismos de aprendizado individualizados para as emoções permitiram que o índice de sucesso aumentasse. Entre os controladores Modificado e Base, isto é, que usam a estratégia de seleção Boltzmann-Gibbs, praticamente não há diferenças e o índice de agentes que concluíram com sucesso a tarefa de aprendizado é aproximadamente o mesmo: 85%. A Tabela 4.3 sumariza a taxa de sucesso dos controladores testados.

	P-Base	Base	P-Modificado	Modificado
Sucesso	≅70%	85%	≅ 100%	85%

TABELA 4.3 – Taxa de sucesso do conjunto de controladores testados.

4.2.3 Avaliação do Desempenho: Colisões

Considerou-se a porcentagem média, por período, em que os trinta diferentes agentes de uma mesma arquitetura colidiram em um obstáculo, porém sem que a colisão fosse tal que interrompesse a simulação. Colisão pode ser tanto aquela necessária para que a fonte libere energia, como colisão em um obstáculo qualquer. Um dos resultados do aprendizado deve ser o evitar colisões desnecessárias e, uma vez que tenha colidido, aprender a se desvencilhar do obstáculo; mas permitir aquelas que ativem a fonte de energia. No caso do controlador

Referência (Figura 4.2 a), não há aprendizado e geralmente colide-se apenas para obter energia ou em situações em que se desenvolveu alta velocidade, mas apresenta rápida atuação do comportamento Desviar de Obstáculos (orientação para outra direção) e ativação de um comportamento diferente.

Na Figura 4.2 é ainda possível perceber que todos os controladores (exceto o Referência, que não aprende) diminuem a porcentagem de colisões de acordo com o aprendizado. Como os controladores que se utilizam da seleção *p-greedy* (Figura 4.2 b) perscrutam mais a escolha do comportamento, naturalmente têm uma menor incidência inicial de colisões.

a) b)

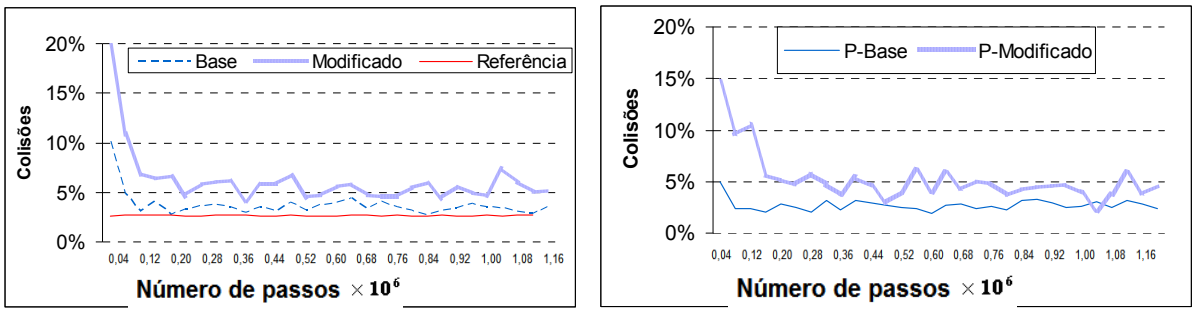


FIGURA 4.2 Gráfico da porcentagem de colisões ao longo do tempo/ aprendizado: a) Obtido nos experimentos com seleção de comportamento Boltzmann-Gibbs. O controlador Referência aparece apenas para comparação; b) Seleção de Comportamento *p-greedy*.

4.2.4 Avaliação do Desempenho: Energia

Considerou-se o nível médio de energia dos trinta diferentes agentes de uma mesma arquitetura durante o período. Uma simulação mal sucedida é aquela cujo agente não conseguiu aprender a adquirir energia. O agente Referência (Figura 4.3 a), por desenvolver velocidades mais altas e percorrer mais o seu ambiente, encontra mais rapidamente as fontes de energia, mas também gasta a sua energia prontamente, mantendo o menor nível médio de energia dentre os controladores.

a) b)

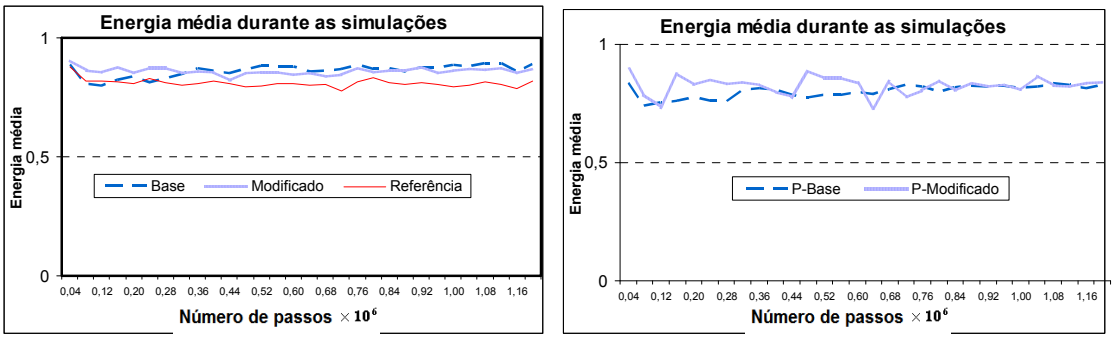


FIGURA 4.3 Gráficos do nível médio de energia ao longo do tempo. a) Obtido nos experimentos com seleção de comportamento Boltzmann-Gibbs. O controlador Referência aparece apenas para comparação; b) Seleção de Comportamento *p-greedy*.

Os controladores Modificado e Base (Figura 4.3 a) têm quase a mesma curva de energia, porém, o Modificado aprende mais rapidamente a adquiri-la, como já era esperado, visto se beneficiar do aprendizado categorizado (a utilidade de cada comportamento de acordo com a emoção dominante). A curva do controlador Referência é a mais baixa porque, como atinge grandes velocidades sem focar em guardar energia, gasta mais recursos do que os outros, embora adquira energia quando encontra uma fonte.

Como o P-Base é o que apresenta o maior número de insucessos (aproximadamente 30% das simulações falharam), espera-se que seja o controlador com mais dificuldades para aprender a sua tarefa. A curva de energia (Figura 4.3 b) mostra isso. O P-Modificado é o que apresenta curva mais irregular, mas é, também, o controlador com maior número de sucessos, ou seja, quase 100%, dado que apenas uma simulação falhou - devido a uma colisão. Já a taxa de sucesso dos controladores Base e Modificado é a mesma: 85 %.

4.2.5 Avaliação do Desempenho: Reforço Médio

O reforço médio recebido significa a emoção dominante média dos trinta diferentes agentes sob uma mesma arquitetura durante o mesmo período. O agente deve aprender a maximizar recompensas.

Como várias simulações do controlador P-Base não aprendem a tarefa, os que conseguem são os que aprendem mais rapidamente, perscrutando desde o início os comportamentos (Figura 4.4 c), e recebendo reforços maiores. Já os controladores Base e Modificado (Figura 4.4 b), como exploram mais os seus comportamentos (para exploração X perscrutação, ver Seção 4.2.7), recebem sinais de reforço menores, porém, aumentam a taxa de sucesso (Seção 4.2.2). A (Figura 4.4 a) adaptada de (GADANHO 1999) aparece para comparação e validação dos experimentos.

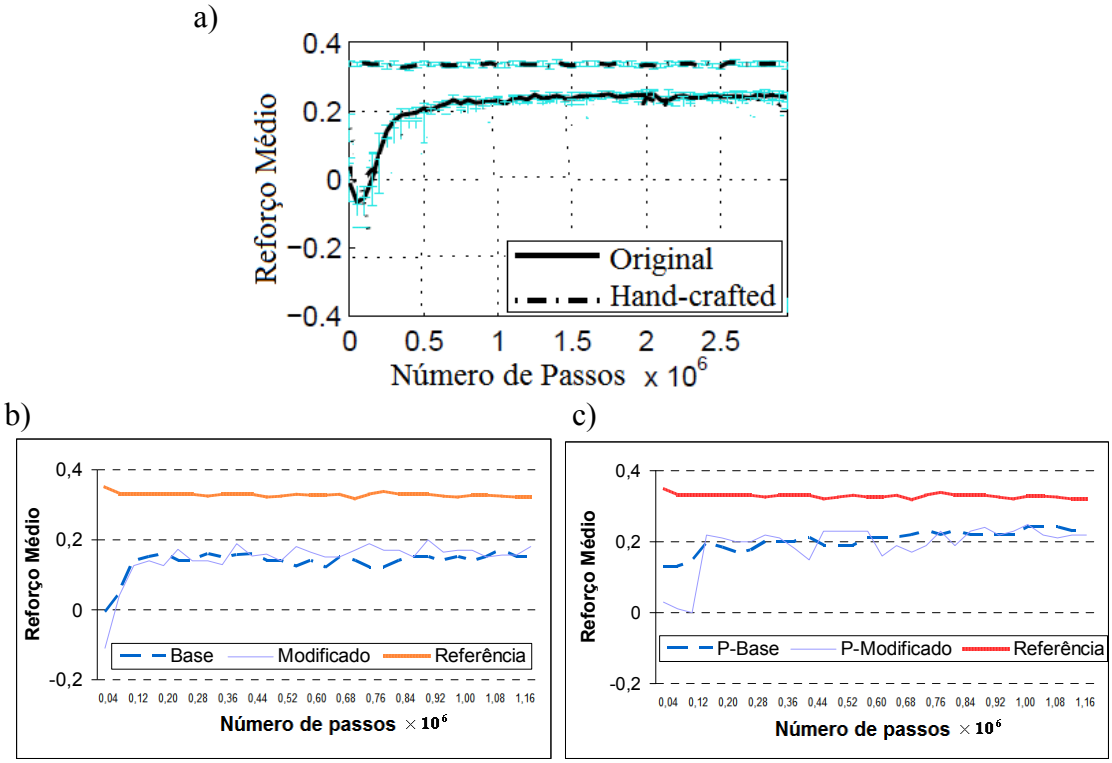


FIGURA 4.4 Gráficos de Reforço Médio ao longo do tempo: a) Adaptado de (GADANHO 1999); b) Obtido nos experimentos com seleção de comportamento Boltzmann-Gibbs; c) Seleção de Comportamento *p-greedy*. O controlador Referência aparece apenas para comparação.

4.2.6 Avaliação do Desempenho: Emoção Dominante

Considera-se a média dos estados emocionais durante a simulação completa: durante o processo de aprendizado, o agente deve aprender a coordenar os seus comportamentos para que obtenha recompensas (emoção dominante Felicidade), mas enquanto tal aprendizado ocorre, qualquer uma das emoções pode ser dominante. Para comparação com a distribuição

obtida nos experimentos (Figuras 4.5 b e 4.5 c), apresenta-se a distribuição da emoção dominante (Figura 4.5 a) adaptada de (GADANHO 1999). Os resultados são qualitativamente similares, embora os simuladores e ambientes utilizados tenham sido diferentes.

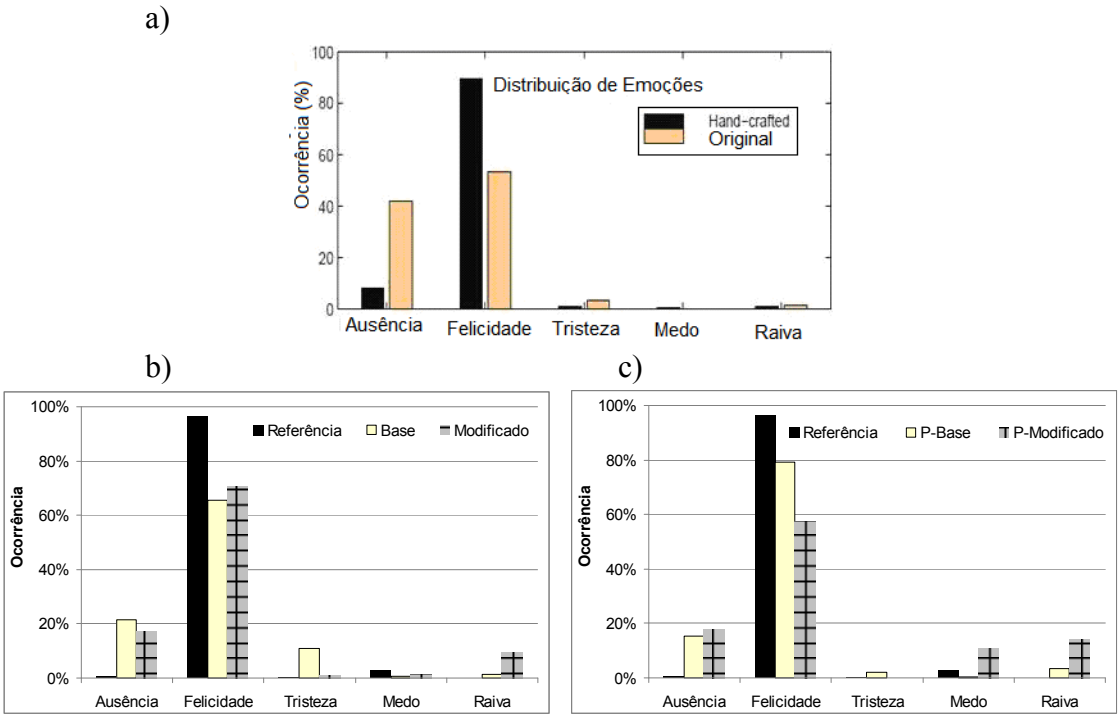


FIGURA 4.5 Gráficos da porcentagem de ocorrência das emoções como dominantes: a) Adaptado de (GADANHO 1999); b) Obtido nos experimentos com seleção de comportamento Boltzmann-Gibbs; c) Seleção de Comportamento *p-greedy*. O controlador Referência aparece apenas para comparação.

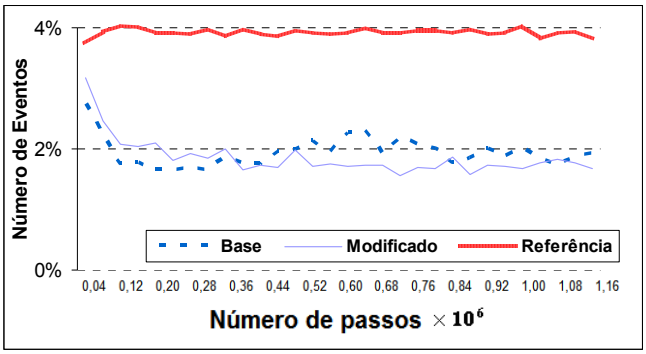
O controlador P-Base é o que adquire os maiores reforços médios (Figura 4.4 c) e índices de Felicidade (Figura 4.5 c), pelas razões descritas na Seção 4.2.5. É natural que os controladores modificados diversifiquem mais seus índices de emoções dominantes porque, ao terem processo de aprendizado inicialmente mais lento (dado serem doze redes neurais artificiais para terem seus pesos corrigidos), no início, acabam experienciando mais as emoções com conotação negativa, por exemplo, como o Medo, por ainda não saberem evitar colisões (Seção 4.2.3), ou a Raiva, devido ao aumento da Ansiedade por permanecerem em uma mesma pequena região por muito tempo ao não identificarem ainda uma coordenação correta dos comportamentos.

4.2.7 Avaliação do Desempenho: Eventos, Reforços por Eventos e Exploração

Na Figura 4.6 pode ser visto o número de eventos (ou seja, número de instâncias em que ocorre aprendizado) por controlador, durante cada período. Sendo a expectativa a de que os agentes aprendam a maximizar recompensas, espera-se que a variação emocional diminua com o passar do tempo (significando diminuir as variações decorrentes do aprendizado), e que os agentes aprendam a manter o estado emocional Felicidade. Na Figura 4.6, pode-se ver que o número de eventos de todos os controladores diminuiu com o próprio aprendizado e, como os controladores *p-greedy* perscrutam mais, tendem a detectar menos eventos com o passar do tempo, ao manterem um estado emocional mais estável. Na Figura 4.4 pode-se ver que estes controladores têm uma curva de reforço médio um pouco mais acentuada em relação aos que se utilizam da estratégia Boltzmann-Gibbs.

Na Figura 4.7 estão as médias dos valores emocionais que ativaram o Detector de Eventos por período (em cada controlador). Na verdade esses valores emocionais funcionaram como mecanismos de atenção; assim, espera-se que, com o aprendizado, variações emocionais com conotação negativa passem a dar lugar às variações emocionais com conotação positiva. Espera-se que o agente, que aprendia a não repetir situações negativas (visto as emoções com conotações negativas serem mais numerosas do que as positivas), passe a aprender a repetir as positivas - muito embora, assim como (GADANHO1999) pondera, situações que inspirem respostas emocionais com conotações negativas sempre possam surgir.

a)



b)

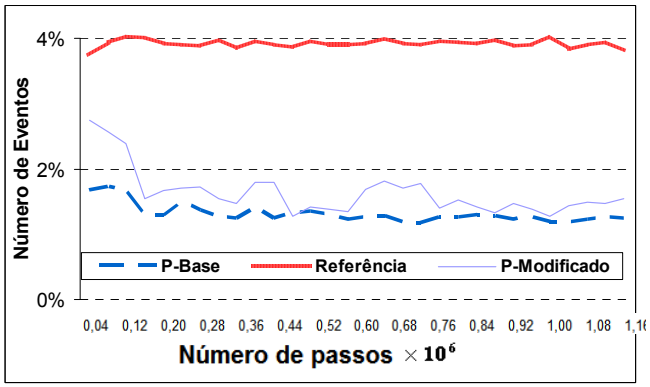
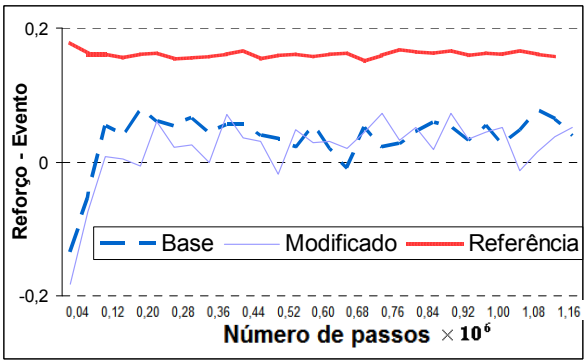


FIGURA 4.6 Gráficos da porcentagem de aprendizado: a) Obtido nos experimentos com seleção de comportamento Boltzmann-Gibbs; b) Seleção de Comportamento *p-greedy*. O controlador Referência aparece apenas para comparação.

a)



b)

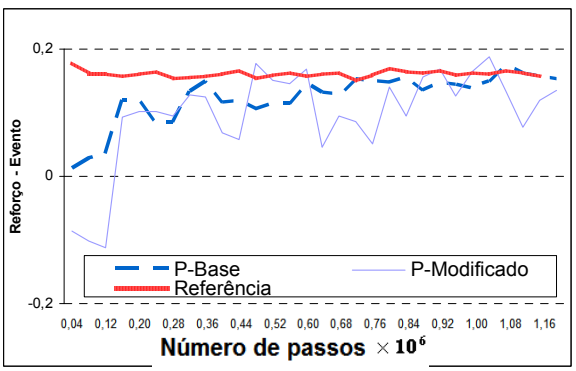


FIGURA 4.7 Gráficos do Reforço Médio apenas das Emoções que ativaram o Módulo de Aprendizado: a) Obtido nos experimentos com seleção de comportamento Boltzmann-Gibbs; b) Seleção de Comportamento *p-greedy*. O controlador Referência aparece apenas para comparação.

4.2.8 Comportamento por Emoção Dominante

Após o desenvolvimento dos controladores Modificados, passou a fazer sentido a divisão de preferências por comportamento segundo a emoção dominante, pois o treino se dá na rede associada a ambos, à emoção dominante e ao comportamento quando houve a detecção de um evento imediatamente anterior ao evento atual. Com a coleta do comportamento por emoção dominante, é possível mapear a preferência por um ou outro comportamento de acordo com uma emoção dominante específica. A seguir faz-se uma pré-análise da relação entre as emoções dominantes e os comportamentos e, na Figura 4.9, os resultados experimentais confirmando tal relação. Embora apenas os controladores Modificados tenham redes neurais artificiais associadas à emoção dominante, as preferências dos controladores Base e P-Base também foram mapeadas para comparação.

- Figura 4.8 a) Ausência de emoção dominante: nenhum evento está ocorrendo; assim, o melhor comportamento a ser acionado é o Siga Paredes porque, além de possibilitar que alguma fonte de energia seja encontrada, o simples movimento já aumenta a emoção Felicidade. Porém, se ativar o comportamento Evite Obstáculos e, depois, o Siga Paredes, o agente apenas mudará a direção do seu movimento – quando isso ocorre, o agente não recebe variação no sinal de reforço. Devido a isso e à estratégia de seleção Boltzmann-Gibbs, que permite uma probabilidade aproximadamente igual de seleção entre comportamentos com previsões de utilidade similares, os controladores Base e Modificado têm maior incidência do comportamento Evite Obstáculos (Figura 4.8 a), do que os controladores *p-greedy* (Figura 4.8 b).
- Figuras 4.8 c) e 4.8 d) Tristeza: quanto mais rapidamente se locomove, mais o agente gasta a sua energia. Porém, se não se mover, não encontrará fontes de energia. Assim, deverá haver uma combinação entre os comportamentos para que o agente encontre energia. Uma vez encontrada, deve ser ativado o Busque por Luz e, quando o Sistema

Homeostático passar a ser influenciado por ter achado a fonte de energia, a emoção dominante do agente mudará para Felicidade.

- Figura 4.8 e) e 4.8 f) Raiva: se o agente ficar parado, o sentimento Ansiedade aumentará, consequentemente aumentando a intensidade da emoção dominante Raiva. O comportamento ideal, então, é o Siga Paredes, que fará com que o robô se locomova e diminua o sentimento Ansiedade. Porém, quanto mais se locomove, mais obstáculos encontra, devendo desviar-se deles para seguir em frente. Eventualmente, pode se deparar com uma fonte de energia.
- Figura 4.8 g) e 4.8 h) Medo: ocorre quando o agente está colidindo em algum obstáculo ou preso em alguma região; dessa forma, o comportamento mais coerente com a situação é o Desvie de Obstáculos.
- Figura 4.8 i) e 4.8 j) A emoção Felicidade ocorre quando o agente está se locomovendo bastante ou detectando energia, mas, principalmente, quando a está obtendo. O agente se locomove (Segue Paredes), encontra uma fonte de energia, se aproxima desta (Busque por luz) e, através do Desvie de Obstáculos, se vira para ela: durante 200 iterações precisa receber luminosidade sobre seus sensores traseiros para obter energia. O comportamento Evite Obstáculos também se faz necessário quando, uma vez em alta velocidade, o agente se depara com um obstáculo e ainda não houve resposta do Sistema Homeostático para mudar para uma emoção dominante coerente com a colisão, ou para que o agente adquira energia.

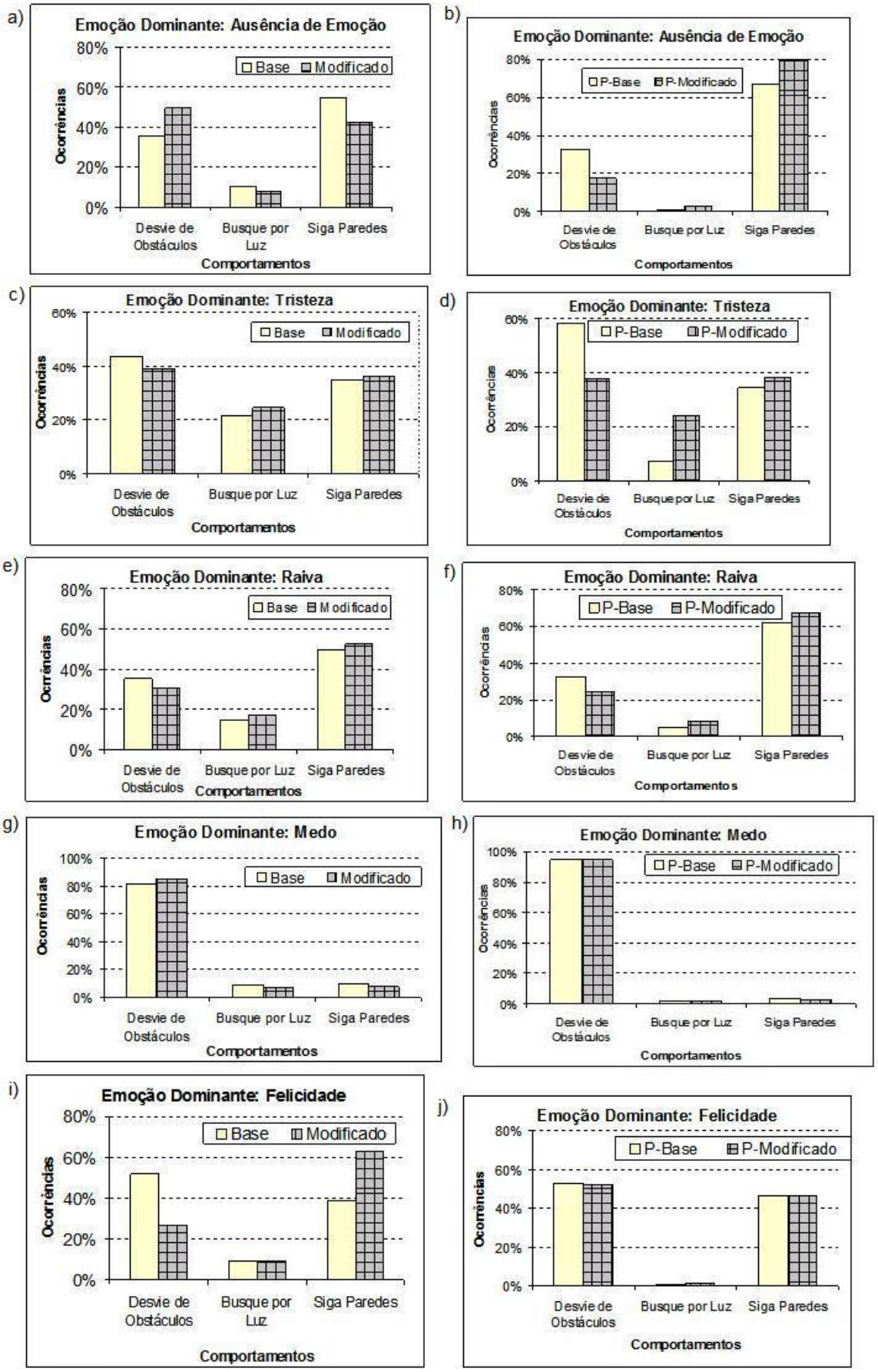


FIGURA 4.8 Gráficos da Preferência por um ou outro comportamento de acordo com a emoção dominante.

A coleta do comportamento por emoção dominante, que tornou possível o mapeamento da preferência por um ou outro comportamento de acordo com uma emoção dominante específica, evidencia que as arquiteturas parecem prover estados emocionais e seleções de comportamento coerentes com a situação corrente observada.

4.2.9 Aprendendo a Coordenar os Comportamentos com o Controlador Referência

Assumindo que o controlador Referência (o controlador que não aprende, pois já foi completamente projetado antes da simulação) coordena os comportamentos corretamente em resposta ao ambiente, por que não usá-lo para treinar as redes neurais artificiais e, após o treino, inseri-las ao Submódulo Adaptativo da arquitetura do controlador Base? Tal indagação foi concretizada da seguinte maneira: como o controlador Referência possui o mesmo Sistema Homeostático que os outros controladores, seus sentimentos e bias passaram a alimentar as redes, uma para cada comportamento, para calcular a função de utilidade e, sempre que um comportamento diferente é selecionado, a rede responsável pelo comportamento executado na iteração imediatamente anterior $n-1$ retorna aos valores desta iteração, recebendo os sentimentos da iteração anterior $n-1$ e também a resposta emocional, emoção dominante, da iteração n e, com o valor alvo, os pesos da rede são corrigidos. O procedimento é o mesmo que aquele descrito no Capítulo 3. O controlador Referência interage em seu ambiente durante 100 mil iterações, período em que há o treinamento das três redes, uma por comportamento. Após este período de treino das redes, o controlador Base recebe as três redes pré-treinadas (ao invés de redes com pesos de saída com valores pequenos e aleatórios, e com os pesos de entrada zerados) e dois diferentes experimentos são executados:

- Na simulação completa não há aprendizado. O objetivo é testar se, com as redes treinadas a partir de uma coordenação de comportamentos considerada referência, o

controlador conseguirá coordenar os comportamentos e executar a sua tarefa. Este controlador foi denominado Sem Aprendizado;

- Simulação igual àquela do controlador Base, ou seja, há aprendizado, mas as redes iniciais são as pré-treinadas. Este controlador foi denominado Com Aprendizado.

Foi feita a mesma coleta do reforço médio que a da Seção 4.2.5, com a exceção de também aparecerem, na Figura 4.9 a seguir, o reforço inicial (zero) e o reforço após cem iterações, visto ser dentro deste intervalo (*para* os experimentos executados, sendo este valor variável) que os reforços entre os controladores *Com Aprendizado* e *Sem Aprendizado* começam a destoar. No início das simulações, com nível de energia no nível máximo e todo o Sistema Homeostático zerado, o sinal de reforço tende a ser bem maior do que a média. Na Figura 4.9 fica claro que o controlador que inicia a simulação com as redes pré-treinadas, Figura 4.9 a) Com Aprendizado, tem um desempenho melhor do que o que inicia com as redes não treinadas, Figura 4.9 b) Base, porém, aprender a coordenar os comportamentos a partir de um treinamento quando do desempenho do controlador Referência não funciona, Figura 4.9 a) Sem Aprendizado, porque as redes não são treinadas com situações realmente ruins, dado que o controlador Referência não permite que estas ocorram, como, por exemplo:

- ficar quase sem energia;
- permanecer por um longo período em uma mesma pequena região sem que seja para adquirir energia;
- não ser persistente o suficiente no comportamento Desvie de Obstáculos até que tenha se desvencilhado de um obstáculo.

Ambos os controladores com as redes pré-treinadas iniciam a simulação já obtendo energia, isto é, quando encontram uma fonte de energia, se movem em sua direção e rapidamente direcionam seus sensores traseiros, sendo também esta a razão para os altos valores de reforço no início (a posição inicial do agente fica em frente a uma fonte de energia

e, também, em espaço livre de obstáculos); porém, parecem não ter aprendido adequadamente a desviar de obstáculos.

a) b)

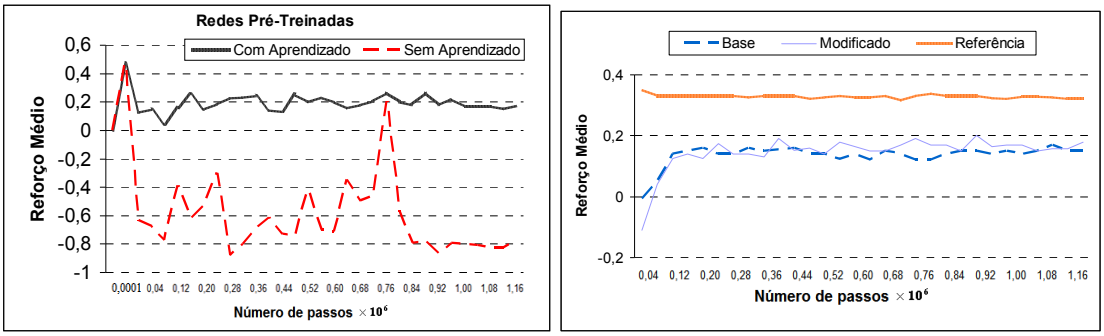


FIGURA 4.9 a) Gráficos do Reforço médio obtido pelo controladores que iniciam a simulação com redes neurais artificiais pré-treinadas. b) Anteriormente apresentado na Seção 4.2.5, aparece para comparação.

Na Figura 4.10 pode-se perceber, para o controlador *Sem Aprendizado*, como fica a seleção de comportamento com as redes pré-treinadas, evidenciando-se que as redes ficaram subtreinadas ao experienciarem padrões tendenciosos (situações ambientais melhores proporcionadas pelo controlador Referência) e não conseguiram generalização suficiente para dar boas soluções, seleções de comportamento, para situações ambientais ruins. O conjunto de eventos experimentados pelo controlador Referência acaba sendo muito diferente do dos outros controladores.

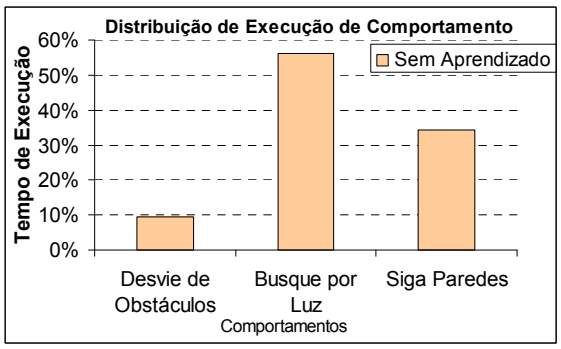


FIGURA 4.10 Gráfico da porcentagem de tempo de execução dos comportamentos do controlador que inicia e termina a simulação com as mesmas redes neurais artificiais já treinadas.

Para confirmar se os controladores com as redes pré-treinadas realmente não conseguem desviar de obstáculos adequadamente, um gráfico com a porcentagem de colisão

(Figura 4.11 a), mostra que, já nas primeiras iterações de aprendizado, o controlador *Com Aprendizado* aprende a desviar de obstáculos; já o outro controlador, como não aprende, fica muito tempo selecionando comportamentos errôneos, fazendo com que o agente fique parado em um mesmo lugar por muito tempo, especialmente diante de obstáculos e gastando toda a sua energia. Ao ficar sem energia, o controlador falhou (além de ter falhado, o sinal de reforço diminuiu drasticamente); porém, a simulação continuou mesmo assim apenas para obtenção deste gráfico.

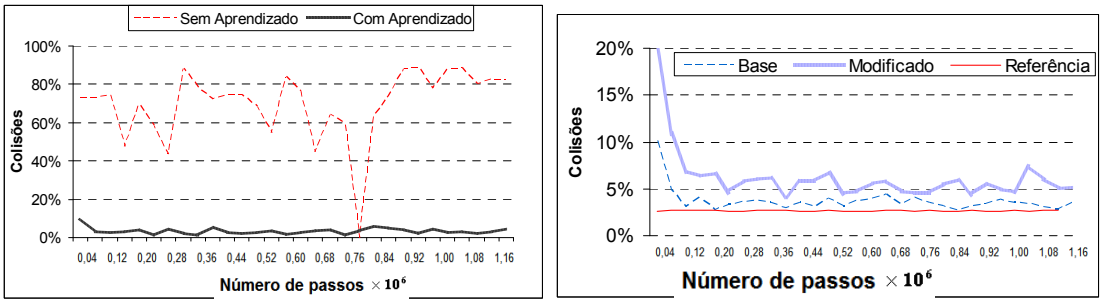


Figura 4.11 a) Gráfico da porcentagem colisões ao longo do tempo para os Controladores Com Aprendizado e Sem Aprendizado. b) Gráfico apresentado na Seção 4.2.3, aparece apenas para comparação.

4.3 Considerações

Assim como em (GADANHO 1999), os experimentos mostraram que as emoções do Sistema Homeostático podem ser usadas como mecanismos de atenção na tarefa de aprendizado por reforço: tornando mais evidentes os aspectos relevantes do ambiente; na função de reforço ao atribuir valores a diferentes situações ambientais; e determinando a ocorrência de eventos significativos através de mudanças repentinas no estado emocional. O agente tinha objetivos homeostáticos definidos em sua tarefa, ou seja, manter seu nível de energia, se movimentar pelo ambiente e evitar obstáculos. A arquitetura permitia reações rápidas ao mesmo tempo em que providenciava significado para o agente se adaptar ao seu ambiente.

Assim como nos experimentos relatados em (GADANHO 1999), as redes neurais da arquitetura tendem a perder experiências que foram pequenas em número, porém importantes,

além de demonstrar dificuldades em diferenciar corretamente experiências diferentes; assim, se o agente for colocado em um ambiente diferente após o aprendizado, se adaptará ao novo ambiente mas, durante este processo, esquecerá o aprendizado específico daquele ambiente anterior. O fato de a aquisição de política ser indistinguível do modelamento do mundo tem a desvantagem de requerer nova aquisição de política sempre que a tarefa do agente muda.

Assim como já considerado em (GADANHO 1999), a função de reforço dependente de emoção tem a característica de somente considerar no máximo uma emoção em cada iteração e ignorar a informação de reforço provida por todas as emoções não dominantes, introduzindo maior especificidade no cálculo da função de reforço. O Controlador de Eventos contribui para que a tarefa de aprendizado seja resolvida porque, além de interromper a dinâmica dos comportamentos menos frequentemente do que os experimentalmente melhores mecanismos de interrupção baseados em intervalos, tem sua taxa de interrupção variável de acordo com o aprendizado e associada aos reforços recebidos pelo agente.

A arquitetura Original é extremamente dependente da estratégia de exploração adotada: a estratégia *p-greedy* não é uma boa opção neste caso. Porém, a modificação da arquitetura, de modo a possibilitar uma categorização de acordo com a emoção dominante, permitiu que o aprendizado fosse mais lento, mas, no caso da estratégia *p-greedy*, mais efetivo e, além disso, mais flexível em relação à estratégia de exploração, visto que a arquitetura Modificada obteve sucesso para os dois tipos de seleção de exploração.

5 Conclusões

Quando se procura simular algum tipo de inteligência, é imprescindível que se conheçam diferentes teorias para o problema mente-cérebro e, da mesma forma, que se tenha uma posição objetiva acerca de qual teoria se está a admitir como base para o modelo computacional que se está a desenvolver. Os mecanismos de atenção, memória, os modos de filtrar os dados para que o agente artificial saiba lidar com a abundância de informações do seu ambiente... se se pretende realmente reproduzir algo do humano, deve-se ter fundamentada a concepção de mente e cérebro em que se está a pautar, porque é a partir dela que se interpretam as emoções, a consciência, o processamento do que chega a partir dos órgãos dos sentidos, etc.

Como enfatizado no Capítulo 2, nesta tese, o posicionamento acerca das emoções seguiu aquele de (DAMÁSIO 1994), inspiração inclusive para a arquitetura de (GADANHO 1999) através da hipótese do Marcador Somático. Em estudos mais recentes (ANDERSON et al. 2004), a hipótese do Marcador Somático ainda explica casos reais de pacientes que sofreram danos cerebrais; em (BECHARA et al. 2000), é explicada a idéia central desta hipótese: "*decision making is a process that is influenced by marker signals that arise in bioregulatory processes, including those that express themselves in emotions and feelings. This influence can occur at multiple levels of operation, some of which occur consciously and some of which occur non-consciously. (...) The somatic marker hypothesis proposes that individuals make judgements not only by assessing the severity of outcomes and their probability of occurrence, but also and primarily in terms of their emotional quality.*"

Assumidas como premissas a existência de uma exterioridade e, também, a possibilidade de percebê-la, que as sensações transmitem algo dessa exterioridade, que as emoções são processos cognitivos e não parte de uma alma imaterial e passíveis de reprodução e modelamento artificial, teve-se por objetivo a reprodução, o teste e modificação da arquitetura de (GADANHO 1999) em que há seleção autônoma de comportamento

baseada em simulação computacional de emoções e processos hormonais. As denominadas emoções artificiais foram usadas na Função de Reforço; os sentimentos providos pelo Sistema Homeostático alimentaram as redes do Módulo de Aprendizado (Submódulo Adaptativo); as emoções também serviram como mecanismo de atenção através do Módulo de Detecção de Eventos. As arquiteturas consistiram em uma simplificação: existe apenas uma emoção (a Emoção Dominante), por iteração, para o aprendizado e para o mecanismo de chamada de atenção (Detector de Eventos), ou seja, as emoções artificiais restantes são ignoradas para o processo de tomada de decisão.

Os resultados seguiram o padrão observado em (GADANHO 1999), validando a reprodução. O aprendizado categorizado de acordo com o estado emocional do agente consistiu na modificação do Modelo de (GADANHO 1999) e proporcionou considerações interessantes acerca de diferentes estratégias de exploração.

Em (GADANHO 1999) o objetivo era testar a influência de emoções artificiais sobre o controle, sem que se priorizasse um comportamento ótimo. Sendo assim, as técnicas que teriam se mostrado mais simples e bem sucedidas teriam sido escolhidas. Os pontos positivos desta arquitetura seriam o favorecer o agente com reações rápidas, enquanto permitiria que este aprendesse sua política para agir no ambiente. A generalização sobre os dados de entrada permitiria a aceleração do processo de aprendizado.

(GADANHO 1999) pondera que a opção pela junção das técnicas que se mostraram mais simples trouxe algumas desvantagens, as quais foram também observadas durante o desenvolvimento desta tese:

1. A capacidade de aprendizado do agente não ocorre de modo sofisticado o suficiente para que haja alto grau de autonomia, pois, para tal, seria preciso uma decomposição autônoma de comportamentos;

2. O agente pode ter limitações sobre suas capacidades de comportamento desde que possui um número restrito de ações discretas;
3. Sendo a aquisição de política indiferenciável do modelamento do ambiente, haveria a desvantagem de uma nova aquisição de política sempre que o objetivo do agente é alterado.
4. As habilidades de navegação providas pela arquitetura seriam pobres porque o agente não conhece sua localização no espaço.

O agente desenvolvido com a arquitetura de Módulo de Aprendizado, assim como o de (GADANHO 1999), parece prover estados emocionais e seleções de comportamento coerentes com a situação corrente observada (preferência por comportamento: Seção 4.2.8).

A vantagem da arquitetura Modificada se deu em relação à taxa de sucesso quando do uso da estratégia *p-greedy* para seleção de comportamentos (Tabela 4.3). As adições feitas sobre a arquitetura original não melhoraram as políticas de ações já bem sucedidas (Figura 4.4), mas aumentaram a probabilidade de ocorrência de políticas de sucesso (Tabela 4.3).

O assunto desta pesquisa pode ser relevante em diferentes áreas, particularmente no desenvolvimento de robôs de resgate. Adicionalmente, provê questões filosóficas sobre o problema mente-corpo, aprendizado, mecanismos de atenção, temperamento e livre-arbítrio. Quando alguém tenta simular emoções humanas através de um modelo artificial, já está assumindo as emoções não como parte imaterial da alma, mas como elemento de processos cognitivos.

5.1 Trabalhos Futuros

5.1.1 Sobre a Mesma Arquitetura Descrita Nesta Tese

Para trabalhos futuros pode-se testar, assim como mencionado em (GADANHO 1999), uma taxa de aprendizado pautada na emoção dominante, assim, situações ligadas a Emoções mais intensas terão uma taxa de aprendizado maior.

Também poderia ser considerado com maior atenção o sentimento Ansiedade: a finalidade seria a de eliminar a necessidade de reiniciá-lo sempre que o Módulo de Aprendizado é ativado. Além, disso, os próprios sentimentos, sensações e emoções poderiam ser diferentes.

Seria interessante fazer uma análise de sensibilidade dos pesos do Sistema Homeostático fornecidos por (GADANHO1999), os quais também poderiam ser alterados: poderia ser construído um módulo que, a partir das dependências que se quisesse testar entre os sentimentos e as emoções (dados de entrada), fornecesse como saída os pesos do Sistema Homeostático. Adicionalmente, poderia ser mapeado o quanto os comportamentos são influenciados pelos pesos da Tabela 2.4, e como essa influência se propaga no robô. Além disso, os pesos do Sistema Homeostático poderiam ser alterados ao longo do tempo à medida que se quisesse explorar influências sociológicas (do meio) sobre o agente, e as consequentes adaptações e mudanças de comportamento deste.

5.1.2 A partir do Conhecimento Adquirido ao Modelar a Arquitetura Descrita

Através de uma arquitetura como Clarion (SUN et al 1998, 2001; SUN 2002), em que há a possibilidade de modelar processos cognitivos em um sentido psicológico, pode-se simular a interação social entre agentes sob políticas de ação diferentes, em que a escolha varie de acordo com a moralidade, ou seja, para sujeitos morais e imorais há uma barreira

ligada à seleção de ações, sendo que o primeiro, ao contrário do segundo, não a ultrapassa; já o segundo, ultrapassa-a propositadamente. Um terceiro tipo de agente, o amoral, aprenderia, através de observação e imitação, processo este também possível através da arquitetura Clarion, a simular empatia e a se passar por moral ou imoral.

Uma simulação social com foco moral explorará áreas multidisciplinares e, a partir das premissas inerentes e necessárias a esta mesma simulação, assim como dos resultados experimentais, testará teorias acerca do conhecimento, de fatores ambientais e culturais, como em (SHAMAY-TSOORY, TOMER, AHARON-PERETZ. 2005a) ao embasar, após pesquisas, que, para alguém ser capaz de entender o sarcasmo, são necessárias a habilidade de entender a crença do falante acerca da crença do ouvinte, e a habilidade de identificar emoções.

Referências

ARISTÓTELES. **Ética a Nicômaco**. São Paulo: Pensadores, Abril Cultural, 1984.

ARISTÓTELES. **Metafísica I, II e VI**. São Paulo: Os pensadores, Abril Cultural, 1984.

ARISTÓTELES. **Tópicos V1 e VIII**. São Paulo: Pensadores, Nova Cultural, 1991.

ARISTÓTELES. **De Anima I-III**. Tradução Lucas Angioni. Campinas: Série Textos Didáticos, IFCH/UNICAMP, 1999.

ARISTÓTELES. **Física I e II**. Tradução Lucas Angioni. Campinas: Série Textos Didáticos, IFCH/UNICAMP, 1999.

BARRETO, A. M. S. **Soluções aproximadas para problemas de tomada de decisão Sequencial**. Rio de Janeiro: UFRJ, 2008. 240f. Tese (Doutorado em Engenharia Civil). Programa de Pós-Graduação em Engenharia Civil, Faculdade de Engenharia Civil, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2008.

BAYLISS, L. E. **Living control systems**. Londres: English University Press, 1966.

BERKELEY, G. [1710]. **A Treatise concerning the principles of understanding**. Oxford: Oxford University, 1998.

BERNARD, Claude. **Introduction à l'étude de la médecine expérimentale.** 1865. Disponível em: <http://classiques.uqac.ca/classiques/bernard_claude/intro_etude_medecine_exp/intro_etude.html> Acesso em: julho 2010.

BERTSEKAS, D. ; TSITSIKLIS, J**. Neuro-Dynamic programming**. Cambridge, Massachusetts: Athena Scientific, 1996.

BOTELHO, L. M. ; COELHO, H. Adaptive agents: emotion learning. In: 5TH INTERNATIONAL CONFERENCE OF THE SOCIETY FOR ADAPTIVE BEHAVIOR, 1998, Zürich. Disponível em: <http://www.ai.univie.ac.at/~paolo/conf/sab98/sab98ws.html> Acesso em: julho 2009

CAÑAMERO, D. Modeling motivations and emotions as a basis for intelligent behavior. In: PROCEEDINGS OF THE FIRST INTERNATIONAL SYMPOSIUM ON AUTONOMOUS AGENTS. AA'97. The ACM Press, 1997.

COSTA, N. C. A. **O conhecimento científico**. São Paulo: Discurso Editorial, 1999.

COPELAND, L. **A practitioner's guide to software test design**. Boston, MA: Artech House Publishers, 2007.

DAMÁSIO, A. **O Erro de Descartes**: emoção, razão e cérebro humano (Título original: Descartes' Error - Emotion, Reason and the Human Brain, 1994, tradução: Dora Vicente e Georgina Segurado), Portugal: Fórum da Ciência, Publicações Europa-América, 1995,

DAMASIO H. ; DAMASIO A ; BECHARA A. Emotion, decision making and the orbitofrontal cortex. **Cerebral Cortex**, Oxford University Press, Oxford, março, 2000. Vol. 10, Nº 3, p. 295–307.

DAMASIO, A. ; DAMASIO H ; ANDERSON, S. W. A neural basis for collecting behaviour in humans. **Brain,** Oxford University Press, Oxford, novembro, 2004. Vol. 128, Nº 1, p. 201-212.

DAWKINS, R. **The selfish gene**. Oxford: Oxford University Press, 1976.

DEL NERO, H. S. **O sítio da mente**. São Paulo: Collegium Cognitio, 1997.

DENNETT, D. **Consciousness explained**. New York: Back Bay Books, 1991.

EKMAN, P. An argument for basic emotions. In: **Cognition and Emotion**, 1992. Vol. 6 Nº3 e 4, p.169–200.

FERGUSON, I. A. **Touring machines:** An architecture for dynamic, rational, mobile agents. Cambridge: University of Cambridge, 1992. 206f. PhD, Departamento de Filosofia, Cambridge University, Cambridge.

FOLIOT, G. ; MICHEL, O. Learning object significance with an emotion based process. In: 5TH INTERNATIONAL CONFERENCE OF THE SOCIETY FOR ADAPTIVE BEHAVIOR, Zürich, 1998. Disponível em:

<http://www.ai.univie.ac.at/~paolo/conf/sab98/sab98ws.html> Acesso em: julho 2009.

FREGE, G. [1918] **Investigações lógicas e outros ensaios**. São Paulo: Cadernos de Tradução *Humanitas*, 2001.

FRIJDA, N. H. Moods, emotion episodes, and emotions. In Lewis, M. & Haviland, J.M. (eds.). **Handbook of emotions** Nova York e Londres: The Guilford Press, 1993. P. 381-403.

GADANHO, S. **Reinforcement learning in autonomous robots**: An Empirical Investigation of the Role of Emotions. PhD, Edinburgh University, Edinburgh, 1999.

GADANHO, S. ; HALLAM, J. (2001a) Emotion-triggered Learning in Autonomous Robot Control. **Cybernetics and Systems: an international journal**. Julho, 2001, V. 32, p. 531-559.

GADANHO, S. ; HALLAM, J. (2001b) Robot Learning driven by emotions. **Adaptive Behaviour**, Vol. 9 Nº 1, 2001.

GADANHO, S. ; CUSTÓDIO, L. **Learning behavior-selection in a multi-goal robot task.** Techinical Report RT-701-02, Instituto de Sistemas e Robótica, IST, Lisboa, Portugal, 2002a.

GADANHO, S. ; CUSTÓDIO, L. Asynchronous Learning by Emotions and Cognition. In: FROM ANIMALS TO ANIMATS VII, PROCEEDINGS OF THE SEVENTH INTERNATIONAL CONFERENCE ON SIMULATION OF ADAPTIVE BEHAVIOR (SAB'02), Edimburgo, UK, 2002b.

GADANHO, S. Learning Behavior-Selection by Emotions and Cognition in a Multi-Goal Robot Task. **Journal of Machine Learning Research**, JMLR, julho 2003, Nº 4, p.385-412.

GUDWIN, R. R. Novas fronteiras na inteligência artificial e na robótica. In: 4º CONGRESSO TEMÁTICO DE DINÂMICA, CONTROLE E APLICAÇÕES, UNESP, Bauru, 2005.

HAACK, S. **Philosophy of logics**. Cambridge: Cambridge University Press, 1978.

HALLAM, B. ; Hayes, GM. Comparing robot and animal behaviour. In*:* SIMULATION OF ANIMAL BEHAVIOUR, Havaí, Dezembro, p. 6-11, 1992, Paper #598.

HAYKIN, S. **Neural Nertworks:** A comprehensive foundation. Prentice Hall, 1999.

HEGEL, G. W. F. [1807] **Fenomenologia do espírito:** I; II. Petrópolis, RJ: Editora Vozes, 1988.

HERTZ, J. A. ; KROGH, A. ; PALMER, R. G. **Introduction to the theory of neural computation**. Redwood City, USA: Addison-Wesley Publishing Company, 1991.

HOFSTADTER, D. R. [1979] **Um entelaçamento de gênios brilhantes Gödel Escher Bach.** São Paulo, Imprensa Oficial do Estado, 2001.

HUSBANDS, P. ; SMITH, T. ; JAKOBI, N. ; O'SHEA, M. Better living through chemistry*:* evolving GasNets for robot control. **Connection Science**, Vol. 10 Nº3 e 4, p. 185-210, 1998.

KANER, C ; FALK, J. ; NGUYEN, H. Q. **Testing computer software**. Nova York, NY: John Wiley and Sons, 1999.

KRIPKE, S. **Naming and necessity**. Cambridge, Massachusetts: Harvard University Press, 1980.

LAUDAN, L. **Science and values:** the aims of science and their role in scientific Debate. Berkeley: University of California press, 1984.

LAUDAN, L. **Science and Relativism***:* Chicago, University of Chicago press, 1990.

LAUDAN, I., & HERNÁNDEZ, J. A. Testing the role of attribution and appraisal in predicting own and other's emotions. **Cognition and Emotion**, 12, p. 27-43, 1998.

LEDOUX, J. **The emotional brain**: The Mysterious Underpinnings of Emotional Life. New York, NY: Simon & Schuster, 1996.

LIN, L. J. **Reinforcement learning for robots using neural networks**. PhD, Carnegie Mellon University. Technical report CMU-CS-93-103, 1993.

MACEDO, B. Projeto educativo de escola: do porquê construí-lo à gênese da construção. **Inovação**, 4, p. 127 – 139, 1991.

MARICONDA, P. R.; PLASTINO, C. E. Filosofia das Ciências Naturais. In: Marilena de Souza Chaui. (Org.). **Primeira Filosofia - lições introdutórias**. 7 ed. São Paulo: Brasiliense, 1984, p. 196-217.

MARICONDA, P. R. O Diálogo e a condenação. In: **Galileu Galilei. Diálogo sobre os dois máximos sistemas do mundo**. Tradução, introdução e notas de P. R. Mariconda. São Paulo, Discurso Editorial/FAPESP, 2001, p. 15-70.

MARICONDA, P. R. ; LACEY, H. A águia e os estorninhos: Galileu e a autonomia da ciência. **Tempo social**, Vol. 13, Nº 1, p. 49-65, Maio, 2001.

MCFARLAND, D. Animal robotics - from self-sufficiency to autonomy. Gaussier, P. e Nicoud, J.-D. (eds), Proceedings FROM PERCEPTION TO ACTION, 1994, p. 47 – 54.

MINSKY, M. **Societies of Mind**. Nova York, NY: Simon & Schuster, 1986.

MOIOLI, R. ; VARGAS, P. ; VON ZUBEN, F. ; HUSBANDS, P. Evolving an artificial homeostatic system. **Advances in Artificial Intelligence** - SBIA 2008, LNAI 5249*,* Springer, 2008, p. 278-288.

MOIOLI, R. ; VARGAS, P. ; VON ZUBEN, F. ; HUSBANDS, P. Towards the Evolution of an Artificial Homeostatic System. Proceedings of the IEEE WORLD CONFERENCE ON COMPUTATIONAL INTELLIGENCE 2008, WCCI'2008, Hong Kong, China, p. 4024-4031.

MONDADA, F. ; FRANZI, E. ; IENNE, P. Mobile robot miniaturization: A tool for investigation in control algorithms. Yoshikawa, T. and Miyazaki, F. (eds), Experimental Robotics III, **Lecture notes in Control and Information Sciences**. Londres: Springer-Verlag, 1994.

NAGEL, E. **Ciência: natureza e objetivo**. São Paulo: Morgenbesser, Sidney (org.). Filosofia da Ciência. Editora Cultrix/Editora da Universidade de São Paulo, 1975.

NAGEL, E. **The structure of science***.* Londres: Routledge & Kegan Paul, 1979.

NORVIG, P. ; RUSSELL, S. **Artificial intelligence***:* a modern approach. Prentice Hall, 2002.

PANKSEPP, J. The emotional brain and biological psychiatry. **Advances in biological Psychiatry**, 1995, Nº 1, p. 263 - 288.

PERRETTA, S. J. ; Gallagher, J. C. **WSU Khepera Simulator**. Wright State University, 2003. Disponível em:<http://carl.cs.wright.edu/page11/page11.html> Acesso: julho 2007.

PETTA, P. ; STALLER, A. Towards a Tractable Appraisal-Based Architecture for Situated Cognizers. In: 5TH INTERNATIONAL CONFERENCE OF THE SOCIETY FOR ADAPTIVE BEHAVIOR, Zürich, 1998.
Disponível em:<http://www.ai.univie.ac.at/~paolo/conf/sab98/sab98ws.html> Acesso: agosto 2009.

PICARD, R. W. **Affective computing**. Cambridge, MA: MIT Press ,1997.

PICARD, R. W. Affective computing. **MIT Media Laboratory Perceptual Computing Section**, Technical Report Nº 321, 1995.

POPPER, K. [1963] **Conjecturas e refutações**. Brasília: UnB, 1972.

POPPER, K. [1982] **O Realismo e o objetivo da ciência**: Pós-Escrito à Lógica da Descoberta Científica. Lisboa: Dom Quixote, 1987.

PORCHAT, O. [1967] **A ciência dialética em Aristóteles.** São Paulo: Editora Unesp, 2000.

PUTNAM, H. **What is mathematical truth**. Cambridge: Cambridge University Press, Mathematics, Matter and Method. Philosophical Papers, v.1, 1975.

PUTNAM, H. **Meaning and the moral sciences**. Boston: Routledge & Kegan Paul, 1978.

PUTTINI, R. F. ; JÚNIOR, A. P. Além do mecanicismo e do vitalismo: a "normatividade da vida" em Georges Canguilhem. **PHYSIS: Revista Saúde Coletiva**, Vol. 17, Nº3, p.451-464, 2007.

REEKUM, C. M. van, & Scherer, K. R. Levels of processing for emotion-antecedent appraisal. G. Matthews (ed.), **Cognitive science perspectives on personality and emotion**. Amsterdam: Elsevier Science, 1997, p. 259-300.

RIBEIRO, C. H. C. ; COSTA, A. H. R. ; Romero, R. A. F. Robôs móveis inteligentes: Princípios e técnicas. In: **Anais do XXXI Congresso da SBC** (JAIA 2001). Fortaleza, 2001.

ROSEMAN, I.J; ANTONIOU, A.A.; JOSE, P.A. Appraisal determinants of emotions: constructing a more accurate and comprehensive theory. **Cognition and Emotion**, Vol. 10 Nº 3, p. 241-277, 1996.

RUEBENSTRUNK, G. **Emotional computers. computer models of emotions and their meaning for emotion-psychological research**. Novembro 1998. Disponível em: <http://www.ruebenstrunk.de/emeocomp/content.HTM> Acesso: Maio 2007.

SACKS, O. **The man who mistook his wife for a hat.** New York: Simon & Schuster publisher, 1985.

SCHACHTER, S. The interaction of cognitive and physiological determinants of emotional state*s*. In Berkowitz, L. ed. **Advances in Experimental Social Psychology.** Nova York: Academic Press, Nº 1, p. 49–80, 1964.

SHAMAY-TSOORY, S. G. ; TOMER, R ; AHARON-PERETZ, J. The neuroanatomical basis of understanding sarcasm and its relationship to social cognition. **Neuropsychology**. 2005a; Nº 19, p. 288–300, 2005.

SIMON, H. A. Motivational and emotional controls of cognition. **Psychological Review**, Nº 74, p. 29 – 39, 1967.

SLOMAN, A. [1987]. Motives mechanisms and emotions. Emotion and Cognition**.** M. A. Boden (ed.), **The Philosophy of Articial Intelligence**, Oxford Readings, Philosophy Series, Oxford University Press, 1990, p. 231-247.

SMART, J.J.C. **Between science and philosophy**. Nova York: Ramdom House, 1968.

SOUSA, J. V. ; CORRÊA, J. **Projeto pedagógico: a autonomia construída no cotidiano da escola**. Vieira, Sofia Lerche (Org.). Rio de Janeiro: DP&A, Gestão da escola: desafios a enfrentar, 2002.

SUN, R. ; PETERSON, T. Autonomous learning of sequential tasks: experiments and analysis. **IEEE Transactions on Neural Networks**, Vol. 9, Nº6, p.1217–1234, novembro 1998.

SUN, R. ; ZHANG, X. Top-down versus bottom-up learning in cognitive skill acquisition. **Cognitive Systems Research**, Vol.5, Nº.1, p.63-89, Março 2004.

SUN, R. The **CLARION cognitive architecture**: Extending cognitive modeling to social simulation. In: Ron Sun (ed.), Cognition and Multi-Agent Interaction. Nova York: Cambridge University Press, 2006.

SUTTON, R. S. ; BARTO, A. G. **Time-derivative models of pavlovian reinforcement. In Learning and Computational Neuroscience.** Foundations for Adaptive Networks. MIT Press, 1990.

SUTTON, R.S. Reinforcement learning architectures. PROCEEDINGS ISKIT'92 INTERNATIONAL SYMPOSIUM ON NEURAL INFORMATION PROCESSING, Fukuoka, Japão, 1992.

SUTTON, R. S. ; BARTO, A. G. Generalization in reinforcement learning: successful examples using sparse coarse coding. In Touretzky, D. S., Mozer, M. C., & Hasselmo, M. E. (eds.) **Advances in Neural Information Processing Systems** 1996, Nº 8: 1038-1044.

SUTTON, R. S. ; BARTO, A. G. **Reinforcement learning**. The MIT Press, 1998.

TODA, M. **The urge theory of emotion and cognition**. Chapter 1 Emotions and urges. SCCS technical report 93-1-01, Chuyko University. Versão inglesa do livro "Kanjo Emotion", 1993.

VELÁSQUEZ, J. D. Modeling emotions and other motivations in synthetic agents. In: PROCEEDINGS OF THE FOURTEENTH NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND NINTH INNOVATIVE APPLICATIONS OF ARTIFICIAL INTELLIGENCE CONFERENCE. Menlo Park, 1997.

VELÁSQUEZ, J. D. A computational framework for emotion-based control. In: SAB'98 WORKSHOP ON GROUNDING EMOTIONS. **Adaptive Systems**, 1998.

YAVNAI, A. Criteria for systems autonomability. Amsterdam: **IAS** 1989, p. 448-458.

WATKINS, C. J. C. H. **Learning with delayed rewards**. Cambridge: Cambridge University, 1989. Ph.D. – Departamento de Psicologia, Cambridge University, Cambridge, 1989.

WATKINS, C. J. C. H. ; Dayan, P. Technical note Q-Learning. **Machine Learning,** 1992, Nº 8, p. 279.

WERBOS, P. J. **The roots of backpropagation: from ordered derivatives to neural networks and political forecasting**. Nova York: Wiley, 1994.

ZINGANO, M. **Razão e sensação em Aristóteles***: um ensaio sobre "De Anima III, 4-5"*. Porto Alegre: L&PM, 1998.

Apêndice A

A.1 Simulador, máquina e linguagem de programação empregados

A máquina usada nos experimentos foi um AMD Athlon (tm) 64 X2 Dual Core Processor 4400 + 2.31 GHz 2,00 GB de RAM; sistema operacional Windows. Dispondo de tais recursos, cada experimento consistiu em trinta intervalos de quarenta mil iterações, sendo necessárias dez horas de simulação para o experimento original e doze para o modificado.

O simulador de robôs khepera utilizado foi o WSU, versão 7.3, março de 2007, do Laboratório de Robótica e Autonomia Computacional da Universidade Wright State em Dayton, Ohio. Embora o simulador empregado tenha sido diferente do simulador utilizado no projeto que inspirou o presente trabalho, ou seja, o simulador *Sim* (Olivier Michel, 1997), o WSU apresenta diversas vantagens.

O simulador WSU poder ser usado nos sistemas operacionais Linux, Windows e Mac OS X; de caráter realístico, tem seu código aberto na linguagem de programação Java; possui razoável documentação explicativa, dispõe de opções para que o agente se comporte de modo semelhante em diferentes máquinas (ou seja, reguladores de velocidade e de precisão dos sensores, tanto em relação à luz quanto à distância) e, além disso, foi atualizado em 2007. Dentre as suas desvantagens, enquadram-se: não apresentar a opção de "simulação rápida", não fornecer as coordenadas do agente ou a identificação dos objetos no ambiente.

Assim como no texto original (GADANHO 1999), as arquiteturas apenas foram testadas em simuladores, e não em robôs físicos; porém, uma característica valorizada do simulador utilizado para a presente tese é o seu caráter realístico.

A.1.1 Tempo Necessário para cada Simulação

No simulador WSU e na máquina utilizada, o controlador Base executa uma iteração a cada 0,03 segundos. Portanto, para alcançar as 1.200.000 iterações de uma simulação completa (Seção 4.2), gasta dez horas porque, uma das limitações do simulador, é a ausência da opção "simulação rápida" (sem uso da interface gráfica), que encurtaria o tempo de execução sem alterar as características da simulação. Apresenta-se na Tabela A.1 o tempo necessário para cada simulação de acordo com o controlador utilizado - deve ser observado que eram feitas três simulações por vez (independentemente do controlador escolhido), pois, frequentemente, ocorriam problemas a partir da quarta simulação executada em conjunto com outras três.

Controlador	Tempo
Base	10 horas
P – Base	10 horas
Modificado	12 horas
P – Modificado	12 horas
Referência	10 horas

TABELA A.1. Tempo necessário para cada simulação de acordo com o controlador utilizado.

A.1.2 Testes sobre o Simulador WSU

Conhecer e utilizar a literatura geral de Teste de S*oftware* permite que o desenvolvedor teste um código de forma embasada e estruturada. Os paradigmas de Teste Exploratório e de Roteiro (COPELAND 2007) são de extrema importância e, se combinados adequadamente, evitam desgaste e perda de tempo. O desenvolvimento do código para esta tese, assim como seu teste e simulação, estiveram, em momentos distintos, sob ambos os paradigmas de Teste.

Os testes feitos sobre o simulador WSU foram do tipo γ, ou seja, feitos pelo usuário, com o código de *software* pronto. Os Testes foram realizados tanto para testar e adquirir familiaridade com o simulador, quanto para adquirir conhecimento sobre seu funcionamento, e seguiram um roteiro:

- Inserir objetos de modo a construir o mundo do agente. Testar as funções do simulador inserir / retirar objeto.
- A função de gravar o mundo construído funciona? E depois de gravado o mundo, este pode ser aberto e rodado normalmente? Na documentação do simulador é documentado que as fontes de energia, quando gravadas, podem não funcionar adequadamente, sendo preferível adicioná-las ao mundo a cada execução.
- A função de inserir o agente em local específico e girá-lo 15º funciona devidamente?
- Os motores direito e esquerdo do agente respondem às diferentes velocidades aplicadas?
- Os índices de identificação, *ID*, dos oito sensores funcionam corretamente para obstáculo? Há leitura devida quando da proximidade ou ausência de obstáculo? Qual a abrangência do ruído?
- Os índices de identificação, *ID*, dos oito sensores funcionam corretamente para luminosidade? Há leitura devida quando da proximidade ou ausência de fonte de energia? Qual a abrangência do ruído?
- O simulador realmente grava e reproduz uma simulação? E os valores dos sensores do agente, são gravados também ou apenas o movimento do agente em seu mundo? Sim.

A resposta do simulador WSU foi adequada para todos os itens. A seguir apresentam-se as considerações mais pertinentes aos experimentos realizados, possibilitadas através dos testes discriminados acima.

No simulador WSU há quatro tipos de objetos para se povoar o mundo do agente: luzes e muros, que são estáticos, e bolas e tampas, as quais podem ser empurradas e/ou carregadas.

Os valores possíveis de velocidade aplicáveis aos motores esquerdo e direito do robô khepera variam entre [-9, 9]. Velocidade positiva sobre o motor esquerdo e velocidade negativa sobre o motor direito: rodopio em sentido horário; velocidade negativa sobre o motor esquerdo e velocidade positiva sobre o motor direito: rodopio em sentido anti-horário; ambos os valores negativos e iguais: o agente se move para trás; ambos os valores positivos e iguais: o agente se move para frente.

Os valores utilizados em (GADANHO 1999) para o cálculo do sentimento Temperatura foram TempRaiseTh (=14) e TempLowerTh (=10). Porém, para que a descrição do sentimento se adequasse ao contexto do simulador utilizado, os valores adotados foram TempRaiseTh (=8) e TempLowerTh (=4).

Quando o agente praticamente encosta-se em um objeto, seu sensor de proximidade correspondente marca valores entre 900 e 1023. Valores baixos, inferiores ou iguais a 10, são os mais ruidosos. Quando há um obstáculo por perto, o sensor correspondente marca acima de 10. O acesso aos valores registrados pelos sensores de proximidade se dá através da chamada getDistanceValue (identidade do sensor correspondente).

Se o sensor de luminosidade indicar menos de 400, há uma fonte de energia por perto; já se marcar menos do que 100, está extremamente próxima. Os valores habituais, quando não há nenhuma fonte sendo detectada, variam até pouco mais de 500. O acesso aos valores registrados pelos sensores de intensidade de luz se dá através da chamada getLightValue

(identidade do sensor correspondente). Na Figura A.1 são apresentados os índices dos sensores de intensidade de Luz e de Proximidade do robô khepera.

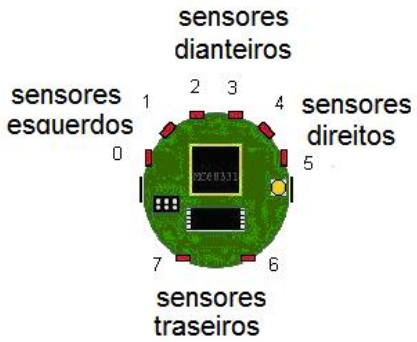


FIGURA A.1 Ordem e identidade dos sensores de intensidade de Luz e de Proximidade do robô khepera.

A.2 Os Três Comportamentos no WSU

Cada comportamento significa uma junção propositada das ações: "vá em frente", "pare", "vire-se para". São brevemente descritos em (GADANHO 1999) os três comportamentos que o agente deverá aprender a coordenar: Evite Obstáculos - se os sensores não detectarem obstáculo por perto, permaneça parado; caso contrário, desvie do obstáculo; Busque por Luz - vá em direção à fonte de energia mais próxima. Se nenhuma fonte de energia for detectada, permaneça parado; Siga Paredes - se não houver paredes por perto, ande em frente em velocidade total. Uma vez que uma parede é detectada, siga-a.

Geralmente, para chegar a cumprir sua função, cada comportamento precisa de algumas iterações, por ex.: em situação de colisão, o agente pode precisar de algumas iterações para se desvencilhar completamente do obstáculo. Porém, uma vez distante de colisão, continuar a execução do comportamento Desviar de Obstáculo é equivocado. Por isso é importante saber quando continuar a executar um mesmo comportamento (por quantas iterações executá-lo) e quando mudar.

Apenas para ilustrar a necessidade da continuidade de um comportamento, para o agente se esquivar de um obstáculo, uma vez estando bem diante dele, precisará fazer um giro de 90º. Aplicando a velocidade máxima possível no simulador de robôs Khepera WSU

(PERRETTA, GALLAGHER 2003), ou seja, de 9 unidades no motor esquerdo e -9 unidades no direito, o agente precisará de oito iterações para se desviar para a direita, percorrendo os 90º necessários (como o obstáculo está em frente, o giro poderia ter sido também para o outro lado). Isso significa que, para esta velocidade, o agente precisaria persistir no comportamento de Desviar de Obstáculos por oito iterações. A figura A.2 mostra o agente em situação de colisão e o giro que deve dar em torno do próprio eixo para se desvencilhar deste.

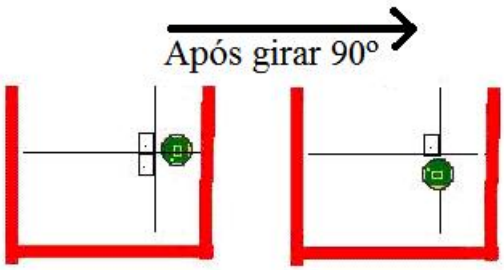


FIGURA A.2 Agente no WSU em situação de colisão e o giro de 90º para evitá-la.

A.2.1 Comportamento Siga Paredes

Se o agente já estiver seguindo uma parede específica e seus sensores de proximidade detectarem a continuidade daquela, deve continuar a segui-la. Neste caso, como saber diferenciar as situações em que o agente está realmente seguindo uma parede daquelas em que está simplesmente andando em linha reta e apenas coincidentemente próximo à parede? Quando o agente está seguindo paredes, seus sensores laterais de proximidade acusam valor alto e contínuo, assim, sua velocidade é ajustada para o valor médio em cada motor; em contrapartida, quando está apenas "seguindo em frente", sua velocidade é máxima: nove unidades em cada motor. Na Figura A.2 está um exemplo de situação em que os dois sensores direitos de Proximidade marcam valor superior a trezentos enquanto os restantes (exceto os traseiros) acusam valor inferior a dez. Na Figura A.3 está representada a situação em que o agente está seguindo paredes.

FIGURA A.3 Agente seguindo paredes no simulador WSU.

A.2.2 Comportamentos Busque por Luz e Desviar de Obstáculos no WSU

Se o agente esbarrar em um obstáculo, a simulação acabará imediatamente, significando que o ativar um comportamento que não o de Desviar de Obstáculo quando da presença do mesmo, interromperá o experimento. Para contornar este problema, um "cinturão de proteção" foi criado, cinturão ilustrativamente indicado em azul na Figura A.4.

FIGURA A.4 Ilustração do cinturão de proteção – em azul.

Quando o agente ativar os comportamentos Busque por Luz ou Siga Paredes e os sensores de proximidade de obstáculo detectarem proximidade perigosa, o agente parará automaticamente e assim permanecerá; o único comportamento capaz de fazê-lo se movimentar, neste caso, é o Desvie de Obstáculos.

Por "proximidade perigosa" quer-se dizer o limite a partir do qual um movimento, que não o giro característico do comportamento Evite obstáculos, interromperá a simulação devido a uma colisão. O cinturão de proteção foi desenhado a partir de sucessivas execuções cuja finalidade seria o delinear as leituras de cada um dos oito sensores para permitir máxima flexibilidade ao agente sem, no entanto, colidir verdadeiramente. Porém, ainda assim, durante os experimentos, houve ocasiões de colisão definitiva.

Sempre que os comportamentos Busque por Luz ou Siga Paredes ativam o cinturão, os motores do agente recebem valor zero e uma colisão é detectada. O cinturão também é

utilizado na identificação da sensação correspondente ao sentimento Dor, ou seja, se há colisão (se o cinturão é acionado), e dependendo do número de sensores de distância envolvidos, isto é, com valores altos, esta sensação é maior ou menor.

Além disso, o agente deve receber reforço negativo sempre que fizer uso do cinturão, visto que deva aprender a evitar situações de colisão. Neste caso, a própria sensação, ativando uma emoção com conotação negativa, pune o agente, contribuindo, assim, para que aprenda a acionar o comportamento Evite Obstáculos nestes cenários.

A.3 Mundo do agente

O mundo do agente utilizado na maioria dos experimentos de (GADANHO 1999) está ilustrado na Figura A.5. Este mundo foi feito e executado no simulador *Sim* (MICHEL, OLIVER 1997), e cada quadrado representa uma fonte de energia conectada a várias paredes, sendo que o agente encontra-se no meio do mundo. Os três quadrados do mapa representam as fontes de energia rodeadas por muros que não impediriam a passagem/ detecção de luz, mas evitariam que o agente ficasse preso a elas (GADANHO 1999).

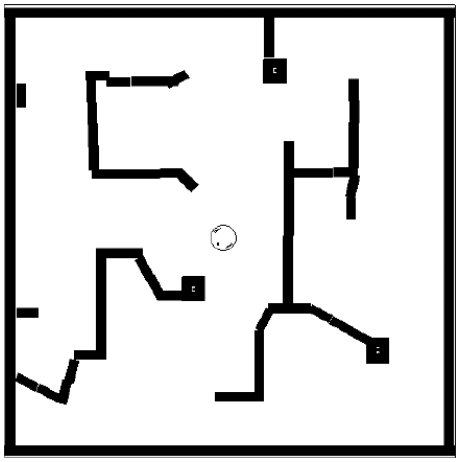
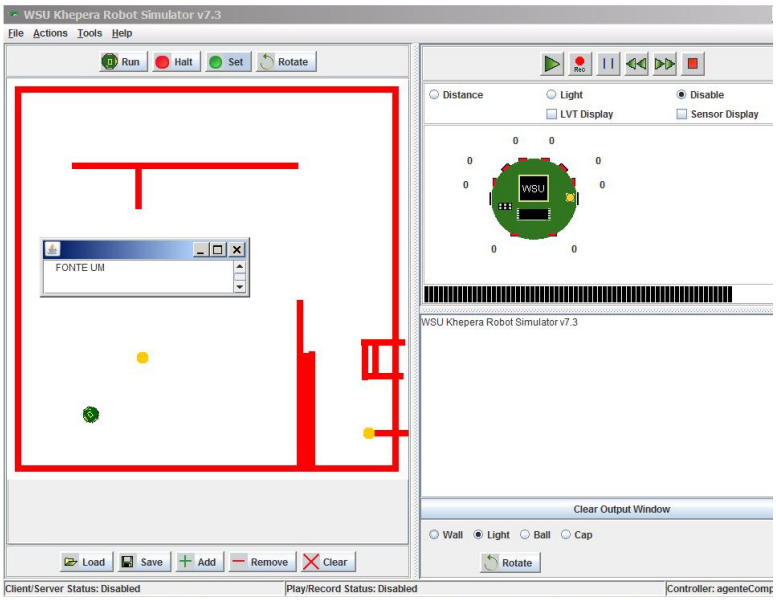


FIGURA A.5 Ambiente padrão em (GADANHO 1999).

Como mencionado anteriormente, devido ao não fornecimento, pelo simulador WSU, das coordenadas do agente e identificação de diferentes objetos de uma mesma categoria, optou-se por trabalhar com apenas duas fontes de energia, significando que o mundo do agente seria diferente do utilizado no trabalho tema. Cada fonte de energia tem a sua quantidade própria de itens de energia; assim, cada uma precisa ser identificada.

Para ser possível a discriminação de cada fonte de energia e ao mesmo tempo evitar erros provenientes de ruído, a diferenciação passou a ocorrer através dos sensores de proximidade de obstáculo e de intensidade de luz, ou seja: uma fonte de energia foi posicionada estrategicamente em um espaço livre para que os sensores não detectassem proximidade de obstáculo (mesmo sob a possibilidade de ruído) e, a outra, cercada por obstáculos. Dessa forma, quando algum sensor de proximidade identifica obstáculo por perto e altos valores de intensidade de luz (energia), está diante da fonte *dois*; se não houver obstáculo, mas intensidade de luz: fonte *um*. As posições das duas fontes de energia utilizadas e sua identificação, feita pelo agente, são mostradas na Figura A.6.

a)



b)

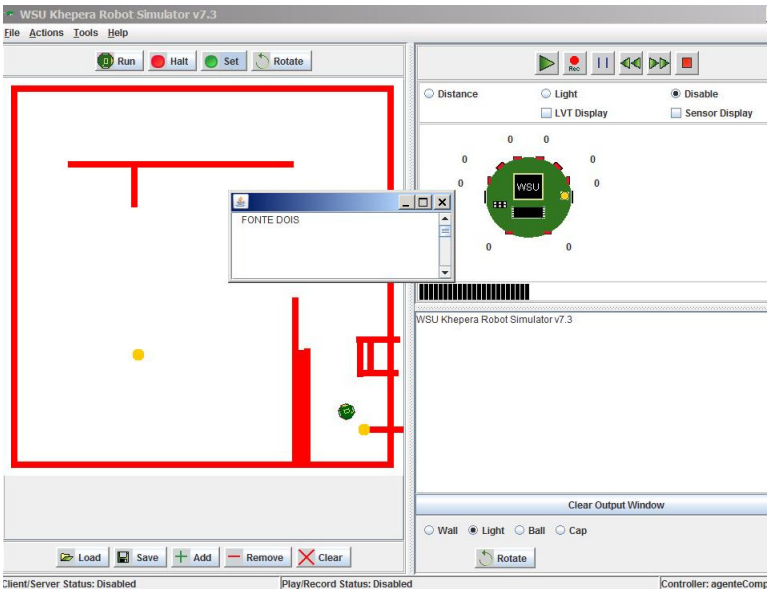


FIGURA A.6 a)Agente identificando a fonte *um*; b) Agente identificando a fonte *dois*.

Pelo fato de a fonte *dois* estar cercada por obstáculos, situações em que o agente fica preso naquela região são comuns. Mas o comportamento Siga Paredes se torna útil nessas ocasiões, devendo o agente aprender a utilizá-lo tempo o suficiente para se locomover até outra região. Quando a fonte de energia em questão se esgota, dificilmente o agente entra nesta região; o que costuma atraí-lo até ela é o comportamento Busque por Luz - o qual

somente ativa os motores se houver fonte por perto que ainda possua itens de energia para liberar.

Quando a fonte não apresenta mais itens de energia, passa a ser um obstáculo qualquer para o agente e, conforme dito anteriormente, o comportamento Busque por Luz, se ativado na ausência de energia, faz com que o agente fique parado.

Por trabalhar-se com uma fonte de energia a menos do que o ambiente do projeto tema, que possuía três fontes com cinco itens de energia cada (*MaxFoodItems* = 5), no início da simulação, as duas fontes do ambiente modificado receberam *sete* itens de energia cada (*MaxFoodItems* = 7). Como mencionado na Seção 3.1, um novo item de energia é criado e direcionado à fonte com menos itens de energia (desde que a fonte tenha menos itens do que *MaxFoodItems*) a cada *x* iterações, sendo este *x* um número aleatório entre *um* e vinte mil.

A.3.1 Controlador Referência

Quando do desenvolvimento deste controlador, os trechos de código mais trabalhosos são aqueles direcionados a evitar que o agente fique rodeando as fontes de energia e, também, impedir colisões que paralisem o agente e interrompam, assim, a simulação. Um problema que se associa a esta dificuldade é o fato de os sensores do robô serem muito ruidosos. Para evitar que o controlador Referência ficasse rodeando a fonte de energia *um*, que está localizada em um espaço completamente livre de obstáculos, colocaram-se duas tampas (Seção A.1.2) bem próximas à fonte. Tal procedimento, além de evitar este problema, não prejudicou na identificação das fontes de energia. A Figura A.7 mostra o mundo utilizado em todos os experimentos, porém, com uma única diferença para o controlador Referência: o uso das duas tampas bem próximas à fonte de energia *um*.

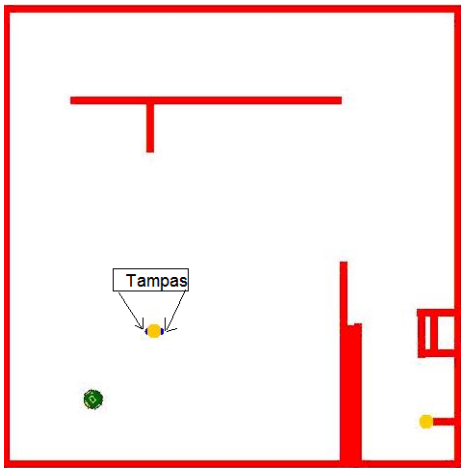


FIGURA A.7 Mundo utilizado para o controlador Referência e identificação das tampas.

Apêndice B

A seguinte publicação foi derivada desta tese:

ELIOTT, F. M. ; RIBEIRO, C. H. Autonomous Behaviour Selection Based on Computational Simulation of Emotions and Hormonal Processes. In: IX SIMPÓSIO BRASILEIRO DE AUTOMAÇÃO INTELIGENTE (SBAI 2009), Brasília, setembro 2009. **Anais do IX Simpósio Brasileiro de Automação Inteligente**, Paper #55681, setembro, 2009.

FOLHA DE REGISTRO DO DOCUMENTO

1. CLASSIFICAÇÃO/TIPO	2. DATA	3. REGISTRO N°	4. N° DE PÁGINAS
DM	12 de julho de 2010	DCTA/ITA/DM-030/2010	126

5. TÍTULO E SUBTÍTULO:

Aprendizado autônomo para robôs móveis baseado em emoções artificiais

6. AUTOR(ES):

Fernanda Monteiro Eliott

7. INSTITUIÇÃO(ÕES)/ÓRGÃO(S) INTERNO(S)/DIVISÃO(ÕES):

Instituto Tecnológico de Aeronáutica - ITA

8. PALAVRAS-CHAVE SUGERIDAS PELO AUTOR:

Modelos computacionais bioinspirados, Robôs autônomos, Aprendizado por reforço.

9.PALAVRAS-CHAVE RESULTANTES DE INDEXAÇÃO:

Robôs, Aprendizagem, Inteligência Artificial, Robótica, Controle, Computação.

10. APRESENTAÇÃO: **X Nacional**

Internacional

ITA, São José dos Campos. Curso de Mestrado. Programa de Pós-Graduação em Engenharia Eletrônica e Computação. Área de Informática. Orientador Carlos Henrique Ribeiro. Defesa em 08/07/2010. Publicada em 2010.

11. RESUMO:

Especialistas da área de neurofisiologia têm proposto a consideração dos sentimentos como parte dos processos cognitivos, e não de uma alma imaterial: tem sido defendido que as emoções não devem mais ser entendidas como opostas às decisões inteligentes, mas sim como parte e elemento decisivo para estas. Consequentemente se tornaram defensáveis a introdução de emoções artificiais no aprendizado de agentes artificiais, bem como a construção de modelos homeostáticos computacionais para estes. Nesta dissertação são relatados experimentos sobre uma arquitetura de controle baseado em comportamento e fundamentada sobre a simulação de processos hormonais e emocionais. São apresentadas e discutidas a arquitetura e modificações sobre esta, ou seja, a separação, da estrutura de aprendizado baseado em emoções, em diferentes redes neurais artificiais, uma rede para cada emoção. Os resultados mostraram que é razoável considerar modelos computacionais para processos emocionais que possam sustentar seleção de comportamento autônomo inteligente.

12. GRAU DE SIGILO:

(X) OSTENSIVO **( ) RESERVADO** **( ) CONFIDENCIAL** **( ) SECRETO**