Métodos de Avaliação de Ensino e Treinamento em Ambientes Virtuais

Mini-curso – ERMAC 2008

Autores:

Ronei Marcos de Moraes

Laboratório de Tecnologias para o Ensino Virtual e Estatística – LabTEVE
Departamento de Estatística – UFPB
ronei@de.ufpb.br

Liliane dos Santos Machado

Laboratório de Tecnologias para o Ensino Virtual e Estatística – LabTEVE
Departamento de Informática – UFPB
liliane@di.ufpb.br

Novembro/2008

Sumário

Este mini-curso apresenta o estado da arte em métodos de avaliação de ensino e treinamento em ambientes virtuais, sejam estes ambientes baseados em realidade virtual ou na Web. Um sistema de avaliação deve monitorar as interações do(s) usuário(s) com o sistema de treinamento e verificar se essas interações são corretas para que este usuário possa ser qualificado a realizar o procedimento real para o qual foi treinado. O mini-curso faz um histórico do ensino e treinamento baseados em realidade virtual e na Web, mostrando as suas várias formas de treinamento e as formas de avaliação. São apresentadas metodologias de avaliação para um usuário em treinamento, bem como as formas de avaliação continuada e avaliação para múltiplos usuários.

Agradecimentos

Este trabalho é parcialmente financiado pelo CNPq (Processo 303444/2006-1) e pela Finep (Convênio 01-04-1054-000) e desenvolvido no Laboratório de Tecnologias para o Ensino Virtual e Estatística (LabTEVE) do Departamento de Estatística da Universidade Federal da Paraíba (UFPB).

Capítulo 1. Ensino e treinamento baseado em Realidade Virtual e na Web

Introdução

A aquisição de conhecimento pode ser realizada de muitas maneiras. Abordagens conhecidas de ensino costumam lidar com a transmissão de conhecimento na forma oral, visual ou experimental. A transmissão oral de conhecimento é provavelmente a forma mais tradicional. Observa-se, entretanto, que a exploração do sentido visual contribui de forma significativa no processo de aprendizado. A experimentação, por sua vez, permite a sedimentação dos conhecimentos e é a base do construtivismo, linha pedagógica na qual o aluno experimenta e constrói seu conhecimento [Salvador et al.,2000].

Com o uso da tecnologia, novos recursos passaram a ser empregados no processo de aprendizado e ensino com o objetivo de incrementar e/ou facilitar a assimilação de conhecimento. Dentre eles podem ser citadas as salas de bate-papo, os portais de apostilas na web, os objetos de aprendizagem e as vídeo-aulas. Como objetivo principal, estes recursos buscaram ampliar o acesso dos alunos à informação de modo que ela pudesse também ser obtida fora do ambiente escolar e em momentos que o aluno considerasse conveniente. Adicionalmente, a comunicação entre o aluno e o professor no processo de aprendizado pode ocorrer de forma síncrona ou assíncrona. Na comunicação síncrona, aluno e professor compartilham o mesmo espaço, seja ele real ou virtual, e trocam informações em tempo-real. Na comunicação assíncrona o aluno geralmente lê, escuta ou assiste a uma aula e remete suas dúvidas ao professor, obtendo a resposta depois de algum tempo. Nesse contexto, o acesso às redes de computadores e a disponibilização de equipamentos com maior capacidade de processamento permitem uma maior exploração da sensação de realidade e presença dos alunos com o tema em estudo, ou seja, permitem oferecer aplicações computacionais que utilizem contextos tridimensionais, interativos e/ou de comunicação em tempo-real. Estas aplicações podem ser caracterizadas como aplicações baseadas em realidade virtual e aplicações para web.

De modo geral, as formas de treinamento podem ser decompostas em quatro categorias principais: treinamento físico, assistido, computacional e pela web. No treinamento físico o estudante tem contato direto com o assunto, podendo manuseá-lo e eventualemente modificar suas propriedades. Este tipo de treinamento permite ao aluno

observar instantaneamente os resultados de suas ações. No treinamento assitido, há também contato físico, mas o aluno é auxiliado por um professor ou tutor durante a execução da tarefa. Já no treinamento computacional o aluno utiliza um programa de computador para simular a tarefa. Quando esta simulação ocorre através da Internet este treinamneto é chamado de treinamento pela Web. Em qualquer um destes tipos de treinamento o objetivo é sempre oferecer ao aluno a oportunidade de desenvolver suas habilidades psicomotoras.

Este capítulo apresentará os conceitos principais necessários à compreensão da realidade virtual, apresentando sistemas voltados ao ensino e treinamento que a utilizam. Em seguida será apresentado como o ensino e o treinamento podem ser empregados na web, utilizando ou não recursos de realidade virtual. Uma vez que a reabilitação envolve a realização de uma atividade, esta será também tratada neste capítulo.

1.1 Realidade Virtual

No ano de 1991, Howard Rheingold escreveu sobre uma tecnologia revolucionária chamada Realidade Virtual [Rheingold,1991]. Essa nova tecnologia referia-se a mundos artificiais gerados por computador e a dispositivos eletrônicos sofisticados e prometia transformar a sociedade. Na década que se seguiu, a Realidade Virtual ganhou fama através dos filmes de ficção científica e manteve sua aplicação restrita a jogos e pesquisas na indústria da guerra, como os simuladores de vôo [Netto et al.,2002]. Com o desenvolvimento tecnológico ela passou a ser utilizada em diferentes áreas do conhecimento e vêm ganhando destaque nas atividades de treinamento.

A Realidade Virtual (RV) pode ser definida como uma ciência que engloba conhecimento de diversas áreas, como a computação, eletrônica, física e cognição, dentre outras, visando oferecer sistemas computacionais que integram características de imersão e interatividade para simular ambientes reais onde os usuários têm estimulados os seus vários sentidos pelo uso de dispositivos específicos. Estes sistemas reagem às ações do usuário em tempo-real, demandando rápido processamento. A utilização de dispositivos de interação e imersão em sistemas de RV está diretamente associada aos aspectos cognitivos do ser humano e pode explorar estímulos visuais, auditivos, táteis, motores e olfativos [Rosemblum et al.,1998]. Esses dispositivos dividem-se em dois grupos principais: entrada e saída. Os dispositivos de entrada são responsáveis pelo envio de informações de interação ou movimentação do usuário para o sistema. Essas informações são utilizadas para o processamento de uma nova cena gráfica, de sons ou de vibrações. Entre os dispositivos de entrada, encontram-se: os sensores de trajetória, responsáveis por detectar a posição do objeto ou parte do corpo ao qual estão acoplados; os dispositivos de interação, como os tradicionais mouse, teclado e *joystick*, e as luvas de dados (utilizadas no reconhecimento dos movimentos da mão); e os sensores biológicos, que usam reconhecimento de voz ou sinais elétricos musculares [Machado e Cardoso,2006]. Os dispositivos de saída, por outro lado, visam o estímulo dos sentidos do usuário, enviando para ele as respostas do sistema. Nesse grupo encontram-se os dispositivos visuais, auditivos, físicos e de locomoção.

Os dispositivos de saída são responsáveis por oferecer os resultados da aplicação computacional e, por esta razão, os principais responsáveis pela sensação de presença e

imersão do usuário no sistema. Os dispositivos visuais são os dispositivos de saída mais comumente utilizados em RV e podem ser monoscópicos ou estereoscópicos permitindo ou não a visualização das imagens destacadas do plano de apresentação (a tela ou o monitor). Quando o sistema oferece visualização estereoscópica o usuário tem a impressão que os objetos da cena estão dispostos no espaço, transcendendo o plano de projeção. Outros dispositivos de saída são: os auditivos, sistemas de som tridimensionais responsáveis pela apresentação dos sons; as plataformas móveis, que interferem no equilíbrio do corpo e são utilizadas em simuladores de vôo e jogos eletrônicos; e os dispositivos físicos. Neste último caso, os dispositivos estimulam as sensações físicas por meio do retorno de forças, de sensações de toque e de temperatura [Netto et al.,2002].

1.1.1 Dispositivos de Interação

Os dispositivos de interação permitem a comunicação do usuário com a aplicação computacional pois permitem capturar as ações do usuário que serão utilizadas como fonte para o processamento em tempo-real de reações, tais quais: um novo ponto de vista da cena, deformação de objetos, sons e vibrações. Apesar de sua pouca intuitividade, o mouse e o teclado são os dispositivos mais tradicionalmente utilizados em sistemas computacionais, além de serem os mais acessíveis economicamente.

O mouse foi criado em 1964 por Doug Engelbart e, por se deslocar no plano, permite o posicionamento em espaços bidimensionais. É considerado intuitivo por relacionar o deslocamento real da mão do usuário com o deslocamento do apontador na interface gráfica. Sua ação no espaço é possível através da combinação do deslocamento com o pressionamento de seus botões. Atualmente, é o dispositivo de interação mais popular. Tal qual o mouse, o teclado também é um importante meio de comunicação entre o usuário e as aplicações computacionais. Entretanto, a movimentação no plano ou no espaço é menos intutitiva com ele, pois ocorre a partir do pressionamento de suas teclas.

Para aplicações em realidade virtual, procura-se empregar dispositivos cuja manipulação seja similar a de ferramentas do mundo real com o objetivo de aproximar a simulação da realidade. Por esta razão uma série de dispositivos chamados de "não-convencionais" são utilizados na realidade virtual. As luvas de dados, por exemplo, permitem que o computador receba informações advindas dos movimentos dos dedos do usuário e que podem ser utilizadas para informar o computador que o usuário está pegando ou tocando um objeto. A Figura 1.1 permite exemplificar este tipo de interação. Nela a luva de dados manipula objetos virtuais e possui uma representação gráfica. À luva de dados geralmente é acoplado um dispositivo de rastreamento com a função de capturar a movimentação espacial da mão. Entretanto, dispositivos de rastreamento podem ser acoplados a qualquer objeto real, permitindo que o usuário manipule-o e o computador receba os dados do deslocamento espacial deste objeto.

Uma categoria específica de dispositivo de interação permite que o computador receba dados e envie informações, realizando a função de um dispositivo tanto de entrada quanto de saída. Os monitores sensíveis ao toque são um exemplo, pois fornecem a visualização e enviam informações ao computador de acordo com a posição tocada pelo usuário. Alguns dispositivos físicos também possuem a capacidade de atuar como dispositivo de entrada e saída e alguns são chamados de dispositivos hápticos.

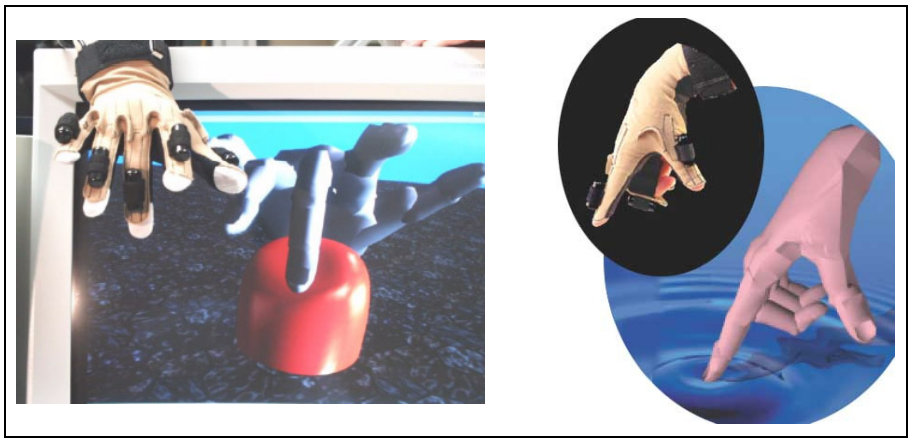

Figura 1.1. Luvas de dados sendo utilizadas para manipular elementos em cenas gráficas. Fonte: http://www.immersion.com

1.1.1.1 Dispositivos Hápticos

Dispositivos hápticos são aqueles que incorporam sensores e atuadores, permitindo o monitoramento das ações do usuário e fornecendo-lhe sensação tátil e/ou de força (*force-feedback*). A sensação tátil está associada à natureza do contato com o objeto, como textura e temperatura, enquanto a sensação de força refere-se ao senso de posição e movimentação junto com as forças associadas ao movimento durante a interação com um objeto [Machado,2007]. Para isso que isso ocorra, atuadores são responsáveis pelo envio das sensações ao usuário.

Os dispositivos hápticos que fornecem sensação de força podem ser divididos em duas categorias básicas: fixos (*ground-based*) ou móveis (*body-based*). Os dispositivos hápticos fixos, como os *joysticks*, são aqueles que estão fisicamente atrelados a uma plataforma ou superfície estável que permite o envio de reações de força ao usuário. Já os dispositivos hápticos móveis utilizam um ponto de conexão do próprio dispositivo para fornecer a reação de força e apresentam a vantagem de poderem ser portáteis na maioria das vezes, como no caso de luvas e exoesqueletos. Estes dispositivos podem permitir movimentos com diferentes graus de liberdade, suportam e reagem com diferentes valores de força, oferecem manipulação em um espaço limitado e utilizam tecnologias diversas.

Os dispositivos hápticos para braços ou pernas visam oferecer, na sua grande maioria, retorno de força. Uma das razões deles não serem tão voltados para as sensações de tato é que as mãos são mais utilizadas e sensíveis para esta atividade. Desse modo, apesar de alguns oferecerem retorno de tato e de força, sua principal função é restringir as ações do usuário, de acordo com suas ações, quando utilizados. Estes dispositivos costumam ter um custo proporcional à sua capacidade de suportar e retornar mais ou menos força.

Uma categoria de dispositivos hápticos fixos utiliza-se de uma espécie de braço mecânico para oferecer sensações de tato e *force-feedback*. Uma das suas vantagens é permitir movimentos espaciais (translações e rotações), bastante adequados a aplicações com visualização tridimensional. O Phantom – *Personal Haptic Interface Mechanism*, da

empresa Sensable, é um deles. Este dispositivo é composto por uma base à qual está atrelado o braço mecânico e cuja extremidade assemelha-se a uma caneta. Disponível em diferentes versões (Figura 1.2), utiliza potenciomêtros para oferecer retorno de força com diferentes intensidades em 3 (translações) ou 6 (translações + rotações) graus [Machado,2007]. Sua popularidade deve-se à sua empunhadura, que permite naturalmente simular diversos instrumentos puntiformes.

Figura 1.2. Dispositivos hápticos baseados em braços mecânicos.

1.1.3 Simuladores e treinamentos baseados em RV

A educação ou o treinamento é provavelmente uma das áreas mais promissoras dentro das aplicações da realidade virtual. Neste contexto encontram-se os simuladores cujo objetivo é permitir a prática de técnicas em ambientes que imitam a realidade de um procedimento. Esses sistemas preparam estudantes tecnicamente e psicologicamente para tarefas reais, bem como oferecem a profissionais meios de treinar novas técnicas simulando diferentes situações genéricas (desassociadas das peculiaridades de uma situação ideal). Na maioria das vezes, sistemas dessa natureza fornecem visualização estereoscópica e preocupam-se em oferecer algum tipo de retorno tátil ou visual.

A área médica é uma das grandes beneficiárias dos sistemas de simulação baseados em realidade virtual voltados ao treinamento. Sua vantagem insere-se na ausência de riscos envolvidos, bem como o grau de realismo que ele pode prover. Com estes sistemas, o usuário pode treinar em um ambiente virtual que imita a realidade do procedimento médico. Isto significa que não há riscos a terceiros ou ao próprio usuário, que pode treinar quantas vezes quiser sem que materiais ou tecidos sejam gastos ou danificados. Geralmente, os simuladores para treinamento médico baseados em realidade virtual reunem visualização tridimensional estereoscópica, imagens realistas, modelos que se deformam interativamente quando tocados e interação com retorno tátil e de força em um único sistema. Entretanto esta é uma tarefa complexa em termos de desenvolvimento e cara computacionalmente devido ao número de eventos a serem tratados. Por essa razão, o desenvolvimento de grande parte dos simuladores atuais encontrados na literatura utiliza apenas algumas dessas características levando em conta o seu grau de impacto no realismo da aplicação. Modalidades contempladas com simuladores para treinamento baseado em realidade virtual são: prática de sutura, laparoscopia, coleta de medula óssea para transplante, prátia de exame ginecológico e reabilitação, dentre outras. Nestes treinamentos, dispositivos hápticos simulam as sensações sentidas durante a manipulação das ferramentas médicas e auxiliam na aquisição de destreza manual.

Procedimentos laparoscópicos tem sido desenvolvidos em simuladores baseados em realidade virtual devido à complexidade de sua realização e por se enquadrarem na categoria de procedimentos minimamente invasivos, ou seja, que podem ser realizados através de pequenas incisões no corpo do paciente. Desse modo, o tempo de recuperação do paciente é menor e os riscos de um procedimento aberto são minimizados. Entretanto, a dextreza manual é fator chave pois os movimentos realizados pelo médico ocorrem no sentido inverso no interior do paciente. Nos simuladores, as pinças utilizadas no procedimento real são dispositivos hápticos que comunicam ao computador os movimentos do usuário. No ambiente de realidade virtual, o usuário visualiza uma representação virtual das pinças e pode manipular os tecidos e verificar os resultados de suas ações. Na Figura 1.3 é possível visualizar as pinças utilizadas na imagem da esquerda e o ambiente de realidade virtual na imagem da esquerda.

Figura 1.3. Simulação de laparoscopia.
Fonte: Basdogan et al. (2004) e Burdea e Coiffet (2003) pg. 300.

O simulador para prática de sutura apresentado na Figura 1.4 oferece um modelo monoscópico da pele adicionado de propriedades físico-elásticas, de forma a oferecer retorno tátil e de força ao usuário durante a manipulação de um dispositivo háptico [Webster et al.,2001]. O objetivo do sistema é ensinar o usuário a realizar uma sutura utilizando uma agulha especial. Durante a simulação o usuário pode perceber as reações elásticas da pele ao puxar a linha, bem como observar o melhor local para inserção dos pontos. Além disso, o simulador permite gravar a execução do procedimento para posterior conferência das ações realizadas.

No Brasil, o primeiro simulador cirúrgico desenvolvido utilizando realidade virtual visou o treinamento da coleta de medula óssea para transplante em crianças [Machado,2003]. Este sistema oferece visualização estereoscópica e manipulação tridimensional de um modelo com múltiplas camadas dotadas de propriedades físico-elásticas individuais percebidas pelo usuário durante a manipulação do dispositivo háptico. Neste simulador, o usuário manipula o dispositivo háptico para identificar o local de inserção da agulha de coleta, bem como para perfurar o corpo virtual e perceber as camadas de pele trespassadas pela agulha até chegar à medula óssea (Figura 1.5). Por visar um procedimento pediátrico, a destreza do médico acelera a sua realização, resultando em um menor desconforto e uma recuperação mais rápida do paciente.

Figura 1.4. Sistema para treinamento de sutura [Webster et al.,2001].

Figura 1.5. Simulador para coleta de medula óssea em execução.

O SITEG - Sistema Interativo para Treinamento em Exames Ginecológicos - é um sistema que tem como finalidade prover um ambiente virtual interativo onde o usuário pode simular um treinamento e ser avaliado na execução de um exame ginecológico do colo do útero. Para isso, ele permite simular diferentes patologias relacionadas ao colo do útero que podem ser identificadas em um exame ginecológico. O treinamento simulado oferecido pelo SITEG obedece às etapas de um exame ginecológico real. Assim, as etapas de pré-diagnóstico abordadas são o exame visual com o instrumento utilizado (espéculo) e o exame de toque onde não há informações visuais da anatomia interna da região. A etapa de toque depende do uso de um dispositivo háptico para palpação das estruturas internas (não-visíveis) do modelo da região envolvida no exame [Souza et al.,2006].

A fisioterapia e a reabilitação têm sido abordadas em sistema de realidade virtual para ensinar o usuário a forma correta de realizar movimentos, bem como avaliar e reabiliar processos cognitivos e meta-cognitivos como a percepção visual, atenção e memória [Pugnetti et al.,1998]. Neste sentido, ela depende do uso de dispositivos que monitoram os movimentos do usuário. Características importantes destes sistemas são o realismo visual e a interação intuitiva que devem propiciar um bom grau de imersão ao usuário, permitindo-lhe concentrar-se na tarefa a ser desempenhada. Segundo Weiss et al.

(2004), a imersão é importante em sistemas que visam treinar ou restaurar habilidades cognitivas e motoras.

Além da medicina, outras áreas são também contempladas com sistemas de treinamentoo e simulação. Dentre elas, pode-se destacar a manutenção equipamentos de licados ou de difícil acesso e o treinamento de pilotos de aeronaves [Machado,2006]. Entretanto, novas aplicações da realidade virtual têm sido desenvolvidas e gradativamente suas vantagens têm sido evidenciadas, principalmente quando altos custos e vidas humanas estão relacionados.

1.1.4 Treinamentos via Web

A educação, especialmente a educação à distância, vem experimentando um grande avanço decorrente da aplicação de novas tecnologias, principalmente da Internet. A educação à distância tem destaque tanto pela sua flexibilidade quanto ao tempo dos usuários, acessibilidade e difusão do conhecimento a locais remotos. Com o advento do ensino a distância, o treinamento via web vem ganhando destaque. Nesse contexto, objetos de aprendizagem (OA) exercem papel importante, na medida em que se constituem em módulos de ensino/aprendizagem reutilizáveis, disponibilizados na Internet, por exemplo, para serem utilizados por quem deles necessitar [Bettio e Martins,2002]. Esses objetos de aprendizagem podem ser simples ou avançados e devem ter características específicas para atingir seus objetivos. Além deles, sistemas variados para treinamento têm sido disponibilizados.

O treinamento via web permite ainda que haja colaboração entre usuários remotos de modo que situações diversas possam ocorrer como, por exemplo: um usuário guiar outros na realização de uma tarefa ou vários usuários interagirem de forma síncrona na realização de uma tarefa. Em qualquer tipo de treinamento pela web, seja colaborativo ou através de objetos de aprendizagem, podem ser utilizados dispositivos convencionais ou não (como os empregados em sistemas de realidade virtual). Entretanto, uma limitação comumente encontrada relaciona-se à velocidade de tráfego das informações na rede, que pode comprometer o treinamento. Esta limitação é particularmente prejudicial se o treinamento realizado for baseado em realidade virtual, cuja execução demanda tempo-real. Observa-se, entretanto, que tal dificuldade tende a diminuir com o avanço das redes e da própria Internet, visto que as aplicações pela web têm sido foco de pesquisa e desenvolvimento em todo mundo.

Referências

Basdogan, C. et al.; Haptics in Minimally Invasive Surgical Simulation and Training. IEEE CG&A, v. 24, n. 2, p. 56-64, 2004.

Bettio, R.W.; Martins, A.; Objetos de Aprendizado: Um novo modelo direcionado ao Ensino a Distância. Congresso da ABED 2002. Online: http://www.abed.org.br/congresso2002/trabalhos/texto42.htm. Acesso em outubro de 2008.

Burdea, G. e Coiffet, P.; Virtual Reality Technology, 2a edição. Wiley-Interscience, 2003.

Machado, L.S.; A realidade virtual no modelamento e simulação de procedimentos invasivos em oncologia pediátrica: um estudo de caso no transplante de medula óssea, Tese de Doutorado, Escola Politécnica da USP. 2003.

Machado, L.S;. Dispositivos Hápticos para Interfaces de Realidade Virtual e Aumentada. In: Claudio Kirner; Robson Siscoutto. (Org.). Realidade Virtual e Aumentada: Conceitos, Projeto e Aplicações, cap. 9, p. 152-167, 2007.

Machado, L.S. ; Cardoso, A.; Dispositivos de Entrada e Saída para Sistemas de Realidade Virtual. In: Romero Tori; Claudio Kirner; Robson Siscoutto. (Org.). Fundamentos e Tecnologia de Realidade Virtual e Aumentada, cap. 2, p. 39-50, 2006.

Netto, A.; Machado, L.; Oliveira, M. Realidade Virtual: Conceitos e Aplicações. Visual Books, 2002.

Pugnetti, L. et al.; Probing memory and executive functions with virtual reality: past and present studies. Cyberpsychology & Behaviour, v. 1, p. 151-162, 1998.

Rheingold, H.; Virtual Reality. New York, Touchstone, 1991.

Rosemblum, L.; Burdea, G.; Tachi, S. VR Reborn. IEEE Computer Graphics and Applications, v.18, n.6, p. 21-23, 1998.

Salvador, C.C. et al.; Psicologia do Ensino. Artmed. 2000.

Souza, D. F. L ; Valdek, M. C. O. ; Moraes, R.M. ; Machado, L.S. SITEG Sistema Interativo de Treinamento em Exame Ginecológico. In: Proc. of Symposium on Virtual Reality, Belém , p. 445-456, 2006.

Webster, R.; Zimmerman, D.; Mohler, B.; Melkonian, M. e Haluck, R.; A prototype haptic suturing simulator, Studies in Health Technology and Informatics, IOS Press, v. 81, p. 567-569, 2001.

Weiss, P.; Rand, D.; Katz, N.; Zikony, R. Video capture virtual reality as a flexible and effective rehabilitation tool. Journal of NeuroEngineering and Rehabilitation, v.1, n. 12, 2004.

Capítulo 2. O Treinamento e suas Formas

Introdução

Desde tempos remotos, procuram-se formas de explicar o contexto de uma atividade utilizando a prática. Esta permite que sejam desenvolvidas habilidades do aprendiz. O treinamento de habilidades pode ser realizado de diferentes maneiras, incluindo ou não recursos tecnológicos, e envolve o aprendizado de procedimentos, sejam estes gerais ou específicos. Segundo Salvador et al. (2000), o treinamento pode ser diferenciado de acordo com sua abordagem e envolve o desenvolvimento de destrezas motrizes ou cognitivas na realização de procedimentos compostos por passos e decisões que podem levar a resultados padronizados ou não.

Quando a tecnologia é utilizada, busca-se, em geral, conseguir benefícios em relação a outros métodos. Estes benefícios podem ser o encurtamento de distâncias entre aluno – professor e aluno – aluno; a redução de custos do treinamento; o não desgaste de materiais; a ausência de riscos ao aprendiz e a terceiros; e a recriação de situação similar à real (exploração dos cinco sentidos). O encurtamento da distância entre alunos e professores pode ser obtido, por exemplo, com o uso de sistemas para web, vídeo-aulas ou salas de bate-papo, ou seja, aplicações que permitem a comunicação via redes de computadores. As redes também podem permitir a troca de informações entre alunos, particularmente em atividades colaborativas. A redução de custos é obtida quando equipamentos e componentes caros são utilizados na realização do procedimento real. Com o uso de sistemas de computadores é possível simular a situação real sem que haja desgaste de materiais. O mesmo se aplica à ausência de risco a terceiros e à possibilidade de reutilização de materiais. Por último, a recriação de situações por meio de simulação pode ser obtida com a aplicação de técnicas de realidade virtual e visam aproximar o aprendiz do assunto e isola-lo do mundo real. Com isso, o treinamento ocorre em condições similares a do procedimento real.

A seqüência deste capítulo irá abordar as formas tradicionais de treinamento, caracterizando-os na área médica e no ensino à distância. Visto que o objetivo do treinamento é capacitar, serão apresentadas formas utilizadas para avaliar a qualidade de treinamentos quando estes são realizados por meio de sistemas de realidade virtual, ou seja, em ambientes virtuais que recriam a realidade, e como pode ela pode ocorrer para aplicações via web.

2.1 Tipos de Treinamento

Há diversas formas de se realizar um treinamento. Em comum, todas visarão sempre o mesmo objetivo: capacitar o aprendiz para realizar uma tarefa específica. Para isto, diversos recursos foram e tem sido empregados e envolvem desde o uso de atlas de imagens a sistemas baseados em realidade virtual.

Particularmente, algumas áreas têm o treinamento como uma etapa crítica na formação de novos profissionais, pois o procedimento real pode envolver equipamentos caros ou oferecer risco à vida. Por esta razão serão apresentadas nas próximas sessões as formas de treinamento comumente utilizadas em medicina, sendo esta a área a ser utilizada em exemplos posteriores desde texto.

2.1.1 Treinamento Médico

O treinamento de futuros médicos e profissionais da área médica visa permitir ao estudante compreender as fases de procedimentos e desenvolver habilidades motoras, bem como a capacidade de tomar decisões diante de situações adversas. As escolas e cursos atuais nesta área constumam utilizar aulas expositivas, modelos plásticos, cadáveres e cobaias para realizar o treinamento. As aulas expositivas procuram apresentar aos alunos os principais conceitos envolvidos no procedimento e a localização das estruturas no organismo. Em alguns casos, são utilizados modelos plásticos compostos por várias peças para apresentar ao aluno como as várias partes de um organismo ou órgão se relacionam e encaixam. Observa-se que tal abordagem é interessante no sentido do conhecimento visual, pois auxiliarão o futuro profissional a reconhecer as partes ou o todo de um organismo. Entretanto, tais objetos não apresentam características compatíveis sob o ponto de vista tátil e, por esta razão, não são sufucientes para ensinar ao aluno quais são as propriedades físicas de tecidos.

O uso de cobaias é um aprática comum no treinamento de procedimentos médicos. Particularmente, porcos-da-índia são utilizados pela similaridade física de suas propriedades com as do corpo humano. Em alguns casos, são utilizados cadáveres para o treinamento de procedimentos. Estas duas formas de treinar têm sido fortemente desestimuladas, seja pela dificuldade de conseguir corpos humanos quanto pelas organizações em defesa dos animais. Algumas disciplinas de prática cirúrgica, entretanto, costumam utilizar cobaias compradas em açougue. Além destas cobais não possuírem as exatas propriedades físicas do corpo humano elas são descartadas após um número de treinamentos, pois após algumas manipulações (cortes, extrações, etc.) do organismo ele deixa de ser útil para novos treinamentos.

Um dos últimos estágios do treinamento tradicional costuma ocorrer através do acompanhamento de procedimentos reais, onde o aluno assiste ao procedimento e posteriormente faz perguntas ao professor. Em alguns casos, o aluno auxilia o professor em procedimentos reais até adquirir prática para realiza-lo por si mesmo.

Observa-se que em nenhuma das formas utilizadas para o treinamento médico é capaz de ensinar de fato o aluno de modo a obedecer todos os seguintes aspectos:

1. conhecimento das estruturas do organismo e suas dependências;
2. conhecimento das propriedades físicas dos tecidos;
3. conhecimento das reações advindas de manipulações ao organismo.

O reconhecimento de algumas patologias, por exemplo, é treinado durante o período de residência médica. Apesar disso, algums patologias podem possuir muitas variações, sendo algumas raras. Isso pode fazer que durante os meses de residência o aluno não entre em contato com nenhum caso real que apresente determinado estágio da doença.

Com o uso de realidade virtual é possível recriar ambientes tridimensionais voltados ao treinamento médico. Estes ambientes são compostos por objetos gráficos similares visualmente a estruturas do corpo humano, bem como programadas para apresentar características e responder fisicamente a manipulações realizadas pelo usuário aprendiz. Desse modo as três situações enumeradas anteriormente podem ser atendidas com o objetivo de recriar um procedimento real em um ambiente virtual de treinamento. O aprendiz poderá visualizar espacialmente estruturas do corpo humano e separá-las para conhecer as dependências entre elas; pode conhecer o material e as propriedades que compõem tecidos; e pode realizar procedimentos e ver interativamente as reações advindas da sua manipulação [McCloy e Stone,2001].

Uma vantagem importante em sistemas de treinamento baseados em realidade virtual é que por estes se tratarem de programas computacionais, eles podem ser programados para monitorar e avaliar as ações do usuário, permitindo verificar as habilidades do usuário e até medir sua evolução. Uma vez que não há desgaste de materiais, o aluno/aprendiz pode realizar o procedimento quantas vezes quiser.

2.1.2 Treinamento à Distância

O treinamento a distância vem sendo utilizado desde tempos remotos com um modo de levar informação para pessoas distantes e oferecer flexibilidade quanto ao tempo e duração dos períodos disponíveis para estudar. A forma mais comum de treinamento à distância é certamente por meio de correspondência. Entretanto, esta atividade de aprendizado também pode utilizar o telefone e, mais recentemente, a Internet.

A Internet é atualmente conhecida como a forma de ensino à distância capaz de oferecer mais recursos ao estudante, pois permite desde o acesso a textos até o uso de laboratórios dispostos remotamente [Machado et al.,2007]. Com a gama de diferentes tipos de aplicações possíveis de serem disponibilizadas em sítios da Internet, o treinamento à distância torna-se particularmente vantajoso em relação às demais formas de treinamento. Alguns exemplos destas aplicações são:

a) Objetos de aprendizagem: o usuário definine parâmetros de condições de realização de uma atividade e observa as conseqüências de suas escolhas em uma animação gerada com os parâmetros escolhidos [Prata e Nascimento,2007];
b) Portais de apostilas: permitem acessar apostilas, documentos e vídeos que descrevem ou apresentam os procedimentos;

c) Laboratórios virtuais: aplicações baseadas em realidade virtual que apresentam um laboratório que só existe como aplicação computacional, mas cujos equipamentos podem ser utilizados para simular experimentos em tempo-real;
d) Salas-de-aula virtuais: sala de chat em que alunos e professores podem se encontrar para discutir via Internet os aspectos de uma técnica ou procedimento;
e) Tutores virtuais: aplicações computacionais dotadas de agentes inteligentes capazes de "conversar" com o aprendiz e sanar suas dúvidas relacionadas a um tema. Esta mesma designação pode ser dada a pessoas que ficam responsáveis por responder perguntas ou conduzir debates relacionados a um tema em fóruns ou salas de bate-papo.

Um dos desafios relacionados ao ensino à distância e suas diversas formas é a avaliação. Sabe-se, entretanto, que ela permitirá conferir o conhecimento adquirido pelo aprendiz, atribuindo-lhe certificações, diplomas ou disponibilizando-lhe tarefas em níveis mais elevados do processo de aprendizado.

2.2 Avaliação de treinamento em simuladores baseados em Realidade Virtual

Uma das grandes vantagens dos simuladores de treinamento baseados em realidade virtual é que estes são aplicações computacionais de tempo-real, que recriam a situação real do problema. Por isto, eles permitem que os dados de interação do usuário sejam coletados e armazenados. Estes dados, e outros do sistema, podem ser utilizados para verificar o grau de acurácia do aluno quando realizou o procedimento virtual [Pugh e Youngblood,2002].

Em vários tipos de sistemas baseados em realidade virtual para a área médica, a execução de um procedimento pelo usuário é gravada em vídeo e analisada posteriormente por especialistas que determinam o seu desempenho. Depois de algum tempo, o usuário recebe a sua avaliação. Entretanto, em vários casos, o desempenho do usuário não pode ser simplesmente classificado como bom ou ruim, devido à complexidade do procedimento. Assim, a existência de uma ferramenta de avaliação incorporada ao sistema de simulação de treinamento baseado em realidade virtual permite a melhoria do aprendizado e uma melhor avaliação do procedimento do usuário.

Para a avaliação de treinamento em simuladores baseados em realidade virtual o sistema deve supervisionar os movimentos do usuário e os parâmetros associados a eles. O sistema pode coletar informações sobre a posição espacial, forças, torques, resistências, velocidade, aceleração, temperaturas visualização e/ou ângulo de visualização, sons, cheiros, etc. A Figura 2.1 apresenta e forma esquemática como a coleta de dados é feita em um simulador baseado em realidade virtual. Nela pode-se notar que os dados relacionados ao ambiente virtual e seus objetos devem ser coletados, assim como as informações da interação do usuário com estes. Elas serão utilizadas para a avaliação, cujo parecer ou resultado permitirá classificar ou quantificar o grau de conhecimento do aprendiz.

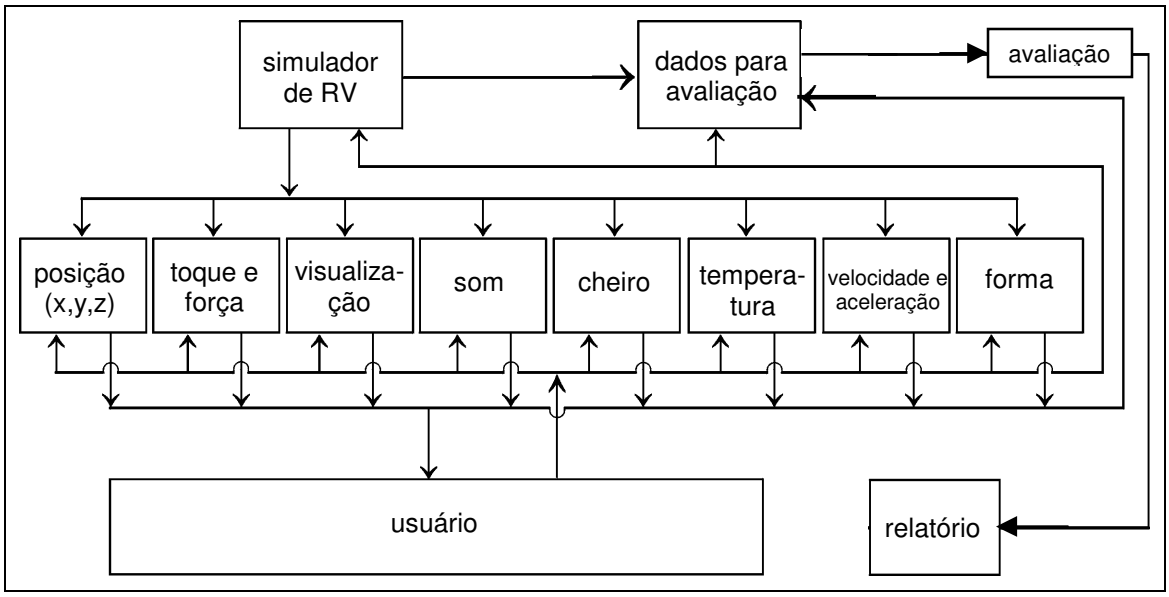

Figura 2.1. Diagrama com as diversas informações que podem ser coletadas para avaliar o usuário em um simulador de realidade virtual para treinamento.

2.3 Avaliação *online* e *offline*

Em vários tipos de sistemas computacionais, a execução de um procedimento pelo usuário é gravada em vídeo e analisada posteriormente por especialistas que determinarão o seu desempenho. Depois de algum tempo, o usuário recebe a sua avaliação. Esse é um problema, visto que em algumas horas o usuário não se lembrará quais foram exatamente as suas ações quando executou o procedimento simulado. Com isso, uma avaliação do seu desempenho sobre aquele procedimento pouco ou nada poderá contribuir para melhorar a sua forma de realizá-lo. Além disso, em vários casos, o desempenho do usuário não pode ser simplesmente classificado como bom ou ruim, devido à complexidade do procedimento. Assim, a existência de uma ferramenta de avaliação incorporada ao sistema de simulação de treinamento permite a melhoria do aprendizado e uma melhor avaliação do procedimento do usuário.

Vários métodos já foram propostos para a avaliação de treinamento. Alguns deles são *off-line* e outros são *on-line*. Os métodos *off-line* avaliam o usuário em um sistema desvinculado da aplicação. Por esta razão, os dados do treinamento devem ser coletados e gravados para uma posterior análise. Já os métodos de avaliação *on-line* estão vinculados à aplicação e são capazes de oferecer os resultados imediatamente após a execução do procedimento simulado. Por estarem acoplados ao sistema de treinamento, os métodos de avaliação *on-line* devem possuir baixa complexidade computacional para não comprometerem o desempenho da simulação do treinamento. Porém, também devem possuir elevada acurácia para não comprometerem a avaliação. Dentre esses, alguns se baseiam em métodos estatísticos como a máxima verossimilhança, regra de bayes *fuzzy*, modelos de mistura gaussiana [Moraes e Machado,2003] ou modelos markovianos escondidos [Moraes e Machado,2004]. Outros usam sistemas especialistas baseados em

regras *fuzzy* ou ainda modelos conexionistas [Machado e Moraes,2004], como as redes neurais ou as redes neurais evolutivas *fuzzy* [Moraes e Machado,2005a].

2.4 Avaliação continuada de treinamento

Mais recentemente, foram propostos métodos para avaliação continuada dos treinamentos realizados por uma mesma pessoa [Moraes e Machado,2005b]. Nesse caso, a metodologia une ferramentas estatísticas e um sistema especialista para construir um perfil individual do usuário. Assim, análises automáticas das informações derivam medidas específicas, gráficos de desempenho, tabelas e modelos estatísticos, que avaliados pelo sistema especialista mostra a evolução ou não do usuário em aspectos do treinamento. Pode-se observar na Figura 2.2 o sistema de avaliação continuada e seus componentes. Dois tipos de relatórios são criados para o usuário: o relatório de avaliação do seu último treinamento e o relatório de avaliação continuada.

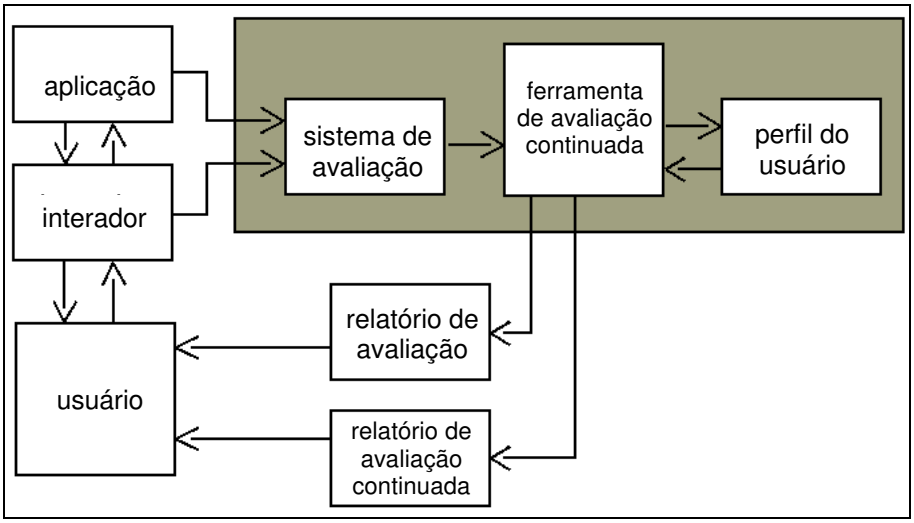

Figura 2.2. Diagrama de um sistema de avaliação continuada.

2.5 Avaliação de treinamento pela Web

A avaliação em treinamentos baseados em sistemas interativos na Web tem particular importância com a expansão do ensino à distância. Uma avaliação deste tipo pode ser realizada por meio de questionários corrigidos por ferramentas inteligentes, ou mesmo com a coleta de dados de interação posteriormente transmitidos para avaliação. Neste caso, é possível que a análise dos dados e o resultado sejam *online* ou *offline*.

Especificamente, o uso de sistemas de realidade virtual na Internet como ferramenta de ensino permitirá oferecer ambientes de treinamento mais realistas, ao mesmo tempo que exigirá maior velocidade do tráfego das informações para evitar interrupções nas reações ao usuário, comprometendo a característica de tempo-real destas aplicações. Neste aspecto, o tempo-real ainda é a maior limitação devido às altas taxas de transferência de informações necessárias para este tipo de aplicação.

Um problema relacionado à avaliação de treinamento pela Web é a garantia de identidade do aluno. Neste caso, há uma série de estudos que visa certificar a identidade do aprendiz/usuário, principalmente quando testes e questionários precisam ser respondidos [Thorpe,1998].

Referências

Machado, L.S.; Costa, T.K.L.; Restivo, M.T.; Moraes, R.M. Improving Interaction in Remote Laboratories Using Haptic Devices. In: REV2007 Conference Proceedings, 2007. CDROM.

Machado, L.S.; Moraes, R.M. Neural Networks for on-line Training Evaluation in Virtual Reality Simulators". Proceedings of World Congress on Engineering and Technology Education (WCETE'2004). Brasil, p. 157-160, 2004.

McCloy, R.; Stone, R.; Virtual reality in surgery. BMJ, v. 323, p. 912-915, 2001.

Moraes, R.M.; Machado, L.S. Fuzzy Gaussian Mixture Models for on-line Training Evaluation in Virtual Reality Simulators. Anals of the International Conference on Fuzzy Information Processing (FIP'2003). China, v. 2, p. 733-740, 2003.

Moraes, R.M.; Machado, L.S. Using Fuzzy Hidden Markov Models for Online Training Evaluation and Classification in Virtual Reality Simulators. Int. Journal of General Systems, v.33, n.2-3, pp. 281-288, 2004.

Moraes, R.M.; Machado, L.S. Evaluation System Based on EFuNN for On-line Training Evaluation in Virtual Reality. Lecture Notes in Computer Science, v. 3773, p. 778-785, 2005a.

Moraes, R.M.; Machado, L.S. Continuous Evaluation in Training Systems Based on Virtual Reality. Proceedings of Global Congress on Engineering and Technology Education (GCETE'2005). Brasil, p.1048-1051, 2005b.

Prata, C.L.; Nascimento, A.C. Objetos de aprendizagem: uma proposta de recurso pedagógico. Organizadoras. Brasília : MEC, SEED, 2007.

Pugh, C.A.; Youngblood, P. Development and validation of assessment measures for a newly developed physical examination simulator. Journal of the American Medical Informatics Association v. 9, p.448-460, 2002.

[Salvador et al.,2000] Salvador, C.C. et al. Psicologia do Ensino. Artmed. 408p.

Thorpe, M.; Assessment and 'Third Generation' Distance Education. Distance Education 19, n. 2, p. 265-286, 1998.

Capítulo 3. Métodos de Avaliação de Treinamento

Introdução

Os atributos encontrados em sistemas de treinamento baseados em RV podem ser ambientes compostos por objetos com topologias e comportamentos similares aos objetos reais. Os modos de interação com esse ambiente, trazem a possibilidade de realizar um treinamento sem riscos ou danos para o usuário/trainee ou para outras pessoas [Burdea, 2003]. Vários simuladores também foram propostos, tais como: exame de próstata [Burdea et al.,1998], cirurgia ocular [Mahoney,1998], laparoscopia [Voss et al.,2000] [McBeth et al.,2002], extração de medula [Machado et al.,2001], cirurgia de medula [Morris, et al.,2006] e exame ginecológico [Machado et al.,2006], entre outros. A meta de muito destes sistemas é prover um ambiente de treinamento similar ao ambiente do procedimento real pelo uso de dispositivos e técnicas que explorem os sentidos humanos. Em várias situações os sistemas podem prover treinamento para um usuário a cada vez, mas alguns procedimentos precisam ser realizados por várias pessoas simultaneamente, tais como cirurgias, ou o controle de um avião civil.

Na literatura, alguns tipos de avaliação tem sido propostos. Gande and Devarajan (2003) usaram uma estação Instrutor/Operador para monitorar os movimentos do usuário para aumentar ou diminuir o grau de dificuldade da simulação. A estação Instrutor estava capacitada a avaliar e monitorar o desempenho do médico residente baseado nas especificações mencionadas para gerar um relatório de efetividade. O Sistema GeRTiSS [Alcañiz et al, 2003] pode também emitir um relatório de avaliação ao final da simulação, o qual contém informações sobre o tempo da intervenção, número de cortes e cauterizações, entre outras. O cirurgião pode usar essas informações para fazer uma auto-avaliação do seu procedimento. O sistema WebSET [John et al.,2001] simula cenários-teste nos quais os usuários podem fazer simulações. Neste sistema, os usuários são avaliados pela velocidade das respostas, número de movimentos do *mouse*, acurácia da resposta, tomadas de decisão e aprendizado.

No caso da avaliação em ensino ou treinamento à distância via Web [Keegan,1991], nem sempre os sistemas possuem alguma forma de avaliação, sendo em grande parte das vezes, sistemas que simplesmente transmitem o conhecimento usando

vários recursos multimídia [Moraes e Machado,2001][Univeridade do Oregon,2006]. Existem ainda os sistemas de treinamento realistas usando gráficos de alta resolução e que permitem ações de simulação pseudo-real sobre mecanismos [Pucel e Anderson,2003] ou sobre cenários [John et al.,2001]. Entretanto, a avaliação sempre foi um dos principais problemas, e muito se discute sobre as vantagens e desvantagens dessa forma de ensino ou treinamento e como introduzir as componentes de avaliação [Hooper,1998]. Existem sistemas de avaliação automáticos e semi-automáticos para análise de arquivos de registro (*log files*) para a criação de um perfil para avaliação [Zaina et al.,2002]. Nesta categoria pode-se encontrar sistemas que rastreiam as ações do estudante/usuário sobre tópicos específicos ou recursos utilizados. Outros sistemas [Pesin et al.,2003] fazem a avaliação usando testes, exercícios, perguntas e questionários. Brusilovsky e Miller (1999) usaram testes locais e alguma inteligência para não repetência dos pré-testes aplicados ou para avaliar as respostas.

3.1 Métodos para Avaliação em Realidade Virtual

Geralmente, um método de avaliação *online* deveria ser capaz de monitorar o sistema de simulação e as interações do usuário com ele. Para isso, é necessário coletar informações sobre posições no espaço, forças, torques, resistência, velocidade, aceleração, temperaturas, visualização e/ou ângulo da visualização, sons, aromas, etc. Essas informações serão usadas para alimentar o sistema de avaliação. Na Figura 3.1, pode-se observar que o simulador em realidade virtual (RV) e o sistema de avaliação são sistemas independentes, mas que ajem de modo simultâneo. As interações do usuário com o sistema são monitoradas e as informações são enviadas para o sistema de avaliação que analisa os dados e emite um relatório sobre o desempenho do usuário ao final do treinamento. Dependendo da aplicação, todas aquelas variáveis ou algumas delas serão monitoradas de acordo com a sua relevância ao treinamento.

Alguns exemplos disponíveis na literatura mostram o uso de análise de movimentos opto-eletrônicos e gravações em vídeo, para adquirir dados de movimentos e posturas de especialistas e médicos residentes em diferentes contextos, usando distribuições de probabilidade e estatística descritiva [McBeth et al.,2002]. Pucel e Anderson (2003) desenvolveram uma simulação oftálmica computacional para testar a habilidade pessoal para desempenhar tarefas usando estatísticas descritivas. Morris et al. (2006) sugerem o uso de modelos de regressão linear para avaliar o progresso do usuário em cirurgias de medula. Entretanto, também existem outras formas de se avaliar o usuário através de técnicas mais sofisticadas e algumas delas são descritas nas próximas seções deste Capítulo. Mas, antes é interessante contextualizar o procedimento da avaliação de desempenho, no universo da simulação.

Na medicina, por exemplo, vários procedimentos são realizados sem informação visual para o médico, como o exame de próstata ou a coleta de medula óssea. Para treinar estes tipos de procedimentos, ambientes virtuais são propícios, pois permitem manipulação ilimitada e eliminam a necessidade de modelos plásticos ou de cobaias. Entretanto, a principal vantagem é a representação das propriedades físicas dos tecidos humanos vivos, dado que organismos mortos apresentam degradação das suas propriedades ao longo do tempo [Holton,2001]. Essa representação é possível usando

dispositivos hápticos que oferecem sensação de toque ao usuário [Mahoney,1997] e nesse caso, os movimentos poderiam ser coletados e utilizados na avaliação do desempenho do usuário [Moody et al.,2002]. Especialistas ainda fazem avaliações analisando vídeos gravados, observando as interações dos usuários durante a execução dos procedimentos [Burdea et al.,1998].

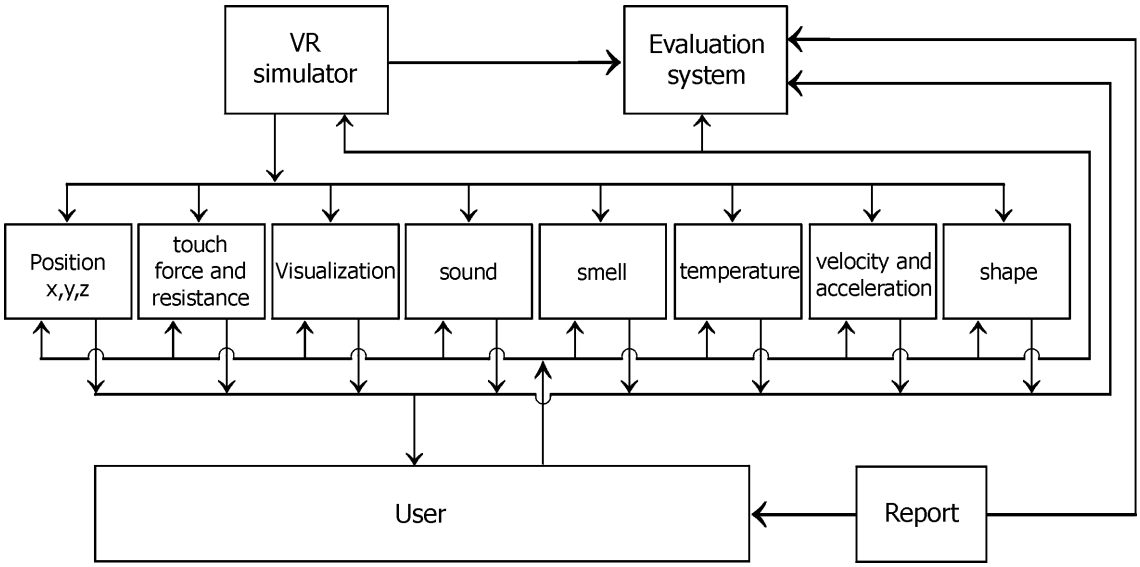

Figura 3.1. Diagrama de um sistema completo: O Simulador em Realidade Virtual e o Sistema de Avaliação.

A ferramenta de avaliação deveria supervisionar os movimentos do usuário e outros parâmetros associados a ele. Porém, existe a necessidade de que a ferramenta seja calibrada para tal. Assim, um ou mais especialistas calibram a ferramenta de avaliação, executando várias vezes o procedimento e rotulando-os segundo as classes de desempenho previstas por ele(s) próprio(s). Por exemplo, para $M=4$ classes, as suas interpretações poderiam ser: "bem qualificado", "necessita de algum treinamento ainda", "necessita de mais treinamento", "novato", etc. Assim, a informação sobre a variabilidade desses procedimentos são adquiridas usando uma metodologia apropriada. Quando um usuário ou *trainee* usa o sistema para o seu treinamento, o seu desempenho é comparado com as M classes de desempenho e é calculada uma medida de compatibilidade entre eles. Ao final, o desempenho do usuário é rotulado pela classe com maior grau de compatibilidade e o usuário recebe um relatório com todas as possíveis classes de desempenho e a sua compatibilidade com cada uma delas.

3.1.1 Métodos Baseados em Sistemas Especialistas

3.1.1.a Usando a Lógica Clássica

Avaliação inteligente usando a lógica clássica é aquela na qual um sistema computacional procura realizar a tarefa de avaliação de modo similar a um ou mais especialistas da área. Em sistemas de avaliação inteligentes, o conhecimento de um ou mais especialistas é armazenado em uma base de conhecimento usando alguma representação lógica, como por exemplo, por regras. O controle da utilização das regras é feito por um sistema de inferência, a aprtir de fatos, que são dados que podem ser adquiridos inicialmente ou posteriormente, durante o processo de dedução [Buchanan e Shortlife,1985]. Usando-se regras e fatos, novos fatos são obtidos utilizando-se uma ou mais formas de raciocínio. A arquitetura formada pela base de conhecimento e pelo sistema de inferência é chamada "sistema especialista" [Rich e Knight,1993]. Nos sistemas especialistas, o conhecimento sobre um determinado procedimento é modelado por regras, que são formulações lógicas do tipo:

SE <condição> ENTÃO <conclusão>.

Uma regra pode ser composta por conjunções e disjunções e pode levar a mais de uma conclusão simultaneamente. Por exemplo:

SE (<condição1> OU <condição2>) E (<condição3>)
ENTÃO (<conclusão1> E <conclusão2>).

A regra acima com mais de uma conclusão é equivalente a um conjunto de regras com apenas uma conclusão:

SE (<condição1> OU <condição2>) E (<condição3>)
ENTÃO <conclusão1>
SE (<condição1> OU <condição2>) E (<condição3>)
ENTÃO <conclusão2>.

Cada regra diz respeito a uma variável de interesse e cada especialista pode ter a sua própria opinião sobre cada fato específico, porém existe a necessidade de que não haja conflito entre essas opiniões para manter a consistência do conhecimento. Isso significa que o sistema não poderá, por exemplo, concluir "sim" e "não" a partir das mesmas informações. Quando os valores das variáveis de interesse são medidos por um subsistema externo e enviados ao sistema especialista, o sistema de inferência dispara regras da base de conhecimento, que avaliam os dados para gerar conclusões.

3.1.1.b Usando a Lógica Fuzzy

Como mencionado anteriormente, em sistemas de avaliação inteligentes, o conhecimento do especialista deve ter algum tipo de representação lógica, como por exemplo, a representação por regras. A arquitetura formada pela base de conhecimento e pelo sistema de inferência é chamada "sistema especialista" [Rich e Knight,1993]. Podem haver casos, nos quais os atributos não podem ser medidos com a exatidão necessária, ou o próprio especialista verifica que o conhecimento não é válido em qualquer situação. Nestes casos, podem ser introduzidos coeficientes de certeza para codificar essa incerteza na conclusão da regra [Moraes,1998][Moraes et al.,2002]. A informação incerta pode ser tratada através de probabilidades [Rich e Knight,1993], crença e plausibilidade [Senne,1988], possibilidades [Dubois e Prade,1996], entre outras.

Para se definir um conjunto "fuzzy", faz-se necessário relembrar a definição de conjunto na teoria dos conjuntos clássica: um conjunto A de um universo X pode ser definido por uma função de pertinência $\mu_A(x)$, com $\mu_A: X \rightarrow \{0,1\}$, onde 1 significa que um dado $x \in X$ está incluído em A e 0 significa que x não pertence a A. Porém, para um conjunto "fuzzy" não se pode afirmar sempre que um dado elemento $x \in X$ pertence ou não ao conjunto A. Dessa forma, cria-se uma nova forma de função de pertinência, $\mu_A(x)$ onde $\mu_A: X \rightarrow [0,1]$, ou seja, podem existir graus de pertinência parciais para elementos de X em A [Zadeh,1965]. Pode-se dizer que um conjunto "fuzzy" A de X é dito ser "preciso" quando $\exists\, c^* \in X$ tal que $\mu_A(c^*)=1$ e $\forall c \neq c^*,\ \mu_A(c)=0$. Um conjunto "fuzzy" A é dito ser exato (ou "crisp"), quando $\forall c \in X,\ \mu_A(c) \in \{0,1\}$.

Quando há subjetividade na base de conhecimento, sua representação pode ser feita por modelos "fuzzy" [Zadeh,1988], onde a subjetividade é modelada por conjuntos "fuzzy", como descrito acima. Em sistemas especialistas "fuzzy", o conhecimento do especialista sobre um determinado procedimento é modelado por regras "fuzzy". Cada regra diz respeito a uma variável de interesse e cada especialista pode ter a sua própria opinião sobre cada fato específico. Quando os valores das variáveis de interesse são medidos por um subsistema externo e enviados ao sistema especialista, o sistema de inferência dispara regras da base de conhecimento, que avaliam os dados para gerar conclusões.

Por exemplo, utilizando um sistema especialista "fuzzy" acoplado a um simulador cirúrgico em RV, especialistas podem definir de modo impreciso as camadas de tecido e regiões do osso nas quais pode-se extrair a medula com segurança e nas quantidades desejadas. Para avaliar o treinamento de um residente, os valores das variáveis são coletados pelo dispositivo que fornece reação tátil Phantom Desktop [Massie e Salisbury,1994] e enviados ao sistema especialista para julgamento. O sistema deve analisar cada passo realizado pelo residente em tempo real e ir armazenando as suas impressões para no final do treinamento classificar o candidato segundo as classes de desempenho. Em Machado et al. (2000) utilizou-se cinco classes descritas pelos seguintes conjuntos nebulosos: "precisa de muito mais treino no procedimento"; "precisa de mais treino no procedimento"; "ainda precisa de treino no procedimento"; "procedimento bem realizado"; "procedimento muito bem realizado", demonstrando a necessidade ou não de mais treino.

As regras do sistema especialista foram modeladas por funções de pertinência de acordo com as especificações de especialistas e foram utilizadas as funções de pertinência trapezoidais [Dubois e Prade,1980], pois há um erro tolerável na posição da agulha em todos os eixos e nos ângulos e nas forças aplicadas que permitirão ao residente obter sucesso na punção, da mesma forma. O sistema de inferência utilizado foi do tipo Mamdani [Mamdani e Assilian,1975]. Um exemplo de regra utilizado neste avaliador inteligente é:

Se Posição_X é *esquerda_do_centro* E Posição_Y é *acima_do_centro* E Posição_Agulha é *aceitável* E Força é *forte* E Extração_Medula é *Sim* Então Classe_residente é *ainda precisa de treino no procedimento*.

onde: Posição_X e Posição_Y são coordenadas onde a agulha toca o corpo do paciente; Posição_Agulha é o ângulo que a agulha entra no corpo do paciente; Força é a força aparente com que o residente realiza a entrada da agulha; Extração_Medula mostra o sucesso ou fracasso do residente em extrair a medula e Classe_residente é a classificação do residente.

Segundo Mandani e Assilian (1975), a avaliação das proposições com conectivo E e OU são realizadas de acordo com as operações sobre conjuntos "fuzzy" de intersecção e união, respectivamente. A intersecção e a união de dois conjuntos "fuzzy" são executadas através do uso dos operadores de *t-normas* e *t-conormas*, respectivamente, as quais são mapeamentos comutativos, associativos e monotonicos de $[0,1]\rightarrow[0,1]$. Além disso, uma t-norma Γ (respec. t-conorm $\perp$) tem 1 (respec. 0) como elemento neutro (por exemplo: $\Gamma=min$, $\perp=max$) [Dubois e Prade,1988]. Assim, a intersecção de dois conjuntos "fuzzy" A e B com funções de pertinência $\mu_A(x)$ e $\mu_B(x)$ é um conjunto "fuzzy" C com função de pertinência dada por:

$$C=A \cap B \Leftrightarrow \mu_C(x)= \Gamma\{\mu_A(x),\mu_B(x)\},\ \forall x\in X.$$

A união de dois conjuntos "fuzzy" A e B com funções de pertinência $\mu_A(x)$ e $\mu_B(x)$ é um conjunto "fuzzy" C com função de pertinência dada por:

$$C=A \cup B \Leftrightarrow \mu_C(x)= \perp\{\mu_A(x),\mu_B(x)\},\ \forall x\in X.$$

O complemento de um conjunto "fuzzy" A em X, denotado por $\neg A$ é definido como:

$$\mu_{\neg A}(x) = n(\mu_A(x)),\ \forall x\in X.$$

onde: $n: [0,1]\rightarrow[0,1]$ é um operador de negação que satisfaz as seguintes propriedades:

- $n(0)=1$ e $n(1)=0$
- $n(a) \leq n(b)$ se $a>b$
- $n(n(a))=a,\ \forall x\in [0,1]$

e uma negação é chamada "estrita" se é continua e também satisfaz:

- $n(a)<n(b)$ se $a>b$.

O principal operador de negação que satisfaz as quarto condições acima é $n(a) = 1-a$.

A função de implicação entre dois conjuntos "fuzzy" A e B com funções de pertinência $\mu_A(x)$ e $\mu_B(x)$ é um conjunto "fuzzy" C com função de pertinência dada por:

$$C=A \Rightarrow B \Leftrightarrow \mu_C(x,y)= \nabla\{\mu_A(x),\mu_B(y)\},\ \forall x\in X,\ \forall y\in Y$$

onde ∇: $[0,1]^2 \rightarrow [0,1]$ é um operador de implicação que obedece as seguintes propriedades: $\forall a, a', b, b' \in [0,1]$:

- Se $b \leq b'$ então $\nabla(a,b) \leq \nabla(a,b')$;
- $\nabla(0,b)=1$;
- $\nabla(1,b)=b$.

As implicações puras obedecem também a:

- Se $a \leq a'$ então $\nabla(a,b) \geq \nabla(a',b)$;
- $\nabla(a, \nabla(b,c))= \nabla(b, \nabla(a,c))$.

Quando se usa uma t-norma para representar o conectivo E e uma t-conorma para representar o conectivo OU, a implicação ∇ é dada por uma t-norma. É importante observar que a escolha dessas normas deve seguir algumas regras [Moraes,1998].

No caso do procedimento de coleta de medula óssea, este é totalmente realizado às cegas e portanto, o especialista avaliador humano só pode avaliar o procedimento através de suas impressões externas ao corpo do paciente, pois é apenas o que ele pode ver. O avaliador proposto por Machado et al. (2000) tem uma posição privilegiada na avaliação do médico novato, pois poderá acompanhar toda a trajetória do procedimento, externa e internamente ao corpo virtual sobre o qual atua, medindo posição e força da agulha transpassando as camadas de tecido e do osso, impossíveis de serem avaliadas pelo observador humano.

3.1.1.c Usando Múltiplos Sistemas Especialistas

Em casos complexos, modelar o conhecimento de um procedimento pode exigir uma abordagem menos conservadora. Retomando os casos médicos, existem exames que são feitos em fases seqüenciais e exigem diferentes conhecimentos para cada uma. Por exemplo, no caso do exame ginecológico, existem diferentes patologias, tais como o caso normal, Herpes ou HPV e casos de inflamação que precisam ser identificados, como descritos na Tabela 3.1. Machado et al. (2006) propuseram um simulador de Exames Ginecológicos (SITEG) que permite o treinamento de exames ginecológicos e simula diferentes fases da patologia de forma aleatória.

Tabela 3.1. Descrição das propriedades visuais e táteis.

	Color	Texture	Viscosity	Cervix elasticity
Normal	rosy	similar to buccal mucous membrane	smooth	Similar to an orthopedic rubber
Herpes / HPV	white with warts	spongy and irregular	with bubbles	very soft
Inflamed	red	similar to buccal mucous membrane	smooth	hard/tense

Os dois estágios do exame real foram divididos para compor a fase visual e a fase tátil do exame. Na fase visual, o usuário deve colocar o espéculo para manter a vagina aberta e deve observar as paredes vaginais e o cólo do útero e notar a sua coloração. Após o exame visual, ele faz o exame de toque notando a textura e buscar a existência de feridas e/ou caroços. A diferença entre o exame real e o simulado é a colocação do especulo que não é simulado. Um dispositivo háptico é usado no primeiro estágio simulando a lanterna usada pelo médico para melhorar a sua visualização do interior da vagina e na segunda fase, este mesmo dispositivo fornece o retorno tátil necessário para o exame de toque (Figura 3.2).

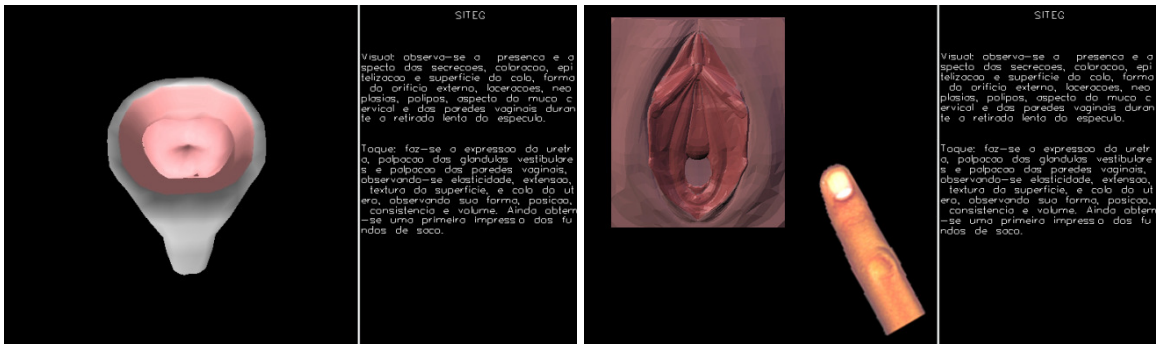

Figura 3.2. Exame visual das paredes da vagina e do colo do útero no SITEG. O objeto em cinza clario é a representação de um espéculo (esquerda); Vista externa od exame unidigital no SITEG (direita).

Neste caso, o sistema de avaliação não é similar àqueles apresentados previamente para procedimentos cirúrgicos e requer uma abordagem diferenciada, baseada no conhecimento que o usuário usou para construir o seu diagnóstico e se este diagnóstico está correto ou não. Burdea et al. (1999) parcialmente usou este mesmo conceito no seu simulador de exame de próstata e isso foi extendido para uma avaliação computacional, a qual permite conhecer informações relevantes que contribuíram para o diagnóstico final e para avaliar este diagnóstico. Assim, o sistema de avaliação deve capturar as informações

relevantes sobre o exame em seus vários estágiso e interagir com o usuário para conhecer as razões do seu diagnóstico final. Este diagnóstico deve ser avaliado em acordo ao caso apresentado pelo simulador, cujas patologias são aleatoriamente apresentadas ao usuário. A avaliação no simulador de exames ginecológicos segue o formato real do exame que o futuro médico executará na sua vida profissional, onde ele deverá reconhecer os diferentes tipos de casos.

O sistema avaliador do SITEG é baseado em dois sistemas especialistas basedos em regras "fuzzy", onde cada um é baseado no conhecimento obtido de especialistas para cada fase, como apresentado na Figura 3.3. O primeiro sistema corresponde à fase visual e o segundo à fase de toque. Na fase visual, o usuário deve identificar a coloração das paredes vaginais e o cólo do útero, de acordo com os diagnósticos apresentados na Tabela 3.1. Cada coloração tem a sua própria função de pertinência "fuzzy": normal (rósea), Herpes ou HPV (esbranquiçada) e inflamada (vermelha).

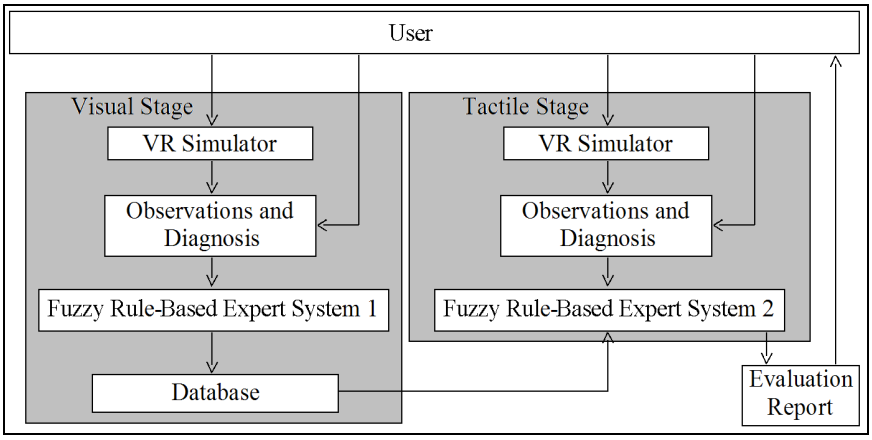

Figura 3.3. Diagrama do Sistema de Avaliação.

Ao final do primeiro estágio, o usuário deve prover ao sistema a sua opinião sobre a coloração coloração das paredes vaginais e o colo e o seu diagnóstico preliminar. O sistema de avaliação compara-o com a informação relacionada ao caso apresentado ao usuário e as armazena em uma base de dados que será utilizada pelo segundo sistema especialista basedos em regras "fuzzy" na segunda fase.

Na segunda fase, o usuário realiza o exame de toque das paredes da vagina e do colo do útero e emite novamente a sua opinião sobre os possíveis diagnósticos. Essa informação é utilizada e armazenada pelo segundo sistema especialista Ao final das fases, o sistema de avaliação requer do usuário um diagnóstico final que deve estar entre:

$$D = \{\text{normal, inflamada, herpes ou HPV, dúvida}\}.$$

A opção de "dúvida" é pertinente devido a combinação de informações durante o exame real eventualmente não serem conclusivas. Isto pode ocorrer devido a contradições nas decisões do médico sobre as condições da paciente entre a primeira e a segunda fase do exame. Internamente, o sistema de avalição executa em tempo-real as regras de cada fase de acordo com o caso apresentado ao usuário. O sistema de avaliação é também capaz de decidir se o treinamento foi realizado com sucesso pela combinação das duas fases e do diagnóstico final com o caso apresentado pelo simulador. Assim, o sistema de

avaliação pode classificar o usuário nas diferentes classes de desempenho escolhidas pelos especialistas. No trabalho de Machado et al. (2006), foram usadas cinco classes de desempenho, com a seguinte interpretação: a) o usuário está qualificado para executar um exame real; b) o usuário está quase qualificado para executar um exame real; c) o usuário ainda precisa de treinamento para executar um exame real; d) o usuário precisa de mais treinamento para executar um exame real; e) usuário é um iniciante e precisa de muito treinamento para executar um exame real.

É importante notar que nos casos de diagnósticos errados, o sistema de avaliação está apto a detectar onde o usuário cometeu o erro pela análise dos graus de pertinência da sua informação. Todas essas informações, incluindo a classe de desempenho do usuário, é fornecida ao usuário através do relatório de avaliação (Figura 3.3). Devido a todos esses atributos, essa metodologia de avaliação também pode ser usada na avaliação contínua do treinamento para o mesmo usuário.

3.1.2 Modelos de Mistura

3.1.2.a Modelos de Mistura Gaussianos

Nos casos onde as distribuições estatísticas dos dados das variávies não seguem uma distribuição normal (univariada ou multivariada), uma possível solução é o uso dos modelos de mistura. Moraes e Machado (2003a) apresentaram uma possível solução usando Modelos de Mistura Gaussianos (GMM) como o cerne de um sistema de avaliação.

As equações para estimação de parâmetros do modelo são apresentados a seguir [Tran e Wagner,1999a]. Seja $X=\{x_1, x_2,...,x_T\}$ o conjunto de T vetores, onde cada um é um atributo d-dimensional extraído pelas T diferentes informações no espaço virtual e obtidas pelo simulador. Essas informações podem ser forças aplicadas, ângulos, posições tri-dimensionais e torque, extraídas em a cada intervalo de tempo pré-definido. Dado que a distribuição estatística desses vetores pode ser desconhecida, ela pode ser modelada de forma aproximada por uma mistura de densidades normais (ou gaussianas), como uma soma ponderada de c densidades, dada pela seguinte equação:

$$p(x_t \mid \lambda) = \sum_{i=1}^{c} w_i N(x_t, \mu_i, \Sigma_i), \quad t=1,...,T \tag{3.1}$$

onde λ denota um protótipo consistente com um conjunto de parâmetros do modelo $\lambda = \{w_i, \mu_i, \Sigma_i\}$, w_i, $i=1,...,c$ são os pesos da mistura e $N(x_t, \mu_i, \Sigma_i)$ são as d-variadas densidades normais, com vetores de media μ_i e matrizes de covariância Σ_i:

$$N(x_t, \mu_i, \Sigma_i)=\exp\{-1/2\}(x_t - \mu_i)' \Sigma_i^{-1} (x_t - \mu_i)\} / (2\pi)^{d/2} |\Sigma_i|^{1/2} \tag{3.2}$$

Para treinar um GMM, os parâmetros são estimados pelo melhor casamento entre os vetores de treinamento e a distribuição. O estimador de Máxima Verossimilhança é largamente usado como método de treinamento. Para uma seqüência de vetores de treinamento X para um λ, a verossimilhança do GMM é dada por:

$$p(X|\lambda) = \prod^{T}_{t=1} p(x_t|\lambda) \quad (3.3)$$

O objetivo da estimação de Máxima Verossimilhança é encontrar um novo modelo $\overline{\lambda}$ tal que $p(X|\overline{\lambda}) \geq p(X|\lambda)$. Como a expressão em (3.3) é uma função não-linear dos parâmetros em λ, a sua maximização direta não é possível. Entretanto, esses parâmetros podem ser obtidos interativamente usando o algoritmo Expectation-Maximization [Dempster et al.,1977]. Nesse algoritmo, usa-se uma função auxiliar Q dada por:

$$Q(\lambda, \overline{\lambda}) = \sum^{T}_{t=1} \sum^{c}_{i=1} p(i|x_t, \lambda) \log [w_i N(x_t, \overline{\mu_i}, \overline{\Sigma_i})] \quad (3.4)$$

onde $p(i|x_t, \lambda)$ é a probabilidade *a posteriori* para a classe de desempenho i, $i=1,...,c$ e satisfaz

$$p(i|x_t, \lambda) = [w_i N(x_t, \mu_i, \Sigma_i)] / \{\sum^{c}_{k=1} w_k N(x_t, \mu_k, \Sigma_k)\} \quad (3.5)$$

O algoritmo Expectation-Maximization é tal que se $Q(\lambda, \overline{\lambda}) \geq Q(\lambda,\lambda)$ então $p(X|\overline{\lambda}) \geq p(X|\lambda)$ [Rabiner,1989]. As seguintes equações de reestimação são encontradas a partir das derivadas da função Q com respeito a $\overline{\lambda}$ tendendo a zero:

$$\overline{w_i} = 1/T \sum^{T}_{t=1} p(i|x_t, \lambda) \quad (3.6)$$

$$\overline{\mu_i} = \sum^{T}_{t=1} [p(i|x_t, \lambda)x_t] / [\sum^{T}_{t=1} [p(i|x_t, \lambda)] \quad (3.7)$$

$$\overline{\Sigma_i} = \{\sum^{T}_{t=1} [p(i|x_t, \lambda)(x_t - \overline{\mu_i}) (x_t - \overline{\mu_i})'\} / [\sum^{T}_{t=1} [p(i|x_t, \lambda)] \quad (3.8)$$

O algoritmo para o treinamento de um GMM é dado por:

1. Gerar a probabilidade *a posteriori* $p(i|x_t, \lambda)$ aleatoriamente e satisfazendo (3.5);
2. Computar os pesos da mistura, o vetor de médias e a matriz de covariâncias seguindo (3.6), (3.7) e (3.8);
3. Atualizar as probabilidades *a posteriori* $p(i|x_t, \lambda)$ de acordo com (3.5) e computer a função Q usandop (3.4);
4. Parar se o incremento do valor da função Q na interação corrente, em relação ao valor da mesma função na interação anterior estiver abaixo de um valor limite escolhido previamente. Caso contrário, retorna ao passo 2.

A classificação pelo GMM

Para executar a classificação pelo GMM, necessita-se várias classes de desempenho λ. Assim, sejam λ_k, $k=1,...,N$, modelos de N possíveis classes de desempenho. Dado um vetor de atributos X, um classificador é delineado para classificar X em cada uma das N classes de desempenho usando N funções discriminantes $g_k(X)$. Computa-se as

similaridades entre o X desconhecido e cada classe de desempenho λ_k e seleciona-se a classe de desempenho λ_{k*} se:

$$k^* = \underset{1 \leq k \leq N}{\arg\max} \; g_k(X) \tag{3.9}$$

No classificador de taxa de erro mínima, a função discriminante é a probabilidade *a posteriori*:

$$g_k(X)=p(\lambda_k|X) \tag{3.10}$$

Pode-se usar a Regra de Bayes

$$p(\lambda_k|X) = [p(\lambda_k)\, p(X|\lambda_k)] \,/\, p(X) \tag{3.11}$$

e pode-se assumir a probabilidade a priori igual para todas as classes de desempenho, ou seja, $p(\lambda_k)=1/N$. Se $p(X)$ é a mesma para todos os modelos de desempenho, a função discriminante de (3.10) é equivalente a {Joh98}:

$$g_k(X)=p(X|\lambda_k) \tag{3.12}$$

Finalmente, usando a log-verossimilhança, a regra de decisão usada para a identificação da classe de desempenho é dada por:

Seleciona-se o modelo de desempenho k se*

$$k^* = \underset{1 \leq k \leq N}{\arg\max} \; \sum^{T}_{t=1} \; p(x_t|\lambda_k) \tag{3.13}$$

onde $p(x_t|\lambda_k)$ é dado por (3.1) para cada k, $k=1,...,N$.

3.1.2.b Modelos de Mistura Gaussianos Fuzzy

Para os casos nos quais há imprecisão nas medidas das variáveis ou existe a possibilidade múltiplas formas de execução correta de um procedimento, pode-se usar uma abordagem "fuzzy" na avaliação, ou mais precisamente para o GMM. Tran e seus colaboradores [Tran et al.,1998] [Tran e Wagner,1998] [Tran e Wagner,1999a] propuseram uma abordagem "fuzzy" para o GMM, denotado por FGMM e esta abordagem foi seguida por Moraes e Machado (2003b). Tran e seus colaboradores usaram uma modificação do algoritmo "Fuzzy C-Means" para estimar os parâmetros do GMM. A principal vantagem dessa abordagem é a habilidade de reconhecimento de diferentes formas geométricas para as distribuições de X usando as matrizes de covariância e os pesos da mistura. A partir da função de soma dos erros quadráticos intra-grupos de Bezdek [Bezdek,1993], usada como função objetivo:

$$J_m(U,\lambda) = \sum^{T}_{t=1} \sum^{c}_{i=1} u^m_{it}\, d^2_{it} \tag{3.14}$$

onde $U=\{u_{it}\}$ é uma partição-c "fuzzy" de X; cada u_{it} representa o grau de pertinência do vetor x_t à i-ésima classe. Para $1 \le i \le c$ e $1 \le t \le T$, tem-se que:

$$0 \le u_{it} \le 1, \qquad \sum^{c}_{i=1} u^m_{it} = 1 \text{ and } \qquad 0 < \sum^{T}_{t=1} u_{it} < T; \tag{3.15}$$

$m \ge 1$ é um expoente de ponderação de cada pertinência "fuzzy" u_{it} e é chamado grau de nebulosidade ("fuzziness"). A medida de dissimilaridade denotada por $d^2_{it} > 0$, é a uma medida de distância na equação (3.14) e é dada por:

$$d^2_{it} = -\log p(x_t, i|\ \bar{\lambda}) = -\log [\ \bar{w}_i N(x_t, \bar{\mu}_i, \bar{\Sigma}_i)] = -\log \bar{w}_i + \tfrac{1}{2} \log (2\pi)^d |\ \bar{\Sigma}_i\ | + \\ + \tfrac{1}{2} (x_t - \bar{\mu}_i)'\ \bar{\Sigma}_i^{-1} (x_t - \bar{\mu}_i). \tag{3.16}$$

Substituindo (3.16) em (3.14), obtém-se:

$$J_m(U,\lambda) = -\sum^{T}_{t=1} \sum^{c}_{i=1} u^m_{it} \log \bar{w}_i - \sum^{T}_{t=1} \sum^{c}_{i=1} u^m_{it} [\ \tfrac{1}{2} \log (2\pi)^d |\ \bar{\Sigma}_i\ |\ ] - \\ - \sum^{T}_{t=1} \sum^{c}_{i=1} u^m_{it} [\tfrac{1}{2} (x_t - \bar{\mu}_i)'\ \bar{\Sigma}_i^{-1} (x_t - \bar{\mu}_i)\ ] \tag{3.17}$$

Para minimizar J_m usa-se o Método do Multiplicador de Lagrange [Tran et al.,1998]:

$$\bar{w}_i = \sum^{T}_{t=1} u^m_{it} \ / \sum^{c}_{i=1} \sum^{T}_{t=1} u^m_{it} \tag{3.18}$$

$$\bar{\mu}_i = \sum^{T}_{t=1} u^m_{it}\ x_t \ / \sum^{T}_{t=1} u^m_{it} \tag{3.19}$$

$$\bar{\Sigma}_i = \{ \sum^{T}_{t=1} u^m_{it} (x_t - \mu_i)'\ (x_t - \mu_i)\ \} / \sum^{T}_{t=1} u^m_{it} \tag{3.20}$$

O algoritmo para o treinamento de um FGMM é dado abaixo [Tran e Wagner,1999a]:

1. Escolher c e m, $2 < c < T$ e $m > 1$. Gerar a matriz U aleatoriamente satisfazendo (3.15);
2. Para i=1,...,c, computar os pesos de mistura "fuzzy", o vetor de medias "fuzzy" e a matriz de covariâncias "fuzzy", seguindo as equações (3.18), (3.19) e (3.20).
3. Computar as medidas d_{it}, $1 \le i \le c$ e $1 \le t \le T$, seguindo (3.16). Atualizar a matriz U usando:

$$u_{it} = [\ \sum^{c}_{k=1} (\ d_{it} / d_{kt}\ )^{(2/(m-1))}\ ]^{-1} \tag{3.21}$$

Se $d_{it} = 0$ para algum t, fazer $u_{it} = 1$ e $u_{is} = 0$, para todo $s \neq t$.
4. Parar se a diferença entre o valor da função objetivo "fuzzy" $J_m(U, \bar{\lambda})$ na interação corrente e o seu valor na interação anterior for menor que um valor limite pré-estabelecido. Caso contrário, retornar ao passo 2.

A classificação pelo FGMM

No algoritmo GMM, dado um vetor de treinamento X, os parâmetros do modelo $\bar{\lambda}$ são determinados de modo que a função $Q(\lambda, \bar{\lambda})$ na equação (3.4) é maximizada e a função de verossimilhança $p(X| \bar{\lambda})$ é maximizada também. Então, pelo uso da log-verossimilhança na função discriminante dada por (3.12), obtém-se a regra de decisão para a identificação da classe de desempenho dada por (3.13). No método FGMM, existem dois casos [Tran e Wagner,1998] para selecionar a função discriminante: a) m é próximo de um (incluindo o próprio) ou b) $m > 1$ e m não é próximo da unidade.

No primeiro caso, minimizar J_m é equivalente a maximizar $Q(\lambda, \bar{\lambda})$ dado por (3.4) no GMM. Assim, a regra de decisão usada para a identificação no FGMM é a mesma aplicada ao GMM e dada por (3.13).

No segundo caso, minimizar J_m não é equivalente a maximizar $Q(\lambda, \bar{\lambda})$. Então é necessário escolher uma função discriminante alternativa:

$$Q_m(U,\lambda) = - J_m(U,\lambda) = \sum^{c}_{i=1} \sum^{T}_{t=1} u^m_{it} \log [\bar{w_i} N(x_t, \bar{\mu_i}, \bar{\Sigma_i})] \qquad (3.22)$$

De acordo com Tran e Wagner (1998), para reduzir os cálculos, a matriz U em $Q_m(U,\lambda)$ é substituída pela expressão (3.21) e a distância por (3.16):

$$Q_m(U,\lambda) = Q_m(\lambda) = \sum^{T}_{t=1} \{ \sum^{c}_{i=1} [- \log p(x_t, i| \bar{\lambda})]^{(1/(m-1))} \}^{1-m} \qquad (3.23)$$

Sejam λ_k, $k=1,...,N$, os modelos de N possíveis classes de desempenho. Dado um vetor X, um classificador é delineado para classificar X em cada uma das N classes de desempenho usando N funções discriminantes $g_k(X)$, tal que:

$$g_k(X) = Q_m(\lambda_k). \qquad (3.24)$$

Finalmente, a regra de decisão usada para a identificação da classe de desempenho é dada por:

Seleciona-se o modelo de desempenho k se*

$$k^* = \arg\max_{1 \le k \le N} \sum^{T}_{t=1} \{ \sum^{c}_{i=1} [- \log p(x_t, i| \bar{\lambda})]^{(1/(m-1))} \}^{1-m} \qquad (3.25)$$

para cada k, $k=1,...,N$.

Deve-se notar que a regra de decisão em (3.25) pode ser aplicada ao primeiro caso também, quando m é próximo da unidade.

3.1.2.c Modelos em Fases

Uma das formas de aperfeiçoar a avaliação é utilizando métodos em fases, nas quais o avaliador seguinte aperfeiçoa a avaliação anterior. Pode-se encontrar na literatura duas

propostas de avaliadores: os Modelos de Mistura Gaussianos com Relaxação e os Modelos de Mistura Gaussianos "Fuzzy" com Relaxação "Fuzzy". Esses modelos são estudados em detalhes a seguir.

Modelos de Mistura Gaussianos com Relaxação

Simulators baseados em RV para treinamento necessitam computadores topo-de-linha, com alta velocidade processamento para prover interação e retorno táteis realistas, visualização estereoscópica de modelos tri-dimensionais e texturas. Por estes motivos, os avaliadores do tipo "online" devem possuir baixa complexidade para não comprometer o desempenho da simulação, mas devem ter alta acurácia para não comprometer a avaliação. Os Modelos de Mistura Gaussianos (GMM) e os Modelos de Mistura Gaussianos "Fuzzy" (FGMM) foram mostrados como uma boa opção para realizar a tarefa. Entretanto, de acordo com Tran et al. (1999b) os Métodos de Relaxação (ou Relaxation Labeling – RL) podem oferecer melhor um desempenho melhorado para problemas deste tipo. Em seu trabalho, um GMM seguido por um RL obteve melhor desempenho na tarefa de classificação que um GMM atuando isoladamente. Moraes e Machado (2003c) propuseram usar a composição GMM-RL para avaliação de treinamento em ambientes baseados em RV.

O Método RL foi introduzido por Rosenfeld et al. (1976) e é uma abordagem interativa para aperfeiçoar as probabilidades de uma classificação prévia. Essa metodologia é empregada com sucesso há algum tempo em classificação de imagens [Fu e Yu,1980]. Seja um conjunto de objetos $A = \{a_1,a_2,...,a_N\}$ e um conjunto de classes $\Lambda=\{\lambda_1,\lambda_2,...,\lambda_N\}$. Uma probabilidade inicial é dada para cada objeto a_t tendo cada um uma classe λ_k, a qual é denotada por $p_t(\lambda_k)$. Essas probabilidades satisfazem a seguinte condição:

$$\Sigma^{N}_{k=1}\ p_t(\lambda_k) = 1, \text{ para todo } a_t \in A \qquad (3.26)$$

O Método RL atualiza as probabilidades $p_t(\lambda_k)$ usando um conjunto de coeficientes de compatibilidade coefficients $r_{tt'}(\lambda_k, \lambda_l)$, onde $r_{tt'}(\lambda_k, \lambda_l)$: $\Lambda\times\Lambda\rightarrow[-1,1]$, cuja magnitude denota o grau de compatibilidade. O significado desses coeficientes de compatibilidade pode ser interpretado como [Tran et al.,1999a]:

a) Se $r_{tt'}(\lambda_k, \lambda_l) < 0$, então λ_k, λ_l são incompatíveis para a_t e $a_{t'}$;
b) Se $r_{tt'}(\lambda_k, \lambda_l) = 0$, então λ_k, λ_l are independentes for a_t e $a_{t'}$;
c) Se $r_{tt'}(\lambda_k, \lambda_l) > 0$, então λ_k, λ_l are compatíveis for a_t e $a_{t'}$;

Para computar esses coeficientes, dois métodos são possíveis e estes métodos empregam conceitos de correlação estatística e informação mútua. Os dois métodos são baseados nos desenvolvidos por Peleg e Rosenfeld (1978). O estimador baseado em correlação dos coeficientes de compatibilidade é definido como:

$$r_{tt'}(\lambda_k)(\lambda_l) = \{\Sigma^{T}_{t=1}\ [p_t(\lambda_k)- p(\lambda_k)][p_{t'}(\lambda_l)- p(\lambda_l)]\}/\{\sigma(\lambda_k)\sigma(\lambda_l)\} \qquad (3.27)$$

onde $p_t(\lambda_l)$ é a probabilidade de $a_{t'}$ tendo o rótulo λ_l e $a_{t'}$ sendo os vizinhos de $a_{t'}$, $p(\lambda_l)$ é a média de $p_t(\lambda_l)$ para todo $a_{t'}$ e $\sigma(\lambda_l)$ é o desvio padrão de $p_t(\lambda_l)$. Para aliviar o efeito de domínio entre os rótulos, os coeficientes modificados são dados por [Tran et al.,1999a]:

$$r^*_{tt'}(\lambda_k)(\lambda_l) = [1 - p(\lambda_k)]\,[1 - p(\lambda_l)]\, r_{tt'}(\lambda_k)(\lambda_l) \tag{3.28}$$

e o estimador de informação mútua do coeficiente de compatibilidad é

$$r_{tt'}(\lambda_k)(\lambda_l) = \log \{T\, \Sigma^T_{t=1}\, p_t(\lambda_k)\, p_{t'}(\lambda_l)\} / \{\Sigma^T_{t=1}\, p_t(\lambda_k)\, p_{t'}(\lambda_l)\} \tag{3.29}$$

Os coeficientes de compatibilidade em (3.29) devem ser escalonados no intervalo [-1;1]. O fator de atualização para a estimativa $p_t(\lambda_k)$ na m-ésima interação é:

$$q_t^{(m)}(\lambda_k) = \Sigma^T_{t'=1}\; d_{tt'}\, [\Sigma^N_{l=1}\; r_{tt'}(\lambda_k)(\lambda_l)\, p_{t'}^{(m)}(\lambda_l)] \tag{3.30}$$

onde $d_{tt'}$ são os parâmetros que ponderam as contribuições para a_t advindas dos seus vizinhos $a_{t'}$ e sujeitas a:

$$\Sigma^T_{t'=1}\; d_{tt'} = 1 \tag{3.31}$$

A probabilidade atualizada $p_t^{(m+1)}(\lambda_k)$ para o objeto a_t é dada por:

$$p_t^{(m+1)}(\lambda_k) = \{p_t^{(m)}(\lambda_k)[1 + q_t^{(m)}(\lambda_k)]\} / \{\Sigma^N_{k=1}\; p_t^{(m)}(\lambda_k)[1 + q_t^{(m)}(\lambda_k)]\} \tag{3.32}$$

O algoritmo RL é apresentado a seguir:

1. Estimar as probabilidades iniciais para cada objeto satisfazendo (3.26)
2. Calcular os coeficientes de compatibilidade usando (3.28) ou (3.29)
3. Calcular o fator de atualização definido por (3.30)
4. Atualizar probabilidades para cada objeto usando (3.32)
5. Repetir os passos 3 e 4 até que a alteração das probabilidades seja menor que um limite escolhido ou igual ao número de interações pré-definido.

A classificação pelo GMM-RL

É óbvio que para o sucesso da classificação pelo GMM-RL, as probabilidades iniciais das classes e os coeficientes de compatibilidade devem ser bem determinados. Estimativas erradas destes parâmetros podem levar a instabilidades no método como um todo. A classificação baseada no GMM determina as probabilidades iniciais (*a priori*) para cada objeto (cada vetor de atributos) e na fase posterior (RL) são definidas as probabilidades finais (*a posteriori*), ainda que isso não possa ser vista do ponto de vista da Inferência Bayesiana. Diferentemente do Método RL para reconhecimento de padrões em imagens, onde a vizinhança m-conectada de "pixels" pode pertencer a diferentes regiões, na identificação do desempenho, todos os vetores de atributos desconhecidos na seqüência $X=\{x_1, x_2,...,x_T\}$ são sabidamente pertencentes a uma certa classe de desempenho λ. Assim, não existe a necessidade de considerar a compatibilidade entre o vetor de entrada e os vetores adjacentes [Tran et al.,1999a]. Isto conduz a:

$$p_{t'}(\lambda_l) = p_t(\lambda_l) \tag{3.33}$$

o que significa que a compatibilidade entre as diferentes classes é comente considerada para o mesmo objeto e portanto a equação (3.32) pode ser re-escrita como [Tran et al.,1999a]:

$$p_t^{(m+1)}(\lambda_k) = \{p_t^{(m)}(\lambda_k)[1- q_t^{(m)}(\lambda_k)]\} / \{\Sigma^N_{k=1}\, p_t^{(m)}(\lambda_k)[1- q_t^{(m)}(\lambda_k)]\} \tag{3.34}$$

O algoritmo GMM-RL para a identificação da classe de desempenho é dada por:

1. Estimar as probabilidades iniciais para cada classe de desempenho usando as probabilidades *a posteriori* dadas em (3.3):

$$p_t(\lambda_k) = p(\lambda_k|x_t)= [p(x_t\,|\lambda_k)\, p(\lambda_k)] / [\, \Sigma^N_{k=1}\;\; p(x_t\,|\lambda_k)\, p(\lambda_k)] \tag{3.35}$$

onde $p(\lambda_k)=1/N$ e $p(x_t\,|\lambda_k)$ é calculado como em (3.1)

2. Calcular os ceoficientes de compatibilidade usando (3.28) ou (3.29), onde $t'=t$ (não são considerados os vizinhos);
3. Calcular o fator de atualização definido por (3.30), onde $t'=t$ e $d_{tt'}=1/T$ por simplicidade;
4. Atualizar as probabilidades para cada classe de desempenho usando (3.32);
5. Repetir os passos 3 and 4 até que a alteração das probabilidades seja menor que um limite escolhido ou igual ao número de interações pré-definido;
6. A probabilidade de cada classe de desempenho $p(\lambda_k)$ depois da execução do algoritmo RL é calculada como:

$$p(\lambda_k) = \Pi^T_{t=1}\;\; p_t(\lambda_k) \tag{3.36}$$

onde $p_t(\lambda_k)$ é a probabilidade *a posteriori* usada in (3.10) ou (3.12). Então, a regra de decisão para a identificação da classe de desempenho é dada por [Tran et al.,1999a]:

Seleciona-se o modelo de desempenho k se*

$$k^* = \underset{1\le k\le N}{\arg\max}\;\; p(\lambda_k) \tag{3.37}$$

Modelos de Mistura Gaussianos "Fuzzy" com Relaxação "Fuzzy"

O método RL normalizado proposto por Rosenfeld et al. (1976) tem sido largamente usado para várias aplicações com bom desempenho [Peleg e Rosenfeld,1978]. Entretanto, vários métodos não-normalizados também foram propostos [Hummel e Zucker,1983] e têm como vantagem permitir a associação de múltiplas classes para um mesmo objeto. Eles são aplicáveis quando pode ser natural atribuir classes não-exclusivas ou não-exaustivas a um vetor de informações. De acordo com Borotschnig e seus colaboradores (1998), dos métodos não-normalizados, o Método de Relaxação "Fuzzy" (FRL) é um dos mais utilizados.

Assim, foi proposto um outro método de avaliação de duas fases, onde o FGMM compõe a primeira fase da avaliação e o FRL compõe a segunda. Como anteriormente, os resultados obtidos pelo método FGMM compõem os graus de pertinência iniciais para cada objeto a_t tendo como classe de desempenho λ_k, as quais são denotadas por $\phi_t(\lambda_k)$ [Bonnet e Cutrona,2001]. O método FRL atualiza os graus de pertinência $\phi_t(\lambda_k)$ e a partir dele, $\phi_t^{(m+1)}(\lambda_k)$ é obtido. Entre os métodos FRL disponíveis na literatura, optou-se por [Borotschnig et al.,1998]:

$$\phi_t^{(m+1)}(\lambda_k) = 1/N \ \Sigma^{N}_{j=1} \ max_T \ [r_{tj} \wedge \phi_t(\lambda_k)] \tag{3.37}$$

onde $\wedge$ significa "E" lógico que pode ser o operador *min* ou o produto dependendo da t-norma utilizada.

O método FRL para quando as alterações nos graus de pertinência ficarem abaixo de um limiar escolhido ou quando esgotar um número de interações pré-definido.

3.1.3 Cadeias de Markov Escondidas

3.1.3.a Caso Clássico Discreto

Em geral, o caso discreto pode ser aplicado a poucos tipos de simulação, principalmente em RV, onde as variáveis aleatórias são predominantemente contínuas. Entretanto, pode-se aplicar essa metodologia sobre variáveis discretas, ou discretizando-se variáveis contínuas a partir de transformações. Para utilizar esta última, deve-se ter em mente que o processo de discretização por melhor que seja, sempre levará a uma perda da informação e isso pode ser decisivo na avaliação do treinamento. Portanto, o uso das cadeias de Markov escondidas discretas para abordar casos contínuos deve ser evitada sempre que possível [Moraes e Machado, 2003]. Apesar disso, pode ser encontrada na literatura casos onde essa forma foi aplicada [Rosen et al.,2000] na avaliação sobre dados contínuos.

A teoria básica das HMMs foi publicada entre o final da década de 1960 e o início da década de 1970 por Baum e seus colaboradores [Baum e Petrie,1966] [Baum,1972]. Nas HMMs clássicas, necessita-se de cinco elementos para especificar uma HMM discreta, ou DHMM [Rabiner,1989]. Seja $\lambda(A,B,\pi)$ uma DHMM, então: (i) o número N de estados do sistema; (ii) o número M de observações distintas de símbolos por estado do sistema, isto é a saída discreta do sistema; (iii) a matriz da distribuição das probabilidades de transição de estados, denotada por $A=\{a_{ij}\}$; (iv) a observação da distribuição das probabilidades dos símbolos no estado j, denotado por $B=\{b_j(k)\}$ e (v) a distribuição do estado inicial $\pi=\{\pi_i\}$, onde $1\leq i,j\leq N$ e $1\leq k\leq M$. Dada a forma de uma DHMM, existem três problemas básicos para a solução do modelo:

1. Avaliação do problema: dada uma seqüência de observações no tempo $O=o_1,o_2,\ldots,o_T$ e um modelo λ, como calcular de modo eficiente $P(O|\lambda)$, a probabilidade da seqüência de observações, dado um modelo?
2. Descoberta da parte escondida do modelo: dada uma seqüência de observações no tempo $O=o_1,o_2,\ldots,o_T$ e um modelo λ, como escolher uma correspondente

seqüência de estados $S=s_1,s_2,\ldots,s_T$ que é ótima em algum sentido (isto é, melhor explica as observações)?

3. Otimizar os parâmetros do modelo: como ajustar os parâmetros do modelo $\lambda(A,B,\pi)$ para maximizar $P(O|\lambda)$?

Para resolver o primeiro problema, pode-se usar o Procedimento "Forward-Backward" [Rabiner,1989]. O segundo problema é um problema de otimização e pode ser resolvido usando-se o Algoritmo de Viterbi [Ryan e Nudd,1993]. O último problema pode ser resolvido por um procedimento interativo como o Método Baum-Welch (ou equivalentemente o Método "expectation-maximization") [Rabiner,1989].

3.1.3.b Caso Clássico Contínuo

Para a maioria das aplicações, as variáveis são vetores de dados quantitativos contínuos. Existe a opção de usar a Quantização Vetorial (VQ) para gerar "codebooks" e fazer uso das Cadeias de Markov Escondidas Discretas (DHMM). Entretanto, para se fazer isso é necessário possuir dados de treinamento para todas as classes de desemmpenho e se por acaso alguma classe for adicionada posteriormente todo o sistema deve ser retreinado do início. Ainda há o problema da perda de informação mencionada na seção anterior. Nas Cadeias de Markov Escondidas Contínuas (CHMM), só seria necessário treinar a nova classe adicionada e não há o problema da perda de informação [Moraes e Machado,2002a] [Moraes e Machado,2003]. As densidades de probabilidade são consideradas como modelos de mistura Gaussianas (GMM), dadas por:

$$b_j(o_t) = P(o_t|\lambda) = \sum^{M}_{k=1} w_{jk} \; N(o_t, \mu_{jk}, \Sigma_{jk}) \tag{3.38}$$

onde o_t, $t=1,\ldots,T$ são vetores de observação sendo modelados, w_{jk}, $j=1,\ldots,N$, $k=1, \ldots, M$ são coeficientes da mistura e $N(o_t, \mu_{jk}, \Sigma_{jk})$ é uma densidade Gaussiana como vetor de médias μ_{jk} e matriz de covariância Σ_{jk} para a k-ésima componente da mistura no estado j, dado por:

$$N(o_t, \mu_{jk}, \Sigma_{jk}) = \{1/[(2\pi)^{N/2} \, |\Sigma_{jk}|^{1/2}]\} \times exp\,\{-1/2\,(o_t - \mu_{jk})'(\Sigma_{jk})^{-1}\,(o_t - \mu_{jk})\} \tag{3.39}$$

e as seguintes restrições são satisfeitas:

$$w_{jk} > 0; \; \sum^{M}_{k=1} w_{jk} = 1; \; \int^{+\infty}_{-\infty} b_j(o_t)\, do_t = 1 \tag{3.40}$$

Um processo de reestimação é usado para encontrar os coeficientes. Rabiner [Rabiner,1989] descreve um procedimento que provê boas estimativas iniciais para os parâmetros do modelo, chamado "K-means segmentado" [Juang e Rabiner,1990]. Este procedimento é uma variação do algoritmo clássico "K-means" para aglomeração de dados [Johnson e Wichern,1998].

Comparação de Modelos de Markov Escondidos

Para fins de avaliação de treinamento, é importante comparar duas CHMMs $\lambda_1(A_1,B_1,\pi_1)$ e $\lambda_2(A_2,B_2,\pi_2)$, medindo-se a similaridade dos dois modelos. Pode-se usar uma medida de distância entre elas, denotada por $D(\lambda_1, \lambda_2)$, definido por:

$$D(\lambda_1, \lambda_2) = 1/T \ [\log P(O^{(2)}| \lambda_1) - \log P(O^{(2)}| \lambda_2)] \tag{3.41}$$

onde $O^{(2)}= O_1,O_2,\ldots,O_T$ é uma seqüência de observações geradas pelo modelo λ_2. A equação (3.41) é a medida de quão bem o modelo λ_1 ajusta as observações geradas por λ_2, relativo a quão bem o modelo λ_2 ajusta as observações geradas por si mesmo [Rabiner,1989]. Infelizmente, essa medida não é simétrica e por isso a expressão natural para tomar essa medida é uma versão simétrica dada por:

$$D_S(\lambda_1, \lambda_2) = (D(\lambda_1, \lambda_2) + D(\lambda_2, \lambda_1)) / 2 \tag{3.42}$$

3.1.3.c Caso Fuzzy

Quando pode haver imprecisão nas medidas das variáveis ou existe a possibilidade múltiplas formas de execução correta de um procedimento, pode-se usar uma abordagem "fuzzy" na avaliação, o que remete a uma versão "Fuzzy" das HMM, que é denotado como FHMM [Moraes e Machado,2002b] [Moraes e Machado,2004]. Neste caso, será utilizado como referência a FHMM proposta por Tran e seus colaboradores [Tran et al.,1999b], ainda que existam outras abordagens "fuzzy" para as HMM, como pode ser observado em Mohamed e Gader (2000).

Seja $u_{ijt} = u_{ijt}(O)$ a função de pertinência, denotando o grau para o qual a seqüência de observações O pertencem ao estado i no tempo t e ao estado j no tempo $t+1$, satisfazendo $0\leq u_{ijt}\leq 1$, $\sum^N_{i=1} \sum^N_{j=1} u_{ijt} = 1$ e $0<\sum^T_{t=1} u_{ijt}<T$. Usando o algoritmo EM [Tran e Wagner,1999b], pode-se mostrar que as equações de reestimação para os coeficientes da mistura de densidades da equação (3.38) dos pesos w'_{jk} , do vetor de médias μ'_{jk} = e da matriz de covariâncias Σ'_{jk}:, respectivamente, são dadas por:

$$\begin{gathered} w'_{jk} = (\sum^T_{t=1} \hat{u}^m_{jkt}) / (\sum^T_{t=1} \sum^M_{k=1} \hat{u}^m_{jkt}) \\ \mu'_{jk} = (\sum^T_{t=1} \hat{u}^m_{jkt}\, o_t) / (\sum^T_{t=1} \hat{u}^m_{jkt}) \\ \Sigma'_{jk} = [\sum^T_{t=1} \hat{u}^m_{jkt} \ (o_t - \mu'_{jk})\,(o_t - \mu'_{jk})^t] / (\sum^T_{t=1} \hat{u}^m_{jkt}) \end{gathered} \tag{3.43}$$

onde x^t denota a transposição do vetor x, $m\geq 1$ é o expoente de ponderação de cada função de pertinência "fuzzy" $u_y(x)$ (e é chamado grau de nebulosidade) e

$$\begin{aligned} &\hat{u}_{jkt} = \{\sum^N_{i=1} \sum^M_{l=1} [d_{jkt} / d_{ilt}]^{2/(m-1)}\}^{-1}, && \text{para } m>1 \\ &\hat{u}_{jkt} = \eta_t(j,k), && \text{para } m=1 \end{aligned} \tag{3.44}$$

onde

$$d^2_{jkt} = d^2_{jkt}(O) = -\log P(O, s_t = j, k_t = k \mid \lambda) =$$
$$= -\log \{ \Sigma^N_{i=1} \ \alpha_t(i)\, a_{ij}\, w_{jk} \ N(o_t, \mu_{jk}, \Sigma_{jk})\, \beta_{t+1}(j)\} \qquad (3.45)$$

e

$$\eta_t(j,k) = P(s_t = j, k_t = k \mid O, \lambda) =$$
$$= [\alpha_t(j)\, \beta_t(j)] / [\Sigma^N_{j=1} \ \alpha_t(j)\, \beta_t(j)] \times$$
$$\times [w_{jk} \ N(o_t, \mu_{jk}, \Sigma_{jk})] / [\Sigma^M_{k=1} \ w_{jk} \ N(o_t, \mu_{jk}, \Sigma_{jk})] \qquad (3.46)$$

onde $\alpha_t(j)$ e $\beta_t(j)$ são respectivamente as estimativas "forward" e "backward" {Rab89}.

Sejam λ_z, $z=1,...,Z$, denotando modelos de Z FHMM. Dado uma nova nova seqüência de vetor de atributos O', pode-se classificar O' em alguma classe de desempenho de duas maneiras diferentes: (1) Para verificar distâncias entre cada par λ_z, e $\lambda_{O'}$ e encontrar a menor [Rabiner,1989], ou ainda (2) Usar o critério discriminante para classificar O' em um dos Z modelos [Tran et al.,1999b].

3.1.4 Redes Neurais

3.1.4.a Multi-layer Perceptron

Um método baseado em uma Rede Neural "Multi-Layer Perceptron" (MLP) para avaliação de treinamento é basicamente, um sistema computacional usando neurônios artificiais os quais são totalmente conectados uns aos outros (sinapses). As principais capacidades das redes neurais são:

- Aprendizado por exemplos;
- Auto-adaptação a outros problemas;
- Organização;
- Generalização;
- Robustêz a ruídos e
- Tolerância a falhas.

Pode ser encontrado na literatura várias aplicações para as MLP com bons resultados em reconheciemento de padrões estatísticos [Sethi e Jain,1991] [Yair e Gersho,1990]. Existem vários tipos de redes neurais, mas várias delas são baseadas no conceito de Perceptron [Simpson,1990]. A Figura 3.4 ilustra este conceito e pode-se observar a combinação linear das entradas x pelos pesos w_{ij} resultando na saída y.

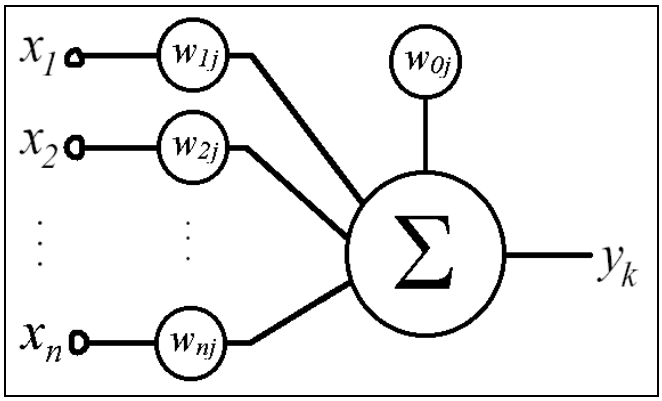

Figura 3.4. Conceito de Perceptron.

A saída y_k é dada por:

$$y_k = w_{0j} + \Sigma^n_{j=1} x_i w_{ij} \tag{3.47}$$

A função de ativação ou função de limiarização mapeia um Perceptron a um intervalo pré-especificado. Quatro funções de ativação são comumente utilizadas: linear, rampa, degrau e sigmoidal. A família das funções sigmoidais é grande, mas a função:

$$S(x) = (1+e^{-x})^{-1} \tag{3.48}$$

é conhecida como função logística e é vista como uma função distribuição acumulada Gaussiana, em termos estatísticos.

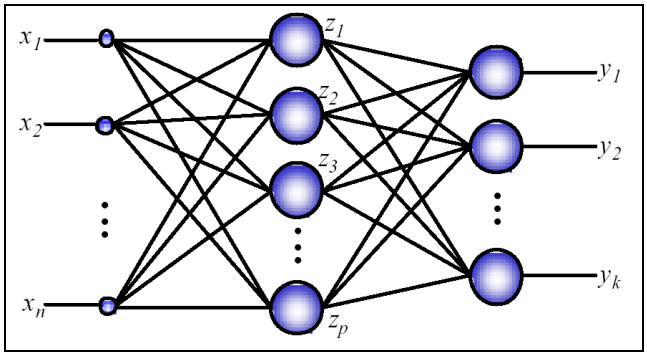

Figura 3.5. Conceito da rede *Multi-Layer Perceptron*.

Machado e Moraes (2004) usaram a MLP, como mostrada na Figura 3.5 como base para um avaliador de treinamento. Existem basicamente três tipos de camadas: a camada de entrada *X*, com vetores de entrada $x_1, x_2, ..., x_n$; a camada intermediária ou escondida, que pode conter um ou mais de um vetor de neurônios e a camada de saída *Y,* com $y_1, y_2, ..., y_k$ saídas. Foi usada da topologia *Feedforward* que permite o fluxo de informações em uma direção. Entretanto, essa topologia tem um inconveniente: não se sabe sobre os erros das camadas escondidas. A solução para este problema é usar o método supervisionado de treinamento de Retropropagação do Erro (ou *Error Backpropagation*) [Simpson,1990] para estimar os pesos do Perceptron.

Para a avaliaçao, um especialista executa o procedimento várias vezes para cada uma das K classes de desempenho disponíveis e a informação da variabilidade desses procedimentos é adquirida pela rede neural MLP através do treinamento por Retropropagação do Erro e usando-se a função de ativação dada por (3.48).

As saídas y_k advindas da rede neural MLP, treinada pelo método da Retropropagação do Erro, são normalizadas e convertidas em escores de desempenho para cada classe de desempenho y_{ko}, onde $y_{ko} \in [0,1]$.

$$y_{ko} = y_k / \{\Sigma^K_{i=1} y_i\}, \quad k = \{1, 2, ..., K\} \tag{3.49}$$

O desempenho do usuário é rotulado e ele recebe um relatório com todas as classes de desempenho possíveis e seu respective escore de desempenho para cada classe, de acordo com a equação (3.49).

3.1.5 Modelos Bayesianos

3.1.5.a Regra de Bayes

Esta seção apresenta um método de avaliação de treinamento baseado no Método da Máxima Verossimilhança (MV) [Moraes e Machado,2005a]. A regra de decisão de MV é uma das mais comuns técnicas usadas na classificação de dados. Formalmente, sejam as classes de desempenho denotadas por w_i , $i=1,...,M$, onde M é o número total de classes de desempenho. Pode-se deteminar a classe mais provável de um vetor de treinamento X, pelas probabilidades condicionais [Johnson e Wichern,1998]:

$$P(w_i \mid X), \text{ onde } i=1,...,M. \tag{3.50}$$

A probabilidade dada por (3.50) fornece a verossimilhança que para um vetor X, a classe correta seja w_i. A regra de classificação é dada de acordo com:

$$X \in w_i \text{ se } P(w_i \mid X) > P(w_j \mid X) \text{ para todo } i \neq j. \tag{3.51}$$

Entretanto, todas as probabilidades dadas por (3.50) são desconhecidas. Assim, se existe informação suficiente disponível para cada classe de desempenho, pode-se estimar essas probabilidades, denotadas por $P(X \mid w_i)$. Usando o Teorema de Bayes:

$$P(w_i \mid X) = [P(X \mid w_i)\, P(w_i)] / P(X),$$

onde

$$P(X) = \Sigma^M_{i=1} P(X \mid w_i)\, P(w_i)]. \tag{3.52}$$

Como $P(X)$ é o mesmo para todas as clases w_i, então não ela não é relevante para a classificação dos dados. Na Teoria Bayesiana, $P(w_i)$ é chamada probabilidade *a priori* para w_i e $P(w_i \mid X)$ é a probabilidade *a posteriori* para w_i onde X é conhecido. Então, a regra de classificação dada por (3.51) é modificada

$$X \in w_i \text{ se } P(w_i \mid X)\, P(w_i) > P(w_j \mid X)\, P(w_i) \text{ para todo } i \neq j. \qquad (3.53)$$

A equação (3.53) é conhecida como a regra de decião de Máxima Verossimilhança ou classificação de Máxima Verossimilhança. Entretanto, pode ser conveniente usar [Johnson e Wichern,1998]:

$$\begin{aligned} g(X) &= ln\, [P(X \mid w_i)\, P(w_i)] \\ &= ln\, [P(X \mid w_i)] + ln\, [P(w_i)] \end{aligned} \qquad (3.54)$$

onde $g(X)$ é conhecida como função de log-verossimilhança e também como função discriminante. Pode-se usar (3.54) para modificar a equação (3.53):

$$X \in w_i \text{ se } g_i\,(X) > g_j\,(X) \text{ para todo } i \neq j. \qquad (3.55)$$

É importante notar que se a distribuição estatística puder assumir uma distribuição Normal multivariada, o uso da equação (3.55) possui interessantes proriedades computacionais [Johnson e Wichern,1998]. Se os dados de treinamento não puderem assumir essa distribuição, a equação (3.55) pode levar a uma redução significativa do custo computacional.

3.1.5.b Regra de Bayes Fuzzy

Moraes e Machado (2006a) propuseram um método de avaliação de treinamento baseado na Regra de Bayes "Fuzzy" [Terano et al.,1987] que tem funcionamento similar ao método da seção anterior, o qual é baseado em probabilidades condicionais.

A probabilidade de um evento "fuzzy" é definida por [Zadeh,1968]: seja (R^n, ϕ, P) um espaço de probabilidades, onde ϕ é uma σ-álgebra em R^n e P é uma medida de probabilidade sobre R^n. Então, um evento "fuzzy"em R^n é um conjunto A em R^n, com função de pertinência $\mu_A(x)$, com μ_A: $X \rightarrow \{0,1\}$ é Borel-mensurável. A probabilidade de um evento A é definido pela integral de Lebesgue- Stieltjes:

$$P(A) = \int_{R^n} \mu_A(x)\, dP = E(\mu_A) \qquad (3.56)$$

em outras palavras, a probabilidade de um evento "fuzzy"A com função de pertinência μ_A é o valor esperado da função de pertinência μ_A.

Novamente, sejam as classes de desempenho dadas por denotadas por w_i , $i=1,...,M$, onde M é o número total de classes de desempenho. Entretanto, agora assume-se que as classes w_i são conjuntos "fuzzy" sobre o espaço de decisão Ω. Seja $\mu_{wi}(X)$ a função de pertinência para cada classe w_i dada por uma fonte de informação "fuzzy"(por exemplo, uma composição de regras de um sistema especialista ou um histograma dos dados amostrais), de acordo com o vetor de dados X. Neste caso, assume-se que a fonte de informação "fuzzy" é um histograma dos dados amostrais.

Pelo uso das probabilidades "fuzzy" e pela regra de Bayes "fuzzy" [Zadeh,1968], sobre a regra de Bayes clássica [Terano et al.,1987], tem-se as probabilidades "fuzzy" da classes w_i dado o vetor de dados X:

$$P(w_i \mid X) = [\mu_{wi}(X)\, P(X \mid w_i)\, P(w_i)] \,/\, P(X), \tag{3.57}$$

com $$P(X) = \Sigma_i\ \mu_{wi}(X) = 1.$$

No entanto, como o denominador é independente, então a regra de classificação Bayes "Fuzzy" é assinalar o vetor de dados de treinamento X do usuário à classe de desempenho w_i, se:

$$X \in w_i \text{ se } \mu_{wi}(X)\, P(w_i \mid X)\, P(w_i) > \mu_{wj}(X)\, P(w_j \mid X)\, P(w_i) \text{ para todo } i \neq j. \tag{3.58}$$

3.1.5.c Naive Bayes

O Método Naive Bayes (NB), também chamado Discreto ou Multinomial NB, é um método robusto para classificação de dados. Formalmente, sejam as classes de desempenho w_i, no espaço de decisões $\Omega=\{1,...,M\}$, onde M é o número total de classes de desempenho. Um avaliador de desempenho baseado em NB [Moraes e Machado,2007] computa as probabilidades condicionais das classes e então prediz a classe mais provável para o vetor de dados de treinamento X, de acordo com os dados amostrais D, onde X é um vetor com n atributos obtidos quando o treinamento é executado, ou seja, $X=\{X_1, X_2, \ldots, X_n\}$. Usando o Teorema de Bayes:

$$\begin{aligned} P(w_i \mid X) &= [P(X \mid w_i)\, P(w_i)] \,/\, P(X) \Leftrightarrow \\ &\Leftrightarrow P(w_i \mid X_1, X_2, \ldots, X_n) = \\ &= [P(X_1, X_2, \ldots, X_n \setminus w_i)\, P(w_i)] \,/\, P(X) \end{aligned} \tag{3.59}$$

Novamente, como $P(X)$ é o mesmo para todas as classes w_i, então ela não é relevante para o processo de classificação dos dados e portanto para a avaliação de desempenho. Então:

$$P(X \mid w_i)\, P(w_i) = P(X_1, X_2, \ldots, X_n \setminus w_i)\, P(w_i) \tag{3.60}$$

A equação (3.60) é equivalente ao modelo de probabilidade conjunta:

$$P(X_1, X_2, \ldots, X_n \setminus w_i)\, P(w_i) = P(X_1, X_2, \ldots, X_n\, , w_i) \tag{3.61}$$

Usando-se sucessivas aplicações da probabilidade condicional sobre a equação (3.61), pode-se obter:

$$\begin{aligned} P(X_1, X_2, \ldots, X_n\, , w_i) &= P(w_i)\, P(X_1, X_2, \ldots, X_n \setminus w_i) \\ &= P(w_i)\, P(X_1 \setminus w_i)\, P(X_2, \ldots, X_n \setminus w_i\, , X_1) \\ &= P(w_i)\, P(X_1 \setminus w_i)\, P(X_2 \setminus w_i\, , X_1)\, P(X_3, \ldots, X_n \setminus w_i\, , X_1, X_2) \\ &\ldots \\ &= P(w_i)\, P(X_1 \setminus w_i)\, P(X_2 \setminus w_i\, , X_1) \ldots P(X_n \setminus w_i\, , X_1, X_2, \ldots, X_{n-1}) \end{aligned} \tag{3.62}$$

O classificador Naive Bayes recebe este nome devido à ingênua hipótese de que cada vetor de atributos X_k é condicionalmente independente de qualquer outro vetor de atributos X_l , para todo $k \neq l \leq n$. Isto significa que o conhecimento da classe é suficiente para determinar a probabilidade de um valor X_k. Essa hipótese simplifica a equação (3.62), devido a:

$$P(X_k \setminus w_i , X_l) = P(X_k \setminus w_i) \tag{3.63}$$

para cada X_k e a equação (3.61) pode ser reescrita como:

$$P(X_1, X_2, \ldots, X_n , w_i) = P(w_i)\, P(X_1 \setminus w_i)\, P(X_2 \setminus w_i) \ldots P(X_n \setminus w_i) \tag{3.64}$$

a menos de um fator de escala S, o qual depende de $X_1, X_2, \ldots, X_n$. Finalmente a equação (3.59) pode ser expressa por:

$$P(w_i \mid X_1, X_2, \ldots, X_n) = (1/S)\, P(w_i)\, \Pi^{n}{}_{k=1} \;\; P(X_k \setminus w_i) \tag{3.65}$$

Então, a regra de classificação Naive Bayes é similar àquela dada por (3.53):

$$X \in w_i \text{ se } P(w_i \mid X_1, X_2, \ldots, X_n) > P(w_j \mid X_1, X_2, \ldots, X_n) \text{ para todo } i \neq j \text{ e } i, j \in \Omega \tag{3.66}$$

e $P(w_* \mid X_1, X_2, \ldots, X_n)$ com $* = \{i, j \mid i, j \in \Omega\}$, é dada por (3.65).

Para estimar os parâmetros para $P(X_k \setminus w_i)$ para cada classe w_i, pode-se usar um estimador de máxima verossimilhança, chamado P_e:

$$P_e(X_k \setminus w_i) = \#(X_k , w_i) / \#(w_i) \tag{3.67}$$

onde $\#(X_k , w_i)$ é o número de casos amostrais pertencentes à classe de desempenho w_i e tendo o valor X_k , $\#(w_i)$ é o número de casos amostrais que pertencem à classe de desempenho w_i.

3.2 Métodos para Avaliação pela Web

3.2.1 Sistemas Especialistas

Seguindo a mesma concepção da avaliação de desempenho mostrada na Seção 3.1.1.b, Moraes e Machado (2005b) propuseram o uso de um Sistema Especialista Baseado em Regras "Fuzzy" para avaliação de treinamento executado à distância pela Web. Para estes casos, uma técnica de avaliação na qual medidas "fuzzy" e informções estatísticas advindas das variáveis de treinamento são usadas como entrada para um sistema especialista baseado em regras "fuzzy" para que seja construído um perfil para o usuário, classificando-o em classes de desempenho.

Um conjunto de medidas estatísticas comumente usadas para propósitos gerais como a media, moda, mediana, desvio-padrão, etc [Tukey,1977] são usadas para medir características da execução do treinamento. Além disso, modelos estatísticos baseados em

modelos de regressão podem modelar seqüências lineares [Draper e Smith,1998] e não-lineares [Seber e Wild,1989] de desempenho na execução de uma tarefa. Pode ser interessante, em alguns casos, usar modelos de series temporais para modelar variáveis no tempo [Box e Jenkins,1994]. Medidas estatísticas e parâmetros estatísticos dos modelos podem ser comparados usando testes de hipóteses apropriados: não-paramétricos [Lehman e D'Abrera,1975] ou paramétricos [Mood et al.,1974]. Como resultados dessas comparações, decisões estatísticas podem ser tomadas sobre a igualdade ou diferença entre esses parâmetros, acompanhados de uma medida de probabilidade de significância.

Como é possível que algumas variáveis no ambiente virtual não apresentem uma correspondência exata com o mundo real, algumas medidas poderão não ser exatas. Desse modo, pode-se usar os conjuntos "fuzzy" para medir essas variáveis [Dubois e Prade,1980]. Essas variáveis "fuzzy" são medidas baseadas em funções de pertinência previamente definidas por especialistas.

Como visto anteriormente na Seção 3.1.1.b, um sistema especialista baseado em regras "fuzzy" combina logicamente todas as informações sobre as variáveis "fuzzy" e não-"fuzzy" para tomada de decisões sobre conjecturas complexas [Turban e Aronson,1998]. A partir de um treinamento realista executado pela Web, pode-se coletar várias informações sobre as interações do usuário. Seqüências de execuções, medidas , etc, podem conter relevantes informações para um sistema de avaliação. No servidor Web, todas as interações podem ser gravadas em arquivos de registro específicos (*log-files*). A partir desses arquivos, informações estatísticas e "fuzzy" podem ser obtidas.

Todas as informações (medidas estatísticas, parâmetros estatísticos, decisões estatísticas, probabilidades de significância e medidas "fuzzy") são usadas como entrada para o sistema especialista "fuzzy". As regras são pré-definidas pelos especialistas com o objetivo de criar um perfil do usuário. Regras "fuzzy" são modeladas por funções de pertinência de acordo com as seguintes especificações e de acordo com os especialistas:

- Seqüência de tarefas executadas;
- Medidas estatísticas da execução de procedimentos corretos;
- Modelos estatísticos da execução do treinamento por especialistas e usuários;
- Decisões estatísticas a partir de testes de hipóteses e a sua respective probabilidade de significância.

Especialistas fornecem os seus conhecimentos para a construção do sistema especialista. Entretanto, algumas características numéricas e formas comparatives são difíceis de serem traduzidas por regras. Assim, a finformação estatística pode ser adquirida das execuções prévias do procedimento por especialistas. Essas medidas estatísitcas e parâmetros são introduzidas no sistema especialista como informações complementares para refinar o conhecimento. Modelos estatísticos das execuções dos especialistas são usados como modelos de comparação com modelos gerados pelo usuário. Testes estatísticos de hipóteses são usados para calcular a probabilidade da diferença entre os modelos dos especialistas e do usuário. Toda essa informação serve como suporte para a decisão final sobre a avaliação do desempenho do usuário no treinamento.

É interessante notar que a combinação de modelos estatísticos e testes de hipótes pode ser considerada uma pré-classificação ou uma pré-avaliação. A partir desta consideração, este sistema de avaliação pode ser considerado similar ao proposto em

[Machado e Moraes,2003]. De fato, ele não é um avaliador em duas fases puro, dado que parte da infomação não é usada pela primeira fase.

Dessa forma, o sistema especialista construirá o perfil do usuário usando informações coletadas de seu treinamento e ele será classificado em classes de desempenho. Moraes e Machado (2005b) usaram cinco classes de treinamento, a saber:

- *Seu treinamento é excelente* – usuário está qualificado para realizar o procedimento real.
- *Seu treinamento é bom* – usuário está quase qualificado para executar o procedimento real. O desempenho é bom, mas poderia ser melhor.
- *Você ainda precisa de treinamento* – usuário precisa treinar para ser qualificado, pois o seu desempenho foi regular.
- *Você precisa de mais treinamento* – desempenho é ruim e o usuário precisa treinar mais para ser qualificado.
- *Você precisa de muito mais treinamento* – usuário é um novato e o seu desempenho é pobre.

A Figura 3.6 mostra o diagrama de blocos do sistema de avaliação. O usuário executa um treinamento realista pela Web interagindo com o sistema. As interações são monitoradas por módulos, os quais podem fazer medidas (estatísticas e/ou "fuzzy"), modelagem e testes de hipóteses O sistema especialista baseado em regras "fuzzy" recebe aquela informação e pode classificar o treinamento de acordo com classes de desempenho pré-definidas. Essa classificação é retornada ao usuário.

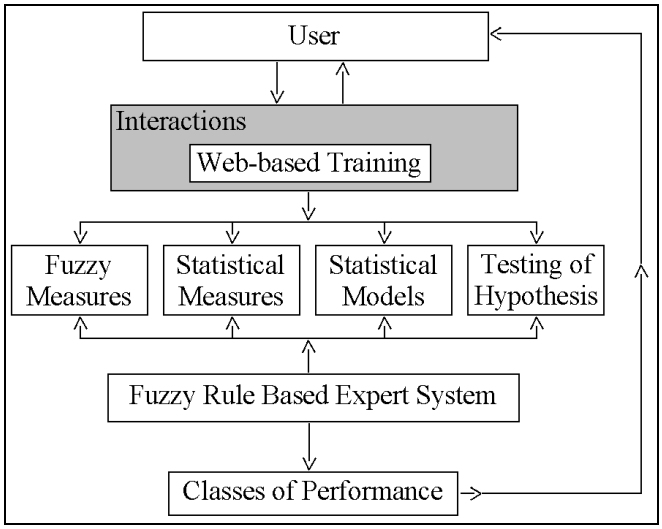

Figura 3.6. Diagrama do sistema de avaliação de treinamento executado pela Web.

3.2.2 Multi-layer Perceptron

A forma apresentada na seção anterior [Moraes e Machado,2005b] é uma forma de avaliação automática, porém o desenvolvimento do sistema especialista depende da obtenção do conhecimento de um ou mais especialistas. Infelizmente, esta tarefa não é simples de ser realizada em algumas situações, ou o conhecimento é muito complexo para ser traduzido por um número reduzido de regras. Além disso, os dados coletados a

partir das interações do usuário podem não se adequar às distribuições estatísticas clássicas, o que pode comprometer as decisões estatísticas baseadas naqueles dados.

Por esses motivos, os mesmos autores [Machado e Moraes,2006] introduziram uma forma alternativa de avaliação automática, onde o conhecimento pudesse ser armazenado por exemplos para executar a avaliação de treinamento pela Web. Além disso, por uma questão de generalização do método, foram suprimidos os testes de hipóteses. Nesta abordagem, as interações do usuário são captadas por medidas estatísticas e parâmetros estatísticos e posteriormente são usadas como entradas em uma rede neural *Multi-Layer Perceptron* (MLP) [Másson e Wang,1990] [Simpson,1990]. As medidas "fuzzy" também foram suprimidas nesta abordagem, devido ao fato de que uma rede neural MLP não é capaz de tratá-las. Mesmo assim, o sistema também pode construir um perfil do usuário e avaliar o seu desempenho no treinamento usando categorias pré-definidas pelos especialistas.

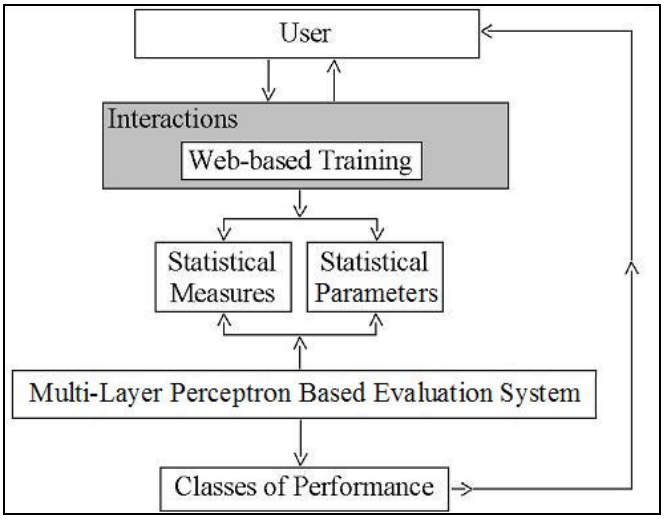

Figura 3.7. Diagrama do sistema de avaliação de treinamento executado pela Web, usando MLP.

A Figura 3.7 mostra o diagrama de blocos deste sistema de avaliação. O usuário realiza seu treinamento pela Web, interagindo com o sistema. As interações são monitoradas pelos módulos, os quais coletam as medidas estatísticas e parâmentros delas. A rede neural MLP recebe essas informações e então classifica o treinamento nas classes de desempenho pré-definidas pelos especialistas. Ao final do treinamento, um relatório com as classes de desempenho do usuário é retornado a ele.

Referências

Alcañiz, M. et al. GeRTiSS: Generic Real Time Surgery Simulation. Studies in Health Technology and Informatics, vol. 94. 2003, pp. 16-18.

Baum, L. E.; Petrie, T., Statistical Inference for Probabilistic Functions of Finite State Markov Chains. Ann Math. Stat., v.37, pp1554-1563, 1966.

Baum, L. E., An Inequality and Associated Maximization Technique in Statistical Estimation for Probabilistic Functions of Markov Processes. Inequalities, v.3, pp.1-8, 1972.

Bezdek, J. C.; A review of probabilistic, fuzzy and neural models for pattern recognition. Journal of Intelligent and Fuzzy Systems, 1(1):1-26, 1993.

Bonnet, N.; Cutrona, J.; Improvement of unsupervised multi-component image segmentation through fuzzy relaxation. IASTED International Conference on Visualization, Imaging and Image Processing, Spain, 2001, 477-482.

Borotschnig, H. et al.; Fuzzy relaxation labeling reconsidered. Proceedings of the 1998 IEEE World Congress On Computational Intelligence, FUZZ-IEEE, 1998, 1417-1423.

Box, G. E. P et al.; Time Series Analysis: Forecasting and Control. Englewood Cliffs: Prentice Hall, 3rd ed., 1994.

Buchanan, B. G.; Shortlife, E. H.; Rule-Based Expert Systems: The MYCIN experiments of the Stanford Heuristic Programming Project. Addison-Wesley, 1985.

Burdea, G. and Coiffet, P. Virtual Reality Technology. 2nd ed., Wiley Interscience, 2003.

Burdea, G. et al.; Virtual Reality Training for the Diagnosis of Prostate Cancer. Virtual Reality Annual International Symposium Proceedings. IEEE, 1998. pp. 190-197.

Burdea, G. et. al, Virtual Reality-Based Training for the Diagnosis of Prostate Câncer. IEEE Trans. on Biomedical Eng., v.46, n.10, p.1253-1260, 1999.

Brusilovsky, P.; Miller, P.; Web-based testing for distance education. In: P. De Bra and J. Leggett (eds.) Proc. WebNet, World Conference of the WWW and Internet, USA, 149-154, 1999.

Dempster, A. P et al.; Maximum Likelihood from Incomplete Data via EM Algorithm. Journal of Royal Statistical Society, Ser. B. 39: 1-38, 1977.

Draper, N. R.; Smith, H.; Applied Regression Analysis. Wiley-Interscience, 3rd Ed., 1998.

Dubois, D.; Prade, H.; Fuzzy sets and systems - theory and applications. Academic Press, 1980.

Dubois, D.; Prade, H.; Possibility theory: An approach to computerized processing of uncertainty. Plenum Press, 1988.

Dubois, D.; Prade, H.; What are fuzzy rules and how to use them. Fuzzy Sets and Systems, v.84, n.2, p.169-185, 1996.

Fu, K. S.; Yu, T. S., Statistical Pattern Classification Using Contextual Information. Research Studies Press, 1980.

Gande, A.; Devarajan, V.; Instructor station for virtual laparoscopic surgery: requirements and design. Proc. of Computer Graphics and Imaging, USA, 2003, pp. 85-90, 2003.

Holton, L. L. H.; Force Models for Needle Insertion Created from Measured Needle Puncture Data. Studies in Health Technology and Informatics: Medicine Meets Virt. Real., Jan., 81:180-186, 2001.

Hopper, M., Assessment in WWW-Based Learning Systems: Opportunities and Challenges. J.UCS: Journal of Universal Computer Science, v.4, n. 4, p. 330-349, 1998.

Hummel, R. A.; Zucker, S. W.; On the foundations of relaxation labeling processes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 5(3), 1983, 267-287.

John, N. W. et al.; Web-based surgical educational tools. Studies in Health Technology and Informatics, v. 81. 2001, pp. 212-217.

Johnson, R. A.; Wichern, D. W., Applied Multivariate Statistical Analysis. Prentice Hall, 1998.

Juang, B-H.; Rabiner, L. R., The Segmental K-Means Algorithm for Estimating Parameters of Hidden Markov Models. IEEE Trans. Acoustics, Speech and Signal Processing. v.38(9), pp1639-1641, 1990.

Keegan, D., Foundations of Distance Education. 2nd Ed. London, Routledge, 1991.

Lehman, E. L.; D'Abrera, H. J. M.; Nonparametrics – Statistical Methods Based on Ranks. San Franciso, Holden-Day, 1975.

Machado, L. S.; Moraes, R. M.; An Online Evaluation Of Training in Virtual Reality Simulators Using Fuzzy Gaussian Mixture Models and Fuzzy Relaxation Labeling. Proc. 2003 IASTED International Conference on Computers and Advanced Technology in Education (CATE'2003). Greece, p. 192-196, 2003.

Machado, L. S.; Moraes, R. M.; Neural Networks for Online Training Evaluation in Virtual Reality Simulators. World Congress on Engineering and Technology Education - WCETE'2004. Brasil, p. 157-160, 2004.

Machado, L. S.; Moraes, R. M.; Evaluation of Training Executed by Web Using Multilayer Perceptron Based Systems. Proceedings of World Congress on Computer

Science, Engineering and Technology Education (WCCSETE'2006). Brasil, p. 228-232, 2006.

Machado, L. S. et al.; A Fuzzy Rule-Based Evaluation for a Haptic and Stereo Simulator for Boné Marrow Harvest for Transplant. Proc. Phantom Users Group – 2000. USA, 2000.

Machado, L. S. et al.; A Virtual Reality Simulator for Bone Marrow Harvest for Pediatric Transplant. Studies in Health Technology and Informatics, v.81, p.293-297, 2001.

Machado, L. S. et al.; Assessement of Gynecological Procedures in a Simulator Based on Virtual Reality. Proceedings of the 7th International FLINS Conference on Applied Artificial Intelligence (FLINS'2006). Italy, p. 799-804, 2006.

Mahoney, D. P.; The Power of Touch. Computer Graphics World, v. 20, n.8, p. 41-48, 1997.

Mahoney, D.P. "The Eyes Have It". Computer Graphics World, v. 21, n. 8, p. 69-70, 1998.

Mamdani, E. H.; Assilian, S.; An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller, International Journal of Man-Machine Studies, v.7, p. 1-13, 1975.

Massie, T. H.; Salisbury, J. K; The PHANTOM Haptic Interface: A Device for Probing Virtual Objects. Proceedings of the ASME Winter Annual Meeting, Symposium on Haptic Interfaces for Virtual Envirronment and Teleoperator Systems, USA, 1994.

Másson, E.; Wang, Y-J. Introduction to Computation and Learning in Artificial Neural Networks. European Journal of Operational Research, v.47, n.1, p. 1-28, 1990.

McBeth, P. B. et al.; Quantitative Methodology of Evaluating Surgeon Performance in Laparoscopic Surgery. Studies in Health Technology and Informatics, v. 85, p. 280-286, 2002.

Mohamed, M. A.; Gader, P., Generalized Hidden Markov Models - Part I: Theoretical Frameworks. IEEE Trans. on Fuzzy Syst., v.8, n.1, pp.67-81, 2000.

Mood, A. M.; Grabyll, F. A.; Boes, D. C.; Introduction to the Theory of Statistics. Tokyo. McGraw Hill, 1974, 3rd. ed.

Moody, L. et al.; Objective Surgical Performance Evaluation Based on Haptic Feedback. Studies in Health Technology and Informatics: Medicine Meets Virt. Real., v. 85, p. 304-310, 2002.

Moraes, R.M.; Uma Arquitetura de Sistemas Especialistas Nebulosos para Classificação de Imagens Utilizando Operadores da Morfologia Matemática, Tese de Doutorado em Computação Aplicada. INPE, S.J.Campos, 1998.

Moraes, R. M. e Machado, L. S. Network Based Education to Statistics Education. Proc. International Conference on Information Technology Based Higher Education and Training (ITHET'2001) [cdrom]. Japan, 2001.

Moraes, R. M.; Machado, L. S.; Hidden Markov Models for Learning Evaluation in Virtual Reality Simulators. Anais da 2002 IASTED International Conference on Computers and Advanced Technology in Education (CATE'2002). Mexico. pg 405-409, 2002

Moraes, R. M.; Machado, L. S.; Fuzzy Hidden Markov Models for Online Training Evaluation in Virtual Reality Simulators, in Computational Intelligent Systems for Applied Research (Da Ran et all. eds.). Proceedings of the 5th International FLINS Conference, p. 296-303, 2002.

Moraes, R. M.; Machado, L. S.; Hidden Markov Models for Learning Evaluation in Virtual Reality Simulators. International Journal of Computers & Applications, v.25, n.3, p. 212-215, 2003.

Moraes, R. M.; Machado, L. S.; Online Training Evaluation in Virtual Reality Simulators Using Gaussian Mixture Models. Studies in Health Technology and Informatics, v. 94, p. 42-44, 2003a.

Moraes, R. M.; Machado, L. S. Fuzzy Gaussian Mixture Models for On-line Training Evaluation in Virtual Reality Simulators. Annals of the International Conference on Fuzzy Information Processing (FIP'2003). China, v. 2, p. 733-740, 2003b.

Moraes, R. M.; Machado, L. S.; Gaussian Mixture Models and Relaxation Labeling for Online Evaluation of Training in Virtual Reality Simulators. Anais da International Conference on Engineering and Computer Education (ICECE'2003). Brasil, 2003c.

Moraes, R. M.; Machado, L. S.; Using Fuzzy Hidden Markov Models for Online Training Evaluation and Classification in Virtual Reality Simulators. International Journal of General Systems. v. 33 n.2-3, p. 281-288, 2004.

Moraes, R. M.; Machado, L. S.; Maximum Likelihood for On-line Evaluation of Training Based on Virtual Reality Proceedings of Global Congress on Engineering and Technology Education (GCETE'2005). Brasil, p.299-302, 2005a.

Moraes, R. M.; Machado, L. S.; Evaluation of Training Executed by Web Using Fuzzy Rule Based Expert Systems Proceedings of Global Congress on Engineering and Technology Education (GCETE'2005). Brasil, p.857-860, 2005b.

Moraes, R. M.; Machado, L. S.; Continuous Evaluation in Training Systems Based on Virtual Reality Proceedings of Global Congress on Engineering and Technology Education (GCETE'2005). Brasil, p.1048-1051, 2005c.

Moraes, R. M.; Machado, L. S.; On-line Training Evaluation in Virtual Reality Simulators using Fuzzy Bayes Rule. Proceedings of the 7th International FLINS Conference on Applied Artificial Intelligence (FLINS'2006). Italy, p. 791-798, 2006a.

Moraes, R. M.; Machado, L. S.; Assessment Based on Naive Bayes for Training Based on Virtual Reality. Proceedings of International Conference on Engineering and Computer Education (ICECE'2007). Brasil, p. 269-273, 2007.

Moraes, R. M. et al.; Expert Systems Architecture for Image Classification Using Mathematical Morphology Operators. Information Science v.142, p. 7-21, 2002.

Morris, et al.; Visuohaptic simulation of bone surgery for training and evaluation. IEEE Computer Graphics and Applications, vol. 26 no. 6, p. 48-57, 2006.

Peleg, S.; Rosenfeld, A., Determining Compatibility Coefficients for Curve Enhancement Relaxation- Process. IEEE Trans. on Systems, Man and Cybernetics, v. 8, n.7, p.548-555, 1978.

Pesin, L. et al.; Assessment Support In Web-Based Authoring And Learning Environment. Proc. 2003 IASTED International Conference on Computers and Advanced Technology in Education (CATE'2003). Greece, p. 186-191, 2003.

Pucel, D. J.; Anderson, L. D. Developing Computer Simulation Performance Tests: Challenges and Criteria. Proc. 2003 IASTED International Conference on Computers and Advanced Technology in Education (CATE'2003). Greece. 2003, pp 170-174, 2003.

Rabiner, L. R., A Tutorial on Hidden Markov Models and Selected Application in Speech Recognition. Proceedings of the IEEE, v.77, n.2, 1989.

Rich, E.; Knight, K.; Artificial Intelligence. New York: McGraw, 1993.

Rosen J. et al.; Hidden Markov Models of Minimally Invasive Surgery, Studies in Health Technology and Informatics, v 70, p. 279-285, 2000.

Rosenfeld, A. et al.; Scene labelling by relaxation operations. IEEE Transactions on Systems, Man and Cybernetics, v. 6, n.6, p. 420-433, 1976.

Ryan, M. S.; Nudd, G. R.; The Viterbi Algorithm, Research Report CS-RR-238, Department of Computer Science, University of Warwick, UK, 1993.

Seber, G. A. F.; Wild, C. J.; Nonlinear Regression. New York: Wiley, 1989.

Senne, E. L. F.; Motor de Inferência Para Sistemas Especialistas - Tese de Doutorado em Computação Aplicada - Instituto de Pesquisas Espaciais, 1988.

Sethi, I. K.; Jain, A. K.; Artificial Neural Networks and Statistical Pattern Recognition. North Holland, 1991.

Simpson, P. K.; Artificial Neural Systems. New York: Pergamon, 1990.

Terano, T. et al., Fuzzy systems theory and it's applications. Academic Press,1987.

Tran, D.; Wagner, M.; Fuzzy Gaussian Mixture Models for Speaker Recognition, in ICSLP98 (the International Conference on Spoken Language Processing) special issue of the Australian Journal of Intelligent Information Processing Systems (AJIIPS), v. 5, n.4, p. 293-300, 1998.

Tran, D.; Wagner, M.; Fuzzy approach to Gaussian mixture models and generalised Gaussian mixture models. in Proceedings of the Computation Intelligence Methods and Applications (CIMA'99). USA, p. 154-158, 1999a.

Tran, D., Wagner, M.; Fuzzy Expectation-Maximization Algorithm for Speech and Speaker Recognition. In Proceedings of 18th International Conference of the North American Fuzzy Information Society (NAFIPS'99). USA, p. 421-425, 1999b.

Tran, D. et al.; Fuzzy Gaussian Mixture Models for Speaker Recognition, in Proceedings of the International Conference on Spoken Language Processing (ICSLP98). Austrália, v. 2, p.759-762, 1998.

Tran, D., et al.; Speaker recognition using Gaussian mixture models and relaxation labeling. in Proceedings of the 3rd World Multiconference on Systemetics, Cybernetics and Informatics/ The 5th Int. Conf. Information Systems Analysis and Synthesis (SCI/ISAS99), v. 6, p. 383-389, 1999a.

Tran, D., et al.; A Fuzzy Approach to Statistical Models in Speech and Speaker Recognition. In Proceedings of the FUZZ-IEEE'99 Conference. Khorea, v.3, p.1275-1280, 1999b.

Tukey, J. W.; Exploratory Dates Analysis. New York. McGraw Hill, 1977.

Turban, E.; Aronson, J. E.; Decision Support Systems and Intelligent Systems. New Jersey, Prentice Hall, 5th ed., 1998.

University of Oregon. Electronic Music Interactive. [online] http://nmc.uoregon.edu/emi/emi.php. Acesso em fevereiro/2006.

Voss, G. et al. “Intelligent Training System for Laparoscopy and Hysteroscopy”. Studies in Health Technology and Informatics, no. 70, pp. 359-364, 2000.

Yair, E.; Gersho, A., Maximum a Posteriori Decision and Evaluation Class Probabilities by Boltzmann Perceptron Classifiers. Proc. of IEEE, v.78, n.10, pp. 1620-1628, 1990.

Zadeh, L. A.; Fuzzy Sets, Information and Control, v.8, p.338-353, 1965.

Zadeh, L. A.;Probability Measures of Fuzzy Events. Journal of Mathematical Analisys and Applications, v.10, p.421-427, 1968.

Zadeh, L. A.; Fuzzy Logic, Computer, v.1, p. 83-93, 1988.

Zaina, L. A. M. et al., Comparative Analysis of Manager Tools to Distance Learning. International Conference on Engineering and Technology Education (INTERTECH’2002). Brazil [cdrom], 2002 [in Portuguese].

Capítulo 4. Outras Formas de Avaliação

4.1 Avaliação Continuada

Apesar das várias metodologias propostas para avaliação de treinamento, a grande maioria delas não prevê uma forma de avaliação continuada para melhoria do desempenho do usuário. A avaliação continuada é uma ferramenta útil usada no ensino presencial e à distância para a construção do conhecimento e o treinamento cognitivo [Aalsvoort et al.,2002][Galef Institute,1995]. Neste caso, a meta é construir um diagnóstico que auxilie o usuário a entender as suas dificuldades, criando um perfil dos seus vários treinamentos e com isso auxiliá-lo a aperfeiçoar as sus habilidades e com isso melhorar o seu desempenho [Arter e Mctighe,2000] [Carlson et al.,2003].

Uma ferramenta para a avaliação continuada deve estar interconectada com uma ferramenta de avaliação (preferencialmente *online*) e deve receber dela as informações de todas as variáveis de interação do usuário em seus diversos treinamentos e criar um perfil desses treinamentos. Essas informações são usadas para avaliar o usuário e melhorar seu desempenho nas tarefas reais [Sternberg e Grigorenko,2001]. A partir dos dados coletados uma base de dados é criada com as variáveis específicas e relevantes à avaliação do treinamento. A cada novo treinamento executado, novas informações são armazenadas na base e uma reavaliação do usuário é executada, levando-se em consideração o seu histórico e também o último treinamento realizado.

4.1.1 Treinamento baseado em RV

As propostas encontradas na literatura procuram construir um perfil do usuário a partir dos seus vários treinamentos visando orientá-lo quanto ao seu progresso. A união de ferramentas estatísticas, para mediada das variáveis observadas durante o treinamento e um sistema especialista para construir o perfil do usuário, compõem a proposta de Moraes e Machado (2005).

As ferramentas estatísticas são programadas para fazer uma análise automática da base de dados e construir medidas estatísticas, tabelas, gráficos e modelos estatísticos (dependentes do tempo ou não). A partir das medidas e parâmetros estatísticos, um sistema especialista baseado em regras pode criar um perfil do usuário e um relatório de avaliação continuada. Este relatório apresenta o perfil do usuário e mostra, com medidas

estatísticas, tabelas, gráficos e modelos, o desempenho do usuário em tarefas específicas. A Figura 4.1 mostra esta metodologia. Pode-se observar que o sistema de avaliação (como o apresentado na Figura 3.1) possui mais componentes. A Ferramenta de Avaliação Continuada e o Perfil do Usuário foram incorporados (area cinza da Figura 4.1). O novo sistema de avaliação cria dois tipos de Relatório: o de Avaliação e o de Avaliação Continuada. O primeiro diz respeito ao desempenho do usuário apenas no ultimo treinamento realizado e o segundo sobre todo o conjunto dos treinamentos realizados por ele.

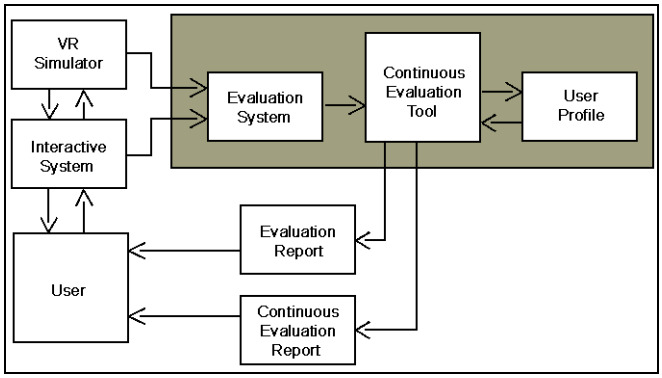

Figura 4.1. Diagrama do sistema de avaliação com a abordagem de avaliação continuada.

Em um primeiro momento, quando o usuário realizou o seu treinamento, o Relatório de Avaliação contém informações sobre o desempenho do usuário ao final do treinamento de acordo com as classes de desempenho previamente definidas pelos especialistas. O Relatório de Avaliação Continuada apresenta informações sobre o desempenho do usuário sobre tarefas específicas lançando mão de métodos ilustrativos, capazes de ressaltar acertos e erros na sua execução (medidas e modelos estatísticos, tabelas e gráficos). Também está incluído neste Relatório informações sobre o último treinamento realizado, porém serão mostradas de forma acumulativa na seqüência dos treinamentos realizados por aquele usuário.

4.1.2 Treinamento baseado na Web

Como mencionado anteriormente, uma ferramenta de avaliação continuada deve estar interconectada com uma ferramenta de avaliação (preferencialmente *online*) e deve receber dela as informações de todas as variáveis de interação do usuário em seus diversos treinamentos e criar um perfil desses treinamentos. A metodologia proposta por Moraes e Machado (2006) para avaliação de treinamento baseado na Web, usa a mesma filosofia de trabalho daquela proposta pelos mesmos autores para RV [Moraes e Machado,2005]. A diferença está nas informações retiradas das interações do usuário com o sistema: o uso de medidas estatísticas, parâmetros de modelos, testes de hipóteses para comparações e medidas “fuzzy” para as variáveis com medida incerta. Então, um sistema especialista baseado em regras “fuzzy” combina logicamente todas as informações disponíveis (“fuzzy” e não-“fuzzy”) e realiza a avaliação de desempenho do último treinamento e constrói o perfil do usuário para a avaliação continuada. Como no

caso da avaliação continuada para treinamentos baseados em RV, dois relatórios são produzidos: o Relatório de Avaliação e o Relatório de Avaliação Continuada.

4.2 Avaliação para Múltiplos Usuários

Sistemas para treinamento baseados em RV para uso simultâneo em ambientes de treinamento complexos tem sido planejados, como salas de cirurgia virtuais para vários propósitos de treinamento. Os avanços em agrupamentos de computadores (*clusters*), melhorias nas placas-de-vídeo, processadores e redes proporcionarão sistemas de treinamento para múltiplos usuários simultaneamente e a baixo custo. Outro importante avanço é na velocidade dos dispositivos de entrada e saída para sistemas de realidade virtual, tais como os sistemas hápticos/táteis. Atualmente, já é possível conectar mais de um sistema de interação em um único computador.

Sistemas computacionais para múltiplos usuários tem sido desenvolvidos desde os anos de 1990, como RB2 [Blanchard et al.,1992], DIVE [Carlsson e Hagsand,1993], MR Toolkit [Shaw e Green,1993], sendo que alguns deles têm suporte para dispositivos hápticos. Sistemas que proporcionam interação entre múltiplos usuários também foram propostos. Baier e seus colaboradores [Baier et al.,1999] propôs um sistema de telepresença usando interações hápticas e visuais. Recentemente, sistemas para colaboração entre múltiplos usuários foram desenvolvidos [Low et al.,2003] [Morris et al.,2004] [Yoshida et al.,2006]. A principal diferença entre os sistemas de treinamento baseados em RV para múltiplos usuários são: aumento da complexidade do sistema de RV: uso de cluster de computadores ou um computador capaz de gerar múltiplas vistas realistas, suporte para as alterações do ambiente virtual para todos os usuários e suporte para sistema de avaliação; rede computacional ponto-a-ponto de alta velocidade para comunicação entre os computadores, sem comprometer a simulação. Eventualmente, haver mais de um dispositivo háptico instalados em um computador e/ou sistemas de rastreamento de posição para cada usuário em treinamento.

O problema mais comum em sistemas distribuídos baseados em rede ou na Web, para interações para múltiplos usuários é a latência [Park e Kenyon,1999]: usuários podem ter diferentes vistas do ambiente compartilhado e isso pode causar diminuição do desempenho dos usuários envolvidos na simulação. Apesar das possibilidades de treinamento que podem ser simuladas por RV, qualquer tipo de treinamento tem pouco valor se o usuário/*trainee* não receber um retorno do seu desempenho. Para a avaliação dos usuários, os principais problemas são a complexidade computacional e a acurácia, mesmo quando há somente um usuário em treinamento. Um sistema de avaliação *online* deve ter baixa complexidade computacional para não comprometer o desempenho das simulações de RV geradas pelo(s) computadore(s), mas ao mesmo tempo, ter alta acurácia para não comprometer a própria avaliação.

Devidos às necessidades específicas para o treinamento de múltiplos usuários, uma abordagem diferente para um sistema de avaliação é requerida. Os requerimentos são: monitorar todos os usuários em treinamento, de acordo com as variáveis relevantes ao treinamento; em um ambiente multi-usuáriom algumas tarefas devem ser completadas por usuários específicos e de acordo com uma agenda específica; levar em conta os detalhes de avaliação (as interações entre os usuários); criar um perfil de usuário e um

perfil de grupo; apresentar baixa complexidade para não comprometer o desempenho das simulações em RV, mas apresentar um alto nível de precisão.

Uma metodologia para avaliação de múltiplos usuários coleta dados das interações de cada usuários e as interações dentro do grupo durante o treinamento para criar um perfil de cada usuário e do grupo. Essas informações são usadas para permitir a melhoria do desempenho do grupo na execução de tarefas reais [Sternberg e Grigorenko,2001]. A metodologia faz a união de ferramentas estatísticas e sistemas especialistas baseados em regras "fuzzy" para construir esses perfis. Como mostrado nos casos da avaliação continuada, as ferramentas estatísticas são programadas para realizarem análises automáticas e produzir medidas, tabelas, gráficos e modelos estatísticos (dependentes do tempo ou não). Essas informações são as entradas do sistema especialista "fuzzy" que gerará os dois tipos de Relatórios: o de Avaliação Individual e o de Avaliação do Grupo. Esses relatórios devem mostrar o desempenho individual ou em grupo da realização de tarefas específicas com medidas estatísticas, tabelas, gráficos, modelos e algumas frases em linguagem pseudo-natural sobre possíveis melhorias e detalhes de procedimentos específicos em tarefas seqüenciais, simultâneas e colaborativas. Esses tipos de tarefas são comuns em salas cirúrgicas e o desempenho do grupo em algumas tarefas podem ser essenciais para a vida do paciente.

A Figura 4.2 mostra a metodologia. Pode-se observar que o Sistema de Avaliação da Figura 3.1 foi alterado e na Figura 4.2 ele é o Sistema de Avaliação Individual. Para construir o Sistema de Avaliação Múltipla, a Ferramenta de Avaliação do Grupo e os Perfis do Usuário e do Grupo foram adicionados. Os *N* usuários que realizam o treinamento usando o Simulador Baseado em RV e interagindo com o Sistema Interativo. Este sistema deve proporcionar as simulações hápticas e visuais para todos os usuários, de acordo com o seu ponto de vista e seus respectivos dispositivos hápticos. A partir dessas informações, medidas estatística e parâmetros são tomados para que o sistema especialista "fuzzy" realize a avaliação individual e do grupo.

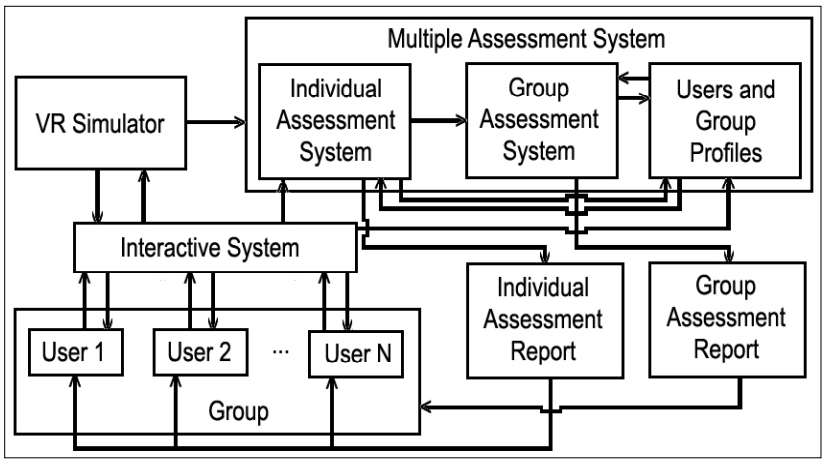

Figura 4.2. Diagrama do sistema de avaliação com a abordagem de Avaliaçao Múltipla.

Essa metodologia de Avaliação Múltipla pode ser usada para vários tipos de treinamento em medicina, como procedimentos em salas cirúrgicas, treinamento de paramédicos em situações de emergência, etc. No entanto, é uma metodologia genérica e pode também ser usada para treinamento em outras áreas como simuladores aeronáuticos,

grupos de manutenção, gerenciamento de usinas nucleares, termoelétricas, hidroelétricas, etc.

Deve-se ressaltar também que um Sistema de Avaliação Múltipla não pode ser classificado como *online* ou *offline*, como nos sistemas de avaliação para um usuário. Como mencionado anteriormente, um sistema de avaliação *online* deve gerar o seu relatório imediatamente após o término da sessão de treinamento. Em alguns casos, os Sistemas de Avaliação Múltipla serão capazes de faze-lo, mas em outros casos não. As principais razões envolvem a configuração de computadores, possíveis atrasos na rede, número de modelos dependentes do tempo necessário para medir algumas variáveis, a complexidade da própria simulação e o número de interações entre os usuários.

Referências

Aalsvoort, G.M. et al.; Learning potential assessment and cognitive training: Actual research and perspectives in theory building and methodology. JAI/Elsevier Science, 2002.

Arter, J. A.; Mctighe, J.; Scoring rubrics in the classroom: using performance criteria for assessing and improving student performance. Corwin Press, 2000

Baier, H. et al.; Distributed PC-based haptic, visual and acoustic telepresence system - experiments in virtual and remote environments. Proc. of IEEE VR, USA, p. 118-125, 1999.

Blanchard, C. et al.; Reality built for two: a virtual reality tool. ACM CG, v. 24, n. 2, p. 35-36, 1992.

Carlsson, C.; Hagsand, O.; DIVE: a platform for multi-user virtual environments. Computers & Graphics, v. 17, n. 6, p. 663-669, 1993.

Carlson, M. O. et al.; Weaving science inquiry and continuous assessment: using formative assessment to improve learning. Corwin Press, 2003

Galef Institute; Assessment: Continuous Learning. Stenhouse Publishers, 1995

Low, K-L. et al. Combining head-mounted and projector-based displays for surgical training. Proc. of the IEEE VR, USA, p. 110–117, 2003.

Moraes, R. M.; Machado, L. S.; Continuous evaluation in training systems based on virtual reality. Proceedings of Global Congress on Engineering and Technology Education (GCETE'2005). Brasil, p.1048-1051, 2005

Moraes, R. M.; Machado, L. S.; Continuous evaluation in training systems based on Web. Proceedings of World Congress on Computer Science, Engineering and Technology Education (WCCSETE'2006). Brasil, p. 223-227, 2006.

Moraes, R. M.; Machado, L. S.; Multiple assessment for multiple users in virtual reality training environments. Lecture Notes in Computer Science. Berlim. v. 4756, p. 950-956, 2007.

Morris, et al.; a collaborative virtual environment for the simulation of temporal bone surgery. Proc. MICCAI, France, pp. 319-327, 2004.

Park, K-S.; Kenyon, R. V.; Effects of network characteristics on human performance in a collaborative virtual environment. Proc. of IEEE Virtual Reality, USA, pp. 104-111, 1999.

Shaw, C.; Green, M.; MR Toolkit peers package and experiment. Proc. IEEE VRAIS, USA, pp. 463-469, 1993.

Sternberg, R. J.; Grigorenko, E.; Dynamic testing: the nature and measurement of learning potential. Cambridge University Press, 2001

Yoshida, S.; Noma, H.; Hosaka, K.; Proactive Desk II: development of a new multi-object haptic display using a linear induction motor. Proc. IEEE VR Conference, USA, pp. 269- 272, 2006.