



INSTITUTO DE ENGENHARIA NUCLEAR

RT-IEN-16/2002

**UM SISTEMA INTELIGENTE FUZZY PARA AVALIAR
INFORMAÇÕES DE USUÁRIOS EM PÁGINAS NA
INTERNET**

por

Antonio César Ferreira Guimarães

Agosto/2002

NOTA
ESTE RELATÓRIO É PARA USO EXCLUSIVO DO
INSTITUTO DE ENGENHARIA NUCLEAR

O direito a utilização de informações relacionadas ao trabalho de pesquisa realizado no IEN é limitado aos servidores da CNEN e pessoal de organizações associadas, nos limites dos termos contratuais que regem os respectivos convênios. O conteúdo dos relatórios não pode ser separado ou copiado sem autorização escrita do IEN



INSTITUTO DE ENGENHARIA NUCLEAR

Título: UM SISTEMA INTELIGENTE FUZZY PARA AVALIAR INFORMAÇÕES DE USUÁRIOS EM PÁGINAS NA INTERNET

Autor(es): Antonio César Ferreira Guimarães, DSc.

e-mail: tony@ien.gov.br

Identificação:
RT-IEN-16-2002

Nº de
páginas:
14

Tipo de Divulgação:
Irrestrita (x) Restrita ()

Divulgar para:
IEN

Localização: DIRE/SETER

Publicação externa associada (congresso/periódico):

Ciência da Informação, ISSN 0100-1965.

Palavras chave: *Lógica Nebulosa, Internet, Páginas WEB, ANFIS.*

Resumo:

Este trabalho consta do desenvolvimento de um Sistema Inteligente na Internet (SII) para indicar o nível de aquisição material dos visitantes em páginas na WWW (*World Wide Web*). A partir do Tempo de permanência (Tp), da Frequência de visitas (Fv) e de um Índice (IDLI) que mede o interesse do internauta na página *web*, o sistema estabelece um padrão existente nestas informações relacionadas com o número de Ordens de Compras dadas por um internauta. Desta forma, as variáveis tratadas como pares de entradas/saídas estabelecem um sistema adaptativo *Fuzzy* que possui a capacidade de aprender com o tempo e a generalizar para definir novos padrões de usuários. Uma aplicação da metodologia proposta foi apresentada para efeito de ilustração. Com os resultados obtidos, pode-se definir um número de ordem de compra para cada item oferecido na página *Web*, estabelecer limites para o estoque adequado, ou estabelecer critérios de exibição de ofertas de produtos na página *Web* da empresa.

Abstract:

This work consists of the development of a Intelligent System in the Internet (SII) to indicate the level of the visitors' material acquisition in pages in WWW (World Wide Web). Starting from the Time of permanence (Tp), of the Frequency of visits (Fv) and of a Index (IDLI) that measures the interest of the swelling in the page web, the system establishes an existent pattern in these information related with the number of Orders of Purchases given by a swelling. This way, the variables as input/output pairs establish an Adaptive Neural Fuzzy Inference System (ANFIS) that possesses the capacity to learn with the time and to generalize to define new patterns of users. For illustration effect an application of the methodology proposal was presented. With the obtained results, can be defined a number of purchase order for each item offered in the page web, to establish limits for the appropriate stock, or to establish approaches of exhibition of offers of products in the page web of the company.

Emissão		Nome	Rubrica	Data
Data: 20/08/2002	Elaboração:	Antonio Cesar Ferreira Guimarães		20/08/2002
Divisão:DIRE		Revisão:	Orlando J. A. Gonçalves Filho	20/08/2002
Serviço:SETER	Aprovação :	Orlando J. A. Gonçalves Filho		20/08/2002

Instituto de Engenharia Nuclear:

Via 5 s/n, Cidade Universitária, Ilha do Fundão, CEP 21945-970, CP 68.550, Rio de Janeiro – RJ - Brasil .

Tel.: 00 55 21 2560-4113 Internet: www.ien.gov.br

1. INTRODUÇÃO

Uma das maiores preocupações de empresas comerciais ligadas a Internet é saber exatamente quais informações mostrar ao internauta e desta forma manter sua atenção em ofertas ou produtos em geral e impedir que o mesmo não abandone a página (*homepage*) *Web*, ou seja, se mantenha navegando no *site* o maior tempo possível. As informações devem ser interessantes, seja no visual ou nas ofertas e produtos oferecidos, de forma que o internauta deseje voltar o maior número de vezes a página *Web* e que o intervalo entre uma visita e outra seja com um intervalo de tempo cada vez menor. Várias empresas comerciais, que utilizam o sistema de *shopping*, fazem de tudo para atrair a atenção dos usuários. Basicamente, um grande número de empresas usa o recurso de um cadastro, onde o usuário após o preenchimento deste cadastro, passando a dispor de um nome-de-usuário (*user name*) e de uma senha (*password*). A partir daí as empresas identificam o cliente e o seu retorno, e desta forma oferece informações personalizadas sobre produtos para novas aquisições, o que é o caso, por exemplo, da livraria virtual *Amazon*, onde novos livros com características semelhantes são ofertados após a confirmação da sua identificação. Os livros que possuem palavras ou contexto semelhantes são oferecidos em sua página personalizada. Existem outras configurações de políticas de amostragem de ofertas, como é o caso da empresa brasileira virtual *Lojas Americanas*, na qual a atualização é feita assim que o internauta abre uma nova página. As ofertas constantes na página principal são atualizadas instantaneamente, e aleatoriamente, com produtos misturados de várias naturezas, toda vez que uma nova página é aberta. Utilizando o ícone de atualização na barra de ferramentas do navegador, novas ofertas são também disponibilizadas aleatoriamente, sem nenhum critério. Utilizando o recurso de abrir uma nova janela pelo navegador, a página da *Lojas Americanas*

também apresenta novas informações se comparadas com aquelas apresentadas em uma página previamente aberta. Esta técnica utiliza recursos de programação apenas e não possui um sistema inteligente que possa filtrar as informações que deverão ser disponibilizadas para o internauta em particular, como no caso da *Amazon*. A diferença entre o que a Amazon oferece de conteúdo e uma página que utilize um sistema inteligente é que a Amazon não possui a preocupação de verificar o tempo em que o internauta permanece em um item de pesquisa, e o potencial de compra através de um indicador de interesse (ILIP – *Interest Level Indicating Point*) oferecendo apenas mercadorias para compra similares baseados em alguma compra feita ou pesquisa realizada.

IP (2000) propôs o desenvolvimento de um sistema (*Intelligent Internet Information Delivery System – IIIDS*) o qual é caracterizado por sua capacidade de “máquina de aprendizado” baseado nos dados de movimentos pontuais da página *Web* gerados pelos usuários e a partir deste ponto avaliar as preferências, os locais mais visitadas e relevantes por meio dos princípios de lógica nebulosa (*Fuzzy Logic*). Neste caso, a metodologia proposta sugere a utilização de um sistema inteligente e utilizando o exemplo da AMAZON.com, se assemelharia a uma situação em que os recursos da *Amazon* oferecidos aos usuários, viesse a oferecer informações personalizadas com as ofertas que mais apontam nesta direção, de acordo com as conclusões feitas em IP (2000).

Apesar da forma inteligente proposta, sempre existe a necessidade de se identificar o usuário para oferecer um conteúdo “personalizado em potencial”, baseado em informações geralmente não consideradas em muitas páginas da *Web*.

A proposta feita neste trabalho é uma contribuição ao já desenvolvido (IP et al., 2000), e se preocupa em apresentar conteúdo atualizado na página sem a necessidade de identificação permanente do internauta. A partir de informações, tais como, tempo de permanência, frequência de visitas a página e o IDLI, desenvolver um sistema que modele uma coleção de dados de

entrada/saída. Permitindo que o sistema tenha a capacidade de se adaptar, fornece ao sistema *fuzzy* desenvolvido a habilidade de aprender o relacionamento embutido na coleção de dados de entrada/saída.

No item 2, será descrito a metodologia envolvida para definição do sistema de inferência nebulosa neural adaptativa (ANFIS – *Adaptive Neural Fuzzy Inference System*). No item 3, será apresentado um caso exemplo para demonstrar a metodologia proposta. No item 4, os resultados foram apresentados e discutidos para este caso exemplo. No item 5, serão apresentadas as conclusões finais sobre o trabalho.

2. METODOLOGIA PROPOSTA

Um sistema de inferência *fuzzy* neural adaptativo (ANFIS – *Adaptive Neural Fuzzy Inference System*) é um sistema de inferência *fuzzy* (FIS – *Fuzzy Inference System*) que pode ser treinado com um algoritmo de backpropagação para modelar um conjunto de dados entrada/saída (Uhrig et al., 1997). O FIS permitindo adaptação contempla o sistema *fuzzy* com a habilidade de aprender inter-relações embutidas nas entrada/saída dos dados utilizados.

A estrutura de rede da ANFIS facilita o processo de computação do vetor gradiente que relaciona a redução de uma função erro para uma modificação nos parâmetros de uma FIS. Assim que o vetor gradiente é obtido, um número de rotinas de otimização podem ser aplicadas para reduzir o erro entre a saída real e a obtida (Uhrig et al., 1997). Na literatura de redes neurais, isto é conhecido com aprender com o exemplo.

O modelo de ANFIS utilizada neste trabalho é do tipo Sugeno (conhecida como *TSK fuzzy model*) proposta por Tagaki & Sugeno (1985) e Sugeno & Kang (1988). Takagi e Sugeno (1985) propuseram utilizar as regras **IF-THEN** do seguinte forma:

$L^{(l)} : \mathbf{IF} \ x_1 \text{ is } F_1^l \text{ and } \dots \text{ and } x_n \text{ is } F_n^l,$

$$\mathbf{THEN} \ y^l = c_0^l + c_1^l x^1 + \dots + c_n^l x_n \quad (1)$$

Onde, F_i^l são conjuntos *fuzzy*, c_i são variáveis reais, y_l é o valor de saída dos sistema devido a regra $L^{(l)}$, e $l = 1, 2, \dots, M$. Ou seja, a parte relacionada **IF** é *fuzzy* mas a parte **THEN** é real (*crisp*) – a saída é uma combinação linear das variáveis de entrada (*input*). Para variáveis de entrada reais cujo vetor $\underline{x} = (x_1, \dots, x_n)^T$, a saída é $y(\underline{x})$ o sistema *fuzzy* do tipo Takagi e Sugeno e uma média dos pesos dos valores de y^l 's:

$$Y(\underline{x}) = (\sum_{l=1}^M w^l y^l) / (\sum_{l=1}^M w^l) \quad (2)$$

Onde os pesos w^l implica no valor verdade da premissa da regra $L^{(l)}$ para as variáveis de entrada e calculada como:

$$W^l = \prod_{i=1}^n \mu_{F_i^l}(x_i) \quad (3)$$

Maiores detalhes sobre *TSK fuzzy model* podem ser encontrados em (Wang, 1993). A configuração do sistema *fuzzy* do tipo Takagi & Sugeno é mostrado na Figure 1.

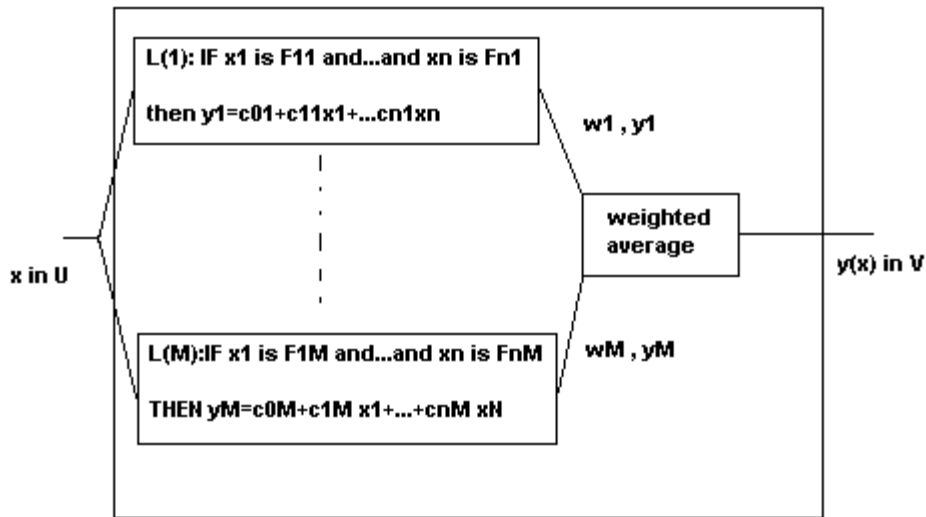


Figura 1 – Configuração básica do sistema fuzzy do tipo Takagi & Sugeno.

O *fuzzy* toolbox do (MATLAB6, 2000), foi empregado para representação do problema.

3. APLICAÇÃO DA METODOLOGIA

Para aplicação da metodologia proposta aqui, as informações e dados de navegação de internauta utilizados em (Ip et al., 2000) foram também empregados nesta aplicação. Inicialmente uma tabela (veja *Table 1* em IP, 2000) foi apresentada contendo informações extraídas de três visitantes da página *Web* de um empresa virtual comercial. Estes dados foram compilados automaticamente e gravados através de um contador utilizando programação em *Visual Basic Script*. Três conjuntos de dados estatísticos são extraídos da informação gravada devido a navegação dos internautas. O tempo total de permanência em cada local, T_s , a frequência de visita de cada local, F_v , e o tempo ocorrido entre duas visitas sucessivas, T_i . Um valor ILIP

(*Interest Level Indicating Point*) para cada local é determinado. Os locais com valores ILIP relativamente altos significou que o visitante tem maior interesse no local. O responsável pelo projeto da página (*Web master*) deverá colocar mais informações sobre estes produtos, e reservar algum espaço na página principal contendo promoções destes produtos para as próximas visitas do mesmo visitante, como é o caso citado anteriormente das “LojasAmericanas.com”, com os produtos exibidos em promoção feito de forma indiscriminada sem um sistema inteligente e da “AMAZON.com” com as informações disponibilizadas personalizadas, somente após a identificação do internauta. É muito possível que a informação em tempo real possa ser obtida conforme o desejado, fornecida por um sistema *fuzzy* trabalhando em conjunto com a página *Web*. Os autores em (IP et al., 2000) declaram que o sistema proposto possui a capacidade de decisão humana relacionada para as vendas de produtos através da Internet sob o conceito da entrega de informação selecionada e desta forma mais sucesso do que a estratégia de entrega de informação massiva tradicional.

Uma segunda tabela (veja *Table 2* em IP, 2000) foi preparada para calcular os valores de ILIP, usando os dados estatísticos da primeira e consta de T_s , F_v , e T_i , para cada internauta. Estes valores foram exibidos em outra tabela (veja *Table 5* e *6* em IP, 2000). Ficou demonstrado que nenhum valor de ILIP pode ser determinado se o internauta tem apenas uma visita em um local da página em um período de navegação. E, que locais com o tempo de permanência grande e visitado em uma única navegação, um ILIP maior é obtido. Outra observação é que deslocamentos entre de locais de pesquisa na página podem ser sempre identificadas.

Em (*Table 6*, IP, 2000) foram compilados para cada visitante, o T_s , o F_v , os valores calculados ILIP, e o número de ordens de compra (*No. of order*) feitas em cada local da página visitada para um período de observação de 30 dias.

Baseado nestes dados foram definidos os pares de entrada/saída para o exemplo da aplicação da metodologia proposta aqui.

Cada visitante foi selecionado como um tipo de conjunto de dados, ou seja, o visitante (Visit 1) de número 1 foi o conjunto de treinamento (*training*). O visitante de número 2 (Visit 2), foi o conjunto de verificação (*checking*), e o visitante de número 3 (Visit 3), o conjunto de teste (*testing*).

Na Figura 2, estes valores foram apresentados, e confirmam os comentários anteriormente apresentados para ordem de compra.

<i>Shopping</i>	<i>Visit 1</i>				<i>Visit 2</i>				<i>Visit 3</i>			
	<i>Ts(min)</i>	<i>Fv</i>	<i>ILIP</i>	<i>Ordem</i>	<i>Ts(min)</i>	<i>Fv</i>	<i>ILIP</i>	<i>Ordem</i>	<i>Ts(min)</i>	<i>Fv</i>	<i>ILIP</i>	<i>Ordem</i>
BE	522	15	85.5	2	210	10	86.1	1	23	3	7.5	0
Com	270	10	77.9	1	15	2	10.1	0	724	21	95.7	3
F/W	103	11	57.2	1	407	15	91.6	3	768	20	94.3	4
ID	440	19	87.7	2	78	6	37.4	0	110	4	19.2	0
Off	653	18	92.3	3	115	8	60.5	0	542	15	81.8	2
SM	221	8	56.1	0	388	13	89.7	3	209	12	44.4	1
Tele	78	5	22.3	0	147	7	60.2	1	441	17	76.8	2

Figura 2 – Dados dos três internautas.

Onde,

BE – *Business Equipament*, Com – *Computer*, F/W – *Food/wine*, ID – *Interior Design*, SM – *Shopping Mall*, e Tele – *Telecommunication*.

Os valores de Ts, Fv, e ILIP, são os dados de entrada, e os valores de Ordem são os dados de saída.

Em seguida a preparação dos dados, a FIS (*fuzzy inference system*) é criada, usando o fuzzy toolbox do MATLAB. Algumas propriedades da FIS Sugeno desenvolvida, contendo o método de implicação e outras, são apresentadas:

```
nome: 'Produto'
type: 'sugeno'
```

```
andMethod: 'prod'  
orMethod: 'probor'  
defuzzMethod: 'wtaver'  
impMethod: 'min'  
aggMethod: 'max'  
input: [1x3 struct]  
output: [1x1 struct]  
rule: [1x27 struct]
```

No próximo item, serão apresentados os resultados exibidos em forma gráfica.

4. RESULTADOS

Na Figura 3, os valores calculados de IDLI em (IP et al, 2000) que foram exibidos na Figura 2, são novamente apresentados em forma gráfica juntamente com os valores estimados com o FIS Sugeno gerado neste estudo, para efeito de comparação.

Os gráficos relacionados a v1c, v2c, e v3c, são referentes aos valores calculados. Os gráficos v1e, v2e, e v3e, são os relacionados aos estimados pela metodologia ANFIS aqui proposta. Na Figura 3, o eixo dos x representa os locais visitados na página Web, e o eixo y, os valores relativos ao número de ordem feita para um determinado produto e para um específico internauta. Para este exemplo, os locais visitados são em número de 7. Conforme descrito anteriormente, e relacionado na parte inferior da Figura 3.

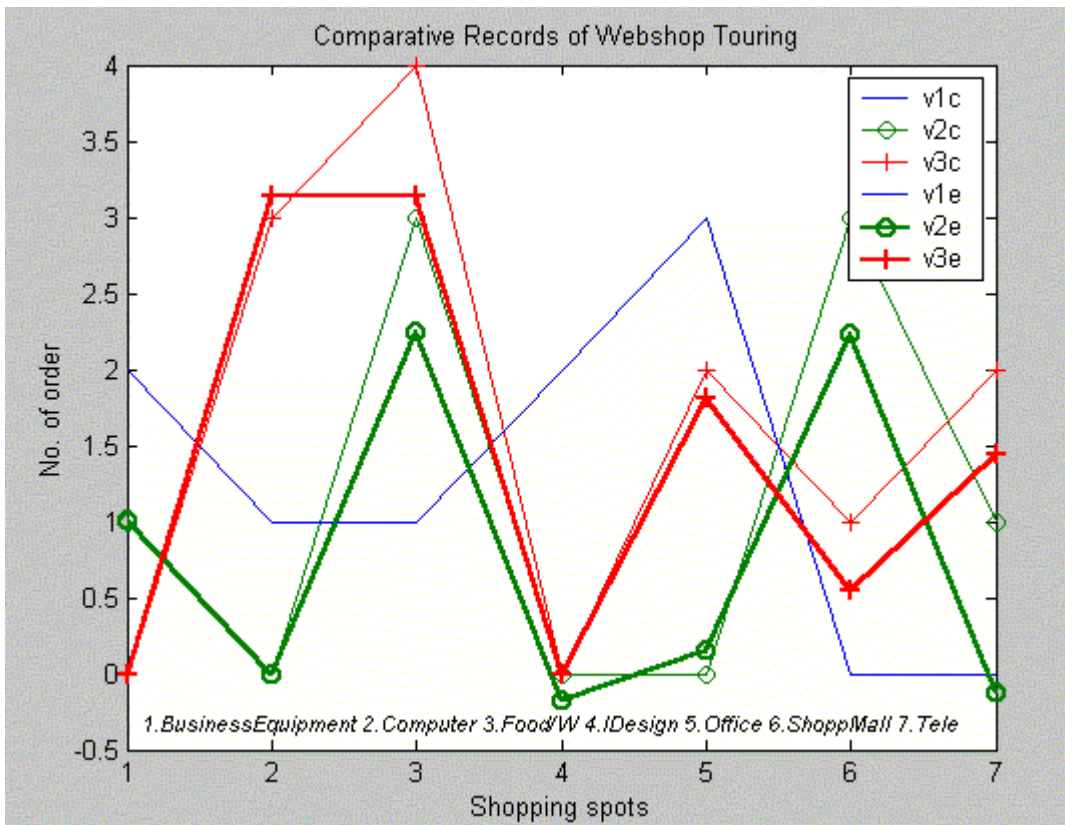


Figura 3 – Dados comparados entre o calculado e o estimado com o método ANFIS.

O primeiro conjunto de dados, o de treinamento é o v1c, utilizado para desenvolver a FIS Sugeno, juntamente com o conjunto de dados v2c, utilizado como conjunto de verificação, e cuja finalidade é o de habilitar o sistema *fuzzy* a generalizar. O conjunto de dados v3c, foi utilizado como conjunto de teste, cuja finalidade é avaliar a capacidade do sistema *fuzzy* a reconhecer algum conjunto novo apresentado ao FIS. É importante ressaltar que este conjunto v3c, não foi utilizado anteriormente para geração da FIS. No primeiro caso, v1c com o gráfico na cor azul, não se verifica diferença entre o calculado e o estimado. No segundo caso, v2c com o gráfico na cor verde, a estimativa pode ser considerada boa. E no terceiro caso, v3c com gráfico na cor vermelha, a estimativa a partir de um conjunto novo, pode ser também considerada boa.

Com relação a aquisição de informações de navegação dos internautas, além do citado anteriormente com a linguagem VBScript, existem dois ambientes computacionais, *Webalizer (2002)* e o *WebTrends (2002)*, para monitoração do arquivo “log”. Este arquivo armazena informações sobre a navegação do internauta durante uma pesquisa na página *Web*, e é encontrado no servidor *WEB* da empresa virtual.

5. CONCLUSÕES

Este trabalho apresentou uma proposta de monitoração das informações registradas em arquivo “Log”, devido a navegação de um internauta em uma página comercial da *Web*, e para tal, um sistema de inferência *fuzzy*, FIS (*Fuzzy Inference System*), foi desenvolvido utilizando a metodologia de um sistema ANFIS (*Adaptive Neural Fuzzy Inference System*) do tipo Sugeno & Tagaki. O autor acredita que se o sistema *fuzzy* for acoplado a página *web* para funcionamento em conjunto, as informações estimadas através do FIS, após um período de observação, podem ser utilizadas em tempo real para auxiliar a disponibilização das informações de ofertas de produtos oferecidos pela empresa virtual, sem a necessidade de identificação do usuário e de maneira mais inteligente e criteriosa.

A base de dados da empresa deverá ser utilizada para se extrair estatísticas para um período de tempo para definição de um padrão, e a partir daí novos dados para serem utilizados na generalização da FIS inicialmente produzida. Um processo contínuo de atualização do padrão.

Este tipo de funcionamento, é uma maneira mais inteligente e mais comercial, se comparada aos dois exemplos citados anteriormente no texto do trabalho, e acredita-se que ocorra em muitos outros endereços (*sites*) comerciais virtuais.

Com a metodologia proposta aqui, uma empresa comercial virtual estaria livre quanto à, e após **identificar o usuário**, coletar informações relevantes que serão disponibilizadas quando do retorno do internauta, **se ele voltar**, como é o caso da *Amazon.com*, e por outro lado, não estaria enchendo a página principal com ofertas indesejáveis, **sem a necessidade de identificação do usuário**, e **não se preocupando se o internauta vai voltar ou não ao site da página**, como é o caso da *Lojas Americanas*.

6. REFERÊNCIAS

IP, R. W. L., H.C.W. Lau, & F.T. Chan 2000, “An Intelligent Internet Information Delivery System to Evaluate Site Preferences.”, *Expert Systems with Applications*, PERGAMON, ELSEVIER, 18, 33-42.

Manual do Fuzzy Toolbox do MatLab6. 2000.

Wang, Li-Xin, ADAPTIVE FUZZY SYSTEMS AND CONTROL – DESIGN AND STABILITY ANALYSIS, University of California at Berkeley. PTR Prentice Hall, Englewood Cliffs, New Jersey 07632, November 1993.

Sugeno, M., & G. T. Kang, “Structure Identification of Fuzzy Models,” *Fuzzy Sets Syst.*, 28, 15 (1988).

Takagi, T., and M. Sugeno, “Fuzzy Identification of Systems and Its Applications to Modeling and Control,” *IEEE Trans. Syst. Man, Cybern.*, SMC-15(1), 116-132 (1985).

Uhrig, R. E., Wrest, D. J., Hines, & J. W. August 1997, “Signal Validation Using An Adaptive Neural Fuzzy Inference System.”, NUCLEAR TECHNOLOGY, VOL. 119.

Webalize, <http://www.mrunix.net/webalizer> , 2002.

WebTrends, <http://www.netiq.com/webtrends/default.asp> , 2002.