

Um sistema difuso inteligente para avaliar informações de usuários na Internet

Antonio Cesar Ferreira Guimarães

Inteligência Artificial – Sistemas Computacionais Inteligentes.
Divisão de Reatores, Instituto de Engenharia Nuclear.
tony@ien.gov.br

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 (IDL) que mede o interesse do usuário na página Web, o sistema estabelece um padrão existente nestas informações relacionadas com o número de Ordens de Compras dadas por um usuário. Desta forma, as variáveis tratadas como pares de entradas/saídas estabelecem um sistema adaptativo difuso 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.

Palavras-chave

Lógica difusa; Internet; Anfis.

Fuzzy intelligent system to evaluate user information on Internet

Abstract

This work consists of the development of an Intelligent System in the Internet (ISI) 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 (IDL) that measures the interest of the users in the page web, the system establishes an existent pattern in these information related with the number of Orders of Purchases given by a user. In 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.

Keywords

Fuzzy logic; Internet; Anfis.

INTRODUÇÃO

Uma das maiores preocupações de empresas comerciais ligadas à Internet é saber exatamente quais informações mostrar ao usuário e desta forma manter sua atenção em ofertas ou produtos em geral, impedindo que o mesmo não abandone a página (*home page*) Web, ou seja, mantenha-se 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 usuário deseje voltar o maior número de vezes à página Web e que o intervalo entre uma visita e outra seja em um intervalo de tempo cada vez menor. Várias empresas comerciais que utilizam o sistema de comércio eletrônico (*shopping*) fazem de tudo para atrair a atenção dos usuários. Basicamente, grande número de empresas usa o recurso de um cadastro, no qual o usuário, após o seu preenchimento, passa 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 oferecem-se informações personalizadas sobre produtos para novas aquisições, como, por exemplo, a livreria Amazon, acessada por meio da URL <http://www.amazon.com>, em que 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, por exemplo, da *Lojas Americanas*, acessada através da URL <http://www.lojasamericanas.com>, na qual a atualização é feita assim que o usuário 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 usuário 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 usuário 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.

Foi proposto em Ip *et al.* (2000) o desenvolvimento de um sistema (*Intelligent Internet Information Delivery System – IIIDS*) 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 visitados e relevantes por meio dos princípios de lógica difusa (*Fuzzy Logic*). Neste caso, a metodologia proposta sugere a utilização de um sistema inteligente e tomando o exemplo da Amazon se assemelharia a uma situação em que os recursos da Amazon oferecidos aos usuários viessem a oferecer informações personalizadas com as ofertas que mais apontam nesta direção, de acordo com as conclusões feitas em Ip *et al.* (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 do usuário. Muitos usuários, durante as visitas, não se identificam e nem fazem cadastro para suas consultas. A partir de informações, tais como tempo de permanência, frequência de visitas à página e o IDLI, deve-se desenvolver um sistema que modele a coleção de dados de entrada/saída, proporcionando adaptação ao sistema e fornecendo a habilidade de aprender o relacionamento embutido na coleção de dados de entrada/saída.

No item 2, será descrita a metodologia envolvida para definição do sistema de inferência difusa neuroadaptativa (*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 relatadas as conclusões finais sobre o trabalho.

METODOLOGIA PROPOSTA

Um sistema de inferência difuso neuroadaptativo (*Anfis – Adaptive Neural Fuzzy Inference System*) é um sistema de inferência difuso (*FIS – Fuzzy Inference System*) que pode ser treinado com um algoritmo de retropropagação para modelar um conjunto de dados entrada/saída (Uhrig *et al.*, 1997). O FIS, permitindo adaptação, contempla o sistema difuso com a habilidade de aprender inter-relações embutidas nas entradas e saídas 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 pode ser aplicado para reduzir o erro entre a saída real e a obtida (Uhrig *et al.*, 1997). Na literatura de redes neurais, isto é conhecido como aprender com o exemplo.

O modelo de *Anfis* utilizado neste trabalho é do tipo Sugeno (conhecido como *TSK fuzzy model*), proposto por Takagi & Sugeno (1985) e Sugeno & Kang (1988). Takagi & Sugeno (1985) propuseram utilizar as regras **IF-THEN** da seguinte forma:

$$\left. \begin{array}{l} L^{(l)} : \mathbf{IF} \ x_1 \text{ is } F_1^l \text{ and } \dots \text{ and } x_n \text{ is } F_n^l \\ \mathbf{THEN} \ y^l = c_0^l + c_1^l x^1 + \dots + c_n^l x_n \end{array} \right| \quad (1)$$

Onde, F_i^l são conjuntos difusos, c_i são variáveis reais, y_l é o valor de saída dos sistemas devido à regra $L^{(l)}$, e $l = 1, 2, \dots, M$. Ou seja, a parte relacionada **IF** é difusa, 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 $x = (x_1, \dots, x_n)^T$, a saída é $y(x)$ o sistema difuso do tipo Takagi e Sugeno e uma média dos pesos dos valores de y^l s:

$$Y(x) = \left(\frac{\sum_{l=1}^M w^l \cdot y^l}{\sum_{l=1}^M w^l} \right) \quad (2)$$

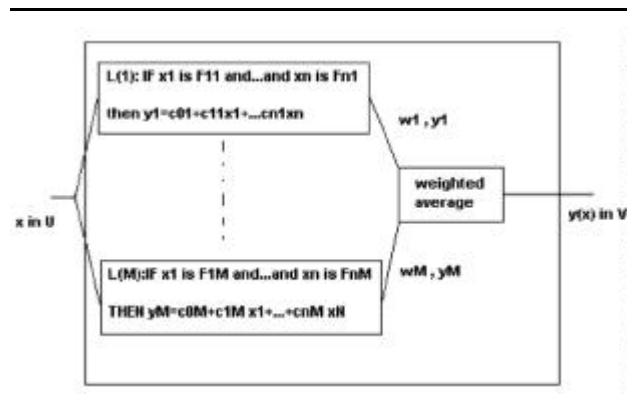
Onde os pesos w^l implicam o valor verdade da premissa da regra $L^{(l)}$ para as variáveis de entrada e calculada como:

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

Mais detalhes sobre *TSK fuzzy model* podem ser encontrados em (Wang, 1993). A configuração do sistema difuso do tipo Takagi & Sugeno é mostrada na figura 1.

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

FIGURA 1
Configuração básica do sistema difuso do tipo Takagi & Sugeno.



APLICAÇÃO DA METODOLOGIA

Para aplicação da metodologia proposta aqui, as informações e dados de navegação de usuários utilizados em (Ip *et al.*, 2000) foram também empregadas nesta aplicação. Serão descritas a seguir algumas etapas de desenvolvimento neste referido trabalho. Foram obtidas inicialmente informações extraídas para três visitantes da página *Web* de um empresa virtual comercial (veja *Table 1* em Ip, 2000). Estes dados foram compilados automaticamente e gravados por meio de um contador utilizando a linguagem de programação em *Visual Basic Script* (VBScript). Maiores informações sobre VBScript podem ser obtidas por meio da URL da Microsoft <http://www.microsoft.com>. Três conjuntos de dados estatísticos são extraídos da informação gravada devido à navegação dos usuários. Dados estatísticos, tais como, o tempo total de permanência em cada local, *Ts*, a frequência de visita de cada local, *Fv*, e o tempo ocorrido entre duas visitas sucessivas, *Ti*, foram retirados da tabela inicial e apresentados em uma segunda tabela para calcular os valores do ILIP (veja *Table 2* em Ip, 2000). Um valor ILIP (*Interest Level Indicating Point*) para cada local visitado é determinado (veja *Table 5* em Ip, 2000). Os locais com valores ILIP relativamente altos significam 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. É muito possível que a informação em tempo real possa ser obtida conforme o desejado, fornecida por um sistema difuso trabalhando em conjunto com a página *Web*. Os autores (Ip *et al.*, 2000) declaram que o sistema proposto possui a capacidade de decisão humana relacionada para as vendas de produtos por meio da Internet sob o conceito da entrega de informação selecionada e, desta forma, tem mais sucesso do que a estratégia de entrega de informação intensiva tradicional.

Os valores dos ILIPs determinados demonstraram que nenhum valor de ILIP pode ser determinado se o usuário tem apenas uma visita em um local da página em um período de navegação e, também, 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 locais de pesquisa na página podem ser sempre identificados.

Em outra tabela (veja *Table 6*, Ip *et al.*, 2000), apresentada aqui em forma de figura, figura 2, foram compilados para cada visitante o *Ts*, o *Fv*, 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 da figura 2, foram definidos então os pares de entrada/saída para o exemplo da aplicação da metodologia Anfis proposta aqui, conforme descrito a seguir.

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*). Os dados foram assim considerados para facilitar a implementação da proposta e não impedem no futuro de se utilizarem dados mais específicos.

Onde,

BE – *Business Equipment*, 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 à preparação dos dados, a FIS (*fuzzy inference system*) é criada, usando o *toolbox* (Matlab, 2000). Algumas propriedades da FIS Sugeno desenvolvidas 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.

RESULTADOS

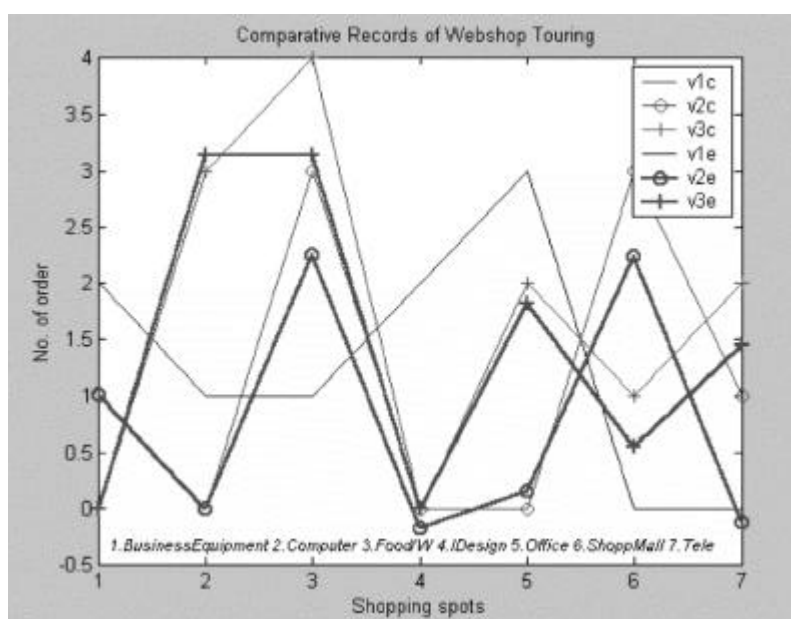
Na figura 3, os valores calculados de IDLI em (Ip *et al.*, 2000), que foram exibidos na figura 2, são novamente apresentados em forma gráfica juntamente com os valores estimados com o FIS Sugeno gerado neste estudo, para efeito de comparação.

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 usuário. Para este exemplo, os locais visitados são em número de 7, conforme descrito anteriormente e relacionado na parte inferior da figura 3.

FIGURA 2
Dados dos três usuários

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

FIGURA 3
Dados comparados entre o calculado e o estimado com o método Anfis



O primeiro conjunto de dados de treinamento é o v1c, utilizado para desenvolver a FIS Sugeno, juntamente com o conjunto de dados v2c, utilizado como conjunto de verificação, cuja finalidade é a de habilitar o sistema difuso a generalizar. O conjunto de dados v3c foi utilizado como conjunto de teste, cuja finalidade é avaliar a capacidade de o sistema difuso 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), não se verifica diferença entre o calculado e o estimado. No segundo caso (v2c), a estimativa pode ser considerada boa. E no terceiro caso (v3c), a estimativa a partir de um conjunto novo pode ser também considerada boa.

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

CONCLUSÕES

Este trabalho apresentou uma proposta inteligente para exibir informações selecionadas na Internet. Para tal, um sistema de inferência difuso, FIS (*Fuzzy Inference System*), foi desenvolvido utilizando a metodologia Anfis (*Adaptive Neural Fuzzy Inference System*) do tipo Sugeno & Tagaki.

Com o "sistema" proposto, pode-se estimar o nível de interesse e potencial de compra de um usuário. O autor acredita que o sistema difuso acoplado à página *Web* possa ser utilizado em tempo real para auxiliar a apresentação das informações sobre produtos oferecidos pela empresa virtual, sem a necessidade de identificação do usuário e de maneira mais inteligente e criteriosa.

O arquivo "log" da empresa deverá ser utilizado para se extrair estatísticas para um período de tempo para definição de um padrão. A partir daí, surgirão novos dados para serem utilizados na generalização da FIS inicialmente produzida, em um processo contínuo de atualização do padrão.

Este tipo de funcionamento é uma maneira mais inteligente e mais comercial, se comparado aos dois exemplos citados anteriormente no texto do trabalho, Amazon e Lojas Americanas, e acredita-se que ocorra em muitas outras empresas comerciais virtuais.

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

Os dados aqui utilizados foram obtidos do exemplo proposto em (Ip *et al.*, 2000) e devem ser considerados apenas importantes do ponto de vista de demonstração. Um conjunto maior de dados deve ser selecionado, de forma a permitir ajustes no sistema Anfis durante a sua fase de treinamento, teste e validação. É importante ressaltar que o *toolbox*, pela sua facilidade de implementação, foi utilizado apenas como um simulador. O sistema pode ser totalmente desenvolvido em uma linguagem apropriada.

Ferramentas automatizadas podem ser empregadas para análise de *log*. Scripts podem ser desenvolvidos para extração de informações do *log*, para em seguida serem incorporados no sistema difuso que vai estimar os níveis de interesse, para em seguida serem repassados à programação *Web*, que disponibilizará informações selecionadas para o usuário.

Artigo aceito para publicação em 22-11-2002

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. [S. l.] : Pergamon, 2000. p. 18, 33-42.
- MANUAL do fuzzy toolbox do matlab 6. [S. l. : s. n.], 2000.
- SUGENO, M.; KANG, G. T. Structure identification of fuzzy models. *Fuzzy Sets Syst.*, v. 28, n. 15, 1988.
- TAKAGI, T.; SUGENO, M. Fuzzy identification of systems and Its applications to modeling and control. *IEEE Trans. Syst. Man, Cybern.*, v. 15, n. 1, p. 116-132, 1985.
- UHRIG, R. E. *et al.* Signal validation using an adaptive neural fuzzy inference system. *Nuclear Technology*, v. 119, Aug. 1997.
- WANG, Li-Xin. *Adaptative fuzzy systems and control : design and stability analysis*. Berkeley : University of California, 1993.
- Webalize*. Disponível em: < <http://www.mrunix.net/webalizer>> . Acesso em: 2002.
- WebTrends*. < <http://www.netiq.com/webtrends/default.asp>> . Acesso em: 2002.