

“Comparação de Estruturas Neuro-Difusas para Identificação de Sistemas”

R. P. Paiva*, A. Dourado*

* CISUC – Centro de Informática e Sistemas da Universidade de Coimbra
Departamento de Engenharia Informática, PÓLO II da Universidade de Coimbra,
Pinhal de Marrocos, P 3030, Coimbra, Portugal
Tel. +351-239-790000, Fax: +351-239-701266, e-mail: {ruipedro, dourado}@dei.uc.pt

RESUMO

Neste artigo apresenta-se uma análise comparativa experimental de diversas estruturas neuro-difusas, nomeadamente estruturas do tipo linguístico e Takagi-Sugeno de ordem 0 e 1. A implementação do modelo é efectuada através do treino de uma rede neuro-difusa, i.e., uma rede neuronal cuja arquitectura representa um sistema difuso. Assim, numa primeira fase a estrutura do modelo é aprendida por meio do algoritmo de agrupamento subtrativo, o qual possibilita a extracção de um conjunto de regras relevantes a partir de um conjunto representativo de amostras de entrada-saída. Na segunda fase, a sintonização dos parâmetros das funções de pertença do modelo é efectuada através do treino de uma rede neuronal. Efectua-se também um estudo comparativo de diversos operadores e funções de pertença Gaussianas simples e generalizadas.

PALAVRAS-CHAVE: identificação de sistemas, redes neuronais, sistemas difusos, redes neuro-difusas, agrupamento de classes.

1 INTRODUÇÃO

Actualmente, assiste-se a um interesse crescente pelas técnicas neuro-difusas na modelização de sistemas. De facto, os modelos difusos gozam de algumas propriedades que os tornam particularmente interessantes, nomeadamente, aproximação universal [1] e possibilidade de interpretação linguística [10], característica esta dificilmente verificada nas redes neuronais MLP (Multi-Layer Perceptron). No entanto, padecem de uma limitação importante, a qual deriva da dificuldade de quantificação dos termos linguísticos associados a cada variável. Neste sentido, surgem as redes neuro-difusas como tentativa de dotar os sistemas difusos de capacidades de aprendizagem.

A metodologia apresentada decorre, essencialmente, em duas fases: na primeira, procede-se à aprendizagem da estrutura do modelo, i.e., à obtenção de um conjunto de regras difusas a incorporar no modelo; na segunda, toma lugar a sintonização dos parâmetros do modelo, i.e., os parâmetros das funções de pertença do sistema difuso.

Com base na metodologia genérica referida, este artigo procura efectuar uma análise relativa aos diversos parâmetros a definir num modelo difuso, e.g., tipo de regras, operadores e funções de pertença. De notar, no entanto, que em termos de funções de pertença, o estudo é restringido a funções Gaussianas simples e generalizadas.

Assim, na Secção 2 o algoritmo de agrupamento subtrativo, utilizado na aprendizagem da estrutura é apresentado, seguindo-se na Secção 3 a estratégia de aprendizagem de parâmetros. Os resultados da aplicação das metodologias descritas à série caótica Mackey-Glass são apresentados na Secção 4. Finalmente, as conclusões deste trabalho são sumariadas na Secção 5.

2 APRENDIZAGEM DA ESTRUTURA

De modo a obter-se um conjunto de g regras condicionais difusas, representativas de um sistema em estudo, os algoritmos de agrupamento de classes afiguram-se particularmente interessantes. Tal circunstância advém do facto de tais métodos possibilitarem um particionamento disperso do espaço de entrada-saída, encontrando-se assim um conjunto de regras relevantes. Relativamente aos métodos de particionamento em grelha, há a vantagem de se evitar a explosão da base de regras. Neste artigo, é aplicado o algoritmo de agrupamento subtrativo de Chiu [2]. Esta metodologia apresenta vantagens interessantes, especialmente num contexto de identificação neuro-difusa. De facto, o algoritmo caracteriza-se pela sua eficiência e adequação à inicialização de procedimentos de optimização iterativa, tal como é o caso.

O método de Chiu insere-se na classe dos métodos de função de potencial, sendo, mais precisamente, uma variação do método da montanha. Nesta classe de metodologias, é definido um conjunto de pontos como possíveis centros de grupo, sendo cada um dos quais visualizado como uma fonte de energia. No caso do agrupamento subtrativo, os candidatos a centros são as próprias amostras de dados. Deste modo, ultrapassa-se a principal limitação do método da montanha, a qual reside no facto de os candidatos serem definidos numa grelha. O algoritmo é descrito detalhadamente em [2].

3 APRENDIZAGEM DE PARÂMETROS

Após a determinação da estrutura do modelo difuso a obter, os seus parâmetros, i.e., os centros e desvios padrões das funções de pertença Gaussianas, devem ser sintonizados. Neste sentido, há que considerar o tipo de modelos difusos utilizados. Assim, em modelos linguísticos, as regras condicionais são da forma (1). Para modelos do tipo Takagi-Sugeno de ordem 1, as regras são do tipo (2), definindo-se $f_{or}(x)$ como em (3). Para modelos de ordem 0, a função anterior é representada por uma constante. Tal como se pode verificar, modelos do tipo Takagi-Sugeno de ordem 1 apresentam uma maior flexibilidade no consequente. Deste modo, estas

estruturas podem ser vistas como comutadores suaves entre modelos lineares locais, o que apresenta vantagens face à estrutura interpolativa presente nos modelos de ordem 0 e linguísticos.

Regra r :

$$\begin{aligned} & \text{SE } (X_1 \acute{e} LX1^{(r)}) \text{ E } (X_2 \acute{e} LX2^{(r)}) \text{ E } \dots \text{ E } (X_m \acute{e} LXm^{(r)}) \\ & \text{ENT\AA O } (Y_1 \acute{e} LY1^{(r)}) \text{ E } (Y_2 \acute{e} LY2^{(r)}) \text{ E } \dots \text{ E } (Y_n \acute{e} LYn^{(r)}) \end{aligned} \quad (1)$$

Regra r :

$$\begin{aligned} & \text{SE } (X_1 \acute{e} LX1^{(r)}) \text{ E } (X_2 \acute{e} LX2^{(r)}) \text{ E } \dots \text{ E } (X_m \acute{e} LXm^{(r)}) \\ & \text{ENT\AA O } [y_1 = f_{1r}(x)] \text{ E } [y_2 = f_{2r}(x)] \text{ E } \dots \\ & \text{E } [y_n = f_{nr}(x)] \end{aligned} \quad (2)$$

$$\begin{aligned} f_{or}(x) &= b_{or0} + b_{or1}x_1 + b_{or2}x_2 + \dots + b_{orm}x_m \\ b_{orj} &\in \mathfrak{R}, \quad j = 1, 2, \dots, m; \quad o = 1, 2, \dots, n; \quad r = 1, 2, \dots, g \end{aligned} \quad (3)$$

Assim, os parâmetros de cada uma das estruturas difusas referidas serão ajustados através do treino de uma rede neuronal difusa, i.e., uma rede neuronal susceptível de representar as funções de um sistema difuso, nomeadamente, fuzificação, implicação e desfuzificação.

Basicamente, as redes apresentadas nos parágrafos seguintes são constituídas por uma camada de entrada, seguindo-se uma camada de fuzificação e só depois a camada de regras. Após as camadas iniciais segue-se, no caso de consequentes do tipo Takagi-Sugeno de ordem 0 ou ordem 1, uma camada linear de saída. Para o caso de consequentes difusos, segue-se uma camada de integração de regras com o mesmo consequente e, finalmente, a camada de saída responsável pela operação de desfuzificação.

Por forma a tornar mais clara a leitura das expressões matemáticas utilizadas nos parágrafos seguintes, optou-se por apresentar, desde já, a notação relativa à activação dos neurónios de cada camada:

- $a_i^{(p2)}$: activação do neurónio i da camada 2, relativamente ao padrão p (i denota um termo de entrada: “input”);
- $a_r^{(p3)}$: activação do neurónio r da camada 3, relativamente ao padrão p (r denota “regra”);
- $a_s^{(p4)}$: activação do neurónio s da camada 4, relativamente ao padrão p (s denota “norma-S”);
- $a_o^{(p5)} = y_o^{(p)}$: activação do neurónio o da camada 5, i.e., saída, relativamente ao padrão p (o denota saída: “output”);
- $y_o^{(p)}$: activação desejada para o neurónio o da camada 5, i.e., para a saída da rede, relativamente ao padrão p .

No caso de consequentes do tipo Takagi-Sugeno, a camada de saída é a quarta, tendo-se:

- $a_o^{(p4)} = y_o^{(p)}$: activação do neurónio o da camada 4, i.e., saída, relativamente ao padrão p (o denota saída: “output”);

Consequentes do tipo Takagi-Sugeno

Assim, do exposto resulta, para estruturas difusas do tipo Takagi-Sugeno, a arquitectura representada na Figura 1. Naturalmente, a rede apresentada serve tanto os modelos de ordem 1 como de ordem 0, bastando para tal considerar pesos constantes ou funções de primeira ordem.

Nesta estrutura e nas seguintes, a *camada de entrada* tem por missão única receber dados do ambiente exterior e passá-los à camada seguinte, não realizando, portanto, processamento útil.

Na segunda camada, a *camada de fuzificação*, cada uma das células equivale a uma função de pertença associada a uma das entradas. Definindo funções de pertença Gaussianas convencionais, a saída de cada um dos neurónios desta camada é dada por (4):

$$a_i^{(p2)} = e^{-\frac{(\|x_j^{(p)} - c_{ij}\|^2)}{2s_{ij}^2}} \quad (4)$$

onde c_{ij} e s_{ij} representam, respectivamente, o centro e o desvio padrão da i -ésima função de pertença associada à j -ésima entrada. Na mesma expressão, $x_j^{(p)}$ denota o p -ésimo padrão associado à entrada j .

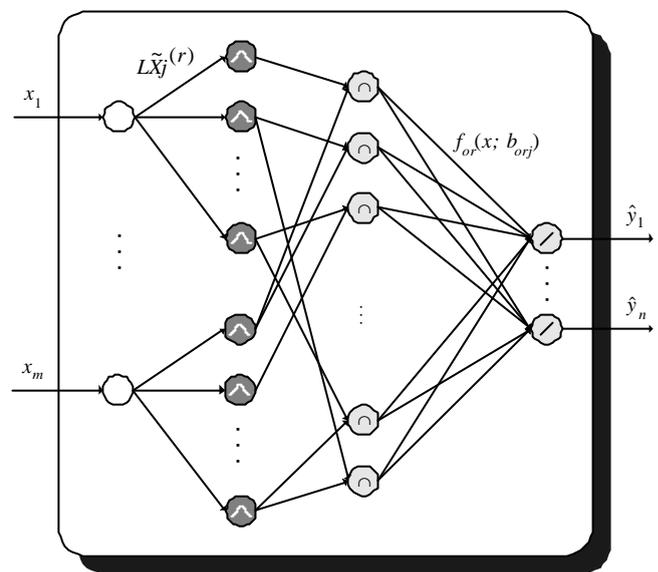


Figura 1. Rede neuro-difusa: consequentes do tipo Takagi-Sugeno.

Alternativamente, poder-se-ão definir funções Gaussianas generalizadas, as quais se caracterizam por poderem ser assimétricas e conter um planalto, constituindo assim uma generalização das Gaussianas convencionais (Figura 2). Como tal, coloca-se a hipótese de se obterem melhores resultados em virtude da maior flexibilidade destas funções.

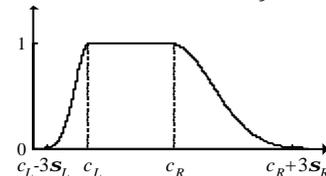


Figura 2. Função de pertença Gaussianas generalizada.

Os parâmetros das funções de pertença constituem os pesos das ligações entre os neurónios da primeira e da segunda camada ($LX_j^{(r)}$) na Figura 1).

Quanto aos neurónios da *camada de regras*, a sua função é, basicamente, efectuar a conjunção dos antecedentes de cada

uma das regras, por meio de uma qualquer norma-T, e.g., produto (5) ou mínimo. O primeiro insere-se na classe dos operadores algébricos e o segundo na classe dos operadores de truncatura.

$$a_r^{[p3]} = \underset{i=1}{\overset{na_r}{\text{norma}}} - T\{a_i^{[p2]}\} = \prod_{i=1}^{na_r} a_i^{[p2]} \quad (5)$$

onde na_r representa o número de entradas que constituem o antecedente da regra r .

No que respeita à *camada de saída*, a sua tarefa consiste no cálculo de saídas reais com base no grau de activação de cada regra. Tal como se referiu anteriormente, em modelos de ordem 0, os pesos desta camada denotam os consequentes do sistema difuso, definidos por constantes. Deste modo, a activação de cada um dos neurónios de saída é definida por (6):

$$\hat{y}_o^{[p]} = a_o^{[p4]} = \frac{\sum_{r=1}^g a_r^{[p3]} \cdot b_{or}}{\sum_{r=1}^g a_r^{[p3]}} \quad (6)$$

Na implementação de regras do tipo Takagi-Sugeno de ordem 1, a rede define um sistema difuso com regras do tipo (2). Deste modo, os neurónios de saída executam uma tarefa em tudo idêntica à do caso anterior, sendo no contexto presente definidos por (7):

$$\hat{y}_o^{[p]} = a_o^{[p4]} = \frac{\sum_{r=1}^g a_r^{[p3]} \cdot \left(\sum_{j=1}^m b_{orj} x_j^{[p]} + b_{or0} \right)}{\sum_{r=1}^g a_r^{[p3]}} \quad (7)$$

Consequentes difusos

No sentido da definição de redes neuro-difusas para a utilização de consequentes difusos, Lin define na sua arquitectura NFCN [5] uma rede composta por cinco camadas. No entanto, a estrutura original foi adaptada neste trabalho, no sentido da utilização de diferentes operadores e funções de pertinência. Relativamente à rede da Figura 1, as três primeiras camadas desempenham um papel igual, estando a diferença nas camadas posteriores.

Assim, a quarta camada, designada por *camada de união*, é responsável pela integração de regras com o mesmo consequente, por meio de uma norma-S, e.g., adição limitada (8) ou máximo. O primeiro insere-se na classe dos operadores algébricos e o segundo na classe dos operadores de truncatura. Af nr_s designa o número de regras que têm o neurónio s por consequente.

$$a_s^{[p4]} = \underset{r=1}{\overset{nr_s}{\text{norma}}} - S\{a_r^{[p3]}\} = \min\left[1, \sum_{r=1}^{nr_s} a_r^{[p3]}\right] \quad (8)$$

Quanto à *camada de saída*, ou *camada de desfuzificação*, os pesos das ligações entre os seus neurónios e os neurónios da camada de união ($LY_o^{(v)}$ na mesma figura) definem os parâmetros das funções de pertinência associadas aos termos de saída. Assim, com base nestas funções de pertinência e na acti-

vação de cada regra, os seus neurónios devem implementar um determinado método de desfuzificação adequado a consequentes difusos, tal como o método apresentado em [5] (9):

$$\hat{y}_o^{[p]} = a_o^{[p5]} = \frac{\sum_{s=1}^{|T(Y_o)|} c_{os} \mathbf{s}_{os} a_s^{[p4]}}{\sum_{s=1}^{|T(Y_o)|} \mathbf{s}_{os} a_s^{[p4]}} \quad (9)$$

onde c_{os} e \mathbf{s}_{os} representam o centro e desvio padrão da s -ésima função de pertinência associada à saída o . No caso de se utilizarem funções Gaussianas generalizadas, vem (10), tal como se define em [9]:

$$\hat{y}_o^{[p]} = a_o^{[p5]} = \frac{\sum_{s=1}^{|T(Y_o)|} \frac{1}{2} c_{osL} \mathbf{s}_{osL} + c_{osR} \mathbf{s}_{osR} \int a_s^{[p4]}}{\sum_{s=1}^{|T(Y_o)|} \frac{1}{2} \mathbf{s}_{osL} + \mathbf{s}_{osR} \int a_s^{[p4]}} \quad (10)$$

Nas expressões anteriores, $|T(Y_o)|$ denota o número de funções de pertinência associadas a cada variável linguística de saída Y_o . A ideia essencial do método de desfuzificação apresentado é pesar a activação de cada regra não só pelos centros, direito e esquerdo, da função, mas também pelos seus desvios padrões. Naturalmente, a expressão (10), reduz-se à expressão (9) no caso de se tratar de uma função Gaussiana usual. Com base na função desempenhada por cada neurónio, as redes linguísticas são treinadas, em modo *batch*, por meio do algoritmo de retropropagação do erro. Relativamente à implementação de modelos difusos Takagi-Sugeno, de ordem 0 ou 1, várias alternativas se afiguram como aplicáveis. Numa primeira hipótese, o treino da rede poderá ser efectuado pela retropropagação. No entanto, em consequência da linearidade da camada de saída, poder-se-á aplicar o estimador dos mínimos quadráticos [6] na forma matricial. Esta estratégia apresenta a vantagem de originar uma redução significativa no número de épocas requeridas. No entanto, o tempo necessário a cada época será mais elevado. A aplicação das metodologias de treino referidas é descrita detalhadamente em [9].

4 RESULTADOS DE SIMULAÇÃO

Um dos casos de estudo mais utilizados na identificação de sistemas consiste na predição da série temporal de Mackey-Glass [7], gerada pela equação diferencial caótica com atraso a qual se descreve pela expressão (11):

$$\dot{x}(t) = \frac{0.2x(t-t)}{1+x^{10}(t-t)} - 0.1x(t) \quad (11)$$

O problema a abordar consiste na predição de valores futuros da série. Utilizam-se os valores $x(t-18)$, $x(t-12)$, $x(t-6)$, $x(t)$ na previsão de $x(t+6)$.

A aplicação das metodologias descritas à série caótica foi conduzida com base nos dados de identificação disponibilizados pelo "IEEE Neural Network Council, Standards Committee, Working Group on Data Modelling Benchmarks", os quais são também utilizados na análise de diversas metodologias. Com recurso às amostras obtidas, foi efectuada a mode-

lização da série caótica, de acordo com o procedimento descrito nas secções precedentes. Neste sentido, definiu-se o parâmetro r_a com o valor 0.5, de onde resultaram 9 regras difusas. Posteriormente, realizou-se o treino da rede neuro-difusa, com 4 entradas e 1 saída.

Deste modo, após a aplicação do conjunto de metodologias descrito, obtiveram-se os resultados da Tabela 1, onde CD, CC e CO1 denotam respectivamente, consequentes difusos, consequentes constantes e consequentes de ordem 1.

Os resultados apresentados sugerem algumas conclusões. Assim, em primeiro lugar a utilização de Gaussianas simples apresenta vantagens no caso de serem utilizados consequentes difusos. No entanto, da maior complexidade das funções Gaussianas generalizadas não resulta um ganho significativo em termos de precisão, pelo que funções Gaussianas simples são preferíveis. Quanto aos operadores difusos, os operadores algébricos originam melhores resultados que os de truncatura. Modelos com consequentes constantes possibilitam melhores resultados do que modelos com consequentes difusos, contudo, modelos do tipo Takagi-Sugeno de ordem 1 são os mais precisos, necessitando de um número significativamente menor de épocas de treino. No entanto, da otimização linear resultam tempos de processamento elevados, os quais poderão levantar problemas em termos de aplicabilidade em tempo real. Deste modo, modelos com consequentes constantes constituem, aparentemente, um bom compromisso entre precisão e eficiência.

Método	Gaus.	Nr. Par.	Op. Dif.	Nr. Ép.	Tempo p/Ép.	RMSE		
						Treino	Teste	
1	CD	Gen.	180	Alg.	2000	0.27s	0.0070	0.0076
2	“	“	“	Trunc.	“	0.24s	0.0111	0.0121
3	“	Simp.	90	Alg.	“	0.26s	0.0066	0.0071
4	CC	Gen.	153	Alg.	1500	0.54s	0.0047	0.0050
5	“	“	“	Trunc.	“	0.52s	0.0097	0.0108
6	“	Simp.	81	Alg.	“	0.53s	0.0050	0.0052
7	CO1	Gen.	189	Alg.	300	4.1s	0.0025	0.0030
8	“	“	“	Trunc.	“	4.3s	0.0038	0.0043
9	“	Simp.	147	Alg.	“	4.0s	0.0030	0.0033

Tabela 1. Série caótica: resultados de treino.

Na Figura 3 apresenta-se graficamente a saída relativa a dados reais e dados de teste para o método 1, a qual demonstra a elevada precisão do modelo.

5 CONCLUSÕES

Neste artigo foi efectuada uma análise comparativa de diferentes parametrizações e estruturas neuro-difusas. Pela aplicação do algoritmo de agrupamento subtractivo obteve-se uma estrutura inicial para o modelo difuso, a qual é utilizada na inicialização de uma rede neuro-difusa. Seguidamente, efectua-se o ajuste dos parâmetros do modelo difuso pelo treino da rede, de acordo com a estrutura definida, i.e., linguística ou Takagi-Sugeno de ordem 0 ou 1.

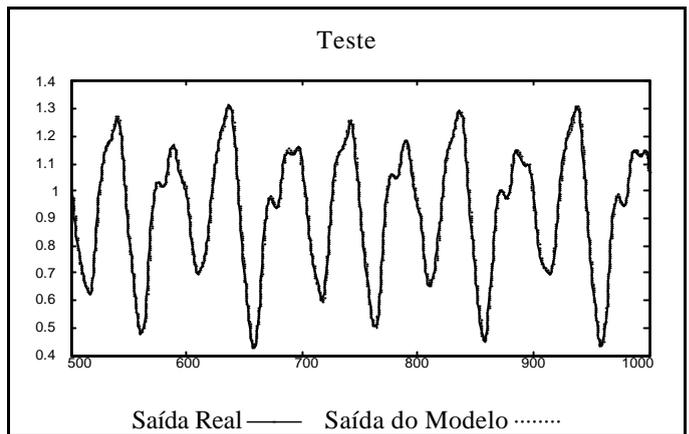


Figura 3. Série caótica: previsão da saída num modelo linguístico com operadores algébricos e funções Gaussianas generalizadas.

Nesse sentido foram utilizados operadores algébricos e de truncatura, bem como funções Gaussianas simples e generalizadas. As metodologias descritas foram aplicadas à série caótica Mackey-Glass, tendo-se concluído que modelos de ordem 1 são os mais precisos. No entanto, modelos de ordem 0 apresentam o melhor compromisso entre eficiência e precisão. Em termos de operadores, os algébricos possibilitaram resultados mais precisos. Quanto ao tipo de funções de pertença, concluiu-se que da complexidade adicional das funções generalizadas não resulta um ganho significativo, sendo preferíveis funções Gaussianas convencionais dada a sua simplicidade.

BIBLIOGRAFIA

- [1] Castro J. L. (1995). “Fuzzy logic controllers are universal approximators”, *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 25, No. 4, pp. 629-635.
- [2] Chiu S. L. (1994). “Fuzzy model identification based on cluster estimation”, *Journal of Intelligent and Fuzzy Systems*, Vol. 2, No. 3, pp. 267-278.
- [5] Lin C.- T. (1995). “A neural fuzzy control scheme with structure and parameter learning”, *Fuzzy Sets and Systems*, Vol. 70, pp. 183-212.
- [6] Ljung L. (1987). *System Identification - Theory for the User*, Prentice Hall, Englewood Cliffs.
- [7] Mackey M. C. e Glass L. (1977). “Oscillation and chaos in physiological control systems”, *Science*, vol. 197, pp. 287-289.
- [9] Paiva R. P. (1999). *Identificação Neuro-Difusa: Aspectos de Interpretabilidade*, Tese de Mestrado, Departamento de Engenharia Informática, Faculdade de Ciências e Tecnologia, Universidade de Coimbra.
- [10] Zadeh L A. (1973). “Outline of a new approach to the analysis of complex systems and decision processes”, *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 3, No.1, pp. 28-44.

AGRADECIMENTOS

Este trabalho foi financiado parcialmente pelo Ministério da Ciência e Tecnologia (MCT), sob o programa PRAXIS XXI.