1° Congresso Brasileiro de Redes Neurais

Escola Federal de Engenharia de Itajuba
Itajuba. 24 a 27 de outubro de 1994

Investigação de Aproximadores de Funções Através de Interpolação, Transformadas Ortogonais e Redes Neurais

João Fernando Marar[1,2]
Edson Costa de Barros Carvalho Filho[2]

[1]UNESP–Universidade Estadual Paulista
Faculdade de Ciências - Departamento de Computação
Bauru - São Paulo - Brasil

[2]UFPE–Universidade Federal de Pernambuco
Cx 7851. 50.732-970. Recife. PE. Brasil
{jfm,ecdbcf}@di.ufpe.br

Sumário. Atualmente. muitos trabalhos têm investigado a utilização de redes neurais multicamadas feedforward como aproximadores universais de funções. Neste artigo abordaremos as técnicas mais usuais em aproximadores de funções. incluindo interpolação. transformadas ortogonais e redes neurais multicamadas.

1 Introdução

Devido às dificuldades na manipulação de certas funções importantes que aparecem em problemas associados a *matemática aplicada*. torna-se necessária sua representação aproximada através de funções mais simples. de maneira que. estas últimas substituam as primeiras. O mesmo procedimento é utilizado quando é necessário trabalhar com funções cuja expressão analítica é desconhecida. *i.e.*. funções as quais se conhece somente alguns de seus valores através de determinações experimentais. Em geral. esta funções aparecem no cotidiano científico sob a forma de sequência de valores (tabelas). Nestes casos. estão inclusos problemas associados a *processamento de sinais: processamento e reconhecimento padrões entre outros*. uma das áreas de grandes aplicações destas técnicas. A utilização dos computadores digitais para o tratamento numérico destes problemas é inevitável. dada a grande massa de dados envolvida e a quantidade enorme de processamento aritmético exigido pelos algoritmos. para esta finalidade. Por exemplo. em síntese de instrumentos de orquestra. uma amostra de 1 segundo de um tom musical. requer acima de 40.000 valores reais [MAR 92].

Quando uma tabela é obtida por meio de uma expressão analítica ou determinada experimentalmente: esta função pode ser representada aproximadamente pela equação 1. por simplicidade estamos nos referindo a funções de uma única variável. contudo o raciocínio se extende para funções de várias variáveis.

$$f(x) \simeq v(x) = \sum_k a_k\, u_k(x) \qquad (1)$$

onde a_k são constantes e u_k podem ser funções tais como polinômiais. exponenciais. trigonométricas ou conjuntos de funções básicas ortogonais.

Caminhando de encontro ao exposto. este trabalho aborda algumas técnicas matemáticas utilizadas como aproximadores de funções. tais como a interpolação. transformadas ortogonais e por fim a apresentação de redes neurais que também são utilizada para este propósito. Este último tópico tem sido muito explorado por uma grande quantidade de pesquisadores. cujo objetivo principal é encontrar meios de formalizar problemas em redes neurais. de maneira os resultados obtidos venham a facilitar as futuras implementacões [FUN 89]. [HEC 89]. [HOR 89].[HAR 90].[LES 93].[MAR 94].

2 Interpolação

A interpolação é uma técnica muito utilizada quando somente possuímos uma tabela com valores de uma certa função para um conjunto de argumentos e, é necessário o cálculo, suficientemente preciso, de algum valor não tabelado, mas que é intermediário a valores consecutivos da mesma.

Como ocorre frequentemente em vários problemas experimentais, deseja-se o valor da função correspondente a um argumento que não figura na tabela nem é intermediário a valores consecutivos da mesma. Neste caso, é necessário fazer uma extrapolação a fim de aproximar o valor desejado. A extrapolação é menos precisa que a interpolação. Mas, ambas constituem, essencialmente, um único algoritmo de aproximação.

De maneira compacta, vamos definir intuitivamente o problema da interpolação. Suponha que seja fornecido uma sequência de valores, correspondentes a uma dada função. Nosso desejo é obter os valores de $f(\bar{x})$, onde $\bar{x}$ é um valor não tabelado. Para isso construímos, a partir desta sequência de valores, uma nova função $\psi(\bullet)$ que interpola $f(\bullet)$, tal que:

$$\forall x_i, x_0 \leq x_i \leq x_n \Rightarrow \psi(x_i) = f(x_i)$$

$$\forall x \in [x_0, x_n] \Rightarrow \psi(x) \cong f(x)$$

A função $\psi(\bullet)$ que interpola $f(\bullet)$ geralmente pertence a uma família de funções, dentre as mais importantes destacamos: as polinômiais, as exponenciais e as trigonométricas.

Um resultado de grande importância para a teoria da aproximação é o teorema de *Stone-Weierstrass*, que diz:

Se $f(x)$ é contínua em $[a, b]$, então para qualquer $\varepsilon > 0$ existe um polinômio $p_n(x)$ de grau n, tal que $| f(x) - p_n(x) | < \varepsilon$ para $a \leq x \leq b$.

A prova deste teorema pode ser vista em Handscomb [HAN 65]. Este teorema constitue uma boa razão para o grande uso dos polinômios, pois estamos interessados em que a aproximação para $f(\bar{x})$, onde $\bar{x}$ não está tabelado, seja suficientemente exata. Embora a prova deste teorema seja construtiva, o polinômio resultante costuma ser de grau elevado, de modo que tal polinômio passa a não ser prático. Além disso, o teorema não diz nada sobre a existência de um polinômio interpolante para um conjunto de dados $\{(x_0, y_0), (x_1, y_1), \ldots, (x_{n-1}, y_{n-1})\}$, qualquer.

Mesmo que decidamos interpolar n pontos pela abordagem polinomial, nem sempre é melhor usar o polinômio global, *i.e.*, de grau $n - 1$. Um bom exemplo dos problemas associados a divergência que ocorrem com interpolação polinômial pode ser visto através da *função de Runge* [ISA 66], [STE 50], [CLÁ 89].

3 Transformadas Ortogonais

As transformadas ortogonais constituem uma poderosa ferramenta matemática para representar funções extremamente complexas através de outras mais simples. Uma importante aplicação desta técnica é a redução de dimensionalidade ou compressão de dados [AHM 75],[WOM 77]. Estas transformadas podem ser divididas em duas classes, segundo as funções básicas utilizadas. Desta forma, temos a classe baseada em funções não senoidais e a de funções senoidais, tendo esta última como única representante a transformada de Fourier.

Em geral, aplica-se esta técnica a um conjunto de dados obtidos experimentalmente, cujo objetivo é encontrar uma lei de formação para a função. Os motivos que nos levam a não usarmos a interpolação são dois, em especial: (1) Devido a erros experimentais no conjunto de dados amostrados, não faz sentido calcular exatamente a função que origina os pontos e por isso, em vez de procurar uma função que interpola os n pontos dados, procuramos uma representação que melhor se ajuste aos pontos dados. Na realidade, o ajustamento traduz um comportamento médio e (2) para um conjunto com muitos elementos, a solução da aproximação dos dados por interpolação pode consumir muita memória e ou tempo de processamento.

Detalhes sobre a representação de funções através desta técnica serão expostas a seguir, bem como apresentaremos a transformada Karhunen-Loève, como uma das inúmeras representante da classe das transformadas ortogonais que são definidas através funções básicas não senoidais.

3.1 Aproximação de Funções via Transformadas Ortogonais

Seja $\Psi = \{\psi_0(t), \psi_1(t), ..., \psi_n(t), ...\}$ um conjunto de funções reais e contínuas,(utilizamos funções reais por conveniência), será dito ortogonal no intervalo $(t_0, t_0 + T)$ se:

$$\int_{t_0}^{t_0+T} \psi_m(t)\psi_n(t)dt = \begin{cases} c_n & \text{se } m = n \\ 0 & \text{se } m \neq n \end{cases} \qquad (2)$$

Para o caso onde $c_n = 1$, o conjunto Ψ é chamado ortonormal. Seja $x(t)$ uma função de valores reais. definida em um intervalo $(t_0, t_0 + T)$, e suponha que $x(t)$ possa ser escrita na forma :

$$x(t) = \sum_{n=0}^{\infty} a_n \psi_n(t) \qquad (3)$$

então os coeficientes a_n podem ser obtidos da seguinte forma : multiplicamos ambos os lados da equação 3.2 por ψ_m e integramos o resultado no intervalo $(t_0, t_0 + T)$, onde obteremos :

$$\int_{t_0}^{t_0+T} x(t)\psi_m(t)dt = \int_{t_0}^{t_0+T} \sum_{n=0}^{\infty} a_n \psi_n(t)\psi_m(t)dt$$

Como ψ_m e ψ_n são ortogonais, temos :

$$a_m = \frac{1}{c_n} \int_{t_0}^{t_0+T} x(t)\psi_m(t)dt, \; m = 0, 1, 2, ...$$

Um conjunto de funções ortogonais é chamado fechado ou completo se for verificada a seguinte condição :

Para qualquer parte contínua de $x(t)$ com :

$$\int_{t_0}^{t_0+T} x^2(t)dt < \infty$$

Qualquer que seja $\epsilon > 0$, existe N tal que seja possível representar uma aproximação de $x(t)$ por uma expansão finita :

$$\int_{t_0}^{t_0+T} | \, x(t) - \hat{x}(t) \, |^2 \, dt < \epsilon$$

onde :

$$\hat{x}(t) = \sum_{n=0}^{N-1} a_n \psi_n(t)$$

Pelo desenvolvimento acima. é visível que por uma expansão em funções ortogonais. sempre será possível representar $x(t)$ por um conjunto infinito,mas enumerável $\{a_0, a_1, a_2, ...\}$. Entretanto. quando Ψ for completo torna-se possível uma aproximação de $x(t)$ através de um conjunto finito $\{a_0, a_1, ..., a_{N-1}\}$.

3.2 Aproximação de Funções via Transformadas Karhunen-Loève

Seja $\{X\}$ um conjunto de vetores. obtidos por amostragem. Um representante de $\{X\}$ é dado por $x_j = (x_{j,1}, ..., x_{j,K})$. A amostra x_j pode ser aproximada por 4:

$$x_j = y_{j,1}\psi_1 + y_{j,2}\psi_2 + ... + y_{j,N}\psi_N = \sum_{i=1}^{N} y_{j,i}\psi_i \;\; N < K \qquad (4)$$

$$y_{j,i} = x_j^t \psi_i \quad i = 1, 2, ..., N \qquad (5)$$

onde K é o número total de componentes da amostra e N é o número de componentes utilizadas na aproximação.

Por definição, o mínimo erro quadrado. ε, é dado pela expressão 6:

$$\varepsilon = (\sum_{i=1}^{K} y_{j,i}\psi_i - \sum_{i=1}^{N} y_{j,i}\psi_i)^2 = \sum_{i=N+1}^{K} \psi_i^t R_X \psi_i \qquad (6)$$

onde R_X é a matriz de covariância do conjunto $\{X\}$.Dada por $R_x = \frac{1}{V}\sum_{j=1}^{V}(x_j - \bar{X})(x_j - \bar{X})^t$. onde V representa o número total de elementos do conjunto $\{X\}$ e $\bar{X}$ é o vetor médio do referido conjunto.

Quando $\{\psi_i\}$ constituem a base ortogonal de Karhunen-Loève, os elementos ψ_i são determinados a partir dos autovetores de R_X, de acordo com a equação 7:

$$R_X \psi_i = \lambda_i \psi_i \qquad (7)$$

O erro de truncamento da equação 6 é minimizado pela equação 8

$$Min(\varepsilon) = \sum_{i=N+1}^{K} \lambda_i \qquad (8)$$

Isto significa que, se utilizarmos apenas N auto-vetores para a representação de funções. o erro de truncamento será mínimo. sendo dado pela equação 8. A equação 4. escrita em termos dos auto-vetores da matriz de covariância, é denominada expansão Karhunen- Loève. A correspondente transformação ortogonal inversa, na equação 5, é chamada transformada Karhunen-Loève (K-L) [CHE 91].

4 Aproximação de Funções via Redes Neurais

Em 1969, Minsky e Papert demonstraram que as redes percepton com uma camada de neurônios com funções de ativação lineares ("*Single Layred Perceptron*"), são incapazes de aproximar quaisquer funções, no máximo elas poderiam mapear uma classe muito pequena de funções especiais, as linearmente separáveis [MIN 69]. Na época, Minsky e Papert sabiam que as redes multicamadas, com arquitetura feedforward eram capazes de proporcionar uma discriminação de funções não lineares, contudo, ainda era desconhecido um algoritmo de aprendizagem eficiente para estas redes. Efetivamente, as pesquisas com redes multicamadas foram reiniciadas nos meados da década de 80, mostrando que estas redes com a arquitetura feedforward podem ser utilizadas como aproximadores universais de funções [HOR 89], [HEC 89], definindo um mapeamento num espaço de dimensão finita para outro, através do algoritmo *Backpropagation* desenvolvido por Rumelhart et.al. [RUM 86].

As redes multicamadas perceptron, conhecidas por *MLP*, são sem dúvida nehuma o tipo mais popular de redes neurais e que podem ser aplicadas em aproximadores de funções. Os algoritmos de aprendizagem para este tipo de rede foram desenvolvidos independentemente por Parker em 1985 e Rumelhart et.al. [RUM 86]. Uma formulação geral, para as redes MLP com apenas uma camada escondida e uma saída, é dada pela equação 9:

$$f(\vec{x}) = h\Big(\sum_{i=1}^{k} c'_i \, h\Big(\sum_{j=1}^{d} c_{i,j} \, x_j + c_{i,d+1} \Big) + c'_{k+1} \Big) \qquad (9)$$

onde $\vec{x}$ é um vetor d-dimensional, correspondente a entrada,$h(\bullet)$ deve ser uma função suave (diferenciável), monotônicamente crescente. c_s e c'_s são coeficientes. k é o número de neurônios escondidos e d é a quantidade de elementos da entrada.

O algoritmo backpropagation é baseado no método do gradiente descendente [RUM 86],[HER 91], [HEC 89], o qual não garante chegar sempre ao erro global mínimo. Contudo, na grande maioria das vezes os resultados apresentados pelo algoritmo são bem próximos aos desejados.

Na prática, as funções de ativação utilizadas para as redes multicamadas são as funções sigmoides [FUN 89],[HEC 89],[SNA 90], tangentes hiperbólicas, gaussianas [HAR 90] e as funções básicas radiais [PAR 91],[SAH 90],[KAD 90], onde a quantidade de neurônios nas camadas escondidas são incrementados até que a precisão desejada seja alcançada na aproximação. Na realidade, a função complexa que uma rede *MLP* aproxima pode ser vista como composições e combinações das funções de ativação. Inclusive, a própria função analítica pode ser escrita, verificando a arquitetura da rede.

Os trabalhos [CUN 87] e [LAP 88], mostram que para adequar a aproximação de uma função desconhecida, utilizando funções do tipo *squashing* é necessário apenas duas camadas. Gallante e White [GAL 88], mostraram que uma particular rede feedforward, com apenas uma camada de funções cossenoides *squashing* é capaz de realizar a *transformada de Fourier*, a qual constrói a aproximação em série de Fourier para uma dada função. Esta rede possue todas as propriedades da representação em séries de Fourier, em particular, são apenas capazes de representar qualquer função quadrado integrável, sobre um conjunto compacto de funções, usando um número finito de unidades escondidas.

Hornik [HOR 89], faz uso do teorema de *Stone-Weirestrass* e as funções cossenos *squashing* de Gallant e White para estabelecer um padrão nas arquiteturas de redes feedforward. Hornik, usando funções *squashing* arbitrárias garante que estas podem aproximar qualquer função com qualquer grau de precisão, desde que exista a quantidade necessária de unidades escondidas disponíveis. Mas não faz mensão nenhuma sobre a quantidade de unidades necessárias para uma dada aproximação. Leshno [LES 93], garante, na mesma linha de raciocínio de Hornik que as funções básicas utilizadas como função de ativação nas redes multicamadas não podem ser polinômios, para que tais redes sejam aproximadores universais de funções.

5 Conclusões

Neste trabalho, apresentamos algumas técnicas mais usualmente utilizadas em aproximação de funções, entre as quais destacamos a interpolação, transformadas ortogonais e redes neurais.

As redes neurais tem mostrado ser uma técnica bastante atraente para a solução de diversos problemas, dentre os quais destacam-se o reconhecimento de padrões, aproximação de funções, entre outros. Embora ainda existam limitações para o seu uso, devido a falta de um formalismo na especificação e na análise dos modelos de redes neurais, *i.e* , a determinação de um modelo adequado para cada tipo de problema, uma vez que para se compreender os mecanismos fundamentais das redes é necessário realizar simulações que na maioria dos casos são tarefas árduas e distantes da realidade do modelo. Essa dificuldade se dá em virtude da complexidade da análise matemática envolvida; o que corresponde em redes neurais a complexidade dos mecanismos de aprendizagem, a estabilidade e a topologia adequada da rede. Esta complexidade matemática inviabiliza decidir qual o modelo mais adequado.

6 Agradecimentos

Ao Departamento de Computação da UNESP-Bauru, pelo afastamento concedido para a realização do programa de doutoramento no DI/UFPE.

À Capes-PICD e CNPq pelo apoio financeiro.

References

[AHM 75] N. Ahmed, K.R. Rao. *Ortogonal Transforms for Digital Signal Processing*, Springer-Verlang, New York, 1.975.

[CHE 91] C. S. Chen, K. S. Huo. *Karhunen-Loève Method for Data Compression and Speech Synthesis*, IEE Proceedings-I, Vol 138.Nro 5, October 1991, p 377-380.

[CLÁ 89] D. M. Cláudio, J. M. Marins, *Cálculo Numérico Computacional*. Editora Atlas S.A., São Paulo, 1.989.

[CUN 87] Y. L. Cun. *Modeles connexionistes de l'apprentissage*. These de Doctorat. Universite Pierre at Marie Curie. 1.987.

[FUN 89] K. Funahashi. *On the Approximate Realization of Continuos Mappings by Neural Networks*, Neural Networks. vol 2. 1.989. p 183-192.

[GAL 88] A. R. Gallant , H. White. *There exists a neural network that does not make avoidable mistables*. In IEEE Second International Conference on Neural Networks. San Diego. 1.988, p 657-664.

[HAN 65] D. C. Handscomb *Methods of Numerical Approximation*. Pergamon Press. Oxford. 1.965.

[HAR 90] E. J. Hartman, J. M. Keeler, *Layered Neural Networks with Gaussian Hidden Units as Universal Approximations*. Neural Computation 2, 1.990, p 210-215.

[HEC 89] R. Hecht-Nielsen *Theory of the Backpropagation Neural Network*, IJCNN 89, june, 1.989, p 593-605.

[HER 91] J. Hertz , A. Krogh , R. G. Palmer *Introduction to the Theory of Neural Computation*. Addison-Wesley Publishing Company. 1.991.

[HOR 89] K. Hornik, *Multilayer Feedforward Networks are Universal Approximators*, Neural Networks, vol 2, 1.989, p 359-366.

[ISA 66] E. Isaacson, H. B. Keller, *Analysis of Numerical Methods*, John Wiley & Sons, Inc. New York, 1.966.

[KAD 90] V. Kadirkamanathan et.al. *Sequential Adapitation of Radial Basis Function Neural Networks and its application to Time-series Prediction*,

[LAP 88] A. Lapedes, R. Faber *How neural networks work*, Tech. Report LA-UR-88-418. Los Alamos National Laboratory, 1.988.

[LES 93] M. Leshno et. al. *Multilayer Feedforward Networks with a Nonpolynomial*

Activation Function can Approximate any Function. Neural Networks, vol 6., 1.993, p 861-867.

[MAR 92] J. F. Marar *Utilização da Transformada Karhunen-Loève em Síntese de Tons Musicais*, Dissertação de Mestrado -USP-São Carlos, 1.992.

[MAR 94] J. F. Marar, E.C.d.B.C. Filho *Aproximadores de Funções* , Relatório Técnico, UFPE-DI, Março,1994.

[MIN 69] M. Minsky , S. Papert *Perceptrons*, MIT Press, 1.969.

[PAR 91] J. Park.I. W. Sandberg *Universal Approximation Using Radial Basis Function*, Neural Computation, nro 3, 1.991, p 246-257.

[RUM 86] D. E. Rumelhart , J. L. McClelland *Parallel Distributed Processing-Vol 1: Foundations*,The MIT Press, Cambridge,1986.

[SAH 90] A. Saha *Oriented Non-Radial Basis Functions for Image Coding and Analysis*, 1.990.

[SNA 90] R. R. Snapp et.al. *Generalizing Smoothness Constraints From Discrete Samples*, Neural Computation 2, 1.990, p 188-197.

[STE 50] J. F. Steffensen *Interpolation*, Chelsea Publishing Company, New York , 1.950.

[WOM 77] M. E. Womble, J. S. Halliday, S. K. Mitter, M. C. Lancaster, J. H. Triebwasser *Data Compression for Storing and Transmitting*, Proceedings of the IEEE, vol 65, Nro 5, May 1977, p 702-706.