

1º Congresso Brasileiro de Redes Neurais

Escola Federal de Engenharia de Itajubá
Itajubá, 24 a 27 de outubro de 1994

Soluções Conexionistas Híbridas para o Reconhecimento de Padrões Unidimensionais e Bi-dimensionais

Solange Sari
Claudio Loesch
Ricardo Miranda Barcia

- Professora do Departamento de Computação na UNISUL - Universidade do Sul de Santa Catarina (Tubarão - SC). e-mail: eps3ssa@brufsc.bitnet.
- Professor do Departamento de Matemática na FURB - Universidade Regional de Blumenau (Blumenau - SC). e-mail : furb@brufsc.bitnet.
- Professor do Curso de Pós-Graduação em Engenharia de Produção na UFSC - Universidade Federal de Santa Catarina (Florianópolis - SC). e-mail: eps1rmb@brufsc.bitnet.

Resumo

Um dos problemas mais genéricos de reconhecimento de padrões envolve a situação em que um ou mais padrões podem simultaneamente encontrar-se contidos dentro de um domínio discreto, gradeado e limitado de \mathcal{R} ou de \mathcal{R}^2 . Este trabalho considera o reconhecimento independente das características de translação, escala ou rotação no plano. A abordagem combinatorial, torna-se impraticável, mesmo para problemas de pequeno porte devido à explosão combinatorial de possibilidades geradas por estas características. Propõe-se explorar a solução do problema através da implementação de uma rede neural híbrida caracterizada por três fases de transformação de dados: segmentação da imagem (com filtragem de ruídos e detecção de agrupamento), processamento do agrupamento enfocado e o posterior reconhecimento do padrão.

Introdução

As redes de alimentação para frente, com múltiplas camadas, e pesos de conexões ajustáveis, comumente via o algoritmo de aprendizagem 'backpropagation', tem sido apontados pela literatura como um potente instrumento para classificação de padrões. Tal arquitetura, com n elementos na camada de entrada, e m elementos na camada de saída pode ser vista como uma aplicação unívoca

$$\phi: \mathcal{R}^n \rightarrow \mathcal{R}^m \quad (1)$$

A capacidade de aprendizagem de uma rede neural, em relação a um padrão, é o quanto a resposta da rede encontra-se próxima da saída desejada (um padrão pré-determinado). Dizer que uma rede foi capaz de aprender um determinado padrão p , na forma de um par $(x_p, d_p) \in \mathcal{R}^n \times \mathcal{R}^m$, significa que, dado um real $\varepsilon > 0$, pode-se tornar $|\phi(x_p) - d_p| \leq \varepsilon$. Normalmente, não pode-se exigir do aprendizado que

$\phi(x_p) = d_p$, porém pode-se arbitrar uma margem de imprecisão pequena o suficiente para propósitos práticos.

Alguns resultados já foram estabelecidos com relação à capacidade de aprendizagem de uma rede. Tem-se como estabelecido o seguinte [1]:

1. Se uma aplicação consiste de uma coleção finita de pontos, uma rede de 3 camadas (1 camada oculta) é capaz de aprende-la.
2. Se uma aplicação é contínua e definida sobre um domínio compacto, uma rede de três camadas é capaz de aprende-la..
3. Sob condições muito gerais, todas as aplicações que podem ser aprendidas por uma rede neural, podem ser aprendidas por uma rede de 4 camadas (2 ocultas).

Estas condições apenas exigem que se tenha um número suficiente de elementos na camada oculta. Além disso, foi verificado que muitas funções que não seguem os critérios acima, podem também ser aprendidas por uma rede de 3 camadas. Em particular, descontinuidades podem ser teoricamente toleradas sob todas as condições prováveis de serem encontradas na prática.

Estas condições cobrem a maioria dos problemas práticos. A grande dificuldade consiste no treinamento para o estabelecimento dos pesos. Embora teoricamente qualquer par de padrões a ser reconhecido possa ser treinado, quando a rede tiver que abstrair muitas características simultaneamente, o aprendizado fica comprometido em termos de eficiência e velocidade. Para o reconhecimento de padrões na forma geral proposta, a rede deve ser invariante a translações, efeitos de escala e rotações no plano. Além disso, em muitos casos, existe a possibilidade do contorno do objeto focalizado não estar bem definido, de

acordo com o enfoque da grade sobre a imagem. A rede deveria também ser invariante a esta possível característica.

A prática tem demonstrado poucos resultados otimistas em relação ao treinamento de todas estas características. Isto exige normalmente um conjunto de treinamento muito grande e assim a velocidade de convergência fica muito comprometida, com grandes possibilidades de cair em mínimos locais ou problemas de paralisia da rede.

Diversos pesquisadores desempenharam esforços no sentido de resolver esta questão, procurando adequar o tratamento do reconhecimento de padrões, aproveitando o lado positivo das capacidades destas redes e embutindo outras formas auxiliares de tratamento que possam amenizar o trabalho de abstrair tantas características simultaneamente. Pao [3], sugere a introdução de redes de ligações funcionais como uma forma alternativa equivalente ao trabalho desempenhado pelas camadas ocultas da rede: o de abstrair características de padrões. A mais competente razão para a escolha deste modelo é a do conhecimento a priori de certas transformações funcionais em que os dados irão expor suas características salientes. Spirkovska e Reid [2] aperfeiçoaram uma arquitetura de redes neurais de ordem mais elevada (HONNs), ou redes pi-sigma, cujo pré-processamento consiste de somas de produtos de elementos de entrada. Restringindo a ordem dos produtos a três, obtiveram sucesso em classificar padrões binários com 100% de acerto, invariantes em translação e rotação e escala, em quadros de pixels de até 128x128.

Filtragem de Ruídos e Detecção de Agrupamentos

A segmentação da imagem, com filtragem de ruídos e detecção de agrupamentos, é

executada por uma rede neural cuja arquitetura consiste de camadas conectadas lateralmente, como mostra a figura 1.

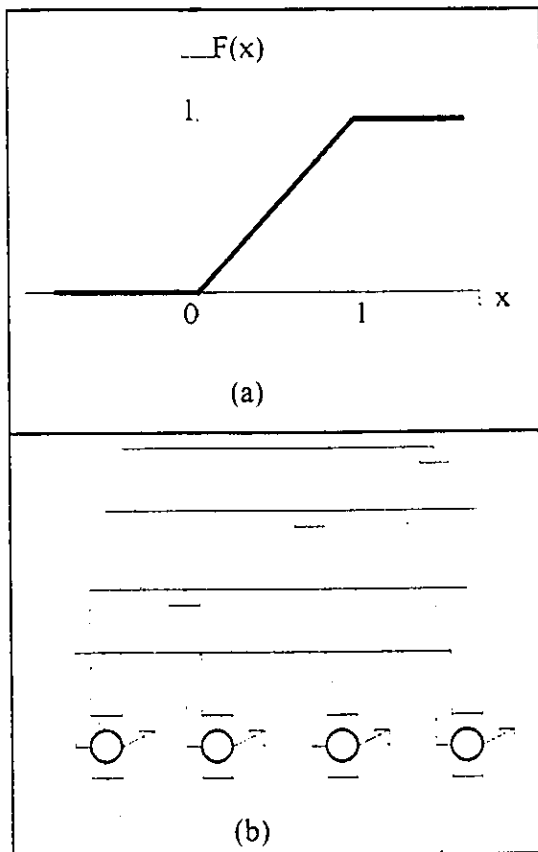


Figura 1. (a) Função de transferência do elemento de processamento. (b) Topologia da rede.

A dinâmica da rede consiste de processar o padrão de entrada, sendo que a saída de cada elemento de processamento alimenta as entradas dos demais elementos. Nesta camada, os elementos de processamento cooperam e inibem entre si os sinais de entrada, de acordo com o sinal algébrico dos pesos das conexões, iterativamente, até alcançarem um estado de estabilidade. A equação de iteração entre os elementos é:

$$x_i^{(k-1)} = F \left(x_i^{(k)} + \frac{\sum_j x_j^{(k)} \cdot w_{ij}}{\sum_j x_j^{(k)}} \right) \quad (2)$$

onde:

k é o índice da iteração;

$x_i^{(k)}$ é o valor do sinal em processamento;

w_{ij} é o peso da conexão da saída do elemento i para a entrada do elemento j;

F é a função de transferência.

Para prevenir crescimentos de valores absolutos numéricos de forma desordenada e não-limitada, escolhem-se para F funções que mantem os limites superiores e inferiores de seus argumentos em um intervalo. Uma solução simples consiste em adotar a equação (3).

$$F(x) = \begin{cases} 0 & \text{se } x < 0 \\ x & \text{se } 0 \leq x \leq 1 \\ 1 & \text{se } x > 1 \end{cases} \quad (3)$$

Escolhendo adequadamente os pesos das conexões, a convergência de (3) fica assegurada para isto. Devem-se escolher pesos que refletem característica de distribuição espacial e características comuns a todos os padrões passíveis de serem classificados. Por exemplo: se os padrões são conexos no domínio, escolhem-se pesos positivos para as conexões de cada elemento com seu espacialmente vizinho e, eventualmente, para os próximos vizinhos. No entanto, pesos negativos de efeito inibitório devem encontrar-se presentes, para que exista efeito competitivo, pois caso contrário, o vetor converge com todos os valores para a unidade.

Desta forma, valores maiores tem a tendência de se fortalecerem ainda mais, e inibir outros eventuais agrupamentos a serem formados. De acordo com a força inibitória, um ou mais agrupamentos pode sobreviver. Estes são identificados como unidades dentro do vetor (que representa o padrão); isto permite isolar um agrupamento para análise individual. Ruídos, eventualmente fracos, deverão ser eliminados pelo processo competitivo.

Processamento do Agrupamento

A transformada de Fourier opera sobre valores contínuos, e permite discretização preservando as propriedades fundamentais dos dados, o que torna viável seu uso computacional. Dado um vetor $x = (x_1, x_2, \dots, x_N)$, a transformada de Fourier discreta de x resulta num vetor X cujas componentes, em igual número, podem ser calculada por (4), ou por métodos mais eficientes, como por exemplo a transformada rápida de Fourier [4]. Observe-se que o vetor da transformada discreta é complexo.

$$X_n = \frac{1}{N} \sum_{k=0}^{N-1} x_k e^{-j2\pi nk/N}, \quad (4)$$

para $0 \leq n \leq N-1$
onde

- j é a unidade imaginária;
- N é o tamanho do vetor;
- x_k é o elemento do vetor de entrada
- X_n é o elemento no vetor de entrada

A propriedade mais importante, neste caso, refere-se ao deslocamento. Se x tem X por transformada, então, para uma translação inteira h , o vetor $y_k = x_{k-h}$ tem por transformada a equação que segue:

$$Y_n = X_n \cdot e^{-j2\pi nh/N} \quad (5)$$

Desta forma, $|Y| = |X|$, ou seja, os vetores possuem a mesma amplitude. Esta propriedade, se aproveitada, confere à rede híbrida o aspecto de "functional link" mencionada por [3].

Classificação do Padrão

A classificação é obtida a partir de uma rede neural, com uma arquitetura de três camadas e alimentação para frente, treinada segundo o algoritmo de

aprendizagem Backpropagation[5]. Treina-se a rede de acordo com as amplitudes das transformadas de Fourier discretas do conjunto de padrões de possível classificação. Desta forma, a rede responde aos dados provenientes do pré-processamento do agrupamento. De acordo com a propriedade de invariância da amplitude da transformada, a rede neural não necessitará aprender a abstrair esta característica.

Exemplificando a dinâmica da rede neural híbrida

Fase 1. Parte-se de um conjunto de padrões bi-dimensionais, com valores entre 0 e 1, os quais representam três classes distintas (figura 2). Estes padrões são apresentados a primeira arquitetura de rede neural, a qual produz padrões de saída mostrados na figura 3.

0.90	0.80	0.20	0.12
0.80	0.19	0.07	0.06
0.42	0.12	0.11	0.03
0.13	0.10	0.09	0.05

0.07	0.05	0.10	0.09
0.11	0.80	0.20	0.19
0.09	0.87	0.18	0.15
0.95	0.20	0.09	0.07

0.88	0.10	0.09	0.17
0.78	0.02	0.21	0.12
0.89	0.07	0.05	0.28
0.99	0.10	0.21	0.08

Figura 2. Padrões que caracterizam três classes diferentes: Classe I, Classe II e Classe III, respectivamente.

1	1	0	0
1	1	0	0
0	0	0	0
0	0	0	0

0	0	0	0
0	1	0	0
1	1	0	0
1	1	0	0

1	0	0	0
1	0	0	0
1	0	0	0
1	0	0	0

Figura 3. Detecção de agrupamentos, conforme padrões da figura anterior.

Fase 2. A área de interesse de cada padrão é obtida através do produto entre do padrão de entrada e o padrão de saída da fase anterior, como mostra a figura 4. Estes novos padrões são agora apresentados a uma camada funcional, a qual produz novos padrões de saída (figura 5).

0.90	0.80	0	0
0.80	0.19	0	0
0	0	0	0
0	0	0	0

0	0	0	0
0	0.80	0	0
0.09	0.87	0	0
0.95	0.20	0	0

0.88	0	0	0
0.78	0	0	0
0.89	0	0	0
0.99	0	0	0

Figura 4. Focalização da área de interesse.

2.69	1.96	0.71	1.96
1.96	1.75	0.61	1.09
0.71	0.61	0.51	0.61
1.96	0.19	0.61	1.75

2.91	2.13	0.83	2.13
1.02	1.94	1.73	0.51
0.99	0.86	0.73	0.86
1.02	0.51	1.73	1.94

3.54	3.54	3.54	3.54
0.21	0.21	0.21	0.21
0.00	0.00	0.00	0.00
0.21	0.21	0.21	0.21

Figura 5. Amplitudes da Transformada de

Fase 3. As amplitudes das transformadas de Fourier de cada padrão são apresentadas a outra arquitetura de rede

neural, a qual encarrega-se da classificação. Como o exemplo, tomamos como conjunto de treinamento os padrões mencionados nas figuras anteriores, e como conjunto de teste os padrões representados na figura 6.

(a)

0.15	0.09	0.07	0.10
0.06	0.11	0.05	0.21
0.09	0.12	0.17	0.83
0.08	0.06	0.78	0.95

0.09	0.12	0.07	0.98
0.07	0.10	0.92	0.92
0.21	0.07	0.87	0.11
0.03	0.06	0.05	0.09

0.03	0.12	0.09	0.79
0.02	0.11	0.12	0.92
0.23	0.12	0.07	0.89
0.05	0.15	0.13	0.77

(b)

0	0	0	0
0	0	0	0
0	0	1	1
0	0	1	1

0	0	0	1
0	0	1	1
0	0	1	0
0	0	0	0

0	0	0	1
0	0	0	1
0	0	0	1
0	0	0	1

(c)

2.73	2.01	0.83	2.01
1.99	1.78	0.68	1.12
0.73	0.62	0.49	0.62
1.99	1.12	0.68	1.78

2.86	2.08	0.72	2.08
1.01	2.12	2.02	0.78
0.84	0.89	0.94	0.89
1.01	0.78	2.02	2.12

3.37	3.37	3.37	3.37
0.18	0.18	0.18	0.18
0.01	0.01	0.01	0.01
0.18	0.18	0.18	0.18

Figura 6. (a) Padrão de entrada da rede neural híbrida (normalizados). (b) Agrupamentos obtidos. (c) Amplitudes para classificação.

San Diego, CA: June 17-21, 1990.
I-21-I-26.

O resultado da classificação é mostrado no quadro 1 a seguir:

Saída Obtida			Saída Desejada		
1.00	0.00	0.00	0.95	0.04	0.02
0.00	1.00	0.00	0.05	0.95	0.02
0.00	0.00	1.00	0.03	0.03	0.95

Quadro 1. Padrões de saída da rede neural de classificação, referentes as três classes distintas: [100] - Classe I, [010] - Classe II e [001] - Classe III.

Conclusões

As literatura atual sobre modelos conexionistas e a necessidade de encontrar uma maneira de reconhecer padrões complexos (uni e bi-dimensionais), levaram a construção de uma rede neural híbrida, ou seja uma rede neural composta de diversas arquiteturas. Com os testes verificou-se que a distribuição de tarefas entre as redes provoca melhorias no reconhecimento de padrões. Como mostra o exemplo, a rede é capaz de reconhecer padrões com 100% de acerto, independente da dimensão e posição espacial.

Referências

[1] MASTERS, Timothy. Practical Neural Networks Recipes in C++. San Diego, CA: Academic Press, 1993.
[2] SPIRKOVSKA, L. e REID, M. Connectivity Strategies for Higher-order Neural Networks Applied to Pattern Recognition. Proc Third Int'l Joint Conf. Neural Networks.

[3] PAO, Yoh-Han. Adaptative Pattern Recognition and Neural Networks. Reading: Addison-Wesley, 1989.

[4] BRIGHAM, E. Oran. The Fast Transform and its Applications. Englewood Cliffs, New Jersey: Prentice-Hall, 1988.

[5] RUMELHART, D. E., HINTON, G. E. e WILLIAMS, R. J.. Learning Internal Representations by Error Propagation. In D.E. Rumelhart and J.L. McClelland (Eds.), Parallel Distributed Processing: Explorations in the Microstructures of Cognition. Vol. 1: Foundations, pp.318-362, MIT Press, Cambridge, CA, 1986.