

Instituto de Ciências Matemáticas de São Carlos

ISSN - 0103-2569



DEDALUS - Acervo - ICMC



30300046469

**RECONHECIMENTO DE SILHUETAS DE
VEÍCULOS USANDO COMBINAÇÃO DE
CLASSIFICADORES**

**PAULO SÉRGIO PRAMPERO
ANDRÉ C.P.L.F. CARVALHO**

Nº 58

RELATÓRIOS TÉCNICOS DO ICMSC

São Carlos
Jul./1997

Class.	RT-SCE-m58
Cutter	Pegg
Tombo	21.563
Sysno	1587071

Reconhecimento de silhuetas de veículos usando combinação de classificadores.

Instituto de Ciências Matemáticas de São Carlos

Universidade de São Paulo

Junho de 1997

Paulo Sérgio Prampero

Andre C. P. L. F. Carvalho

email : prampero@icmasc.sc.usp.br

email : andre@icmasc.sc.usp.br

Resumo

Este artigo investiga a de Combinação de Redes Neurais para Reconhecimento de Padrões. Para isso, três redes neurais foram treinadas para reconhecer veículos utilizando um arquivo contendo 18 características de silhuetas de veículos. Depois das redes treinadas foi obtido os resultados individuais das redes neurais, esses resultados foram utilizados para a combinação dos classificadores. Dois tipos de combinação foram utilizados: combinação hierárquica e combinação pela maioria. Depois de todos os resultados obtidos foi feita uma comparação entre os resultados das estruturas de combinação. Os experimentos foram realizados seguindo a metodologia proposta em Proben1.

Palavras Chave: Reconhecimento de Padrões, Redes Neurais e Combinação de Classificadores.

1. - Introdução

“Re-conhecer” significa, primeiro gerar “conhecimento” a partir de dados, e depois, “re-identificar” dados já conhecidos. Para realizar essas tarefas, muitas pesquisas foram feitas com diferentes técnicas em diferentes abordagens.

Uma dessas técnicas é a Combinação de Classificadores, na qual cada classificador classifica um conjunto de dados através do seu método particular, e depois dos resultados obtidos realiza-se a combinação. Com isso, espera-se que os classificadores escolhidos possam classificar corretamente padrões diferentes. Esta esperança tem fundamento, pois recentemente foi observado que os classificadores de diferentes tipos complementam a performance de outros classificadores [Ho, 94].

A vantagem de classificadores diferentes classificar corretamente padrões diferentes é que através da combinação dessas classificações possamos unir conjuntos corretos complementares de classificação, obtendo assim uma percentual maior de acerto.

Porém a única vantagem concreta da combinação para problemas genéricos é que a combinação de classificadores produz respostas mais estáveis no sistema de classificação. Isso já é uma boa motivação para as pesquisas de combinação.

O problema tratado neste artigo é o reconhecimento de veículos. Esse reconhecimento é baseado nos dados extraídos da silhueta dos veículos. Esta base está disponível na coleção de *benchmarks UCI* (disponíveis no endereço URL: <http://www.ics.uci.edu/ai/ml/mldbrepository.html>) e foi utilizada no projeto *StatLog* (disponível no endereço URL: <http://www.ncc.up.pt/liacc/ml/statlog/>).

O sistema de reconhecimento enfoca a abordagem conexionista utilizando as Redes Neurais Artificiais. Nas Redes Neurais foi escolhida a rede *multi-layer perceptron* através de um teste inicial de desempenho com os dados do problema em questão. Três algoritmos foram

utilizados para o treinamento da rede : *standard backpropagation*, *rpropagation*, *simulated annealing*.

A segunda etapa do sistema é a Combinação dos Classificadores. Dois métodos foram utilizados para a combinação: o método pelo voto e o uma estrutura hierárquica de combinação na qual a combinação é feita por uma outra rede neural.

2 - Redes Neurais Artificiais

Redes Neurais Artificiais (RNA) são estruturas distribuídas formadas por um grande número de unidades de processamento (neurônios), bastante simplificadas e conectadas entre si. Essas conexões são implementadas por canais de comunicação que estão associados a um determinado peso. As unidades realizam operações apenas sobre seus dados locais, que são entradas recebidas através das suas conexões.

As RNAs são novos modelos de computação com características particulares, tais como: Capacidade aprender, generalizar, agrupar ou organizar dados; Não linearidade; Mapeamento entrada saída; Adaptatividade; Respostas com níveis de confiança; Informação contextual; Tolerância a falhas; Uniformidade de análise e *design*; Analogia neurobiológica.

A origem do poder das RNA está baseado na sua estrutura distribuída e massivamente paralela e na sua capacidade de aprender e generalizar. O comportamento inteligente das redes neurais vem das iterações entre as unidades de processamento da rede.

A maioria dos modelos de redes neurais possui alguma regra de treinamento, onde os pesos de suas conexões são ajustados de acordo com os padrões apresentados. Em outras palavras, elas aprendem através de exemplos.

Redes neurais têm um grande potencial de paralelismo e os cálculos realizados por uma unidade dependem apenas dos cálculos feitos pelas unidades antecessoras. Ao invés de executar um programa de instruções sequenciais, como em um computador de *Von Neumann*, uma rede

neural responde às entradas que lhe são apresentadas em paralelo. Os resultados não são armazenados em uma determinada posição de memória, mas são dados pelo estado geral da rede, depois que ela atinge uma condição de equilíbrio, a convergência. Assim, o conhecimento está na forma com que as unidades estão interconectadas e na importância (peso) de cada entrada em uma determinada unidade [Haykin, 94].

Arquiteturas neurais são tipicamente organizadas em camadas, com unidades que podem estar conectadas às unidades das camadas anterior e posterior, como é ilustrado na Figura 1. Usualmente as camadas são classificadas em dois grupos:

- **Camadas Intermediárias ou Escondidas:** onde é feita a maior parte do processamento, através das conexões ponderadas; podem ser consideradas como extratoras de características.
- **Camada de Saída:** onde o resultado final é calculado e apresentado.

Alguns autores mencionam uma camada de entrada, mas como esta camada não realiza nenhum processamento, não será considerada neste trabalho.

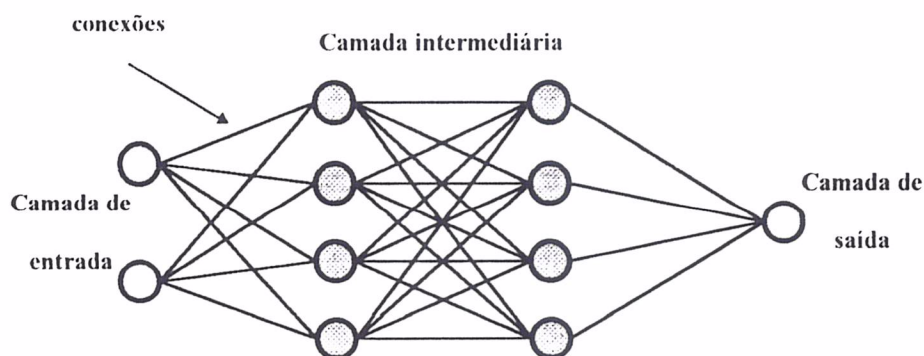


Figura 1 - Organização em camadas.

Uma rede neural é especificada principalmente pela sua topologia, pelas características dos nós e pelas regras de treinamento. O processo de aprendizado varia de acordo com a topologia da rede.

3 - Combinação de Classificadores

Sem analisar a estrutura interna do classificador e decidir sobre qual metodologia ele é baseado, pode-se simplesmente considerar um classificador como uma caixa funcional a qual recebe um exemplo de entrada x e retorna uma classe S , ou seja, $C(x) = S$.

O problema da combinação de classificadores é gerar a saída desta combinação, pois cada classificador individual C_i fornece uma saída S_i para o padrão x , ou seja, $C_i(x) = S_i$. O principal problema é a partir das saídas individuais S_i , que podem ser contraditórias, gerar uma única saída final.

Recentemente foi observado que os classificadores de diferentes tipos complementam a performance de outros classificadores [Ho, 94]. Esta afirmação introduziu idéias para o uso de diferentes tipos de classificadores simultaneamente, o que pode melhorar a exatidão da classificação. Contudo, a combinação de decisões podem ser conflituosas. Este conflito de decisão traz um novo problema à classificação: escolher a decisão correta.

Idealmente, a combinação de classificadores tem vantagens sobre o uso de um classificador individual, pois a classificação é mais estável que aquela fornecida por um classificador individual. Esta estabilidade tem origem na combinação, pois muitas decisões são mais estáveis que apenas uma decisão.

A combinação deve exaltar as boas características de cada classificador e reprimir suas falhas, com isso a performance do sistema poderá ser melhor do que a performance de cada classificador envolvido no sistema.

A principal desvantagem da combinação de classificadores é o aumento da complexidade do problema. No uso de classificadores individuais, não existem os problemas de combinar estruturas de classificação e combinar suas saídas. Outro problema é a soma do erro da combinação, que não deve ser a soma dos erros de cada classificador do sistema, pois desta forma, a combinação estaria exaltando o erro ao invés da boa performance.



4 - Experimento

Os experimentos realizados podem ser decompostos em várias partes. Nas subseções a seguir discutiremos cada uma dessas partes do experimento.

4.1 - Dados

Os dados utilizados nesse artigo são formados por um conjunto de características extraídas de silhuetas de veículos. Cada padrão desse arquivo é formado por um conjunto de 18 atributos, tais como: compactivo (média do perímetro $\sqrt{2}/\text{área}$), circularismo (média do raio $\sqrt{2}/\text{área}$), e outras medidas relevantes.

Para obter essas características foi feita uma captação de imagens. Essa captação foi realizada por uma câmera fixa a uma altura determinada, formando assim um ângulo fixo em relação ao veículo. Várias imagens foram geradas rotacionando os veículos, e a cada rotação do veículo uma nova captação acontecia.

As características foram extraídas das imagens utilizando o HIPS (*Hierarchical Image Processing System*), o qual extrai uma combinação de características utilizando medidas como variância e heurísticas.

Os dados são provenientes da silhueta de quatro tipos de veículos, *um double decker bus*, *um Chevrolet van*, *um Saab 9000* e *um Opel Manta 400*. O número de padrões obtidos foram 946, com 240 padrões do tipo OPEL, 240 padrões do SAAB, 240 padrões do BUS e 226 padrões do VAN.

O conjunto de dados da extração foi dividido em três grupos de conjuntos treinar, validar e testar. Cada grupo treinar_validar_testar foi gerado aleatoriamente à partir do mesmo conjunto inicial de características.

Os resultados apresentados nesse artigo são a média aritmética dos resultados obtidos com cada grupo de arquivo treinar_validar_testar.

4.2 - As Redes

A rede *perceptron* multi-camadas foi escolhida para esse estudo, pois é muito utilizada em problemas de reconhecimento de padrões. A escolha dos algoritmos a serem combinados foi baseada no desempenho e nas características de cada algoritmo. Várias estruturas de rede foram testada para cada algoritmo de treinamento. Todas estruturas possuem 18 entradas e 4 neurônios de saída, as estruturas treinadas foram : 10 neurônios na primeira camada, 20 neurônios na primeira camada, 30 neurônios na primeira camada, 30 neurônios na primeira camada e 5 neurônios na segunda, 30 neurônios na primeira camada e 10 neurônios na segunda e, finalmente, 30 neurônios na primeira camada e 15 neurônios na segunda.

Cada algoritmo de treinamento teve melhor desempenho em uma das estruturas redes anteriores. Todos os algoritmos de treinamento tiveram o mesmo algoritmo de inicialização (*Randomize_Weights* [0.5,-0.5]) e o mesmo algoritmo de atualização (*Topological_Order*). Os resultados abaixo são provenientes do melhor resultado obtido em no máximo 10.000 ciclos.

4.2.1- Standard Backpropagation

O algoritmo *standard backpropagation* é o algoritmo de aprendizado mais comum. Neste algoritmo a saída da última camada é comparada com a saída desejada obtida no arquivo de treinamento. O erro é a diferença entre a saída desejada e a saída obtida. O erro é *propagado*

ajustando os pesos das conexões da rede. A cada novo padrão de treinamento existe um novo cálculo do erro e uma nova *retropropagação* atualizando os pesos das conexões.

A estrutura da melhor rede treinada por este algoritmo tinha 18 entradas, 30 neurônios na primeira camada, 10 neurônios na segunda camada e 4 neurônios de saída. Os resultados estão na tabela abaixo.

	Treinamento	Validação	Teste
Errado	1.18%	7.74%	7.54%
Certo	98.35%	89.26%	90.1%
Desconhecido	0.47%	3%	2.36%

Tabela 1 - Tabela de resultados em porcentagem do algoritmo standard backpropagation.

4.2.2- Resilient Backpropagation

Resiliente Backpropagation, *rprop*, é uma técnica de aprendizado adaptativo local com paradigma supervisionado de aprendizado. O princípio básico do algoritmo *Rprop* é eliminar o dano causado pelo tamanho da derivada parcial sobre o passo de atualização do peso. Como consequência, somente o sinal da derivada é utilizado para indicar a direção de atualização do peso. O tamanho da mudança é exclusivamente determinado pelo valor de atualização [Braun, 93; Riedmiller, 94].

A estrutura da melhor rede treinada por este algoritmo tinha 18 entradas, 30 neurônios na primeira camada, 15 neurônios na segunda camada e 4 neurônios de saída. Os resultados estão na tabela abaixo.

	Treinamento	Validação	Teste
--	-------------	-----------	-------

Errado	0.08%	5.53%	4.40%
Certo	99.13%	88.15%	88.84%
Desconhecido	0.79%	6.32%	6.76%

Tabela 2 - Tabela de resultados em porcentagem do algoritmo rprop.

4.2.3 - Simulated Annealing

Simulated annealing é um dos métodos mais sofisticados para encontrar o mínimo global de superfícies de erro. Nesse algoritmo somente um peso ou *bias* é mudado em cada ciclo. Uma das vantagens do simulated annealing é que o aprendizado não se prende em mínimos locais.

Cada ciclo de treinamento muda-se um peso ou um *bias* randomicamente na faixa [min,max], calcula-se o erro da rede e aceita-se a mudança caso o erro diminuiu.

A estrutura da rede utilizada por este algoritmo foi 30 neurônios na primeira camada, 15 neurônios na segunda camada e 4 neurônios de saída. Os resultados estão na tabela abaixo.

	Treinamento	Validação	Teste
Errado	7.01%	9.64%	7.39%
Certo	42.40%	42.34%	43.71%
Desconhecido	50.59%	48.03%	48.90%

Tabela 3 - Tabela de resultados em porcentagem do algoritmo Simulated annealing.

4.3 - A Combinação

Para a combinação dos classificadores dois métodos foram utilizados, os quais estão descritos abaixo. Uma vantagem da combinação é a flexibilidade de comparação de métodos. Com a combinação pode-se comparar os algoritmos estudados de uma forma mais abrangente do que apenas estudar os algoritmos de treinamento individualmente.

4.3.1- Combinação Hierárquica

No sistema hierárquico a decisão da resposta escolhida foi feita por uma outra Rede Neural. Com saídas das rede descritas anteriormente mais a saída desejada foi gerado o conjunto de treinamento da rede “juiz”. Cada rede inicial tinha uma saída de tamanho 4, portanto a conjunto de treinamento da rede “juiz” tinha 12 neurônios de entrada. A decisão foi supervisionada, pois a saída desejada também está no arquivo de treinamento. Várias estruturas foram investigadas para a rede “juiz”.

A melhor rede “juiz” possui 12 neurônios de entrada 16 neurônios na primeira camada e 4 neurônios na camada de saída. O resultado abaixo apresenta a melhor resposta da rede em 10.000 ciclos.

	Treinamento	Validação	Teste
Errado	0.08%	7.74%	7.23%
Certo	99.92%	89.57%	91.20%
Desconhecido	0.00%	2.69%	1.57%

Tabela 4 - Tabela de resultados em porcentagem da combinação hierárquica.

4.3.2- Paralela

Na combinação em paralelo a escolha da resposta final foi baseada nas resposta de cada rede participante do sistema. Na resposta final cada rede tinha o mesmo poder de voto. Uma resposta era escolhida caso fosse a saída mais votada. Caso todas as redes tivessem saídas diferentes, a resposta final era a resposta da rede com menor erro.

Os resultados estão na tabela abaixo.

	Treinamento	Validação	Teste
Errado	0.00%	0.00%	0.00%
Certo	98.74%	89.42%	90.88%
Desconhecido	1.26%	10.58%	9.12%

Tabela 5 - Tabela de resultados em porcentagem da combinação pela maioria.

5 - Conclusão

O primeiro fato a ser notado nos resultados é que a performance da combinação em paralelo foi inferior à performance do treinamento da rede individual de melhor desempenho mas superior na validação e teste. Com isso pode-se concluir que todos os algoritmos utilizados, independente de sua performance, estão classificando corretamente padrões que pelo menos um dos outros algoritmos classificou incorretamente.

Desta forma é possível utilizar um conjunto de N classificadores cuja a união das saídas corretas pode gerar o conjunto universo das saídas corretas. Assim com uma combinação ideal de classificadores pode-se, teoricamente, produzir 100 % de acerto.

Além disso a combinação em paralelo não classificou de forma errada nenhum padrão. A combinação apenas classificou os padrões como certo e desconhecido. Isto significa que nenhuma rede errou da mesma forma no mesmo padrão produzindo como saída a mesma classe errada.

A combinação Hierárquica possui um agravante em sua estrutura. Os erros das redes iniciais são introduzidos nas entradas das redes, ou seja, o erro da classificação anterior é propagado para a classificação da rede “juiz”, o que deveria aumentar o erro final. Por outro lado, a rede aprende utilizando as saídas desejadas, o que facilita o ajuste dos pesos.

A performance da classificação da combinação em hierarquia foi superior a todas as outras performances. Com isso pode-se concluir que apesar da propagação dos erros das redes, estes

erros foram menores que os erros embutidos no arquivo de treinamento inicial. De outra forma, as redes iniciais trabalharam como filtros dos dados, eliminando ruídos.

A combinação em hierarquia produziu resultados melhores que a combinação em paralelo. Isto aconteceu porque, com o novo treinamento, a rede “juiz” aprendeu novos padrões, ou seja, quanto menor o ruído mais fácil o aprendizado.

Bibliografia

[Beale, 94] R. Beale and T Jackson. "Neural Computing: Na Introduction". IOP Publising Ltd, 1994.

[Bird, 95] Shawn D. Bird and George M. Kasper, "Problem formalization techniques for collaborative systems", IEEE Transaction on systems, Man, and Cybernetics, vol. 25, nº. 2, pg. 231-242, 1995.

[Bishop, 96] *Christopher M. Bishop. "Neural Networks for Pattern Recognition". Oxford University Press, 1996.*

[Braun,93] M. Rieldmiller and H. Braun. A direct adaptive method for faster backpropagation learning: The RPROP algoritmo. In H. Ruspini, editor, Proceedings of the IEEE International Conference on Neural Networks (ICNN), pages 586 - 591, San Francisco, 1993.

[Haykin, 94] *Haykin S. "Neural networks: a comprehensive foundation". IEEE Press, 1994.*

[Ho, 94] Tin Kam Ho, J.J. Hull, and S.N. Srihari, "Decision combination in multiple classifier systems", IEEE trans. on Pattern Analysis Machine Intelligence, vol. 16, nº. 1, pg. 66-75, 1994.

[Jordan, 95] Michael I. Jordan and Lei Xu, "Convergence results for the EM approach to mixtures of experts architectures", *Neural Networks*, vol. 8, nº. 9, pg. 1409-1431, 1995.

[Llinás, 89] *Llinás R.R., editor. "The Biology of the Brain - From Neurons to Networks". W.H. Freeman and Company. 1989.*

[Lu, 96] Yi Lu, "Knowledge Integrations in a Multiple Classifier System", *Applied Intelligence*, vol. 1, nº. 1, pg. 75-86, 1996.

[Riedmiller, 94] M. Riedmiller. Advanced supervised learning in multi-layer perceptrons - from backpropagation to adaptative learning algoritms. *Int. Journal of Computer Standards and Interfaces*, 1994. to appear.

[Ripley, 96] B. D. Ripley, "Pattern Recognition via Neural Networks".

[Rumelhart, 86] *Rumelhart D.E., Hinton G.E., Williams R.J. "Learning internal representation by error propagation. in Parallel Distributed Processing", pp: 318-362, MIT Press, 1986.*

[Smith, 96] Murray Smith, "Neural Networks for Statistical Modeling". International Thomson Computer Press.

[Vasconcelos, 95], Germano C. Vasconcelos, "Redes neurais e Reconhecimento de Padrões". II SBRN - Tutorial 2.

[Xu, 92] Lei Xu, a Krzyzak, and C. Y. Suen, "Methods of combining multiple classifiers and their application to handwriting recognition", IEEE Trans. Systems, Man, and Cybernetics, vol. 22, n^o. 3, pg. 418-435, 1992.