Um Comitê de Redes Neurais para o Reconhecimento de Letras Manuscritas

Viviane Soares Rodrigues Silva e Antonio Carlos Gay Thomé

AEP/NCE - Núcleo de Computação Eletrônica

UFRJ, Caixa Postal 2324, Ilha do Fundão, Rio de Janeiro, RJ, Brasil.

vivianerodrig@yahoo.com.br, thome@nce.ufrj.br

Abstract. The research in the field of the intelligent recognition of handwrite characters has been strongly stimulated by the necessity of several commercial applications such as bank and post office automation. In order to enlarge the research on alternative ways to use Neural Nets engines in such applications, and to establish a comparison with the technique of the Teams of Neural Nets [SILVA 2002], this work proposes the use of a committee formed by a set of MLP neural nets. An approach to previously estimate the committee performance is also explored in this paper. The obtained results with the committee are analyzed and compared with those provided by the team.

Resumo. A necessidade de leitura automática de cheques bancários, endereçamento de cartas, formulários diversos, por exemplo, têm incentivado pesquisas no campo do reconhecimento inteligente de caracteres manuscritos. Este trabalho traz a proposta de um comitê formado por redes neurais do tipo Multi Layer Perceptron (MLP), para o reconhecimento de letras manuscritas. Foram feitos estudos sobre a diversidade e exatidão entre os agentes componentes do comitê a fim de que fosse possível estimar previamente o desempenho a ser alcançado por cada combinação de agentes experimentada. Os resultados de cada tentativa são avaliados individualmente e em seguida são comparados com uma outra técnica de combinação de agentes [SILVA 2002], utilizando a mesma massa de dados.

1 – Introdução

A razão para a persistência de manuscritos numa era digital é a conveniência do papel e da caneta para as diversas situações do dia-a-dia. A automatização da leitura dos formulários para concursos, por exemplo, além de acelerar o processo de cadastramento dos candidatos, contorna de alguma forma a falta de acesso aos meios digitais, ainda uma realidade para um grande número de brasileiros. Porém, de modo geral, a digitalização e a extração de informações de um documento de forma automática possibilitam o acesso ao seu conteúdo de maneira rápida e objetiva, mantendo fidelidade aos documentos originais.

A partir de uma revisão da literatura sobre as técnicas comumente utilizadas para o reconhecimento *off-line* de caracteres manuscritos observa-se que as propostas mais recentes apostam numa combinação destas técnicas. Apoiados no fato de que, de acordo com a complexidade do

problema em questão, o emprego de uma técnica isolada não é capaz de resolvê-lo completamente. Exemplos de técnicas combinadas podem ser encontrados em [HANMANDLU *et al* 2003], [AIRES 2003], [MORITA *et al* 2002], [OH & SUEN 2002] entre outros.

2 - Combinação das Técnicas de Reconhecimento

A fim de aperfeiçoar o processo de reconhecimento e classificação de um sistema, surge a seguinte questão: de que maneira estes métodos poderiam ser combinados de modo que o resultado do reconhecimento melhore significativamente? No intuito de responder a esta pergunta, diferentes estratégias de combinação destas técnicas têm sido estudadas, apresentando resultados promissores [OLIVEIRA *et al* 2002], [OLIVEIRA *et al* 2004], [SHRIDHAR & KIMURA 1991], [ARICA & VURAL 2001] e [OH & SUEN 2002].

2.1 - Métodos de Combinação de Agentes

Os agentes de reconhecimento de padrões podem ser agrupados de modo homogêneo ou heterogêneo. No primeiro caso, todos os agentes escolhidos para compor a combinação empregam a mesma tecnologia. Por outro lado, uma combinação heterogênea é aquela formada por agentes que aplicam tecnologias distintas sobre as características do padrão apresentado.

Outra maneira de classificar uma combinação de agentes de reconhecimento é sob a forma ou arquitetura dos mesmos [ARICA & VURAL 2001]. De acordo com este último método de classificação, existem dois tipos de arquitetura: Serial ou Paralela.

- Arquitetura em Série: a saída de um agente alimenta o próximo agente. Pode-se citar duas disposições que representam uma arquitetura em série: seqüencial ou seletiva.
- Seqüencial: a cada estágio o objetivo é o de reduzir o número de possíveis classes as quais um padrão desconhecido pode pertencer. No início, o padrão desconhecido pode pertencer a qualquer uma das classes existentes e, a cada estágio, o número de prováveis classes as quais o padrão poderia pertencer diminui, até que no estágio final o padrão recebe um rótulo.
- Seletiva: Inicialmente o agente associa o novo padrão apresentado a um grupo com padrões similares entre si. Num estágio posterior, os grupos formados são classificados hierarquicamente como uma árvore. Cada nó filho desta árvore tem semelhanças com o nó pai, de acordo com alguma métrica adotada.
- Arquitetura em Paralelo: combina o resultado de mais de um agente independente. Tais agentes podem seguir técnicas diferentes de classificação, bem como serem treinados com conjuntos de dados diferentes uns dos outros. A combinação empregada pode obedecer a uma de várias estratégias de combinação existentes na literatura, dentre estas pode-se citar: voto, bayesiana e mistura ponderada de agentes.

A seguir, alguns exemplos de técnicas de reconhecimento de padrões, em especial para o reconhecimento de caracteres manuscritos:

Time de Redes Neurais: um exemplo de arquitetura do tipo seqüencial e homogênea pode ser visto no trabalho de [SILVA 2002], que apresenta uma estratégia de combinação de técnicas para o reconhecimento de caracteres manuscritos denominada Time de Redes Neurais. Foram utilizadas exclusivamente redes neurais segundo o modelo MLP para o reconhecimento de letras manuscritas, provenientes de formulários para concurso.

Classe-Modular: um exemplo de arquitetura do tipo paralela utilizando a técnica de voto para combinar as saídas dos agentes está descrito em [OH & SUEN 2002], no qual os autores propõem o conceito de modularidade em classes utilizando redes neurais MLP. Os estudos feitos incluem o reconhecimento de dígitos manuscritos (10 classes), letras maiúsculas (26 classes), pares de dígitos conectados (100 classes) e caracteres do código postal koreano (352 classes).

Two Stage Multi-Network: um bom exemplo de arquitetura serial seletiva e heterogênea é o método proposto em [GOPISETTY et al 1996], no qual um grupo de pesquisadores da IBM combinam Redes Neurais e Template Matching (Cálculo por Similaridade) num sistema de reconhecimento de caracteres denominado TSMN (Two Stage Multi-Network).

3 - Medidas de Eficácia Numa Combinação de Técnicas

Em um artigo relativamente recente [KUNCHEVA & WHITAKER 2003] descrevem medidas de diversidade entre agentes e a relação com a exatidão dos resultados, como uma forma de escolher os agentes para comporem um comitê.

3.1 - Relação entre Diversidade e Exatidão

Segundo [KUNCHEVA & WHITAKER 2003] o grau de similaridade e de discordância nas respostas produzidas por diferentes agentes, de modo individual, possibilita prever de que maneira os agentes devem ser combinados para produzirem um resultado final satisfatório, aumentando desta forma, a exatidão global do sistema. Pode-se entender como exatidão de um sistema ou de um agente individual como sendo a capacidade de classificar corretamente um padrão apresentado para reconhecimento.

No entanto, como explicado em [HADJITODOROV *et al* 2006], para alcançar níveis mais altos de exatidão, a diversidade entre os classificadores que compõem o conjunto deve diminuir de forma a esperar-se um equilíbrio entre diversidade e exatidão. Estes autores afirmam que nenhuma teoria convincente ou estudo experimental já foi feito com o objetivo de sugerir alguma medida que pudesse predizer de modo eficaz o erro de generalização de um conjunto. Contudo, sabe-se através de outros autores [WINDEATT 2005] e [KUNCHEVA 2005], que é necessário achar um ponto de equilíbrio entre diversidade e exatidão.

Através da proposta feita por [HADJITODOROV *et al* 2006], onde afirma que: "A seleção do conjunto através da diversidade mediana permite obter um resultado melhor que a seleção randômica do conjunto ou seleção do conjunto com base na discordância máxima", conclui-se que conjuntos muito diversos são menos exatos que os conjuntos menos diversos. Ou seja, para agrupar-se um conjunto de agentes com respostas muito diferentes entre si, seria necessário selecionar agentes que não apresentassem bom desempenho quando avaliados individualmente. Desta forma, utilizando a máxima discordância, o resultado global do sistema Multi-Agente não seria melhor que o uso de apenas um agente para solução do problema. Portanto, esta hipótese, chamada de regra da correlação média ou *soft-correlation*, sugere que sejam selecionados agentes com discordância média entre si.

3.2 - Medidas de Diversidade

Segundo [KUNCHEVA & WHITAKER 2003], a chave do sucesso de alguns Sistemas Multi-Agentes é que estes são compostos por um conjunto de agentes relativamente diversos. Embora ainda não haja uma descrição formal para o termo diversidade, em [ADEVA 2005] ela é definida como uma medida intuitiva da relação entre os agentes. Neste contexto, pode-se avaliar

o grau de similaridade ou de discordância entre eles. Entende-se que dois agentes são similares entre si quando, construídos sob as mesmas condições ou não, apresentam respostas individuais de classificação muito parecidas quando apresentados aos mesmos padrões de entrada. Ao contrário, um alto grau de discordância entre agentes é verificado quando apresentam respostas distintas sobre a classificação de um mesmo padrão apresentado. Com relação às condições de construção dos agentes, pode-se citar, por exemplo, o conjunto de dados utilizado para o treinamento, a técnica de representação dos dados, ou ainda a estratégia de reconhecimento empregada {VIVIANE}. Nas subseções seguintes, duas formas de cálculo das medidas de diversidade são apresentadas.

3.3 - Matriz de Probabilidade de Discordância

Em [DUIN *et al* 2004] aplica-se o conceito de discordância para medir a diferença entre dois agentes A_a e A_b construídos para resolver um problema de classificação P representado por um conjunto de N amostras. A discordância pontual entre dois agentes quaisquer é dada pela equação 3.1:

$$d_j(A_a, A_b) = \Pr(A_a(x_j) \neq A_b(x_j) \,|\, x_j \in P) \qquad (3.1)$$

Na qual $A_i(x)$ retorna a rotulação dada a um padrão x qualquer pelo agente A_i. A discordância sobre o conjunto de amostras do problema é calculada conforme a equação 3.2:

$$D(A_a, A_b) = \frac{\sum_{j=1}^{N} d_j(A_a, A_b)}{N} \qquad (3.2)$$

M classificadores constituem uma matriz [M x M] de discordâncias D para o problema P, com elementos $D(m,n) = D(A_m, A_n)$.

3.4 - Discordância Através das Matrizes de Confusão

Em [AIRES 2003], a informação contida nas matrizes de confusão de cada agente individual foi utilizada para computar as distâncias que representam as discordâncias entre eles. Denomina-se esta abordagem de Discordância Baseada no Critério da Distância (DD-*based*).

A Matriz de Confusão, figura 1, é uma representação quantitativa do desempenho obtido para cada classificador em termos do reconhecimento de cada classe.

$$\mathbf{M}A_a = \begin{vmatrix} S_{1,1} & S_{1,2} & \dots & S_{1,M} \\ S_{2,1} & S_{2,2} & \dots & \dots \\ \vdots & \vdots & \vdots & \vdots \\ S_{M,1} & \dots & \dots & S_{M,M} \end{vmatrix}$$

figura 1: Matriz de Confusão

Observando-se a matriz de confusão conclui-se que os valores, por exemplo, $S_{\{i,i\}}$ representam as amostras corretamente classificadas na classe "i"; $S_{\{i,k\}}$, $k \neq i$, representam os erros do tipo falso positivo; e $S_{\{k,i\}}$, $k \neq i$, representam os erros do tipo falso negativo para a classe em questão.

A abordagem DD-*based* utiliza a informação contida nas matrizes de confusão, onde para cada par de agentes são computadas as distâncias que representam as discordâncias entre os mesmos. Tais distâncias podem ser obtidas considerando que todas as matrizes de confusão possuem o mesmo tamanho como definido nas equações 3.3 e 3.4.

Sendo MA_a e MA_b as matrizes de confusão geradas pelos agentes A_a e A_b respectivamente, a diferença entre elas resulta na matriz $MCD^{Aa,Ab}$. Esta, por sua vez, é usada no cálculo da distância entre a combinação dos agentes.

$$MCD^{Aa,Ab} = MA_a - MA_b \qquad (3.3)$$

$$D^{A_a,A_b} = \sum_{i=1}^{M}\sum_{j=1}^{M} (MCD_{i,j}{}^{A_a,A_b})^2 \qquad (3.4)$$

4 - Proposta de um Comitê de Agentes Neurais

No intuito de compor um modelo baseado numa combinação de técnicas de reconhecimento conhecida como Máquina de Comitê, baseada em uma Mistura Ponderada de Agentes Especialistas [HAYKIN 2001], idealizou-se o esquema observado na figura 2. Uma mistura de especialistas é composta por K agentes e mais uma rede neural (chamada de rede de passagem) que fornece os fatores g_i de ponderação das saídas dos K agentes. Os K agentes são treinados individualmente e o ajuste dos pesos da rede de passagem é feito com base na retropropagação do erro do comitê.

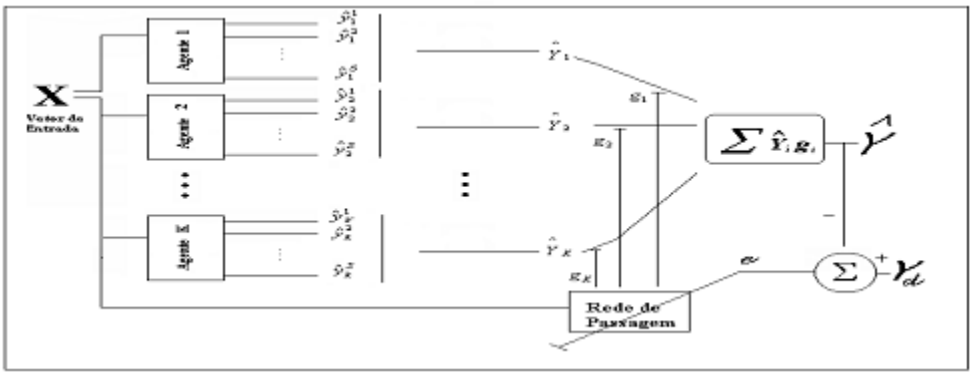

figura2: Modelo de um Comitê Dinâmico composto por Redes Neurais do tipo MLP

Considerando X a variável que representa o conjunto de entrada, a resposta Y_r do comitê para cada padrão apresentado será dada pela equação 4.1:

A rede de passagem possui uma única camada de pesos ajustáveis, conforme a figura 3, suas

$$Y_r = \sum_{i=1}^{K} y_{ri} * g_r \qquad (4.1)$$

saídas são em mesmo número do de agentes do comitê, dada pela equação 4.2, e normalizadas conforme a equação 4.3.

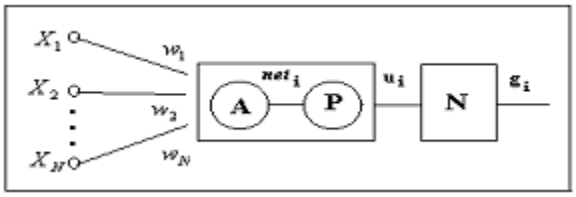

figura 3 - Representação de um Nodo Não-Linear da Rede de Passagem

$$u_k = \log sig(net_k) = \frac{1}{1 + \varepsilon^{-net_k}} \qquad (4.2)$$

$$g_i = \frac{u_i}{\sum_{i=1}^{K} u_i} \qquad (4.3)$$

Cada neurônio da camada de saída da rede de passagem, ver figura 3, é relativo a um agente k, e o ajuste dos seus pesos a cada rodada (época) do algoritmo de treinamento, é dado pela equação 4.4.

$$\Delta W_{lk} = \lambda \sum_{s=1}^{S} ((Y_s - Y_r) * y_{rl}) * f'(net_k) * X_l \qquad (4.4)$$

5 - Estudo de Caso e Avaliação dos Resultados

Os treinamentos dos agentes especialistas e do comitê foram realizados com diferentes conjuntos de dados. Para os testes, os dados foram divididos em quatro conjuntos para que se pudesse avaliar o desempenho médio e a respectiva dispersão de cada agente individualmente e do comitê. Um dos conjuntos de testes foi mantido o mesmo utilizado em [SILVA, 2002] para comparar o comitê e o time de redes. A quantidade total de amostras de imagens de letras maiúsculas (26 classes) utilizada foi de 75729, que foram segmentadas e submetidas ao processo de extração de características segundo a técnica do quadrado rotacionado [RODRIGUES 2003].

Para o treinamento dos agentes especialistas e do comitê (rede de passagem), todo o conjunto de dados, aqueles já utilizados anteriormente por [EUGENIO 2002] e os resultantes de uma nova coleta feita, foram subdivididos em grupos. Onde "Treino Ag." e "Teste Ag." são respectivamente os conjuntos para treinamento e teste dos agentes especialistas. "Treino Co." é o conjunto utilizado para o treinamento do comitê. O conjunto designado para testar o comitê é formado por cinco grupos. O primeiro é o mesmo conjunto utilizado para testar os agentes individualmente, para que fosse possível uma comparação entre o emprego de uma única rede MLP e um comitê na classificação das letras. Os demais, "Teste1", "Teste2" e "Teste3" corroboram para a verificação da estabilidade dos modelos de comitê experimentados. E por fim, "TTime", é o mesmo conjunto utilizado em [EUGENIO 2002], para a proposta apresentada naquele trabalho.

Parágrafo descrevendo a divisão dos dados para treinamento e teste dos agentes e do comitê (rede de passagem).

5.1 - Seleção dos Agentes - Primeira Abordagem

A escolha dos agentes para compor o comitê foi feita com base no desempenho de cada um, selecionando-se os três melhores (tabela 1):

tabela 1: Resultados dos Agentes Individuais

Agente	1	2	3
Resultado	86.83%	86.58%	86.51%

A análise da diversidade e exatidão dos agentes, tabelas 2 e 3, foram calculadas respectivamente, pela Matriz de Probabilidade de Discordâncias e pelas Matrizes de Confusão.

tabela 2 - Matriz de Probabilidade de Discordância

	Agente 1	**Agente 2**	**Agente 3**
Agente 1	0%	7.38%	7.38%
Agente2	7.38%	0%	6.85%
Agente 3	7.38%	6.85%	0%

tabela 3 - Discordância Através das Matrizes de Confusão

	Agente 1	**Agente 2**	**Agente 3**
Agente 1	0	0.0647	0.0656
Agente2	0.0647	0	0.0619
Agente 3	0.0656	0.0619	0

A Matriz de Probabilidade de Discordância revela que a diversidade entre os Agentes 1, 2 e 3, dois a dois, é muito baixa, tendo em vista que a probabilidade de darem respostas diferentes para um mesmo caractere é pequena, em torno de 7%. A Matriz de Confusão reforça esta hipótese mostrando valores baixos e parecidos sobre as distâncias entre os agentes.

Em resumo, os agentes apresentam um baixo grau de diversidade e o resultado do comitê foi inclusive pior que o do melhor agente.

5.2 - Seleção dos Agentes - Segunda Abordagem

Considerando que no contexto manuscrito existe uma probabilidade não desprezível de duas letras pertencentes a classes diferentes terem uma distância espacial menor que duas letras pertencentes a uma mesma classe, resolveu-se treinar os diferentes agentes com diferentes partições do espaço amostral. O espaço de dados de treinamento foi dividido em três a partir da aplicação do algoritmo K-means. Observou-se que todas as letras tinham representação, em diferentes proporções, em todas as partições.

O comitê foi montado com os três agentes assim treinados e um quarto agente (chamado Agente Geral) constituído pela rede neural com melhor desempenho utilizada na primeira abordagem e com índice de acerto de 86,83%, conforme listado na tabela 2. As medidas de diversidade ficaram conforme as tabelas 4 e 5.

tabela 4 - Matriz de Probabilidade de Discordância

	Agente 1	**Agente 2**	**Agente 3**	**Agente G**
Agente 1	0%	47.51%	31.7%	28.85%
Agente 2	47,51%	0%	45.32%	35.51%
Agente 3	31,7%	45.32%	0%	25.4%
Agente G	28.85%	35.51%	25.4%	0%

tabela 5 - Discordância Através das Matrizes de Confusão

	Agente 1	**Agente 2**	**Agente 3**	**Agente G**
Agente 1	0	0.0059	0.0005	0.0017
Agente 2	0.0059	0	0.0064	0.0077
Agente 3	0.0005	0.0064	0	0.0012
Agente G	0.0017	0.0077	0.0012	0

A Matriz de Probabilidade de Discordância revela que a diversidade entre os Agentes 1, 2 e 3, dois a dois, é média, e diminui quando os agentes são comparados com o Agente Geral. Este fato já era esperado tendo em vista os conjuntos distintos de caracteres nos quais os Agentes 1, 2 e 3 e o Agente Geral se especializaram.

A Matriz de Confusão deixa mais claro que os Agentes 1 e 3 tendem a confundir a mesma classe de caracteres pois possuem discordância baixa quando comparada com a relação entre os agentes 1 e 2, e entre os agentes 2 e 3. O Agente 2, por sua vez, apresenta uma discordância maior quando comparado ao Agente Geral.

A tabela 6 resume o desempenho alcançado pelos agentes individualmente e pelos diferentes comitês que foram testados.

tabela 6 - Resultado Final nos Testes dos Agentes para Classificação das Letras

	Agente 1	**Agente 2**	**Agente 3**	**Agente G**	**Co3A**	**Co4A**
T.Ag	72.45%	63.77%	73.46%	86.83%	87.91%	**89.16%**
T1	77.39%	68.35%	83.65%	92.87%	97.48%	**97.57%**
T2	65.06%	65.84%	67.86%	67.4%	**84.02%**	83.33%
T3	67.74%	63.47%	67.61%	68.68%	**84.55%**	83.62%
TTime	77.45%	63.69%	81.09%	91.55%	92.43%	**93.61%**
Md	72.02%	65.02%	74.73%	81.47%	89.28%	**89.46%**
DP	5.6	**2.09**	7.41	12.47	5.08	6.22
Va	31.32	**4.37**	54.88	155.46	32.27	38.69
CV	0.08	**0.03**	0.1	0.15	0.06	0.07

Onde:T.Ag., T1, T2 e T3 ➔ Conjuntos de Testes; TTime ➔ Conjunto de Teste utilizado pelo Time de Redes Neurais; Md ➔ Média dos Índices de Acerto; DP ➔ Desvio Padrão; Va ➔ Variância; CV ➔ Coeficiente de Variância; Co3A ➔ Comitê composto por 3 agentes e Rede de Passagem com 1 Camada; Co4A ➔ Comitê composto por 4 agentes e Rede de Passagem com 1 Camada.

É perceptível o baixo desempenho individual dos agentes especialistas "Agente 1", "Agente 2" e "Agente 3", e com relação ao "Agente G", com exceção dos resultados com os conjuntos de teste "T1" e "Ttime", também observa-se baixos índices de acerto na classificação das letras. Os conjuntos T2 e T3 eram bem mais difíceis que os outros dois, razão pela qual todos os agentes e inclusive os comitês apresentaram uma dispersão relativamente alta no índice de desempenho.

6 – Conclusão

O modelo proposto neste artigo tem por base a composição de um comitê de redes neurais combinados dinamicamente através de uma rede neural adicional conhecida como "rede de passagem".

Um estudo feito sobre o cálculo e o emprego da diversidade entre os agentes como instrumento de estimativa a priori do desempenho e seleção dos melhores candidatos a formar o comitê mostrou-se interessante. Comitês formados por agentes pouco diversos ou muito diversos entre si tendem ao fracasso.

Com a realização de experimentos com diferentes alternativas de combinações, observou-se que a composição de comitês contendo agentes com índices relativamente altos de exatidão individual, e consequentemente pouco diversos entre si, gera pouco ganho no índice de acerto da classificação final. Tal resultado deve-se, possivelmente, ao fato de que os agentes possuem "opinião" muito parecida sobre os padrões apresentados, tanto nas classificações corretas quanto nas incorretas, não contribuindo, desta forma, para o aumento do índice global de acerto.

Inúmeros experimentos foram realizados, sobre um conjunto de 75.729 e a média dos resultados obtidos por agentes neurais isoladamente, cerca de 86,00%, ficou 8 pontos percentuais abaixo daquele provido por um comitê composto por agentes neurais treinados com diferentes partições do espaço de padrões de entrada. O comitê também apresentou ganhos em relação ao Time de Redes Neurais proposto em [SILVA 2002].

Referências

ADEVA, J.J.G (2005) "Accuracy and Diversity in Ensembles of Text Categorisers", CLEI Electronic Journal, vol 9, pp 1-12, http://www.clei.cl/cleiej/paper.php?id=109, maio 2006.

AIRES, S,B,K (2003) "Reconhecimento de Caracteres Manuscritos Baseado em Regiões Perceptivas", Dissertação (Mestrado em Informática Aplicada), PUCPR/PPGIA, Curitiba, Paraná, 82f.

ARICA, N. ; VURAL, F. T. Y (2001)"An Overview of Character Recognition Focused on Off-Line Handwriting" Ieee Transactions on Systems, Man, and Cybernetics - Part C:Aplications and Reviews, vol. 31, no 2, pp. 216-233, maio.

DUIN, A P.; et al (2004) "The Characterization of Classification Problems by Classifier Disagreements" IN: ICPR, pp.140-143, Cambridge, UK.

FRANCO, C. R (2002) "Novos Métodos de Classificação Nebulosa e de Validação de Categorias e Suas Aplicações a Problemas de Reconhecimento de Padrões", Dissertação (Mestrado em Informática) – Universidade Federal do Rio de Janeiro, IM/NCE, 133f.

GOPISETTY, S. et al (1996) “Automated Forms - Processing Software and Services”. IBM J. Res. Develop., V. 40, No 2, pp. 211-230, Março.

HADJITODOROV, S.T. et al (2005)“Moderate Diversity for Better Cluster Ensemble” http:www.informatics.bangor.ac.uk/Kuncheva/recent_publications.htm. ,Maio 2006

HANMANDLU, M. et al.(2003) “Unconstrained Handwritten Character Recognition Based on Fuzzy ogic”, Pattern Recognition, vol 36, no 3, pp 603-623.

HAYKIN, S. (2001) “Redes Neurais: Princípios E Prática”, 2a. Ed. Porto Alegre: Bookman, 900 P. Il.

KUNCHEVA, L.I. & WHITAKER, C.J.(2003) “Measures of Diversity in classifier Ensembles and Their Relationship Whit The Ensemble Accuracy”, Machine Learning, Vol 51, No 2, pp. 181-207.

KUNCHEVA, L.I. (2005) “Measures of Diversity”, Machine Learning, Vol 51, No 2, pp. 181-207.

MORITA, M. et al (2002) “An HMM-MLP Hybrid System to Recognize Handwriting Dates”, Neural Networks, IJCNN'02. Proceedings of the 2002, V.1, Pp. 867-872.

OH, I. & SUEN, C.Y. (2002) “A Class-Modular Feedfoward Neural Network For Handwriting Recognition”, Pattern Recognition, V. 35, Pp. 229-244.

OLIVEIRA, L.S., et al (2002). “Automatic Recognition Of Handwritten Numerical Strings: A Recognition E Verification Strategy” IEEE Transactions On Pattern Analysis And Machine Intelligence, V. 24, N. 11, Pag. 1438-1454, Novembro.

OLIVEIRA, L.S., et al (2004). “Handwritten Month Word Recogniton Using Multiple Classifiers” XVII Brazilian Symposium on Computer Graphics ansd Image Processing (SIBIGRAPI), Curitiba, Brasil, Outubro.

RODRIGUES, J.R., (2003) “Segmentação E Extração de Características para Reconhecimento Automático de Caracteres - Estudo e Propostas”, Dissertação (Mestrado Em Informática) – Universidade Federal Do Rio De Janeiro, Rio De Janeiro, 169f.

SHRIDHAR,M.; & KIMURA,F. (1991) “Handwritten Numerical Recognition Based On Multiple Algoritms”, Pattern Recognit., Vol 24, No. 10, pp. 969-983.

SILVA, E. (2002) “Reconhecimento Inteligente de Caracteres Manuscritos” Dissertação (Mestrado em Sistemas e Computação) - Instituto Militar de Engenharia, Rio de janeiro, 168f.

SILVA, V.S.R. (2006) “Um Comitê de Redes Neurais para Reconhecimento de Caracteres Manuscritos”, Dissertação (Mestrado Em Informática) – Universidade Federal Do Rio De Janeiro, Rio De Janeiro, 121f.

WINDEATT,T., (2005) “Diversity Measures for Multiple Classifier System Analysis and Design” Information Fusion, Vol 6, No 1, pp. 21-36.