

Cr terios de Divis o Espacial para a Rede Neural T-CombNET aplicada ao Reconhecimento de Gestos

Marcus V. Lamar¹, Md. Shoaib Bhuiyan² e Akira Iwata³

¹Departamento de Engenharia El trica

Universidade Federal do Paran  – UFPR

²Information Processing Center

Suzuka University of Medical Science and Technology

³Department of Electrical and Computer Engineering

Nagoya Institute of Technology

Emails:lamar@eletrica.ufpr.br, iwata@elcom.nitech.ac.jp

Abstract

A neural network structure dedicated to the time series recognition, T-CombNET, is presented. The model is developed from the large scale Neural Network CombNet-II, designed to deal with very large vocabulary for Chinese and Japanese character recognition. Our specific modifications of the original CombNet-II model allows it to do temporal analysis and to be used in large set of human movements recognition systems. In T-CombNET structure, one important setting is to define the input spaces that optimize the network performance. In this work a Biologically Motivated Criterion and an Interclass Distance Measurement based criterion are analyzed and applied to select the input spaces in the T-CombNET. The obtained results show that the use of JIDM based criterion improves the classification capability of the network. The T-CombNET is applied in a Japanese finger spelling recognition system, obtaining 99.4% in a user dependent hand gesture recognition system and 96.5% for a user independent.

1. Introdu o

O reconhecimento de gestos   um promissor campo de estudos na  rea de vis o computacional. Muitos pesquisadores t m devotado esfor os em desenvolver t cnicas que permitam aos computadores compreenderem gestos humanos. Devido a um paralelismo existente com reconhecimento de voz, o uso de t cnicas como *Hidden Markov Models* (HMM) e *Dynamic Time Wrapping* (DTW) [1] t m dado bons resultados em aplica es de reconhecimento autom tico da linguagem dos sinais. O uso de redes neurais nestas aplica es, tamb m n o   novidade[2], por m o desenvolvimento de novas estruturas capazes de tratar naturalmente este problema continua sendo amplo objeto de estudos. A T-CombNET   um promissor exemplo. O ponto principal desta rede   a divis o do espa o de entrada em dois subespa os que s o

trabalhados por est gios diferentes da rede. Um bom crit rio para a sele o dos espa os   fundamental para o sucesso da T-CombNET.

Neste trabalho, apresentamos o uso de um crit rio de divis o espacial baseado na Medida da Dist ncia Entre-classes (*Interclass Distance Measurement – IDM*). O IDM   uma medida muito popular utilizada em problemas de *Sele o de Caracter sticas* a partir de um vetor de medidas de entrada. Baseados no IDM, criamos o crit rio IDM Conjunto (*Joint IDM – JIDM*), que   dedicado a dividir um espa o de entrada em dois subespa os complementares. Deste modo, os vetores de entrada projetados nesses subespa os, podem ser mais facilmente analisados por classificadores multi-est gios, como a T-CombNET.

2. A Rede Neural T-CombNET

A rede neural T-CombNET (*Temporal-CombNET*) apresentada neste trabalho foi inspirada na rede neural CombNET-II proposta por Iwata[3]. A CombNET-II, mostrada na Figura 1,   uma rede neural de 4 camadas com uma estrutura em pente, composta por uma rede Tronco (*Stem Network*) e v rias redes Galhos (*Branch Networks*). A rede Tronco divide o espa o de entrada atrav s de uma rede de Quantiza o Vetorial (LVQ) em v rias regi es. Cada neur nio de sa da da rede Tronco   associado a uma rede Galho. As redes Galhos consistem de redes Perceptron Multicamadas (MLP) de 3 camadas, os quais fazem uma classifica o refinada do vetor de entrada em um espec fica regi o do espa o de entrada.

A principal vantagem da estrutura CombNET-II   a simplifica o do est gio de treinamento. Uma vez que o espa o de entrada   pr -classificado, um grande problema   dividido em problemas menores, reduzido o n mero de m nimos locais nas redes Galhos, facilitando o treinamento pelo algoritmo de *Backpropagation*. Esta filosofia permite que um problema com grande n mero de classes seja resolvido facilmente, reduzindo o tempo de treinamento e permitindo que uma boa solu o seja encontrada, aumentando a taxa de reconhecimento. A rede CombNET-II tem sido aplicada com sucesso em

problemas de reconhecimento de caracteres dos idiomas chinês e japonês escritos manualmente e impressos[3].

$$S_0 \subset S \text{ e } S_1 \subset S \mid S = S_0 \times S_1 \quad (1)$$

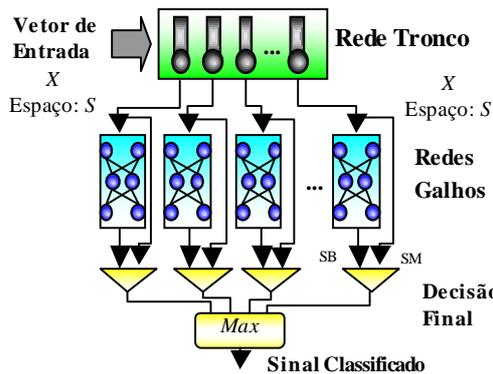


Figura 1: A Rede Neural CombNET-II

Tomando estas vantagens e transportando-as para um problema onde os vetores são variantes no tempo, apresentamos a T-CombNET, mostrada na Figura 2.

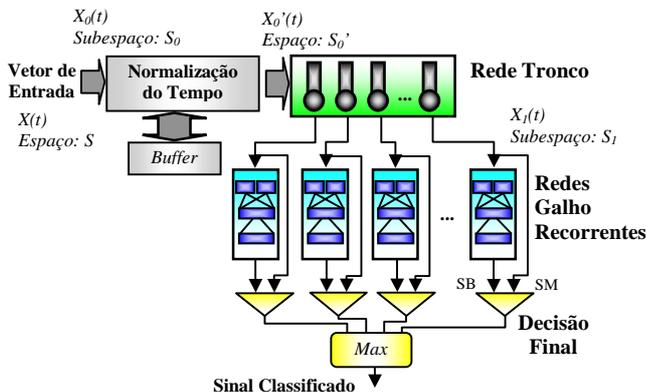


Figura 2: A Rede Neural T-CombNET

A T-CombNET é composta de uma etapa de pré-processamento, que realiza uma normalização temporal, seguida por uma Rede Tronco (*Stem Network*), várias Redes Galhos (*Branches Networks*) e um procedimento para realizar a Decisão Final. A Rede Tronco consiste de uma rede neural baseada na rede LVQ (*Learning Vector Quantization*) e as Redes Galhos consistem de redes neurais parcialmente recorrentes de Elman.

O espaço de entrada S é dividido em dois subespaços S_0 e S_1 , que definem os espaços de entrada para as redes Tronco e Galhos respectivamente. Este procedimento permite que cada rede analisar um específico espaço de características.

Dado X uma seqüência temporal de vetores $\vec{x}(t)$ d dimensionais pertencentes ao espaço S , de duração $T(X)$ amostras. Definimos $\vec{x}_0(t)$ como um sub-vetor de $\vec{x}(t)$, com dimensão d_0 , tal que $\vec{x}_0(t) \in S_0$; e definimos $\vec{x}_1(t)$ um sub-vetor de $\vec{x}(t)$, com dimensão d_1 , e $x_1(t) \in S_1$, tal que

assim, o espaço S pode ser definido como o produto cartesiano dos subespaços S_0 e S_1 .

A seleção dos subespaços S_0 e S_1 , a partir do espaço de entrada S , necessita ser cuidadosamente realizada. Enfocaremos este problema a seguir.

O procedimento de normalização do tempo, objetiva ajustar a série temporal X_0 de dimensão d_0 , presente no subespaço S_0 , a um vetor de entrada \vec{X}_0' de dimensão fixa, independente do tempo, pertencente a um espaço S_0' , uma vez que em várias aplicações, não apenas os vetores $\vec{x}(t)$, mas a duração temporal $T(X)$ do sinal de entrada pode ser uma variável aleatória. Este procedimento é necessário devido à rede LVQ que compõe a Rede Tronco, requerer na sua entrada um vetor de dimensão fixa não dependente do tempo. Então, é necessário a criação de um espaço S_0' capaz de descrever completamente as variações temporais existentes nos vetores pertencentes ao subespaço S_0 . Caso a informação da duração de uma determinada seqüência $T(X_n)$ seja uma característica importante para a correta classificação da seqüência X_n , a mesma deverá se analisada pelas redes Galhos.

Detalhes sobre o funcionamento e treinamento da T-CombNET podem ser encontrados em [4].

3. Seleção dos Subespaços

O uso da divisão espacial permite que cada rede neural se especialize na análise de um determinado conjunto de características de entrada. De modo que características de entrada não correlacionadas possam ser analisadas separadamente pelas redes, reduzindo o tempo de treino e melhorando a qualidade da solução. O método usado para definir os subespaços S_0 e S_1 dado um conjunto de treino no espaço S , é um ponto muito importante para o sucesso da rede. A seguir serão apresentadas duas alternativas para o problema da seleção dos subespaços, em um sistema de reconhecimento de gestos baseado em visão computacional.

3.1. Critério Biologicamente Motivado(CBM)

Uma possível solução para o problema da seleção dos subespaços S_0 e S_1 a partir de um espaço S é através do uso da experiência do projetista. Neste trabalho temos um espaço de entrada de 22 dimensões, com vetores variantes no tempo, definido por

$$S = \left\{ \begin{array}{l} P_v^x, P_v^y, \tau_0^x, \tau_0^y, \zeta_0, \theta_0, \tau_1^x, \tau_1^y, \zeta_1, \theta_1, \\ \tau_2^x, \tau_2^y, \zeta_2, \theta_2, \tau_3^x, \tau_3^y, \zeta_3, \theta_3, \tau_4^x, \tau_4^y, \zeta_4, \theta_4 \end{array} \right\} \quad (2)$$

obtido a partir de um sinal de vídeo contendo o gesto, utilizando o método de extração de características apresentado em [5]. Estudos têm demonstrado que o cérebro humano processa as informações de forma em uma região do córtex distinta da região responsável pelo processamento da informação de movimento [6]. Assim, dividimos o espaço de entrada em dois subespaços, o espaço responsável pela informação de movimento corresponde a um plano 2D S_0 , e o espaço 20D responsável pela descrição da postura da mão S_1 . Chamaremos este de Critério Biologicamente Motivado (CBM). Deste modo temos:

$$S_0 = \{P_v^x, P_v^y\} \quad (3)$$

$$S_1 = \left\{ \begin{array}{l} \tau_0^x, \tau_0^y, \zeta_0, \theta_0, \tau_1^x, \tau_1^y, \zeta_1, \theta_1, \tau_2^x, \tau_2^y, \zeta_2, \theta_2, \\ \tau_3^x, \tau_3^y, \zeta_3, \theta_3, \tau_4^x, \tau_4^y, \zeta_4, \theta_4 \end{array} \right\} \quad (4)$$

No contexto da T-CombNET, isto significa que teremos na entrada da rede Tronco, um vetor correspondente à trajetória normalizada no tempo, sendo então esta rede especializada em análise de trajetórias temporalmente normalizadas. As redes Galhos ficarão responsáveis pela análise das variações temporais das posturas manuais, para uma dada trajetória.

3.2. Critério Baseado na Maximização da JIDM

Em aplicações onde as componentes do espaço de entrada possuem um significado físico, o uso da experiência do projetista vem a ser a forma mais simples para realizar a seleção dos subespaços. No entanto em aplicações onde este significado físico não é muito claro, ou mesmo inexistente, torna-se necessário o uso de ferramentas matemáticas. Em várias aplicações, a seleção das características relevantes a partir de um vetor de medidas pode ser uma tarefa bastante complexa, e muito métodos têm sido propostos na literatura [7]. Uma forma de realizar a *Seleção de Características* é baseada no critério de separabilidade das classes, avaliado para todas as possíveis combinações das características de entrada. A Eq.(5) apresenta um critério baseado em distâncias, que define a Medida de Distância Entre-Classes (*Interclass Distance Measurement - IDM*) [8]

$$J_{\delta} = \frac{1}{2} \sum_{i=1}^m P(\omega_i) \sum_{j=1}^m P(\omega_j) \frac{1}{N_i N_j} \sum_{k=1}^{N_i} \sum_{l=1}^{N_j} \delta(\vec{\xi}_{ik}, \vec{\xi}_{jl}) \quad (5)$$

onde m é o número de classes, $P(\omega_i)$ é a probabilidade da i -ésima classe, N_i o número de vetores pertencentes a classe ω_i , e $\delta(\vec{\xi}_{ik}, \vec{\xi}_{jl})$ uma medida de distância do k -ésimo vetor candidato da classe i ao l -ésimo vetor

candidato da classe j , neste trabalho definida como a Distância Euclidiana. A maximização da IDM J_{δ} em relação ao vetor candidato $\vec{\xi}$, define o conjunto de medidas a serem utilizadas. O critério IDM estima a distância entre as classes através do cálculo das mesmas em um conjunto estatisticamente representativo.

No contexto da T-CombNET, aplicamos o método de *Seleção de Características* baseado no IDM para dividir o espaço de entrada S . Logo é necessário maximizar a *separabilidade conjunta* das classes da rede Tronco e das redes Galhos. A partir da Eq.(5), os valores da probabilidade da classe $P(\omega_i)$ são possíveis de serem estimados somente para a saída das redes Galhos. Para a rede Tronco estes valores não são conhecidos *a priori*, pois não conhecemos quantos neurônios serão necessários para fazer uma classificação ótima dado um subespaço S_0 definido. Além disso, o critério de classificação não é baseado na já conhecida probabilidade das classes finais, mas sim baseada na probabilidade do erro obtido para as redes Galhos. Isto pode ser modelado definindo a IDM Conjunta (*Joint IDM - JIDM*) J_T

$$J_T = J_{\delta S} + \frac{1}{M} \sum_{p=1}^M J_{\delta Bp} \quad (6)$$

onde M é o número total de redes Galhos ou número de neurônios da rede Tronco, $J_{\delta S}$ é a IDM para a rede Tronco $J_{\delta Bp}$ é a IDM para a p -ésima rede Galho.

A formulação original da IDM apresentada na Eq.(5), inclui não apenas as distâncias *Interclasses* (distâncias entre vetores pertencentes a duas classes distintas), mas também as distâncias *Intraclasses* (distâncias dos vetores dentro de uma mesma classe). Podemos separar estas duas distâncias e, deste modo, definir $J_{i\delta S}$ como a distância *Interclasses* e $J_{o\delta S}$ a distância *Intraclasse* para a rede Tronco como:

$$J_{i\delta S} = \sum_{i=1}^{m-1} P(\varphi_i) \sum_{j=i+1}^m P(\varphi_j) \frac{1}{N_i N_j} \sum_{k=1}^{N_i} \sum_{l=1}^{N_j} \delta(\vec{x}_{0ik}, \vec{x}_{0jl}) \quad (7)$$

$$J_{o\delta S} = \sum_{i=1}^m P(\varphi_i)^2 \frac{1}{N_i^2 - N_i} \sum_{k=1}^{N_i-1} \sum_{l=k+1}^{N_i} \delta(\vec{x}_{0ik}, \vec{x}_{0il}) \quad (8)$$

onde $P(\varphi_i)$ é a probabilidade *a priori* obtida para a i -ésima pseudo-classe projetada pelo algoritmo de treinamento da rede Tronco, \vec{x}_{0ik} é o vetor no espaço S_0' correspondente a k -ésima série temporal normalizada que pertence a i -ésima pseudo-classe, m número total de pseudo-classes gerados pela rede Tronco e N_i o número total de vetores pertencentes a pseudo-classe i .

Analogamente, podemos escrever as distâncias *Interclasses* $J_{i\delta Bp}$ e *Intraclasses* $J_{o\delta Bp}$ para a p -ésima rede Galho:

$$J_{i_{\delta S}} = \sum_{i=1}^{m_i-1} P(\omega_{ip} | \varphi_p) \sum_{j=i+1}^{m_i} P(\omega_{jp} | \varphi_p) \cdot \frac{1}{N_i N_j} \sum_{k=1}^{N_i} \sum_{l=1}^{N_j} \delta(\vec{x}_{1ik}, \vec{x}_{1jl}) \quad (9)$$

$$J_{o_{\delta S}} = \sum_{i=1}^{m_i} P(\omega_{ip} | \varphi_p)^2 \frac{1}{N_i^2 - N_i} \sum_{k=1}^{N_i-1} \sum_{l=k}^{N_i} \delta(\vec{x}_{1ik}, \vec{x}_{1il}) \quad (10)$$

onde $P(\omega_{ip} | \varphi_p)$ é a probabilidade condicional da i -ésima classe da p -ésima rede Galho associada ao neurônio p da rede Tronco, \vec{x}_{1ik} é o k -ésimo vetor alocado pela rede Tronco para a classe i projetado em um espaço normalizado no tempo S_1' ; e m_p é o numero de classes alocados ao neurônio p da rede Tronco.

Deste modo, as medidas de distâncias para as redes Tronco e Galhos podem ser escritas como funções das parcelas das distâncias *intra* e *interclasses*:

$$J_{\delta S} = f(J_{i_{\delta S}}, J_{o_{\delta S}}) \quad (11)$$

$$J_{\delta Bp} = f(J_{i_{\delta Bp}}, J_{o_{\delta Bp}}) \quad (12)$$

Substituindo as Eq.(11) e (12) na Eq. (6), obtemos a função J_T a ser maximizada com respeito aos vetores \vec{x}_0 e \vec{x}_1 . Definindo deste modo os subespaços S_0 e S_1 .

A formulação da IDM original apresentada na Eq.(5) usa como função f a seguinte combinação linear

$$J_{\delta S} = J_{i_{\delta S}} + 0.5J_{o_{\delta S}} \quad (13)$$

$$J_{\delta Bp} = J_{i_{\delta Bp}} + 0.5J_{o_{\delta Bp}} \quad (14)$$

Maximizando a Eq.(6), estaremos maximizando não apenas as distâncias *Interclasses*, mas também as distâncias *Intraclasses*. Logo uma melhor escolha para a função f poderia ser

$$J_{\delta S} = J_{i_{\delta S}} \quad (15)$$

$$J_{\delta Bp} = J_{i_{\delta Bp}} \quad (16)$$

onde se considera apenas as distâncias *Interclasses*. Outras alternativas poderiam ser:

$$J_{\delta S} = J_{i_{\delta S}} - \gamma J_{o_{\delta S}} \quad (17)$$

$$J_{\delta Bp} = J_{i_{\delta Bp}} - \gamma J_{o_{\delta Bp}} \quad (18)$$

ou

$$J_{\delta S} = J_{i_{\delta S}} / J_{o_{\delta S}} \quad (19)$$

$$J_{\delta Bp} = J_{i_{\delta Bp}} / J_{o_{\delta Bp}} \quad (20)$$

Nestas duas alternativas, as distâncias *Intraclasses* contribuem inversamente, isto é, maximizando a Eq.(6), estaremos maximizando as distâncias entre as classes e minimizando as distâncias *Intraclasses*, tornando os vetores dentro de uma mesma classe menos dispersos.

O algoritmo adotado para treinar a rede Tronco, necessário para a estimativa das probabilidades $P(\omega_{ip} | \varphi_p)$ no processo de otimização descrito, é o proposto para a rede CombNET-II, um algoritmo de auto-crescimento baseado em um limiar de similaridade e limiar de divisão de neurônio, descrito em [3]. Para

fazer uma classificação temporal refinada, as redes Galhos são compostas por redes neurais parcialmente recorrentes de Elman [9] treinadas pelo algoritmo de *Backpropagation*. Kremert [10] estudou o problema da capacidade de modelagem da rede neural de Elman e demonstrou que a estrutura é capaz de simular, em princípio, qualquer Máquina de Estados Finitos, mas se ela consegue fazer esta simulação na prática poderia depender do método de treinamento. Neste contexto, a estrutura T-CombNET apresentada faz com que o treinamento da rede de Elman seja mais fácil usando o clássico algoritmo de *Backpropagation*, devido à pré-análise realizada pela Rede Tronco, que reduz a complexidade do problema. Então também podemos esperar que a rede T-CombNET seja geral e facilmente treinável.

4. Resultados

Esta seção apresenta os resultados dos experimentos realizados usando uma estação Silicon Graphics, 130MHz, com uma câmera de vídeo colorida e placa de aquisição Viro, usando quadros de tamanho 160×120 pixels e 24 bits/pixel.

O alfabeto manual japonês é formado por 42 posturas estáticas da mão e 34 diferentes movimentos manuais. Os movimentos presentes, são basicamente, os apresentados pela Figura 3.

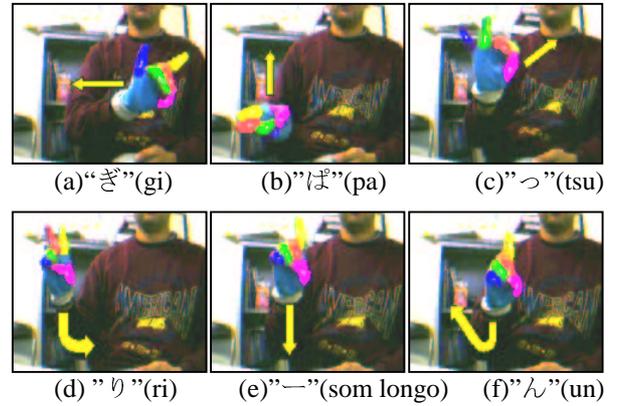


Figura 3: Tipos de movimentos presentes no alfabeto manual japonês

No alfabeto manual japonês existem 20 gestos com o movimento horizontal da esquerda para direita, mostrado na Figura 3(a), diferenciando-se entre si apenas pela postura da mão; Cinco gestos com o movimento vertical de baixo para cima mostrado em 3(b); Cinco gestos com o movimento 3(c), que corresponde a um deslocamento da mão se afastando da câmera; Dois gestos com o movimento 3(d) e um gesto para cada movimento 3(e) e 3(f), totalizando 34 diferentes gestos.

Um banco de dados foi gerado capturando os gestos realizados por cinco usuários. Cada pessoa foi convidada a repetir cada gesto 5 vezes, gerando 170

sinais de vídeos por pessoa. Em um sistema independente do usuário, o conjunto de treino é composto pelos gestos de 4 usuários, reservando-se os gestos do 5º usuário para ser usado como conjunto de teste. Preparou-se também um experimento com um sistema dependente do usuário, considerando apenas os dados de uma única pessoa para criar os conjuntos de treino e teste.

4.1. Avaliação do Critério Biologicamente Motivado (CBM)

Inicialmente avaliou-se o desempenho do uso da T-CombNET, com os espaços S_0 e S_1 definidos como descrito na seção 3.1, em um sistema dependente do usuário. A estrutura T-CombNET foi comparada com a rede LVQ1 de Kohonen, a rede recorrente de Elman, a rede recorrente de Jordan, uma rede MLP de 3 camadas treinada com *Backpropagation* e a estrutura CombNET-II. As redes LVQ, MLP e CombNET-II possuem estrutura não-recorrente, então elas necessitam um estágio de pré-processamento pelo já descrito algoritmo de normalização do tempo para obter uma versão “especializada” da informação temporal. O espaço de entrada neste caso consiste de todas as 22 dimensões do vetor de entrada. As redes de Elman e Jordan, por serem recorrentes, aceitam em suas entradas seqüências temporais, logo as seqüências de vetores de dimensão 22D foram diretamente aplicadas.

Tabela 1: Comparação com Estruturas Clássicas Sistema Dependente do Usuário

Rede Neural	Estrutura	Taxa de Reconhecimento
Jordan	22×272×34+34	92.9%
LVQ1	220×136×34	97.0%
MLP	220×136×34	97.6%
CombNET-II	Tronco 220×4 Galhos: 220×100×7 220×100×9 220×100×23 220×100×4	98.2%
Elman	22×272+272×34	98.8%
T-CombNET	Tronco 16×3 Galhos: 20×92×23 20×44×11 20×36×9	99.4%

A Tabela 1 apresenta a melhor estrutura obtida para cada modelo e a respectiva taxa de reconhecimento atingida. Analisando esta tabela vemos que o uso da rede de Elman é mais eficiente que o uso de redes que necessitem de pré-processamento através da “especialização” do tempo. Porém o uso combinado de

redes recorrentes e técnicas de “especialização”, na estrutura T-CombNET, fornece melhores resultados.

Aplicando-se a T-CombNET em um sistema independente do usuário obtivemos apenas 90.5% de taxa de reconhecimento. Este resultado pode indicar que, ou a estrutura tem um baixo poder de generalização, ou o CBM não divide o espaço de maneira ótima. Para tanto aplicou-se o método de maximização visto na seção 3.2 para a escolha dos subespaços.

4.2. Avaliação do Critério Baseado na JIDM

A aplicação do critério divisão espacial baseado na maximização da JIDM apresentado neste trabalho, requer um elevadíssimo tempo de processamento para obter uma solução ótima. Isto é devido ao fato de todas

as $n = \sum_{i=1}^{d-1} C_i^d$ (no nosso caso $d = 22$ e $n = 4.194.303$)

possíveis combinações das dimensões do espaço S necessitarem serem avaliadas através da Eq.(6). Este procedimento torna-se impraticável quando notamos que, na estrutura T-CombNET, para cada possível divisão espacial candidata, a rede Tronco necessita ser treinada para estimarmos as probabilidades condicionais $P(\omega_p | \varphi_p)$. Métodos de busca rápida para achar uma solução sub-ótima têm sido propostos na literatura Neste trabalho adotou-se o método de busca rápida *Sequential Forward Selection* apresentado em [7].

Aplicamos o processo de maximização a três diferentes definições do JIDM, apresentadas na seção 3.2, em um sistema independente do usuário. A Tabela 2 apresenta o valor da JIDM obtida e o subespaço S_0 definido, utilizando o CBM e o Critério de Maximização da JIDM apresentados na seção 3.

Tabela 2: JIDM: Biológico × Maximização

S_0	Tipo 1	Tipo 2	Tipo 3
CBM	4.514 $\{P_v^x, P_v^y\}$	3.661 $\{P_v^x, P_v^y\}$	4.181 $\{P_v^x, P_v^y\}$
JIDM	5.318 $\{P_v^y, P_v^x, \tau_2^y, \tau_2^x\}$	4.774 $\{P_v^y, P_v^x, \tau_3^x\}$	4.185 $\{\tau_0^x, P_v^y, \zeta_0\}$
Taxa	91.8%	96.5%	92.9%

O cálculo da JIDM foi realizado com base em três abordagens para a definição da função f nas Eq.(11) e (12). Na Tabela 2, o Tipo 1 corresponde ao uso do critério originalmente proposto através das Eq.(13) e (14), o Tipo 2 utiliza apenas as distâncias *Interclasses*, de acordo com as Eq.(15) e (16), e o Tipo 3 faz uso das Eq.(17) e (18). Comparando a taxa de reconhecimento de 90.5% obtida anteriormente com o uso do CBM, vemos que o uso do Tipo 2 para a definição da JIDM resulta em um aumento significativo da taxa de

reconhecimento para 96.5%. As dimensões escolhidas para a definição de S_0 pelo método de maximização, mantém a informação de direção do movimento $\{P_v^x, P_v^y\}$ e acrescentando apenas a informação relativa à coordenada horizontal da centróide do dedo anular τ_3^x . A Tabela 3 apresenta as melhores estruturas obtidas para as redes clássicas e suas respectivas taxas de reconhecimento.

Tabela 3: Comparação com Estruturas Clássicas Sistema Independente do Usuário

Rede Neural	Estrutura	Taxa de Reconhecimento
Jordan	22×272×34+34	75.9%
Elman	22×272+272×34	86.5%
MLP	176×136×34	94.7%
LVQ1	176×136×34	94.7%
CombNET-II	Tronco 176×3 Galhos 176×50×20 176×50×9 176×50×11	95.3%
T-CombNET	Tronco 24×5 Galhos 19×32+32×8 19×32+32×8 19×60+60×15 19×88+88×22 19×100+100×25	96.5%

Vimos no experimento anterior que a rede de Elman apresentava um bom poder de processamento quando comparadas com as outras estruturas, no entanto de acordo com as resultados da Tabela 3, seu desempenho foi baixo neste experimento, indicando que o poder de generalização desta rede fica abaixo das redes que utilizam a “espacialização” do tempo, como LVQ1, MLP e CombNET-II. No entanto, o uso conjunto de redes recursivas e “espacialização” do tempo, na T-CombNET, continua fornecendo resultados superiores. Notemos que os resultados apresentados pela T-CombNET, não são garantidos serem os ótimos, uma vez que usamos um método de busca rápida para reduzir a complexidade computacional da etapa de maximização do JIDM, sendo a partição do espaço obtida uma solução sub-ótima.

5. Conclusões

Este trabalho apresentou a rede neural T-CombNET, uma estrutura dedicada ao processamento de informações temporais. Esta rede é composta de uma rede Tronco, baseada na LVQ, e várias Redes Galhos recorrentes. Uma definição muito importante para o sucesso da T-CombNET é a correta divisão do espaço

de entrada, de modo que a Rede Tronco e as Redes Galhos, analisem de forma eficiente e sem redundância, os vetores de entrada.

Apresentamos duas abordagens para o problema da divisão do espaço de entrada em um sistema de reconhecimento de gestos baseados em visão computacional. Primeiramente apresentamos um Critério Biologicamente Motivado na observação de como o cérebro humano processa a informação de um objeto em movimento. A seguir apresentamos um método matemático baseado na maximização da Medida das Distâncias *Interclasses* Conjunta (JIDM).

Os resultados mostram que o uso do CBM apresenta superior taxa de reconhecimento de 99.4% em um sistema dependente do usuário. A aplicação do método de maximização da JDIM aumenta para de 90.5% para 96.5% a taxa reconhecimento em um sistema independente do usuário. Comparações mostraram que em ambos os casos o desempenho da T-CombNET foi superior aos apresentados por estruturas neurais clássicas. A técnica de divisão espacial baseada na maximização da JIDM apresentada neste trabalho é de uso geral, podendo ser aplicada a qualquer tipo de redes neurais compostas ou cascata de classificadores.

Referências

- [1] T. Starner and A. Pentland, Visual Recognition of American Sign Language Using Hidden Markov Models, in *Int. Workshop Aut. Face and Gesture Rec.*, Zurich, 1995.
- [2] S. Tamura and S. Kawasaki, Recognition of Sign Language Motion Images, in *Pattern Recognition*, Vol.21, No 4, pp. 343-353, 1988
- [3] A. Iwata, Y. Suwa, Y. Ino, and N. Suzumura, Hand-Written Alpha-Numeric Recognition by a Self-Growing Neural Network : CombNET-II, in *Proc. of the Int. Joint Conference on Neural Networks*, Baltimore, 1992.
- [4] Lamar, M. V. *Hand Getsure Recognition using T-CombNET: A Neural Network dedicated to Temporal Information Processing*, Ph.D. Thesis, Nagoya Institute of Technology, Japan, 2001.
- [5] M. V. Lamar, Md. S. Bhuiyan, A. Iwata, Hand Alphabet Recognition using Principal Component Analysis and Neural Networks, in *Proc. of Int. Joint Conf. on Neural Networks*, Vol.4, pp.2839-2844, Washington, USA, 1999
- [6] K. Fukushima, Active and Adaptive Vision: Neural Networks Models, *Biologically Motivated Computer Vision, Lectures Notes in Computer Science 1811*, pp. 623-634, Springer-Verlag Berlin-Heidelberg, Seoul, Korea, 2000
- [7] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, 1990
- [8] T. Y. Young and K. S. Fu, *Handbook of Pattern Recognition and Image Processing*, Academic Press, 1986.
- [9] J. L. Elman, Finding Structure in Time, in *Cognitive Science*, 14, pp.179-221, 1990.
- [10] S. C. Kremer, On the Computational Power of Elman-Style Recurrent Networks, in *IEEE Trans. on Neural Networks*, Vol. 6, No. 4, pp.1000-1004, 1995.