

PROGRAMA DE PÓS-GRADUAÇÃO DOUTORADO EM CIÊNCIA DA COMPUTAÇÃO
ASSOCIAÇÃO UFMA/UFPI
Disciplina: Redes Neurais (DCC40)
Professor: Areolino de Almeida Neto
Aluno: Raimundo Vale

Trabalho sobre Reconhecimento de Fonemas via RNA

Atividade do trabalho

Realizar o treinamento de uma RNA cooperada para aprender a reconhecer as sílabas das palavras DIREITA e ESQUERDA. Utilizei duas RNAs do tipo **MLP**. Foram utilizadas 119 amostras de cada sílaba. As amostras foram gravadas pelos alunos da disciplina em diferentes entonações, velocidades e pronúncias para propiciar mais diversidade. As amostras foram separadas em duas partes: treino e validação.

Proposta para a solução

Os arquivos foram gerados e agrupados conforme as sílabas das palavras DIREITA e ESQUERDA. Cada arquivo de sílaba foi lido e tratado via FFT obter a representação das senoides. Os dados obtidos após o tratamento com FFT resultava em arquivos muito grandes pelo que optei por agrupar as frequências em tamanhos iguais e iniciar o aprendizado. As sílabas das palavras a serem reconhecidas depois de processadas pela FFT geraram um padrão base que estão detalhados no Anexo I deste relatório. O detalhamento da execução dos programas Mathlab está detalhado no Anexo II deste relatório. Ao longo do aprendizado vários cenários (conjunto das variáveis importantes do problema) foram testados para avaliar qual obtinha os melhores resultados. As variáveis que julguei importantes e que foram modificadas nos cenários foram: quantidade de amplitudes agrupadas para formar uma média, quantidade de médias fixa dos arquivos pós-tratamento com FFT, tamanho das bases de validação e treino, taxa de aprendizagem, função de ativação, quantidade de épocas. O cenário escolhido para apresentação final foi:

Quantidade de amplitudes por amostra: 160

Tamanho da camada de entrada: 160

Tamanho da camada oculta: 12

Tamanho da camada de saída: 1

Eta: 0,01

k: 1

Função de ativação da rede 1: sigmoide

Função de ativação da rede 2: tangente hiperbólica

Tamanho da base de treinamento: 70% (1344 amostras)

Tamanho da base de validação: 30% (576 amostras)

Total de épocas por treinamento: 50 por treinamento (50 na rede 1 e 50 na rede 2)

Número de treinamento: 10

Detalhamento dos resultados

MSE de treinamento por épocas

O Gráfico 1 mostra a variação do MSE na fase de treinamento em cada época. Vemos o que o MSE começa alto e vai diminuindo na rede 1 (sigmoide), quando entra a rede 2 (tangente hiperbólica) o MSE dá um salto e não diminui na mesma proporção da rede 1, mas diminui mais devagar e fica um pouco acima do mínimo da rede 1. O padrão se repete a cada 100 épocas sendo que as 50 primeiras épocas são referentes à rede 1 e as 50 épocas seguintes são referentes à rede 2 e assim sucessivamente.

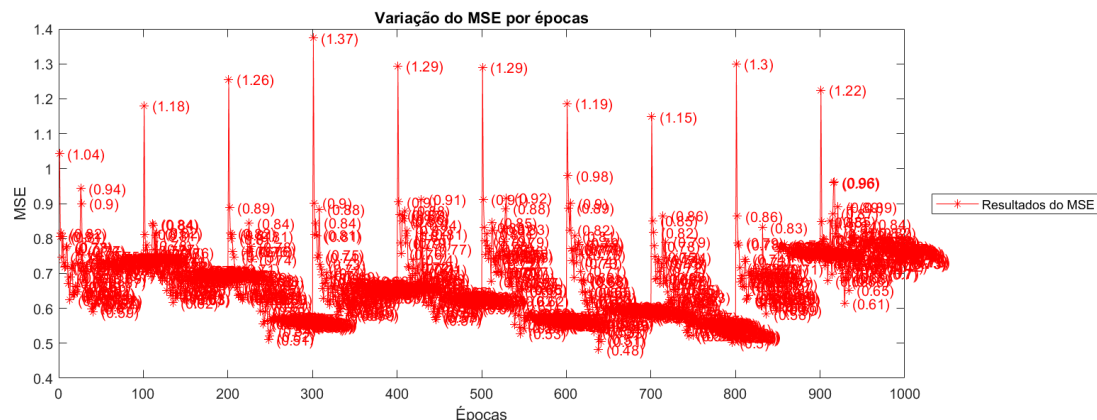


Figura 1: Gráfico 1 – MSE de 10 treinamentos da RNA cooperada. Cada treinamento consistiu em 50 épocas da rede 1 seguidas de 50 épocas da rede 2

A fim de explicar melhor o comportamento das redes fiz nova execução gerando o Gráfico 2 - com um único treinamento com 500 épocas de cada rede em sequência. Vemos que a rede 1 tem mais variação tanto para mais e quanto para menos. Já a rede 2 varia menos e estabiliza o MSE em 0,47. Os dados da execução (valores de todas as variáveis) estão no arquivo “MelhorRede.mat”.

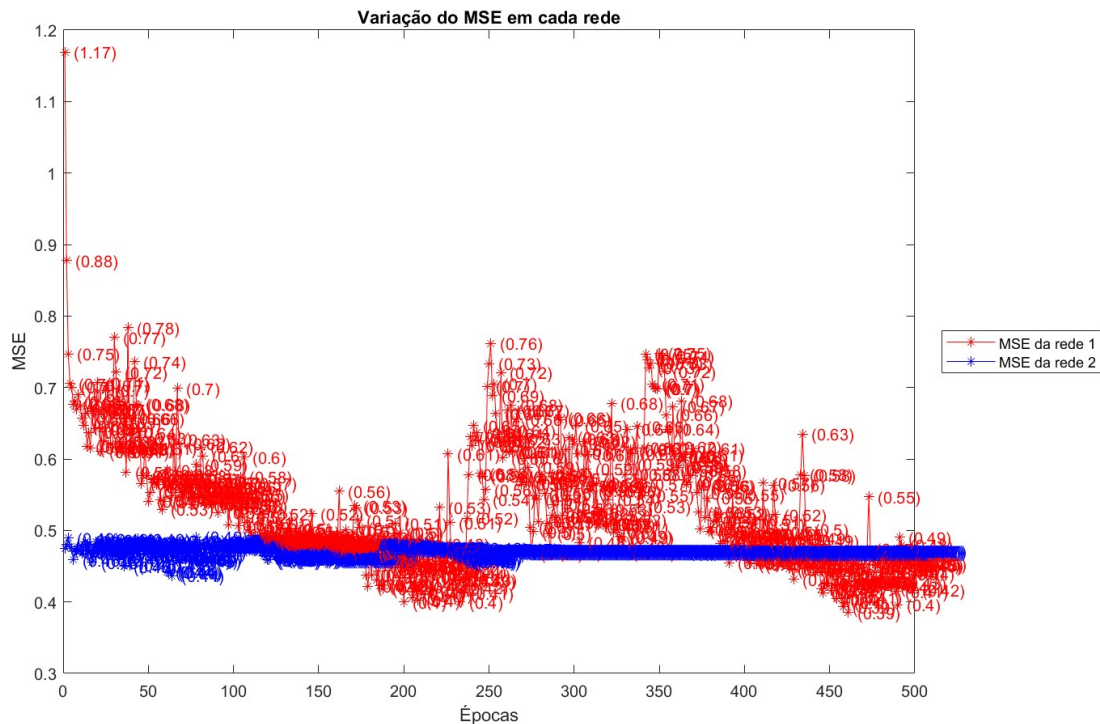


Figura 3: Gráfico 2 - Comparativo do MSE de cada época na fase de treinamento. Este treinamento consistiu 1 treinamento de 500 épocas de cada rede sucessivamente.

Acurácia

A média de acertos variou por treinamento no pior caso foi de 68,4% e no melhor caso 76,22% (treinamento 4). Além da sensível melhora em relação aos resultados obtidos sem redes cooperadas (melhor resultado foi de 22,54%). Além das alterações implementadas – melhoria no tamanho da amostra e junção da validação e teste num único processo – os resultados foram bem superiores.

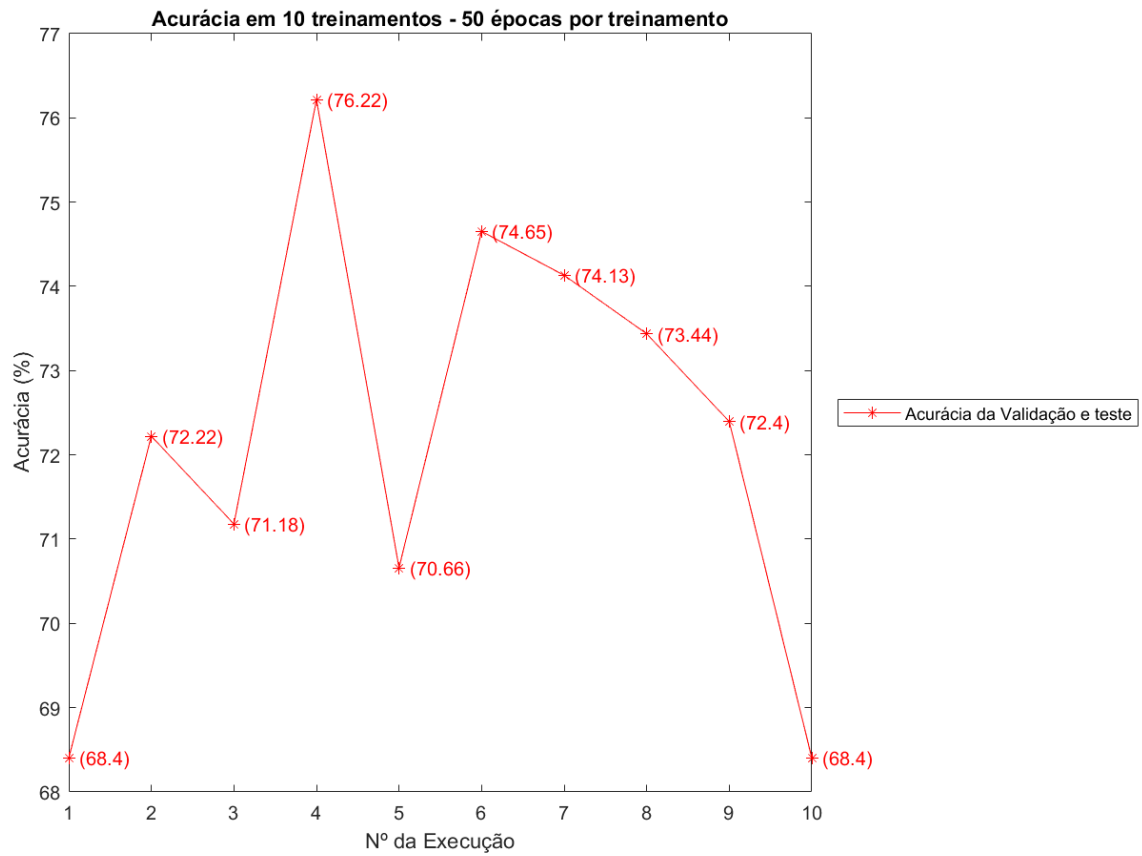


Figura 3: Gráfico 3 - Acurácia da validação em cada treinamento. Cada treinamento consistiu em 50 épocas da rede 1 seguidas de 50 épocas da rede 2

MSE médio por treinamento

O MSE médio variou no pior caso de 3,85 até no melhor caso de 3,47 (treinamento 10). Os testes revelaram melhoria em relação ao modelo anterior em que o melhor caso foi um MSE médio de 3,8.

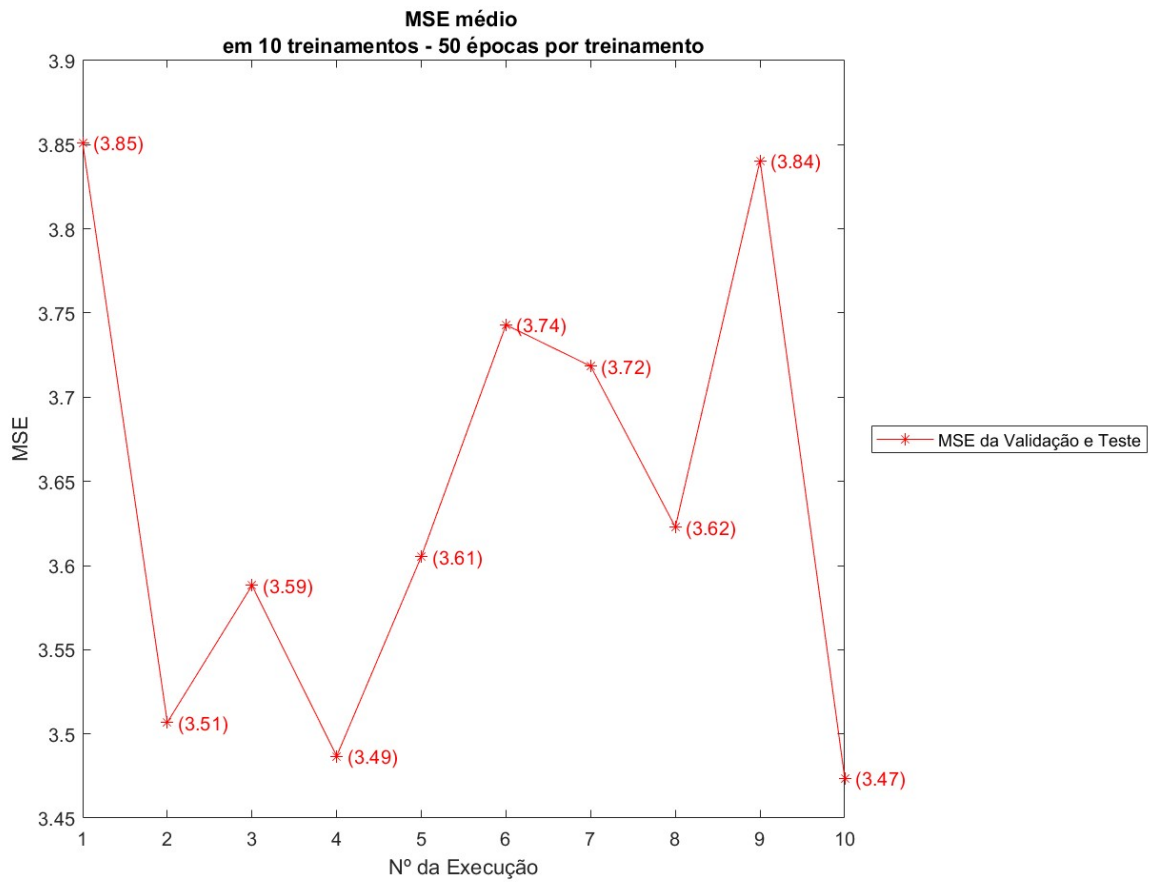


Figura 4: Gráfico 4 - MSE da validação em cada treinamento. Cada treinamento consistiu em 50 épocas da rede 1 seguidas de 50 épocas da rede 2

Acertos vs Erros

A relação Acertos vs Erros ao longo dos treinamentos também melhorou bastante em relação ao treinamento sem cooperação em que a quantidade de erros era maior que a quantidade de acertos. No modelo anterior tanto os erros da validação e do teste foram maiores do que os acertos. Nesta nova rede com cooperação a quantidade de acertos foi bem maior que a quantidade de erros.

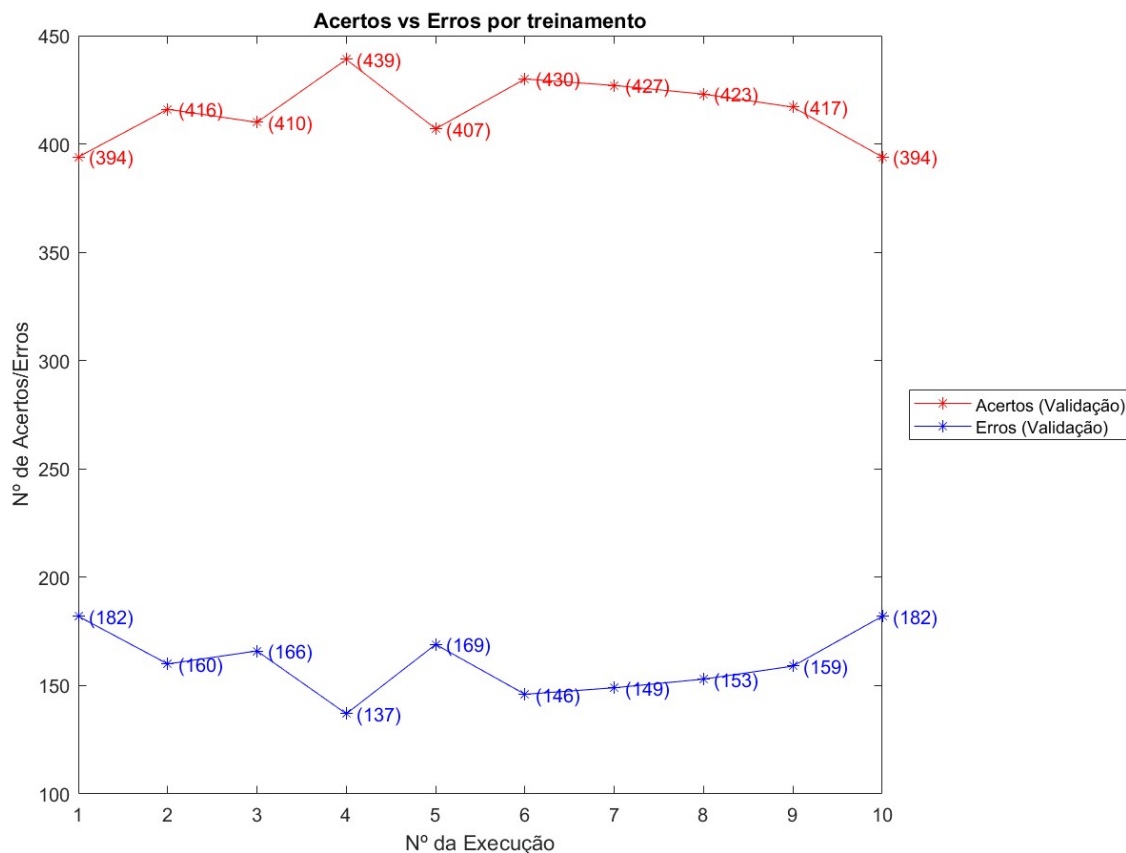


Figura 5: Gráfico 5 - Acertos e erros em cada treinamento. Cada treinamento consistiu em 50 épocas da rede 1 seguidas de 50 épocas da rede 2

Conclusão

1. Das alterações feitas na modelagem das redes e código em relação à atividade anterior, entendo que foram importantes:

- a alteração na quantidade de grupos utilizados nas amostras passando de 40 para 160;
- alteração no código de aplicação da FFT;
- alteração nas quantidades de neurônios das camadas de entrada e oculta;
- tamanho das bases foi ampliada e a fase de validação e teste foi agrupada em uma única fase.

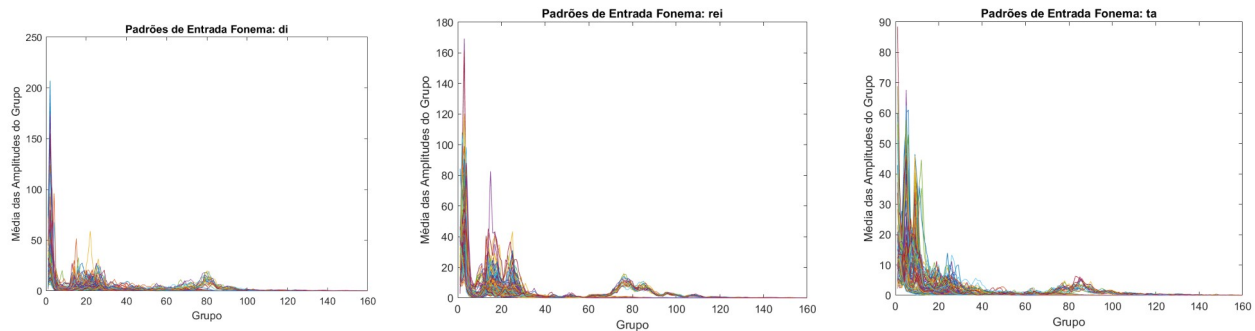
2. Sobre os resultados, a virada positiva em que resultados passaram a ser consistentes com a teoria exposta nas aulas foi bastante animadora. As redes cooperadas além de entregar resultados válidos, os resultados foram bem superiores a implementação anterior de RNA do tipo MLP.

3. Devido ao prazo curto após a finalização do código não consegui melhorar a transição entre as redes cooperadas a fim de diminuir o salto do MSE ao passar de uma rede para outra. Pelas mesmas razões não consegui colocar redes de tipo diferente como uma RNA do tipo RBF para cooperar na resolução do problema.

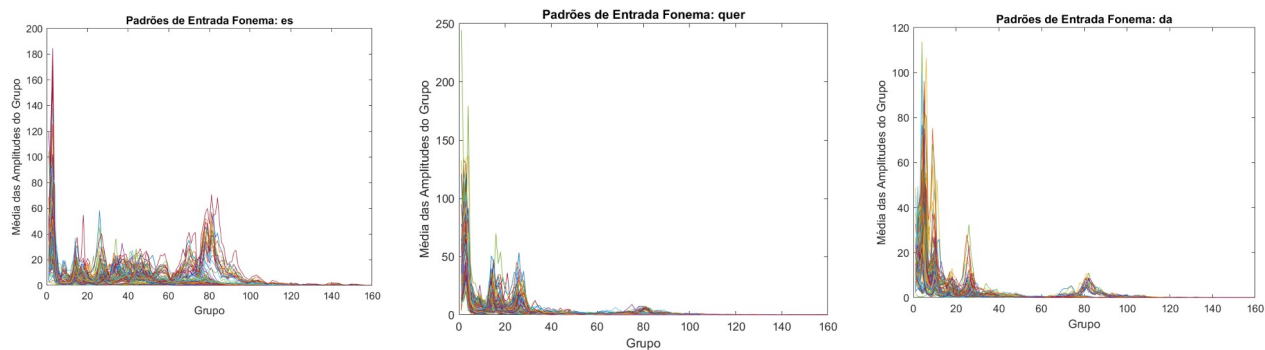
Anexo I – Padrão por sílaba

Através da aplicação da FFT nos arquivos de áudio foram gerados um gráfico para cada sílaba.

Palavra: DIREITA



Palavra: ESQUERDA



Anexo II - Detalhamento da execução dos programas Matlab

A partir do código da atividade anterior, fiz adaptações para RNAs cooperadas utilizando método MLP. A solução consiste em:

Módulo	O que faz	Saídas
preProc_Fonemas.m	1. Lê os arquivos *.wav dentro da pasta fonemas; 2. Define as variáveis: quantidade de amplitudes agrupadas por média (padraoTam); quantidade fixa de padrões por fonema (padraoPorFonema); 3. Aplica FFT nas amostras; 4. Divide as bases para serem processadas.	arquivo: preProc.mat Padrões de cada sílaba do modelo. Um padrão para cada sílaba.
training_Fonemas.m	1. Define variáveis: eta, função de ativação, épocas, ; 2. Chama em um laço na quantidade dos treinamentos as funções: mlp_fit (treinamento da rede); predict (validação). 3. Gera os gráficos resultantes 4. Armazena os dados das variáveis no arquivo MelhorRede.mat	arquivo: MelhorRede.mat Gráficos com os resultados
mlp_fit.m	Na quantidade épocas definidas aplica o cálculo dos pesos, dos erros, utiliza a função de ativação e a taxa de aprendizagem, faz propagação dos pesos nas camadas sobre a base de treinamento e acumula o erro quadrático. Se cooperada inicializa os pesos e o bias da camada de saída com zeros.	
predict.m	Na base de validação compara os resultados de saída com os de referência e assim acumula os resultados de acurácia, acertos e erros.	
grafico.m	Gera os gráficos dos dados acumulados nos processamentos anteriores.	