

Introdução

Leandro Bezerra Marinho

Disciplina: Aprendizagem de Máquina

Programa de Pós-Graduação em Ciência da Computação (PPGCC)

Instituto Federal do Ceará (IFCE)

6 de março de 2015

Resumo

- 1 Objetivos
- 2 Motivações
- 3 Aprendizagem de Máquina
 - Tipos de Aprendizado
 - Redes Neurais Artificiais
 - MLP
 - RBF
 - ELM
- 4 MLM
 - MLM-NN
- 5 Metodologia
 - Base de Dados
 - Experimentos
- 6 Resultados
- 7 Conclusões

- Verificar a aplicação da MLM no reconhecimento de atividades.
- Comparar o desempenho da MLM com as redes MLP, ELM e RBF.
- Selecionar o conjunto de atributos que contribuam com a tarefa de identificação de atividades.
- Analisar a performance da MLM com a distância Euclidiana e Manhattan.
- Mostrar uma nova abordagem da MLM baseada nos vizinhos mais próximos.

- O RAH pode ser usado na segurança, marketing, entretenimento, saúde, entre outros.
- Presença de sensores em dispositivos móveis.
- Mais acessível e menos intrusivo.
- AM permite construir modelos baseados em pouco conhecimento prévio sobre a tarefa de interesse.

Mitchell (1997) define AM como:

“A capacidade de melhorar o desempenho na realização de alguma tarefa por meio da experiência.”

Tipos de Aprendizado

- **Supervisionado:** O conjunto de exemplos possui entrada e saída.
- **Não-Supervisionado:** Agrupa-se um conjunto de exemplos não classificados por regularidades.
- **Por reforço:** Dado sucesso ou insucesso global em uma sequência de ações, determina-se qual é a mais desejável em cada situação.

Redes Neurais Artificiais

- Funcionamento inspirado no neurônio biológico.
- Perceptron
 - Só resolve problemas linearmente separáveis
- MLP, ELM e RBF.

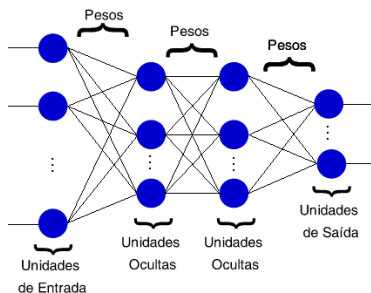


Figura: Modelo de uma Rede Neural Artificial.

MLP (Perceptron de Múltiplas Camadas)

- Pode ter uma ou mais camadas intermediárias.
- A entrada da função de ativação é o produto interno dos vetores de entrada.
- Algoritmo de treinamento *Backpropagation*
 - *Forward*
 - *Backward*
 - Foi adotado como condição de parada o número máximo de iterações.
- Separa padrões de entrada com hiperplanos.
- Foi utilizada a tangente hiperbólica como função de ativação.

RBF (Redes de Funções de Base Radial)

- Funções de Base Radial
 - Função gaussiana.
- Possui apenas uma camada oculta.
- A entrada da função de ativação é a distancia euclidiana entre os vetores de entrada e de pesos.
- Separa padrões de entrada com hiper-elipsóides.

ELM (Máquina de Aprendizado Extrema)

- Regra simples de aprendizado.
- Número de neurônios na camada oculta é maior que a dimensão do vetor de entrada.
- Pesos da camada oculta são gerados aleatoriamente.
- Pesos da camada de saída são ajustados para minimizar o EQM.
- Foi utilizado a tangente hiperbólica como função de ativação.

Mapeamento entre configurações geométricas dos pontos no espaço de entrada e saída.

- Formulação
- Regressão entre Distâncias
- Estimativa da saída
- Algoritmos de Treinamento e Teste
- MLM-NN

Dado N pontos de entrada \mathbf{X} e as saídas correspondentes \mathbf{Y} , queremos estimar f a partir dos dados com o modelo.

$$\mathbf{Y} = f(\mathbf{X}) + \mathbf{R}.$$

onde \mathbf{R} é a matriz de resíduos.

Regressão entre Distâncias

- Para uma seleção de K pontos de referência dos dados de entrada e suas respectivas saídas.
- $\mathbf{D}_x \in \mathbb{R}^{N \times K}$ contêm as distâncias entre os N pontos de entrada e os K pontos de referência.
- $\Delta_y \in \mathbb{R}^{N \times K}$ contêm as distâncias entre os N pontos de saída e as saídas dos K pontos de referência.
- O modelo para estimar g é portanto
$$\Delta_y = g(\mathbf{D}_x) + \mathbf{E},$$
onde \mathbf{E} é a matriz de resíduos.

- Assumindo que g é linear, o modelo torna-se

$$\Delta_y = \mathbf{D}_x \mathbf{B} + \mathbf{E}.$$

As colunas da matriz \mathbf{B} correspondem aos coeficientes para as K respostas.

- A matriz \mathbf{B} pode ser estimada através do método dos mínimos quadrados:

$$\hat{\mathbf{B}} = (\mathbf{D}_x^T \mathbf{D}_x)^{-1} \mathbf{D}_x^T \Delta_y.$$

- Dado um ponto de teste \mathbf{x} e o vetor de distância $\mathbf{d}(\mathbf{x}, \mathcal{R})$ entre este ponto e os K pontos de referência.
- As distâncias correspondentes entre a saída desconhecida \mathbf{y} e as saídas conhecidas dos pontos referências são estimadas por

$$\hat{\delta}(\mathbf{y}, \mathcal{T}) = \mathbf{d}(\mathbf{x}, \mathcal{R})\hat{\mathbf{B}}$$

Regressão entre Distâncias

- O vetor $\hat{\delta}(\mathbf{y}, \mathcal{T})$ fornece uma estimativa da configuração geométrica entre \mathbf{y} e o conjunto de pontos de referência \mathcal{T} .

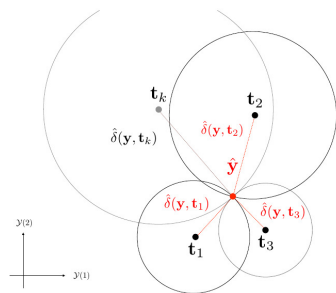


Figura: Procedimento para obter estimativa

- A estimativa da saída \mathbf{y} pode ser formulado como um problema de otimização.
- A solução pode ser alcançada usando o método gradiente descendente ou o algoritmo Levenberg-Marquardt sobre a equação:

- $$\hat{\mathbf{y}} = \arg \min_y \sum_{k=1}^K ((\mathbf{y} - \mathbf{t}_k)' (\mathbf{y} - \mathbf{t}_k) - \hat{\delta}^2(\mathbf{y}, \mathbf{t}_k))^2. \quad (1)$$

Algorithm 1 Algoritmo de Treinamento da MLM

Require: Conjunto de treinamento \mathcal{X} , \mathcal{Y} e K .

Ensure: $\hat{\mathbf{B}}$, \mathcal{R} e \mathcal{T} .

- 1: Aleatoriamente selecione K pontos de referência, \mathcal{R} , de \mathcal{X} e suas saídas correspondentes, \mathcal{T} , de \mathcal{Y} ;
 - 2: Calcule \mathbf{D}_x : A matriz de distância entre \mathcal{X} e \mathcal{R} ;
 - 3: Calcule $\mathbf{\Delta}_y$: A matriz de distância entre \mathcal{Y} e \mathcal{T} ;
 - 4: Calcule $\hat{\mathbf{B}} = (\mathbf{D}_x^\top \mathbf{D}_x)^{-1} \mathbf{D}_x^\top \mathbf{\Delta}_y$;
-

Algorithm 2 Algoritmo de Teste da MLM

Require: $\hat{\mathbf{B}}, \mathcal{R}, \mathcal{T}$ e \mathbf{x} .

Ensure: $\hat{\mathbf{y}}$.

- 1: Calcule $\mathbf{d}(\mathbf{x}, \mathcal{R})$;
 - 2: Calcule $\hat{\delta}(\mathbf{y}, \mathcal{T}) = \mathbf{d}(\mathbf{x}, \mathcal{R})\hat{\mathbf{B}}$;
 - 3: Use \mathcal{T} e $\hat{\delta}(\mathbf{y}, \mathcal{T})$ para encontrar uma estimativa para \mathbf{y} . Isto pode ser conseguido pelo algoritmo Levenberg-Marquardt sobre a função custo na equação 1;
-

- A estratégia de encontrar a saída através do procedimento de otimização pode ser computacionalmente “pesada”.
- **Nova abordagem:** Estimar a saída \mathbf{y} baseado em V vizinhos mais próximos (*nearest neighbors*).
- Uma vez que as distâncias $\hat{\delta}(\mathbf{y}, \mathcal{T})$ tenham sido estimadas, utilizar as classes dos V pontos de referências mais próximos no espaço de saída a \mathbf{y} para escolher a classe do padrão de teste.
 - Moda

- Repositório UCI (Reconhecimento de Atividades Humanas Usando Smartphones)
- Utilizou-se sinais tratados do acelerômetro e giroscópio de um *Samsung Galaxy S II*
- **Treino:** 7352 amostras e 561 características.
- **Teste:** 2947 amostras e 6 classes.

Modelagem dos Dados para Classificação

Classe	Representação
Caminhando	1 0 0 0 0 0
Subindo Escada	0 1 0 0 0 0
Descendo Escada	0 0 1 0 0 0
Sentado	0 0 0 1 0 0
Em pé	0 0 0 0 1 0
Deitado	0 0 0 0 0 1

Tabela: Modelagem das Classes

Experimentos

- Atributos no domínio do tempo, frequência e ambos
- **Validação cruzada:** 10-*folds*
- MLP
 - **Taxa de aprendizagem:** 0.1
 - **Iterações:** 100
- RBF
 - **Gama:** 1, 0.1, 0.01 e 0.001
- MLM
 - **Distâncias:** Euclidiana e Manhattan
 - **K:** (10%, 20%, 30%, ..., 90%, 100%) dos dados de treinamento
- MLM-NN
 - **K:** (10%, 20%, 30%, ..., 90%, 100%) dos dados de treinamento
 - **Vizinhos:** 9, 15 e 25.

Sequência de neurônios da camada oculta para as RNAs.

Métodos	Número de Neurônios
MLP	10 : 50 : 960
ELM	300 : 50 : 2500
RBF	10 : 50 : 960

Tabela: A segunda coluna mostra o início, variação e fim para uma sequência de neurônios.

Os experimentos foram executados 10 vezes.

Métodos	Média	Desvio Padrão
ELM	0.9425	6.800×10^{-3}
RBF	0.9629	3.400×10^{-3}
MLP	0.9080	5.600×10^{-3}
MLM-E	0.9676	3.800×10^{-3}
MLM-M	0.9528	2.700×10^{-3}
MLM-NN-9	0.9654	0
MLM-NN-15	0.9654	0
MLM-NN-25	0.9654	0

Tabela: Taxa média de acerto e desvio padrão com atributos no domínio do tempo.

Métodos	Média	Desvio Padrão
ELM	0.9209	2.600×10^{-3}
RBF	0.9532	8.700×10^{-3}
MLP	0.8889	1.700×10^{-2}
MLM-E	0.9541	5.900×10^{-3}
MLM-M	0.9339	5.700×10^{-3}
MLM-NN-9	0.9535	2.341×10^{-16}
MLM-NN-15	0.9535	2.341×10^{-16}
MLM-NN-25	0.9535	2.341×10^{-16}

Tabela: Taxa média de acerto e desvio padrão com atributos no domínio da frequência.

Métodos	Média	Desvio Padrão
ELM	0.9369	7.500×10^{-3}
RBF	0.9658	3.200×10^{-3}
MLP	0.9027	1.940×10^{-2}
MLM-E	0.9683	7.800×10^{-3}
MLM-M	0.9617	3.500×10^{-3}
MLM-NN-9	0.9715	1.170×10^{-16}
MLM-NN-15	0.9715	1.170×10^{-16}
MLM-NN-25	0.9715	1.170×10^{-16}

Tabela: Taxa média de acerto e desvio padrão utilizando **todos os atributos**.

Resultados

Taxa média de acerto com atributos no domínio do tempo.

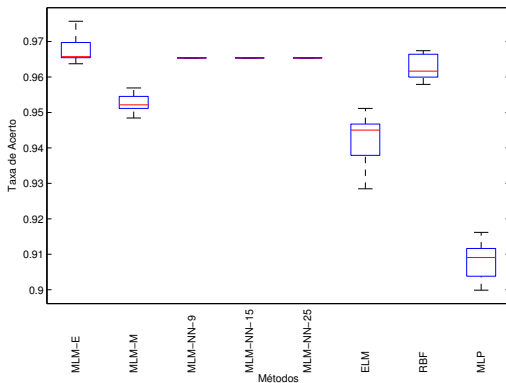


Figura: Taxa média de acerto com atributos no domínio do *tempo*.

Resultados

Taxa média de acerto com atributos no domínio da *frequência*.

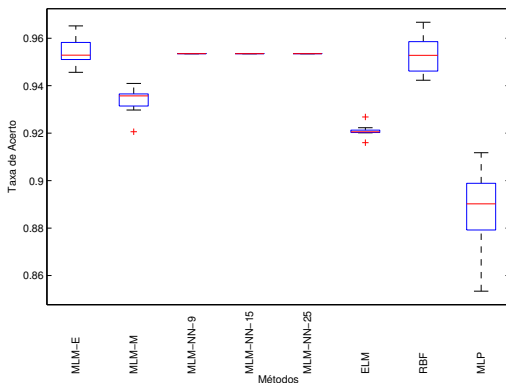


Figura: Taxa média de acerto com atributos no domínio da frequência.

Resultados

Taxa média de acerto com *todos os atributos*.

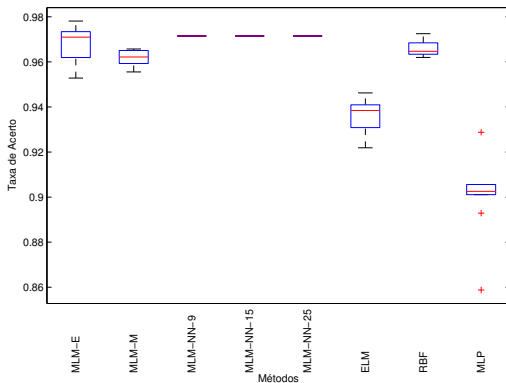


Figura: Taxa média de acerto com todos os atributos.

- Constatou-se que atributos no domínio do tempo foram mais importantes como discriminantes que os correspondentes no domínio da frequência.
- No geral, a MLM atingiu as maiores taxas de acerto.
- A RBF com seleção aleatória dos centros a partir dos dados, produz resultados bem próximos ou equivalentes aos alcançados pela rede MLM.
- Resultados obtidos com o uso da distância Manhattan foram levemente inferiores aos obtidos com distância euclidiana.
- A MLM-NN apresentou resultados promissores, principalmente considerando o aumento da velocidade na etapa de teste.

Referências Bibliográficas I



CARVALHO, André Carlos, P.L.F. de; BRAGA, Antônio P.; LUDEMIR, T. B.

Redes Neurais Artificiais: Teoria e aplicações
LTC Editora, 2007.



CORMEN, Thomas H.; STEIN, Clifford; RIVEST, Ronald L.; LEISERSON, Charles E.

Introduction to Algorithms
3. ed. The MIT Press, 2009.



HAYKIN, SIMON S.


Redes Neurais: Princípios e Práticas
2. ed. Bookman Companhia ED, 2002.




MITCHELL, Thomas M.


Machine Learning
1. ed. McGraw-Hill, 1997.

Referências Bibliográficas II

 RUSSEL, Stuart; NORVING, Peter.
Inteligência Artificial um enfoque Moderno.
2. ed. Pearson Education, 2004.

 VALENÇA, M. J.
Fundamentos das Redes Neurais: exemplos em Java
Editora Livro Rápido, 2010.

 ALLEN, Felicity R; AMBIKAIRAJAH, Eliathamby; LOVELL, Nigel H;
CELLER, B. G.
Classification of a known sequence of motions and postures from
accelerometry data using adapted gaussian mixture models 2006.

 ANGUIA, Davide; GUIO, Alessandro; ONETO, Luca; PARRA,
Xavier; R., Jorge L.
Human Activity recognition on smartphones using a multiclass
hardware-friendly support vector machine. 2012.

Referências Bibliográficas III



BAO, Ling; INTILLE, Stephen S.

Activity recognition from user-annotated acceleration data 2004.



HUANG, Guang-Bin; ZHU, Qin-Yu; SIEW, Chee-Kheong.

Extreme learning machine: Theory and applications., 2003.



HUYNH, Tâm; BLANKE, Ulf; SCHIELE, Bernt.

Scalable recognition of daily activities with wearable sensors., 2007.



SILVA, Vin De; TENENBAUM, Joshua B.

Global versus local methods in nonlinear dimensionality reduction., 2003.



SOUZA JUNIOR, A. H.; CORONA, F.; MICHÉ, Y.; LENDASSE, A.; BARRETO, G.; SIMULA, O.

Minimal learning machine: A new distance-based method for supervised learning., 2013.

 TAMURA, Shin'ichi; TATEISHI, Masahiko.

Capabilities of a four-layered feedforward neural network: four layers versus three. 1997.

Obrigado pela atenção!