

Relatório do Trabalho 1 de Aprendizado de Máquina

Fábio Beranizo Fontes Lopes
Escola de Artes, Ciências e Humanidades (EACH)
Universidade de São Paulo (USP)
Email: f.lopes@usp.br

Este relatório tem como objetivo mostrar os procedimentos e resultados do Trabalho 1 da disciplina de Aprendizado de Máquina. O propósito dessa tarefa foi testar diferentes algoritmos em atividades de reconhecimento biométrico. Duas técnicas de aprendizado supervisionado foram escolhidas para comparação: Multilayer Perceptron (MLP) e Support Vector Machines (SVM).

As seções a seguir explicam as atividades de pré-processamento do conjunto de dados, o treinamento dos modelos e uma discussão dos resultados obtidos.

I. BASE DE DADOS

A base de dados utilizada neste trabalho foi a Homologous Multi-modal Traits Database da Shandong University (SDUMLA-HMT) [1]. Este conjunto de dados contém faces de 106 indivíduos em 7 diferentes ângulos. Estas imagens foram capturadas com diferentes poses, expressões faciais, acessórios e iluminações, assemelhando-se a uma aplicação real.

Cada amostra deste conjunto é rotulada com o indivíduo que aparece na foto, tornando possível aplicar técnicas de aprendizado supervisionado para classificação.

A. Pré-processamento

Originalmente, as imagens do SDUMLA-HMT possuem resolução 480x640 e estão no formato RGB. Antes de realizar a extração de características das imagens, a região de interesse foi segmentada aplicando-se o Algoritmo Viola-Jones. Em uma segunda etapa, as imagens foram convertidas para escala de cinza a sua dimensão alterada para 128x128. A figura 1 ilustra uma das imagens após pré-processamento.

B. Transformada Wavelet

A extração de características das imagens utilizou a transformada Wavelet. Esta técnica de processamento de sinal realiza a análise de um dado sinal no domínio do tempo e frequência [2]. Uma função Wavelet pode ser descrita como uma função que apresenta média zero e obedece a Equação:

$$\int_{-\infty}^{\infty} \Psi(t) dt = 0$$

Para utilizar a transformada Wavelet em imagens 2D as linhas e colunas são tratadas como sinais independentes. Ao final da decomposição Wavelet 2D são geradas 4 subimagens. Três delas (HL, LH, HH) contém os detalhes horizontais,

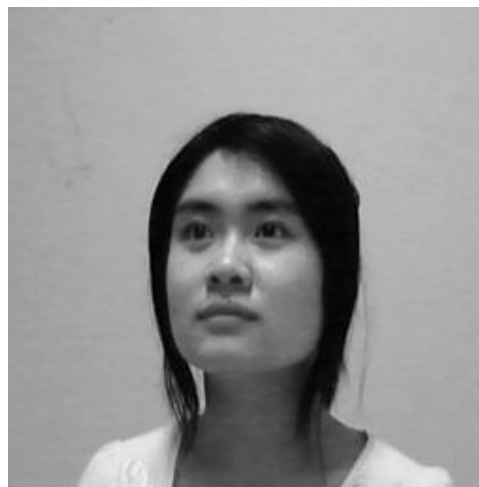


Fig. 1. Amostra de imagem de SDUMLA-HMT após pré-processamento.

verticais e diagonais da imagem original. Uma outra (LL) é uma aproximação da imagem original.

Ao final deste processo são gerados coeficientes que podem ser utilizados para representar a imagem original. Os coeficientes da imagem LL são geralmente os mais utilizados, pois apresentam menor dimensão e boa representatividade [3].

II. CLASSIFICADORES

Duas técnicas de aprendizado supervisionado foram implementadas para este trabalho: MLP e SVM. Estas técnicas não serão descritas em detalhe, pois já foram apresentadas em aula.

A. MLP

Multilayer Perceptron é um tipo de rede neural artificial feedforward. Neste trabalho, o treinamento das redes foi realizado utilizando a estratégia de backpropagation. O método do gradiente conjugado (de Polak-Ribière) foi o escolhido para otimizar os modelos, com a taxa de aprendizado definida pelo método de bisseção. Na seção V será mostrado que a utilização deste método permitiu acelerar a convergência dos treinamentos quando comparado à taxa de aprendizado fixa.

O critério de parada do treinamento da MLP foi dado pelo erro de validação. Esta estratégia reserva algumas amostras não utilizadas em treino, e computa o erro de validação destes dados a cada época. Quando o valor do erro de validação não

diminuir por 5 épocas¹ seguidas, o treinamento é interrompido. Trata-se de uma estratégia para minimizar overfitting (pouca capacidade de generalização dos modelos).

MLPs são sensíveis à alta-dimensionalidade dos dados. Portanto, antes do treinamento foi necessário aplicar uma Análise de Componentes Principais (PCA) aos coeficientes obtidos da transformada Wavelet. Este procedimento selecionou um número de componentes de forma que a variância dos dados fosse pelo menos 90%¹ do valor original. O resultado final definiu aproximadamente 70 valores para utilização no treinamento, variando de acordo da função wavelet, sub-bandas e nível de decomposição.

Uma das vantagens da MLP é o fato de lidar naturalmente com problemas multiclass. Neste trabalho, dada uma imagem era necessário definir a qual dos N indivíduos ela pertence. Na rede neural isso se reflete com a existência de N neurônios na camada de saída, em que a ativação de cada um representa a identificação de um indivíduo.

B. SVM

Máquinas de Vetores Suporte são classificadores que efetuam a separação de amostras em um espaço, geralmente fazendo uso de uma função Kernel para realizar projeção de dados não linearmente separáveis no espaço original. Neste trabalho foi implementado o algoritmo Sequential Minimal Optimization (SMO), conforme visto em aula.

Diferentemente da MLP, SVMs são insensíveis à alta-dimensionalidade e não necessitam do uso de PCA. Por outro lado, este algoritmo apresenta o problema de ser originalmente para classificação binária, portanto, foi necessário o uso de estratégia de classificação multi-classe, neste caso: One-vs-all. Neste método é preciso treinar um classificador para cada um dos N indivíduos, classificando-os em +1 (é o indivíduo) e -1 (não é o indivíduo). O critério de decisão foi dado pela maior distância da amostra ao hiperplano de separação.

Três tipos de função de kernel foram implementados para este trabalho: linear, gaussiano de base radial (rbf) e polinomial. Os parâmetros C, gamma e degree (grau do polinômio) foram ajustados durante o treinamento por meio de uma busca em grid e k-fold cross-validation.

III. AVALIAÇÃO

A avaliação deste trabalho utilizou técnicas comuns à área de aprendizado de máquina: validação cruzada K-folds e um teste não paramétrico para comparação dos classificadores. As métricas utilizadas são: precisão, revocação e medida F1, adequadas à avaliação de problemas de classificação. Além disso, a métrica de acurácia foi utilizada na execução do teste não paramétrico, conforme será explicado na seção III-B.

A. Validação cruzada K-folds

Na validação cruzada K-folds² o conjunto de dados é dividido aleatoriamente em k subconjuntos de aproximadamente mesmo tamanho. O modelo é treinado com amostras de $k - 1$

TABLE I
VALORES CRÍTICOS PARA TESTE DE FRIEDMAN

n	k = 3	
	0,05	0,01
2	-	-
3	6,000	-
4	6,500	8,000
5	6,400	8,400
6	7,000	9,000
7	7,143	8,857
8	6,250	9,000
9	6,222	9,556
10	6,200	9,600

subconjuntos e validado no subconjunto não utilizado no treino (este processo se repete k vezes). A medida de acurácia dos modelos é obtida dividindo-se o número de classificações corretas de validação pelo total de amostras no conjunto de dados.

B. Teste de Friedman

O Teste de Friedman é um teste não paramétrico para comparar repetidas observações em mesmo objeto (ex: modelos computacionais) [4], [5]. A estatística de Teste para Friedman é um χ^2 com k-1 graus de liberdade. Quando o p-valor para este teste é pequeno ($< 0,05$), há evidência para rejeitar a hipótese nula. Para este trabalho, a hipótese nula declara que todos os modelos são equivalentes.

O método desse teste inicia-se a partir de uma matriz $X_{n \times k}$, em que n é o número de conjuntos de teste, k é o número de modelos computacionais, e cada elemento dessa matriz é um valor de acurácia para um modelo em um dos conjuntos.

No segundo passo, calcula-se uma matriz $r_{n \times k}$, que mostra o rank de um modelo para um dado conjunto (maior acurácia possui rank 1, segundo maior rank 2, e assim sucessivamente até o rank k). Os cálculos a seguir mostram como chegar à estatística de teste:

$$\bar{r}_j = \frac{1}{n} \sum_{i=1}^n r_{ij}$$

$$\bar{r} = \frac{1}{nk} \sum_{i=1}^n \sum_{j=1}^k r_{ij}$$

$$SS_t = n \sum_{j=1}^k \bar{r}_j - \bar{r}$$

$$SS_e = \frac{1}{n(k-1)} \sum_{i=1}^n \sum_{j=1}^k (r_{ij} - \bar{r})^2$$

$$Q = \frac{SS_t}{SS_e}$$

A Tabela I exhibe os valores críticos de Q para que a hipótese nula seja rejeitada.

¹Valor arbitrário, baseado em empirismo

²Em inglês, K-folds cross-validation

TABLE II
RANK DOS ALGORITMOS SVM LINEAR, SVM RBF E MLP PARA
COEFICIENTES WAVELET DA FAMÍLIA DAUBECHIES

Indivíduos	SVM Linear	SVM RBF	MLP
1-10			
11-20			
21-30			
31-40			
41-50			
51-60			
61-70			
71-80			
81-90			
91-100			
Média			

TABLE III
RANK DOS ALGORITMOS SVM LINEAR, SVM RBF E MLP PARA
COEFICIENTES WAVELET DA FAMÍLIA SYMLET

Indivíduos	SVM Linear	SVM RBF	MLP
1-10			
11-20			
21-30			
31-40			
41-50			
51-60			
61-70			
71-80			
81-90			
91-100			
Média			

IV. EXPERIMENTO E RESULTADOS

Para gerar diferentes conjuntos de dados e acelerar o tempo de treinamento dos modelos as imagens de SDUMLA-HTM foram divididas em 10 conjuntos mutuamente exclusivos de 10 indivíduos.

Os resultados estão separados por função wavelet mãe (Daubechies e Symlets). Em todos os experimentos foi utilizado o nível de decomposição 3 e sub-banda LL, que produzem menor número de componentes e possuem boa representatividade [3].

Embora o kernel polinomial tenha sido implementado, resultados preliminares não foram satisfatórios. Assim o treinamento foi interrompido e os resultados não serão exibidos.

A. Daubechies

A função db2 do MATLAB foi o wavelet da família Daubechies adotado para os resultados apresentados a seguir. A Tabela II mostra o rank dos algoritmos SVM (kernel linear e RBF) e MLP nos 10 conjuntos de teste. Este rank é utilizado no teste não paramétrico, conforme explicado na seção III-B.

O valor da estatística de Friedman para estas medidas foi $Q =$, abaixo do valor crítico 6,2³. A seção V apresenta algumas considerações sobre este resultado.

Com o objetivo de ajudar a avaliar o desempenho dos modelos, precisão, revocação e medida-F1 são apresentadas na tabela V. Para cada conjunto de teste é apresentada a média das 3 medidas para as 10 classes (indivíduos), além do valor médio para todos conjuntos.

B. Symlets

A função sym5 do MATLAB foi o wavelet da família Symlets adotado para os resultados apresentados a seguir. A Tabela II mostra o rank dos algoritmos SVM (kernel linear e RBF) e MLP nos 10 conjuntos de teste. Este rank é utilizado no teste não paramétrico, conforme explicado na seção III-B.

O valor da estatística de Friedman para estas medidas foi $Q =$, abaixo do valor crítico 6,2. A seção V apresenta algumas considerações sobre este resultado.

³Obtido da Tabela I

TABLE IV
PRECISÃO, REVOCACÃO E MEDIDA F1 PARA COEFICIENTES WAVELET DA FAMÍLIA DAUBECHIES

Indivíduos	SVM Linear			SVM RBF			MLP		
	Precisão	Revocação	Medida-F1	Precisão	Revocação	Medida-F1	Precisão	Revocação	Medida-F1
1-10	0,61	0,65	0,61	0,43	0,48	0,42	0,32	0,33	0,29
11-20	0,53	0,60	0,53	0,39	0,46	0,39	0,01	0,08	0,02
21-30	0,68	0,73	0,68	0,33	0,36	0,32	0,21	0,20	0,15
31-40	0,50	0,55	0,50	0,42	0,46	0,41	0,40	0,42	0,47
41-50	0,40	0,48	0,41	0,13	0,20	0,14	0,03	0,11	0,04
51-60	0,11	0,20	0,12	0,01	0,11	0,02	0,13	0,21	0,14
61-70									
71-80									
81-90									
91-100									
Média									

TABLE V
PRECISÃO, REVOCACÃO E MEDIDA F1 PARA COEFICIENTES WAVELET DA FAMÍLIA SYMLETS

Indivíduos	SVM Linear			SVM RBF			MLP		
	Precisão	Revocação	Medida-F1	Precisão	Revocação	Medida-F1	Precisão	Revocação	Medida-F1
1-10	0,72	0,78	0,73	0,30	0,37	0,30	0,18	0,22	0,19
11-20	0,64	0,66	0,62	0,32	0,36	0,30	0,30	0,35	0,29
21-30									
31-40									
41-50									
51-60									
61-70									
71-80									
81-90									
91-100									
Média									

V. DISCUSSÃO

Os resultados exibidos na seção IV mostram que, em geral, as implementações dos algoritmos SVM e MLP não apresentaram bom desempenho. O algoritmo que obteve melhores resultados foi o SVM de kernel linear, que consistentemente obteve precisão, revocação e medida-F1 superior aos outros algoritmos.

A diferença entre os resultados do SVM Linear e SVM RBF possivelmente deve-se ao fato deste último depender do ajuste de mais parâmetros (C e gamma) que o primeiro (C). O kernel RBF é sensível a escolha do gamma, e este foi testado com valores muito baixos (1/número de componentes). Esta escolha pretendia diminuir o *bias* do modelo, no entanto deve ter causado aumento da variância.

O mau desempenho da MLP pode ter relação com a decisão de adicionar tempo de processamento como um dos critério de parada. Resultados preliminares mostraram que utilizar apenas o erro de validação como critério de parada poderia tornar o treinamento inviável para o hardware e tempo disponíveis.

REFERENCES

- [1] M. Lab. Sdumla-hmt. [Online]. Available: <http://mla.sdu.edu.cn/sdumla-hmt.htm>
- [2] D. M. M. d. Costa, "Ensemble baseado em métodos de kernel para reconhecimento biométrico multimodal," Ph.D. dissertation, Universidade de São Paulo, 2011.
- [3] C. S. Burrus, R. A. Gopinath, and H. Guo, "Introduction to wavelets and wavelet transforms: a primer," 1997.
- [4] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *Journal of Machine learning research*, vol. 7, no. Jan, pp. 1–30, 2006.

- [5] N. Settouti, M. E. A. Bechar, and M. A. Chikh, "Statistical comparisons of the top 10 algorithms in data mining for classification task," *International Journal of Interactive Multimedia and Artificial Intelligence, Special Issue on Artificial Intelligence Underpinning*, vol. 4, pp. 46–51, 2016.