A Comparative Analysis of Attribute Reduction Algorithms applied to Wet-Blue Leather Defects Classification

Willian Paraguassu Amorim
Statistics and Computer Science Dept.
UFMS
São Carlos, Brazil
Email: paraguassuec@gmail.com

Hemerson Pistori
Biotechnology Dept.
UCDB
Campo Grande, Brazil
Web page: www.gpec.ucdb.br/pistori

Manuel Antonio Chagas Jacinto
Embrapa Pecuária Sudeste
São Carlos, Brazil
Email: jacinto@cppse.embrapa.br

***Abstract*—This paper presents an attribute reduction comparative study on four linear discriminant analysis techniques: FisherFace, CLDA, DLDA and YLDA. The attribute reduction has been applied to the problem of leather defect classification using four different classifiers: C4.5, KNN, Naive Bayes and Support Vector Machines. Results and analyses on the performance of correct classification rates as the number of attributes were reduced are reported.**

***Keywords*-attributes reduction; linear discriminant analysis; defect detection.**

I. INTRODUCTION

Wet-blue is the name given to the bovine leather after the first stage of the tanning process. The inspection of this leather, usually visual, is crucial in determining the destination of the leather and its price. A computer system is being developed to automate this process, and in this paper some preliminary results related to the attribute reduction module of this system are presented. As class information is available for this problem, this work concentrates on Fisher Linear Discriminant Analysis (LDA) based approaches. In previous studies, this problem also was shown to be prone to the singularity in the inter-class spreading matrix, which further let us to sharpen our choice of techniques that can handle this problem.

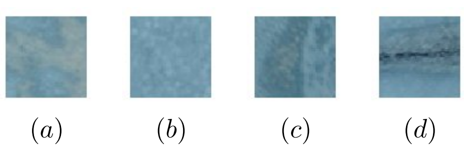

Figure 1. Examples of wet-blue leather defects: (a) scabies (b) ticks (c) hot-iron marks and (d) cuts

The defect detection has been modelled as a supervised learning problem, with 8 classes and 160 attributes. The classes correspond to 6 types of defect, some of them illustrated in Figure 1, one class for no-defect and another class for the leather background. The attributes were selected based on previous study and include co-occurrence matrices, interaction maps, Gabor filter banks and two different color space based approaches. Four attribution reduction techniques: FisherFace, CLDA, DLDA and YLDA, and four classifiers were evaluated. The results show that, for this specific problem, a good trade-off can be achieved with only 24 attributes, CLDA reduction and C4.5 classification. A 91.47% classification rate was reached using this configuration. The next sections present a brief literature review, experimental setup, results, analysis, conclusion and future studies.

II. LITERATURE REVIEW

Fisher Linear Discriminant Analysis (FLDA) has been extended in many different ways during the last decades, including approaches that enable non-linear analysis through the use of the Kernel trick [1]. Kernel based LDA approaches have not been used in this work due to its inherently high computational cost, when compared to the more traditional approaches. In [2], the performance of CLDA, DLDA and YLDA has already been compared; however, the problem studied was face recognition.

The feasibility of FLDA for defect detection problems that depends on texture analysis has been studied in [3]. Extension of FLDA that can handle singular covariance matrices have also been compared in [4].

III. EXPERIMENTAL SETUP

The images used in this experiment were all taken in real-world situations from Brazilian tanneries. Ground-truth classifications of defective regions were provided by field specialists. In total, 50 different wet-blue leather pieces, from Nelore and Hereford cattle, were used to construct the training and testing dataset. For each of the 8 classes: background, no-defect, hot-iron marks, ticks, open cuts, closed cuts, scabies and botfly larvae, 2,000 samples, consisting of a 40x40 pixel windows, were collected. From each sample, 160 texture and color based attributes were extracted: 12 from color (HSB and RGB space), 7 from Interaction Maps, 126 from co-ocurrence matrices and 15 from Gabor filter banks. As for the classifiers, the default parameters for

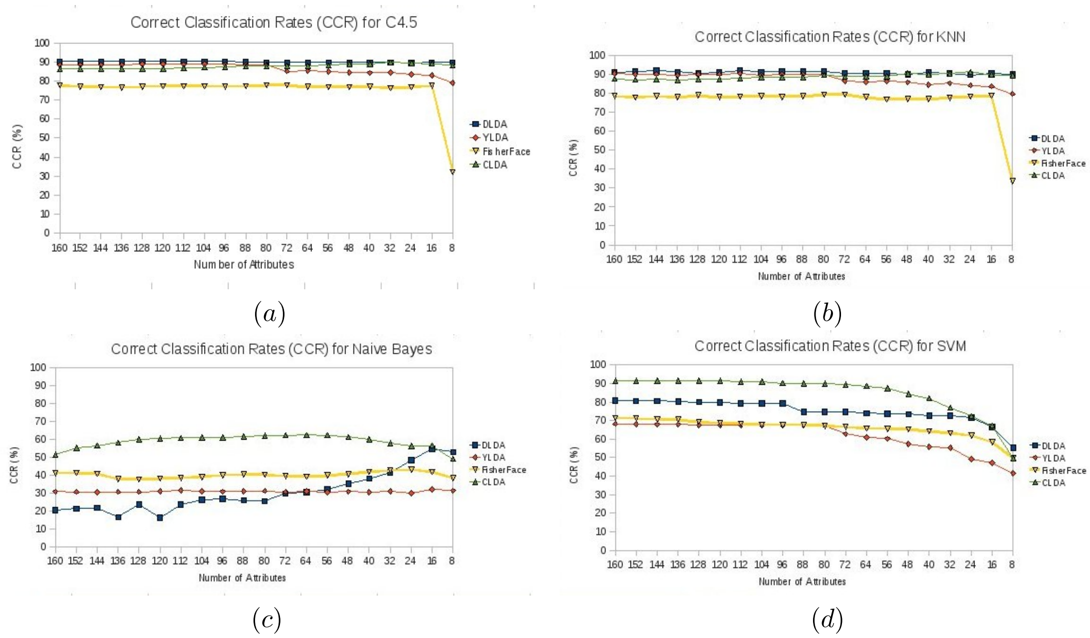

Figure 2. Correct Classification results for C4.5 (a), KNN (b), Naive Bayes (c) and SVM (d)

the C4.5, K-Nearest Neighbors (KNN), Support Vector Machines (SVM) and Naive Bayes implementations available in the Weka [5] software were used.

Starting at 160, the attributes were progressively reduced in 4% at a time, using each of the attribute reduction techniques: FisherFaces, CLDA, KLDA and YLDA. A 3-fold cross-validation approach, with two repetitions, were used to produce correct classification rates for C4.5, KNN, SVM and NaveBayes.

A. *Results and Analysis*

Figure 2 presents the correct classification rates (CCR) results for each classifier. The largest CCR was 92.33%, using 144 attributes, KNN and DLDA. Except for Nave Bayes, all classifiers could reach CCRs above 90% using as many as 96 attributes. Support Vector Machine performance starts to degrade much faster than KNN and C4.5 and presents better CCRs when coupled with CLDA reductions; in contrast, the other two perform better when DLDA is used. In all cases, except for Naive Bayes, Fisherfaces and YLDA presented inferior results. With as little as 8 attributes, KNN and C4.5 could reach 90.34% and 89.69% CCR, respectively.

IV. Conclusion

A comparison using four LDA reduction strategies and four classifiers applied to the problem of wet-blue defect detection was provided. The results pointed to better performance of CLDA and DLDA, in keeping correct classification rates as attributes were reduced. KNN reached the best CCR, using both the maximum and the minimum amount of attributes, in the test. However, classification times for KNN are known to be much higher than C4.5, which could justify the utilization of C4.5 for this problem. It is worth noting that for the system under development, training time is not as critical as the classification time. Future studies include the comparison to Kernel based LDA and the use of different attribute sets in order to measure how they affect the results.

Acknowledgment

This work has received financial support from UCDB, EMBRAPA and FINEP. Some of the authors have received scholarships from CNPQ and CAPES.

References

[1] H. Wu and D. Huang, “Kernel fisher discriminant analysis using feature vector selection for fault diagnosis,” in *Work. on Intell. Informat. Tech. Applications*, 2008.

[2] C. Thomaz, E. Kitani, and D. Gillies, “A Maximum Uncertainty LDA-based approach for Limited Sample Size problems - with application to Face Recognition,” *Journal of the Braz. Comp. Soc.*, vol. 12, no. 1, June 2006.

[3] M. Fukumi, S. Karungaru, and Y. Mitsukura, “Feature generation by simple-flda for pattern recognition,” in *Int. Conf. on Computational Intelligence for Modelling*, 2005.

[4] J. Ye and T. Xiong, “Null space versus orthogonal linear discriminant analisys,” in *Proc. Int. Conf. on Machine learning, Pittsburgh, Pennsylvania*, 2006.

[5] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*. San Francisco: Morgan Kaufmann, 2005.