On the design of a CADS for Shoulder Pain Pathology

K. López de Ipiña, M.C. Hernández, E. Martínez, C. Vaquero

Computational Intelligence Group, Txagorritxu Hospital Rehabilitation Service and Osakidetza Basque Health Service, LEIA Foundation CDT

5th International Conference on Hybrid Artificial Intelligence Systems, Donostia-San Sebastián, June 2010

Introduction

- A musculoskeletal disorder is a condition of the musculoskeletal system which consists in that part of it is injured continuously over time.
- In Primary Health Care Services, shoulder disorders are very common.
- These disorders cause pain and limit the ability to perform many routine activities and affect about 15-25% of the general population.

Medical Method

- Several clinical tests (e.g. Hawkins, Neer, Yergason, Speed) have been described to aid diagnosis of shoulder disorders.
  - However, research acknowledges a lack of concordance in clinical assessment, even between musculoskeletal specialists.
- It seems clear that the main reasons are mechanical and the factors of risk are:
  - Age, Feminine sex, Physical repetitive work, Psychological factors, ...
- The clinical history is not very effective in the accomplishment of the diagnosis.

Medical Method

- Four differential diagnoses of the shoulder pain (*Duke's University of Sports Medicine*):
  1. Impingement of the rotator cuff
  2. Frozen shoulder
  3. Glenohumeral instability
  4. Degenerative osteoarthritis

Goal

- A common problem in medical decision support system design is the limited number of samples that causes numerical problems:
  - The singularity of the covariance matrices.
- In this paper we apply robust covariance estimators to overcome this problem.

Database

- A sample of 200 patients (Rehabilitation Service for SP, Txagorritxu's Hospital).
- Exploration of the shoulder:
  - Empty can test, Hawkins-Kennedy test, crossed adducing, apprehension test and subescapular test.

Database

- Each data item has 30 variables and the collected information is:
  - Age
  - Sex
  - Arm dominance
  - Features of the pain and its intensity by an analogical visual scale
  - Received treatments
  - Limitation of the mobility and its measure by goniometry
  - Existence of previous history
  - Type of work
  - Sports activity
- These variables have been parameterized to be measured and analyzed.

Database

- For validation purposes, we have split the data sample into:
  - Training set (100 samples)
  - Test set (70 samples)

Database

- We have 6 classes:
  - Three main categories:
    - Frozen shoulder
    - Subacromial Impingement Syndrome
    - Rotator Cuff Anomaly
  - One mixed category: Subacromial Syndrome and Rotator Cuff Anomaly
  - Two new categories established by the specialist

Database

- There have been excluded:
  - Fractures
  - Luxations
  - Degenerative glenohumeral osteoarthritis (radiological diagnosis)

Classification and clustering methods

- Several classification methods:
  - MLE (Maximum Likelihood Estimation)
  - MDM (Minimum Distance to Means)
  - FIS (Fisher's Linear Discriminant)
  - QDF (Quadratic Discriminant Function)
  - KNN (K-Nearest Neighbor)
  - DT (Decision Trees)
  - Two different approaches of SVM (Support Vector Machines), SVM1 and SVM2
- Unsupervised clustering: K-means algorithm
- Feature space dimension reduction: Principal Component Analysis (PCA) .

Regularized Covariance Matrix estimation methods

- Some classification methods use covariance matrices to build the discriminant functions.
- When the number of training samples is small, the covariance matrices of the classes are singular and it is impossible to compute the inverse matrix.
- To deal with this problem, we must obtain non singular matrices using one of the following:
  - matrices with common covariance, if most of the classes have similar covariance matrices, or
  - diagonal matrices obtained from the covariance matrices of each class

Regularized Covariance Matrix estimation methods

- Regularized Discriminant Analysis (RDA)
- LOOC estimation methods (LOOC)
- Bayesian LOOC (BLOOC) methods
- Mixed LOOC methods

Regularized Covariance Matrix estimation methods

- The distribution of the vectors in each class is given by a p-dimensional Gaussian distribution:

$$f\{x_{ij}|m_i, \Sigma_i\} = \frac{1}{\sqrt{(2\pi)^p \cdot |\Sigma_i|}} exp\left(-\frac{1}{2}(x_{ij} - m_i)^{\mathrm{T}} \cdot \Sigma_i^{-1} \cdot (x_{ij} - m_i)\right)$$

where

$$m_i = \frac{i}{n_i} \sum_{j=1}^{n_i} x_{ij}$$

$$\Sigma_i = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (x_{ij} - m_i) \cdot (x_{ij} - m_i)^{\mathrm{T}}$$

Regularized Covariance Matrix estimation methods

- The common covariance matrix is:

$$S = \frac{1}{L}\sum_{i=1}^{L}\Sigma_i$$

- The pooled covariance matrix is:

$$S = \frac{1}{n-L}\sum_{i=1}^{L}(n_i - 1)\cdot\Sigma_i \quad n = \sum_{i=1}^{L} n_i$$

- The trace matrix and the diagonal matrix are:

$$mTraza(\Sigma_i) = traza(\Sigma_i)\cdot I \qquad mDiag(\Sigma_i) = diag(\Sigma_i)\cdot I$$

Estimator	Covariance Matrix Estimation
RDA	$C_I^{RDA}(\lambda,\gamma) = (1-\gamma)\cdot C_i(\lambda) + \frac{\gamma}{p}\cdot mTraza(C_I(\lambda))$
LOOC1	$C_i(\alpha_i) = \alpha_{i1}\cdot mDiag(\sum_i) + \alpha_{i2}\cdot\sum_i + \alpha_{i3}\cdot S + \alpha_{i4}\cdot mDiag(S)$
LOOC2	$C_i(\alpha_i) = \begin{cases} (1-\alpha_i)\cdot mDiag(\sum_i) + \alpha_i\cdot\sum_i & 0\le\alpha_i\le 1 \\ (2-\alpha_i)\cdot\sum_i + (\alpha_i-1)\cdot S & 1\le\alpha_i\le 2 \\ (3-\alpha_i)\cdot S + (\alpha_i-2)\cdot mDi & 2\le\alpha_i\le 3 \end{cases}$
BLOOC1	$C_i(\alpha_i) = \begin{cases} (1-\alpha_i)\cdot\frac{mTraza(\sum_i)}{p} + \alpha_i\cdot\sum_i & 0\le\alpha_i\le 1 \\ (2-\alpha_i)\cdot\sum_i + (\alpha_i-1)\cdot S_p(t) & 1\le\alpha_i\le 2 \\ (3-\alpha_i)\cdot S + (\alpha_i-2)\cdot\frac{mTraza(\sum_i)}{p} & 2\le\alpha_i\le 3 \end{cases}$
BLOOC2	$C_i(\alpha_i) = \begin{cases} (1-\alpha_i)\cdot mDiag(\sum_i) + \alpha_i\cdot\sum_i & 0\le\alpha_i\le 1 \\ (2-\alpha_i)\cdot\sum_i + (\alpha_i-1)\cdot S_p & 1\le\alpha_i\le 2 \\ (3-\alpha_i)\cdot S + (\alpha_i-2)\cdot mDi & 2\le\alpha_i\le 3 \end{cases}$
Mixed-LOOC1	$C_i(\alpha_i) = \alpha_{i1}\cdot\frac{mTraza(\sum_i)}{p} + \alpha_{i2}\cdot mDiag(\sum_i) + \alpha_{i3}\cdot\sum_i + \alpha_{i4}\cdot\frac{mTraza(S)}{p} + \alpha_{i5}\cdot mDiag(S) + \alpha_{i6}\cdot S$
Mixed-LOOC2	$C_i(\alpha_i) = \alpha_i\cdot A + (1-\alpha_i)\cdot B$

Experimentation

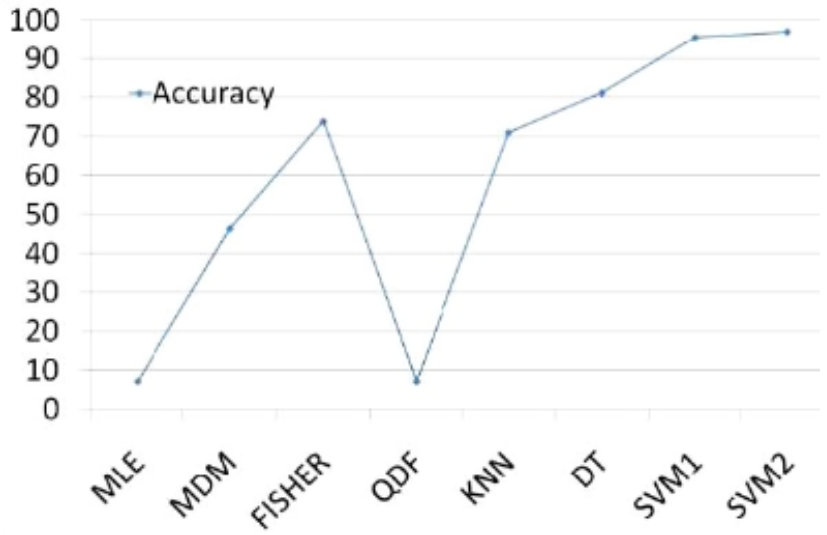

Figure 1: Accuracy for classification methods

Experimentation

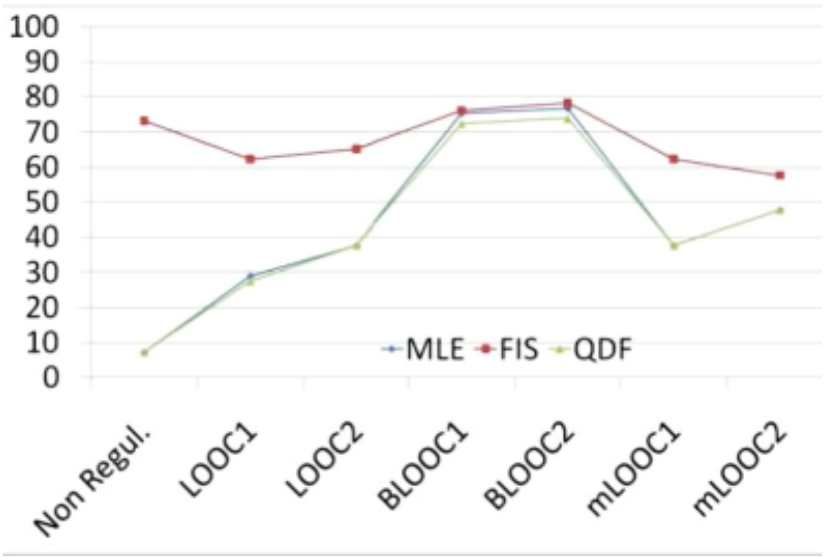

Figure 2: Accuracy of regularized Covariance Matrix estimation methods

Experimentation

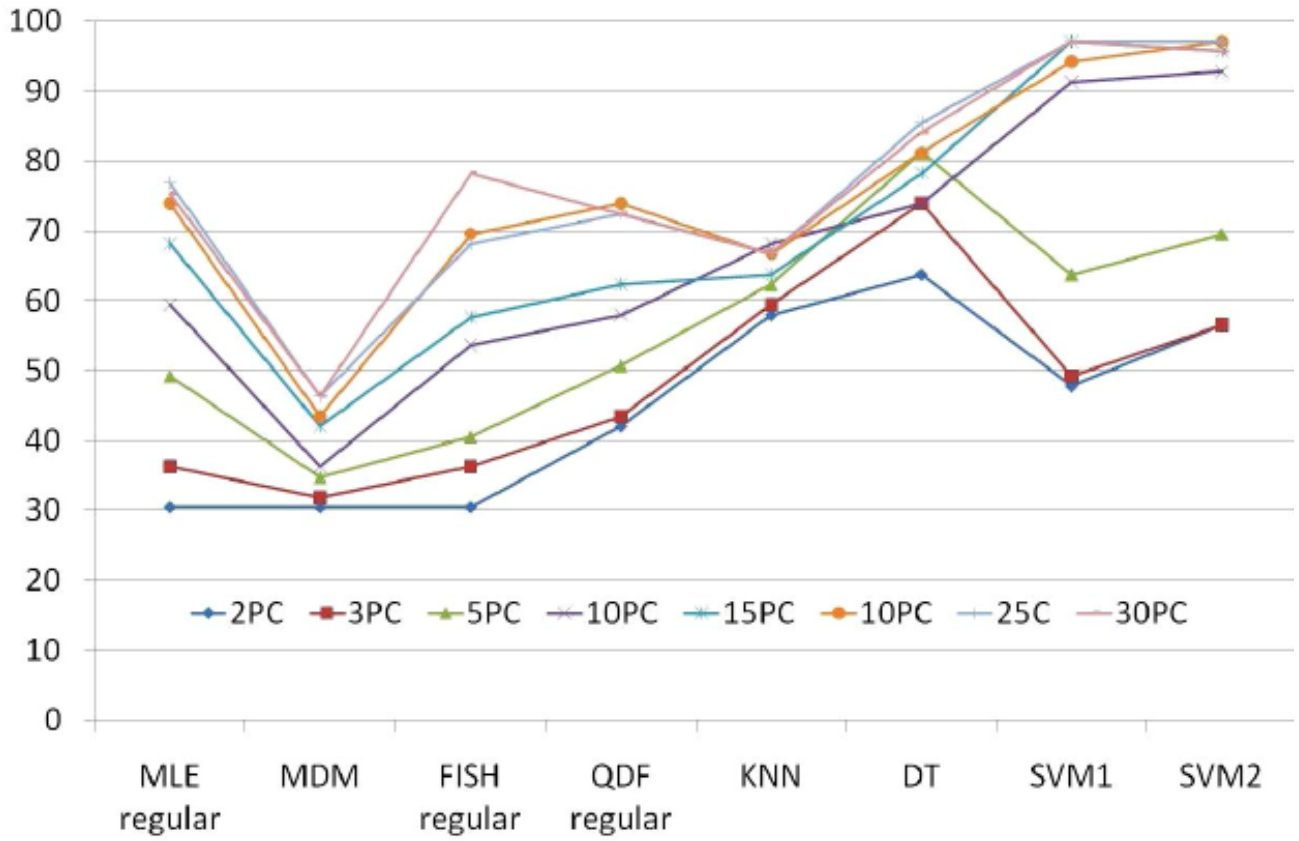

Figure 3: Accuracy of all methods and different number of Principal Components

Experimentation

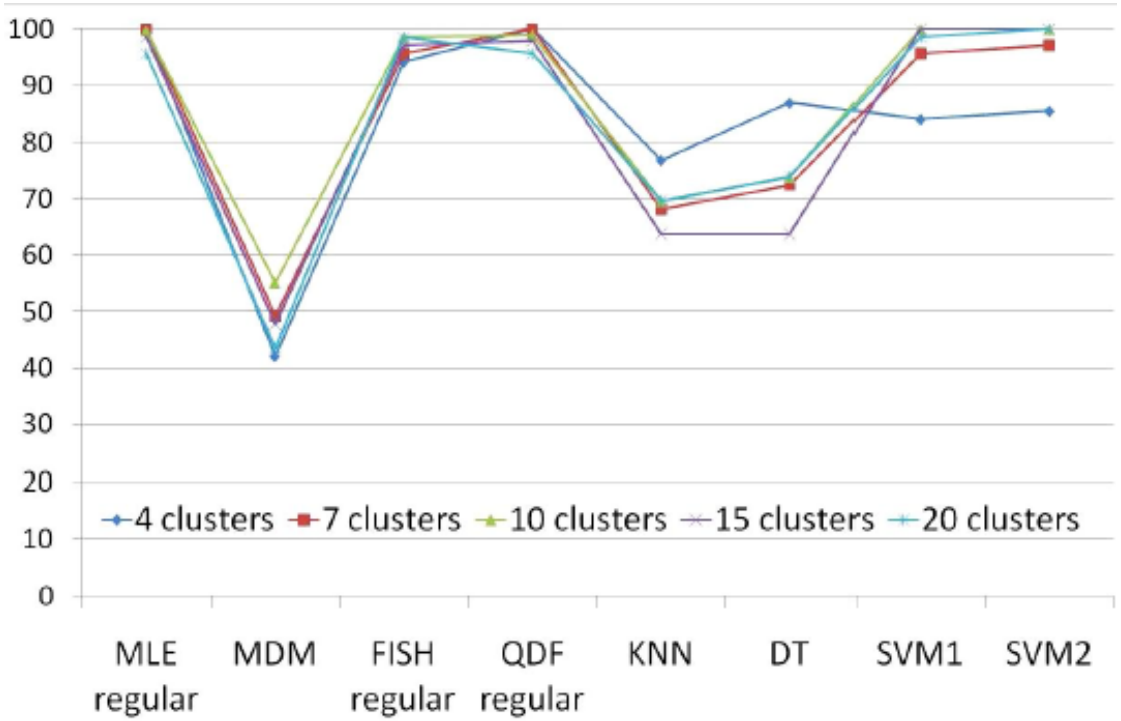

Figure 4: Accuracy of unsupervised classification

Discussion

- Finally, a diagnosis evaluation was carried out with the developed system based on the optimal classification method.
  - In this evaluation, the independent test (70 cases) was used.

Discussion

- The obtained results were supervised by the medical specialist who demonstrates a high degree of satisfaction not only regarding to the visual information provided by the automatic support system but also for the diagnosis results over the test cases.
    - The system has obtained good performance and results are very close to the diagnosis of the specialist.

Discussion

- The specialist prefers in any case the results obtained in supervised mode but new encouraging strategies for the diagnosis have been proposed based on the results of the unsupervised mode.

Conclusions

- The work presents a CADS for Shoulder Pain Pathology.
  - Medical method
  - Several classical classification paradigms improved by covariance estimation methods
- The system obtains good performance and results are very close to the diagnosis of the specialist.
- In future outlines, new covariance estimation methods will be developed and the system will be also improved with new supervised and non-supervised methods.

Acknowledgements

- The involved work has received financial support from the Spanish *Ministerio de Ciencia y Tecnología*, IBEROEKA 2005 plan and the Spanish *Ministerio de Ciencia y Tecnología*, project TSI2006-14250-C02-01.

Thanks