Large Margin Filtering for Signal Sequence Labeling

Rémi Flamary, Benjamin Labbé, Alain Rakotomamonjy

LITIS EA 4108, INSA-Université de Rouen
76800 Saint Etienne du Rouvray, France

Signal Sequence Labeling

Problem: Obtaining a label for each sample of a signal while taking into account the sequentiality of the samples.
Current approaches:
- Hidden Markov Models [1], Conditional Random Fields [2].
- Segment the signal (change detection [3]) and decide the label of the regions afterwards.

Our approach
- Take into account the temporal neighborhood of the sample in the decision (time-delay embedding).
- Jointly learn a temporal filter with the classifier: adapt to noise and delay.

Filter-SVM
- We jointly learn a sample classifier \((w, w_0)\) and a filtering \(F\) of the channels.
- Decision function for the \(j\)th sample of \(X\):
  \[ g_{ij}(X) = \sum_{m=1}^{d} w_{ij, m} X_{i, m} + w_0 \]  \( (4) \)
  where \(w\) and \(w_0\) are the parameters of the linear SVM classifier corresponding to a weighting of the channels.
- Optimal function \(g_{ij}(\cdot)\) obtained by minimizing:
  \[ J_{\text{SVM}}(F) = \frac{1}{2} \left\| \begin{array}{c} w \end{array} \right\|^2 + C \sum_{i=1}^{n} H(Y_i, g_{\cdot i}) + \frac{1}{2} \left\| f \right\|^2 \]  \( (5) \)
  w.r.t. \((F, w, w_0)\) where \(\lambda\) is a regularization term.

Filter-SVM Solver
- Cost non-convex but convex w.r.t. \(w\) and \(w_0\) when \(F\) is fixed.
- We define \(J(F)\) that is differentiable [4]:
  \[ J(F) = \min_{w} \frac{1}{2} \left\| \begin{array}{c} w \end{array} \right\|^2 + C \sum_{i=1}^{n} H(Y_i, g_{\cdot i})^2 \]
- We minimize:
  \[ J(F) + \frac{1}{2} \left\| f \right\|^2 \]
  w.r.t. \(F\) using a gradient descent along \(F\) and a line search to find the optimal step.

Definitions
- \(X \in \mathbb{R}^{N \times d}\) Feature matrix, \(d\) channels and \(N\) samples.
- \(x_{ij}\) value of channel \(i\) for the \(j\)th sample.
- \(y_i\) label of the \(j\)th sample.
- Filtering \(X\) w.r.t. \(F\)
  \[ \tilde{x}_{ij} = \sum_{m=1}^{d} F_{ij,m} x_{i, m} \]  \( (1) \)
  \( \tilde{F} \in \mathbb{R}^{d \times d}\) Filter matrix, \(d\) filters of length \(f\)
- \(n_0\) filter delay
- \(\parallel . \parallel\) is the Frobenius norm of a matrix.

Numerical Experiments on toy dataset
- \(n\) relabel discriminative signals with a switching mean \((-1, 1)\) among \(n\) relabel \(-d\).
- \(\sigma\) Gaussian noise and time-lags applied to the channels.
- \(f = 21\) and \(n_0 = 11\) corresponding to a good average filtering.
- \(\text{Test error is the average of 10 runs.} \)

Figures
- Figure 1: Sample classification vs Time window classification (at time \(t\)).
- Figure 2: Toy example for \(n\) relabel - \(n\) relabel - 1.
- Figure 3: Histograms of the samples for the 2 possible labels of a 1D signal (with / without filtering).
- Figure 4: Fourier Transform of the discriminative information and of the impulse response of learned filters.
- Figure 5: Test error for different \(v\) values and for different number of channels \(n\) relabel.
- Figure 6: Coefficients of \(W\) (left) and coefficients \(F\) weighted by \(w\) (right).

Window-SVM
- We learn a classifier \((W, w_0)\) for a window of samples (Time-Delay Embedding).
- Decision function for the \(j\)th sample of \(X\):
  \[ g_{Wij}(X) = \sum_{m=1}^{d} W_{ij,m} x_{i, m} + w_0 \]  \( (2) \)
  where \(W \in \mathbb{R}^{d \times d}\) and \(w_0 \in \mathbb{R}\) are the classification parameters and \(f\) is the size of the time-window.
- Optimal function \(g_{Wij}(\cdot)\) obtained by minimizing:
  \[ J_{\text{WSVM}}(W) = \frac{1}{2} \left\| W \right\|^2 + \sum_{i=1}^{n} H(Y_i, \tilde{g}_{\cdot i})^2 \]  \( (3) \)
  w.r.t. \((W, w_0)\) with \(H(Y_i, \tilde{g}_{\cdot i}) = \max(0, 1 - y_i g_{\cdot i}(X))\).

Conclusion
- On-line sequence labeling classifier.
- Better variable selection than classical SVM.
- Visualization of space/time discriminative maps.

Future works
- Non-linear SVM (kernel).
- Multi-task approach for \(F\).

Bibliography
- D. Capel, E. Moulines, and T. Ryden. 
  Arinfections in Hidden Markov Models. 
  Spriger, 2005.
- Y. Lavie, A. McCallum, and F. Pereira. 
  Conditional random fields: Probabilistic models for segmenting and labeling sequence data. 
- D. Denisy, M. Darg, and C. Duran. 
  An online kernel change detection algorithm. 
- JF Bonnans and A. Shapiro. 
  Optimization problems with perturbation : A guided tour. 

http://www.litislab.eu

{remi.flamary,benjamin.labb,b_alain.rakoto} @ insa-rouen.fr