



Time series gene expression data classification via L1-norm temporal SVM

Carlotta Orsenigo, Carlo Vercellis

**Dep. of Management, Economics and Industrial Engineering
Politecnico di Milano**

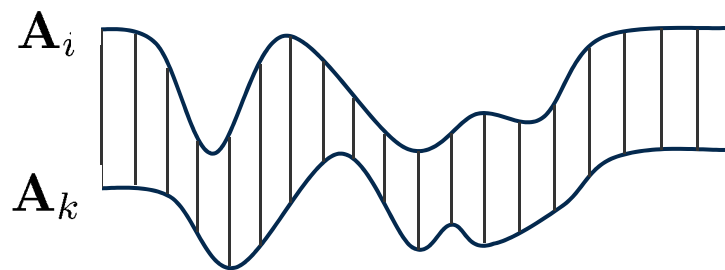


- Time series classification
- Dynamic time warping
- L1-norm support vector machines
- L1-norm temporal support vector machines
- Computational results

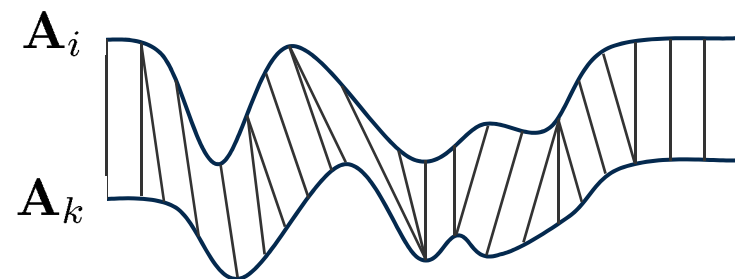


- Main motivation: growing number of experiments aimed at collecting and analyzing time series gene expression data
- Two examples:
 - ▣ categorization of genes based on their temporal evolution in the cell cycle
 - ▣ prediction of the clinical response to a drug
- Time series classification is a supervised learning problem aimed at labeling temporally structured univariate or multivariate sequences
- Several approaches have been proposed, based on
 - ▣ a two-stage procedure (most common paradigm)
 - ▣ the notion of time warping distance

- Time warping distance is an effective measure of similarity between pairs of time series
- It has proven to be more robust and versatile than the Euclidean distance:
 - ▣ it copes with sequences of variable length
 - ▣ performs shifts in the sequences to identify similar profiles with different phases
- The warping distance is usually evaluated by a dynamic optimization algorithm

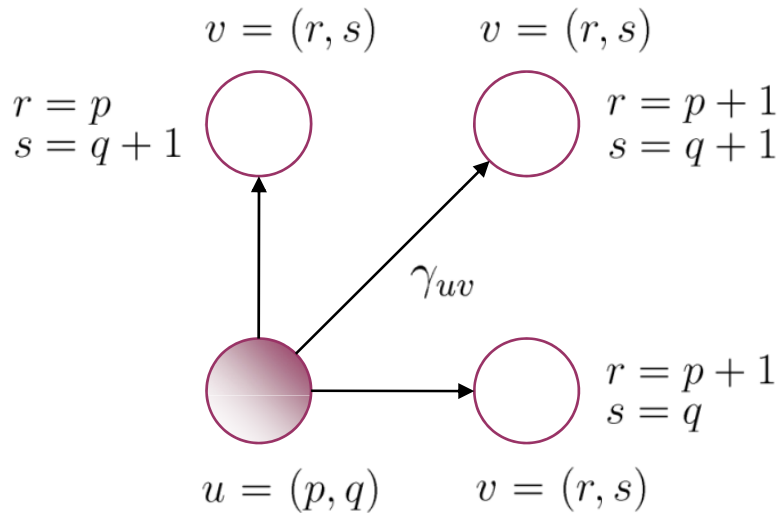


Euclidean distance

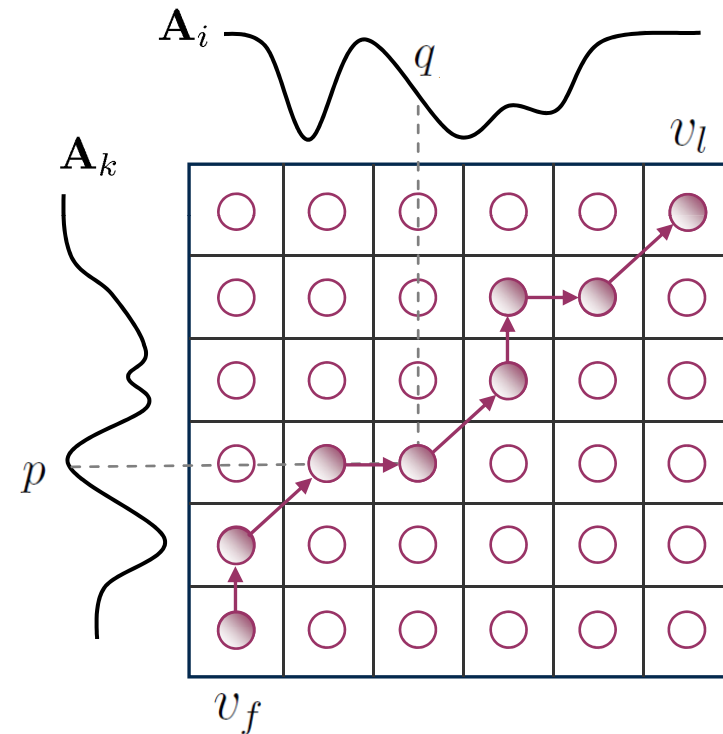


Warping distance

- The warping distance between A_i and A_k is defined as the length of the shortest *warping path* in a directed graph



Length of (u, v) : $\gamma_{uv} = (a_{i1r} - a_{k1s})^2$





- Let the input dataset be represented by a $m \times n$ matrix in which row $\mathbf{x}_i \in \mathbb{R}^n$ represents time series \mathbf{A}_i

L2-norm SVM

$$\min \quad \frac{1}{2} \|\mathbf{w}\|_2 + C \sum_{i=1}^m \xi_i$$

$$\text{s. t.} \quad y_i (\mathbf{w}' \mathbf{x}_i - b) \geq 1 - \xi_i \quad i \in \mathcal{M}$$

$$\xi_i \geq 0 \quad \forall i; \quad \mathbf{w}, b \text{ free}$$

L1-norm SVM

$$\min \quad \|\mathbf{w}\|_1 + C \sum_{i=1}^m \xi_i$$

$$\text{s. t.} \quad y_i (\mathbf{w}' \mathbf{x}_i - b) \geq 1 - \xi_i \quad i \in \mathcal{M}$$

$$\xi_i \geq 0 \quad \forall i; \quad \mathbf{w}, b \text{ free}$$

- We introduce the binary variables $p_i = \begin{cases} 0 & \text{if } \mathbf{w}' \mathbf{x}_i - b \geq 1 \\ 1 & \text{otherwise} \end{cases}$

- The following mixed-integer optimization problem can be formulated, where d_{ik} is the warping distance between \mathbf{A}_i and \mathbf{A}_k :

$$\min \quad \sum_{j=1}^n u_j + C \sum_{i=1}^m \xi_i + \delta \sum_{i=1}^m \sum_{k=i+1}^m d_{ik} r_{ik} \quad (L_1\text{-TSVM})$$

$$\text{s. t.} \quad y_i (\mathbf{w}' \mathbf{x}_i - b) \geq 1 - \xi_i \quad i \in \mathcal{M}$$

$$-u_j \leq w_j \leq u_j \quad j \in \mathcal{N}$$

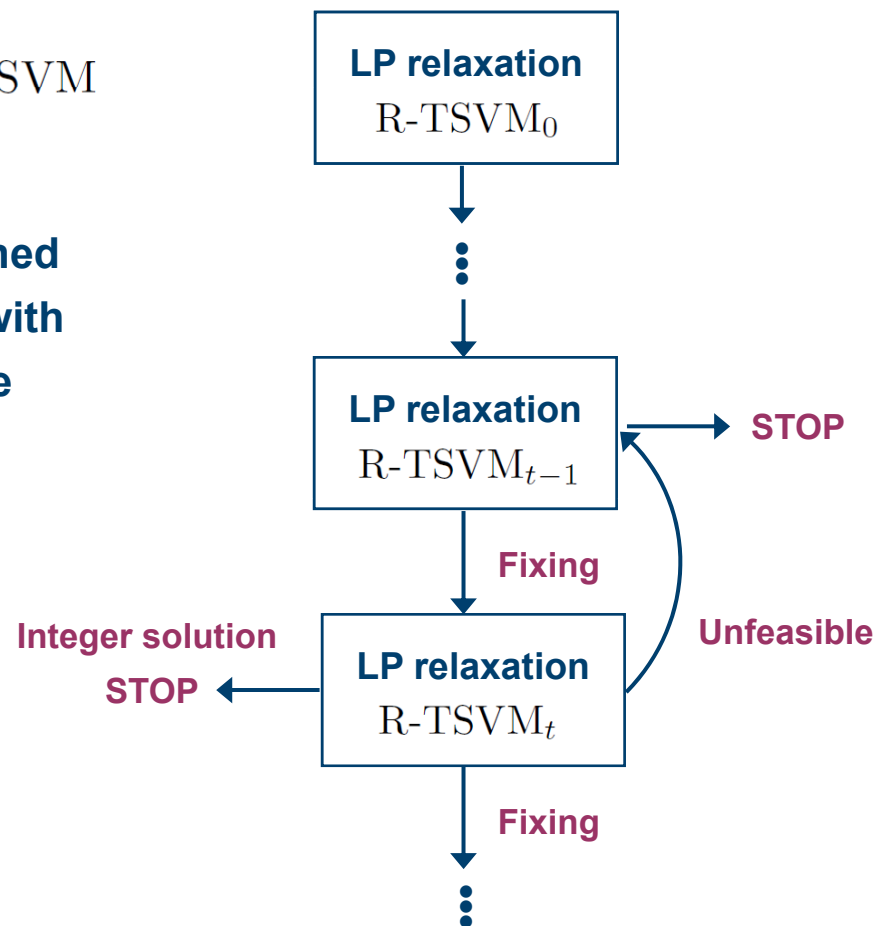
$$\frac{1}{S} \xi_i \leq p_i \leq S \xi_i \quad i \in \mathcal{M}$$

$$-r_{ik} \leq y_i (2p_i - 1) + y_k (2p_k - 1) \leq r_{ik} \quad i, k \in \mathcal{M}, i < k$$

$$u_j, \xi_i, r_{ik} \geq 0 \quad \forall i, j, k; \quad p_i \in \{0, 1\} \quad \forall i; \quad \mathbf{w}, b \text{ free}$$

- A feasible suboptimal solution to model L_1 -TSVM can be obtained by an approximate procedure, in which:

- the relaxation $R\text{-TSVM}_0$ of L_1 -TSVM is considered
- each problem $R\text{-TSVM}_t$ is obtained by fixing to 0 the binary variable with the smallest fractional value in the optimal solution of $R\text{-TSVM}_{t-1}$



- Two microarray time series gene expression datasets were considered:

Summary of gene expression time series datasets

Summary	Dataset	
	<i>Yeast</i>	<i>MS-rIFNβ</i>
Examples	388	52
Classes	Early G1 (67), Late G1 (136), S (77) G2 (54), M (54)	Good responder (33), Poor responder (19)
Time series length	17	[5,7]

- Five methods:

- L_1 -SVM
- SVM_{RBF}
- SVM_{DTW}
- k -NN_{EucL}
- k -NN_{DTW}

- Accuracy evaluation:

- five times 4-fold cross-validation
- on each training set 3-fold cross-validation for parameters tuning

Parameters values tested

Method	Parameters values
k -NN _{EucL}	$k = 2, 4, 6, 8, 10$
k -NN _{DTW}	
SVM _{RBF}	$C = 10^j, j \in [-1, 3]$
SVM _{DTW}	$\sigma = 10^j, j \in [-4, 2]$
L_1 -SVM	$C = 10^j, j \in [-1, 3]$
L_1 -TSVM	$\delta = 10^j, j \in [-1, 1]$



Classification accuracy (%) on the gene expression datasets

Dataset	Method					
	k -NN _{Eucl}	k -NN _{DTW}	SVM _{RBF}	SVM _{DTW}	L_1 -SVM	L_1 -TSVM
<i>Yeast</i>	68.5	51.8	73.3	73.7	72.4	73.9
<i>MS-rIFNβ</i>						
$t \in [0,1]$	83.8	76.9	82.7	84.2	76.9	80.8
$t \in [0,2]$	81.9	78.9	82.7	84.6	80.0	85.4
$t \in [0,3]$	82.7	75.0	81.9	75.4	78.5	83.8
$t \in [0,4]$	76.9	73.1	76.9	71.2	79.2	80.0
$t \in [0,5]$	75.8	69.2	71.5	78.5	79.6	80.8
$t \in [0,6]$	71.2	66.9	68.5	70.8	76.5	78.8



• Empirical remarks:

- 📖 L_1 -TSVM vs L_1 -SVM \Rightarrow increase in accuracy in [0.8%, 5.4%]
- 📖 on *MS-rIFN β* dataset the use of the warping distance appears promising (milder decrease in accuracy for L_1 -TSVM)
- 📖 on *Yeast* dataset L_1 -TSVM and SVM_{DTW} provided comparable results



- A new supervised learning method for time series gene expression classification has been proposed
- It relies on a mixed-integer optimization formulation which aims at improving the discrimination capability in time series classification problems
- Experiments performed on two datasets showed the effectiveness of the proposed method and the usefulness of the warping distance
- Future extensions:
 - 📖 test the novel technique on a wider range of time series datasets
 - 📖 investigate other time series similarity measures
 - 📖 study alternative heuristic procedures



- Baranzini, S., Mousavi, P., Rio, J., Caillier, S., Stillman, A., Villoslada, P., Wyatt, M., Comabella, M., Greller, L., Somogyi, R., Montalban, X., Oksenberg, J.: *Transcription-based prediction of response to IFN β using supervised computational methods*. PLoS Biology 3, 166-176 (2005)
- Cho, R.J., Campbell, M.J., Winzeler, E.A., Steinmetz, L., Conway, A., Wodicka, L., Wolfsberg, T.G., Gabrielian, A.E., Landsman, D., Lockhart, D.J., Davis, R.W.: *A genome-wide transcriptional analysis of the mitotic cell cycle*. Molecular Cell 2, 65-73 (1998)
- Cuturi, M., Vert, J.P., Birkenes, O., Matsui, T.: A kernel for time series based on global alignments. In: Proc. of ICASSP, pp. 413-416 (2007)
- Kadous, M.W., Sammut, C.: *Classification of multivariate time series and structured data using constructive induction*. Machine Learning 58, 179-216 (2005)
- Keogh, E., Ratanamahatana, C.A.: *Exact indexing of dynamic time warping*. Knowledge and Information Systems 7, 358-386 (2004)
- Mangasarian, O.L.: *Exact 1-norm support vector machines via unconstrained convex differentiable minimization*. Journal of Machine Learning Research 7, 1517-1530 (2006)
- Orsenigo, C., Vercellis, C.: *Combining discrete SVM and fixed cardinality warping distances for multivariate time series classification*. Pattern Recognition 43, 3787-3794 (2010)
- Vapnik, V.: *The nature of statistical learning theory*. Springer Verlag, New York, USA (1995)