Large-scale Time-series Clustering with k-ARs

Zuogong Yue[†]

Victor Solo[†]

[†] University of New South Wales, Australia

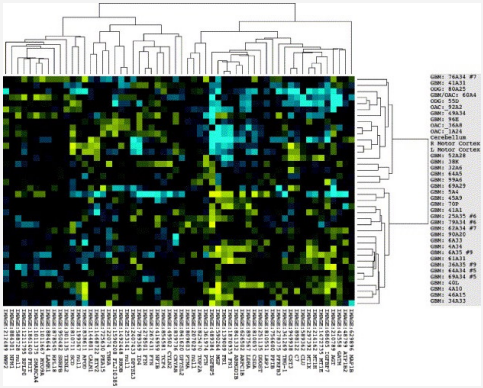
4-8 May · ICASSP 2020 · Virtual Conference

Outline

Motivation

Example: analysis of microarray gene expression data

clustering: grouping genes with similar profiles

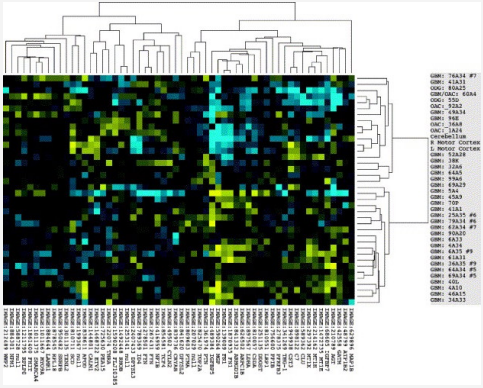


From Gollub and Sherlock (2006)

Motivation

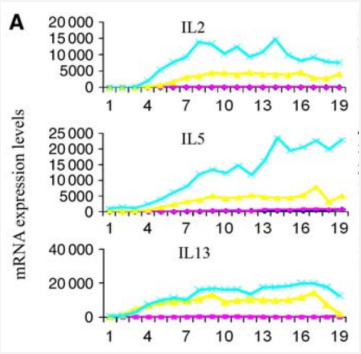
Example: analysis of microarray gene expression data

clustering: grouping genes with similar profiles



From Gollub and Sherlock (2006)

time series in microarray analysis:



From He et al. (2012)

Introduction: time-series clustering

Problem: clustering

- a set of N time series $\mathbb{X} \triangleq \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$, into
- K clusters $\mathbb{C} = \{C_1, \ldots, C_K\}$ $(K < N)$.

Introduction: time-series clustering

Problem: clustering

- a set of N time series $\mathbb{X} \triangleq \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$, into
- K clusters $\mathbb{C} = \{C_1, \ldots, C_K\}$ $(K < N)$.

Literature:

- survey: Liao (2005), Aghabozorgi et al. (2015), etc.;
- Mixture AR model for multi-modal signals: Wong and Li (2000)
- Mixture ARMA for time-series clustering: Xiong and Yeung (2002, 2004)

Outline

Preliminaries: AR models

Given $\mathbf{x} = \{x_1, x_2, \ldots, x_T\}$, the AR($p$) model is

$$x_t = \phi_0 + \sum_{j=1}^{p} \phi_j x_{t-j} + e_t, \quad t = 1, \ldots, T.$$

And the log-likelihood is

$$\ln P(\mathbf{x} \mid \Phi, \sigma^2) \propto -\frac{T^*}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \|e\|^2.$$

Preliminaries: mixture AR models

Mixture AR models (MxARs):

$$\ln P(\mathbf{x} \mid \Phi, \Sigma, \Pi) = \sum_{k=1}^{K} \pi_k \ln P(\mathbf{x} \mid \Phi_k, \sigma_k^2),$$

where

- mixing coefficient $\sum_{k=1}^{K} \pi_k = 1$;
- binary $z_k \in \{0, 1\}$ and then $P(z_k = 1) = \pi_k$.

Preliminaries: mixture AR models

Mixture AR models (MxARs):

$$\ln P(\mathbf{x} \mid \Phi, \Sigma, \Pi) = \sum_{k=1}^{K} \pi_k \ln P(\mathbf{x} \mid \Phi_k, \sigma_k^2),$$

where

- mixing coefficient $\sum_{k=1}^{K} \pi_k = 1$;
- binary $z_k \in \{0, 1\}$ and then $P(z_k = 1) = \pi_k$.

For the whole data set $\mathbb{X}$:

$$\ln P(\mathbb{X} \mid \Phi, \Sigma, \Pi) = \sum_{n=1}^{N} \ln P(\mathbf{x}_n \mid \Phi, \Sigma, \Pi).$$

Preliminaries: EM algorithm for MxARs

Maximize $\ln P(\mathbb{X} \mid \Phi, \Sigma, \Pi)$:

- E-step:

$$\begin{aligned}
\mathbf{e}_{nk} &= \mathbf{y}_n - \mathbf{X}_n \Phi_k, \\
r_{nk} &= \frac{\pi_k P(\mathbf{x}_n \mid \Phi_k, \sigma_k^2)}{\sum_{k=1}^{K} \pi_k P(\mathbf{x}_n \mid \Phi_k, \sigma_k^2)}, \\
P(\mathbf{x}_n \mid \Phi_k, \sigma_k^2) &= \left(2\sigma_k^2\right)^{-T^*/2} \exp\left[-\|\mathbf{e}_{nk}\|^2 / \left(2\sigma_k^2\right)\right];
\end{aligned}$$

- M-step:

$$\begin{aligned}
\pi_k &= (1/N) \textstyle\sum_{n=1}^{N} r_{nk}, \\
\Phi_k &= \left(\textstyle\sum_{n=1}^{N} r_{nk} \mathbf{X}_n^T \mathbf{X}_n\right)^{-1} \left(\textstyle\sum_{n=1}^{N} r_{nk} \mathbf{X}_n^T \mathbf{y}_n\right), \\
\sigma_k^2 &= \left(\textstyle\sum_{n=1}^{N} r_{nk} \|\mathbf{e}_{nk}\|^2\right) / \left(T^* \textstyle\sum_{n=1}^{N} r_{nk}\right).
\end{aligned}$$

Outline

k-ARs: EM algorithm

Motivated by the relation between GMMs and k-Means, taking $\sigma_k^2 \to 0$, it yields

$$r_{nk} \to \begin{cases} 1 & \text{if } k = \arg\min_{k'} \|\mathbf{e}_{nk'}\|^2, \\ 0 & \text{otherwise.} \end{cases}$$

k-ARs: EM algorithm

Motivated by the relation between GMMs and k-Means, taking $\sigma_k^2 \to 0$, it yields

$$r_{nk} \to \begin{cases} 1 & \text{if } k = \arg\min_{k'} \|\mathbf{e}_{nk'}\|^2, \\ 0 & \text{otherwise.} \end{cases}$$

The index set of cluster k can be introduced as

$$\mathcal{I}_k = \{n : r_{nk} = 1\}.$$

k-ARs: EM algorithm

- E-step:

$$\mathbf{e}_{nk} = \mathbf{y}_n - \mathbf{X}_n \Phi_k,$$

$$r_{nk} = \begin{cases} 1 & \text{if } k = \operatorname{argmin}_{k'} \|\mathbf{e}_{nk'}\|^2, \\ 0 & \text{otherwise.} \end{cases}$$

- M-step:

$$\pi_k = (1/N) \sum_{n=1}^{N} r_{nk} = |\mathcal{I}_k|/N,$$

$$\Phi_k = \left(\sum_{n \in \mathcal{I}_k} \mathbf{X}_n^T \mathbf{X}_n \right)^{-1} \left(\sum_{n \in \mathcal{I}_k} \mathbf{X}_n^T \mathbf{y}_n \right),$$

$$\sigma_k^2 = \left(\sum_{n \in \mathcal{I}_k} \|\mathbf{e}_{nk}\|^2 \right) / \left(T^* |\mathcal{I}_k| \right).$$

k-ARs: initialization

- Φ_k: randomly choose K time series and estimate their subject-specific AR(p) coefficients;
- σ_k^2: all set to 1;
- π_k: all set to $1/K$.

k-ARs: stopping criterion

A straightforward criterion could be

$$\max\left\{\|\Phi^{+}-\Phi\|_{F}, \|\Sigma^{+}-\Sigma\|_{2}, \|\Pi^{+}-\Pi\|_{2}\right\} < \epsilon.$$

k-ARs: fast computation

Consider QR decomposition: $\mathbf{X}_n = Q_n R_n$ with Q_n: $T^* \times p$, R_n: $p \times p$

Let $\mathbf{y}_{Q_n} \triangleq Q_n^T \mathbf{y}_n$, then

- $\Phi_k = \left(\sum_{n \in \mathcal{I}_k} R_n^T R_n\right)^{-1} \left(\sum_{n \in \mathcal{I}_k} R_n^T \mathbf{y}_{Q_n}\right),$
- $\|\mathbf{e}_{nk}\|^2 = \|\mathbf{y}_n\|^2 - \|\mathbf{y}_{Q_n}\|^2 + \|\mathbf{y}_{Q_n} - R_n \Phi_k\|^2.$

Outline

Numerical examples

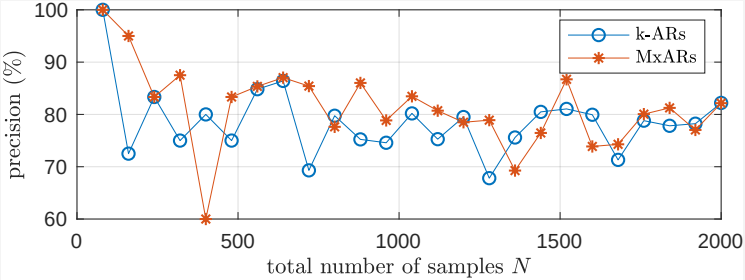
Problem setup based on a real brain imaging application:

- the order of the AR(p) models: $p = 5$,
- the length of time series: $T = 250$,
- the number of groups: $K = 160$,
- the number of time series per cluster: $N_c = \sim 180$,

where the total number of time series is $N = KN_c$.

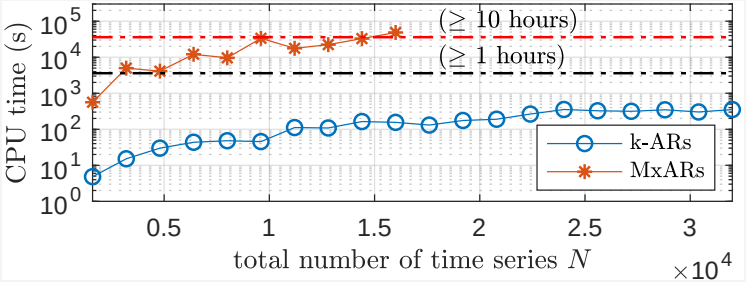
Numerical examples

Comparative clustering precision of MxARs and k-ARs methods:



Numerical examples

Comparative CPU times of MxARs and k-ARs for different problem sizes:



Conclusions

Summary of contributions:

- k-ARs is much faster than MxARs;
- k-ARs has no underflow issues;
- k-ARs is capable to handle large-scale problems.

Thank you!

Bibliography

Aghabozorgi, S., Shirkhorshidi, A. S., & Wah, T. Y. (2015). Time-series clustering–A decade review. *Information Systems*, *53*, 16–38.

Gollub, J. & Sherlock, G. B. T. .-.-. M. i. E. (2006). [10] Clustering Microarray Data. In *Dna microarrays, part b: Databases and statistics* (Vol. 411, pp. 194–213). Academic Press. doi:https://doi.org/10.1016/S0076-6879(06)11010-1

He, F., Chen, H., Probst-Kepper, M., Geffers, R., Eifes, S., del Sol, A., ... Balling, R. (2012). PLAU inferred from a correlation network is critical for suppressor function of regulatory T cells. *Molecular Systems Biology*, *8*(1).

Liao, T. W. (2005). Clustering of time series data—a survey. *Pattern recognition*, *38*(11), 1857–1874.

Wong, C. S. & Li, W. K. (2000). On a mixture autoregressive model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *62*(1), 95–115.

Xiong, Y. & Yeung, D.-Y. (2002). Mixtures of ARMA models for model-based time series clustering. In *2002 ieee international conference on data mining, 2002. proceedings.* (pp. 717–720). IEEE.