A Bayesian Hybrid Approach to Unsupervised Time Series Discretization

Yoshitaka Kameya
Tokyo Institute of Technology

Gabriel Synnaeve
Grenoble University

Andrei Doncescu
LAAS-CNRS

Katsumi Inoue
National Institute of Informatics

Taisuke Sato
Tokyo Institute of Technology
Outline

• Review: Unsupervised discretization of time series data
  – Preliminary experimental results
• Hybrid discretization method based on variational Bayes
• Experimental results
• Summary and future work
Discretization

• ... converts numeric data into symbolic data

• ... is a \textit{preprocessing} task in knowledge discovery

• ... may lead to noise reduction and a good data abstraction
  – We wish to have \textit{interpretable} discrete levels

• ... may help \textit{symbolic} data mining
  – Frequent pattern mining
  – Inductive logic programming

[Fayyad et al. 1995]
Unsupervised discretization of time series data

Common strategy:
- **Smoothing** at the time (x) axis
- **Binning** or **clustering** at the measurement (y) axis

- **Binning:**
  - Equal width binning
  - Equal frequency binning
  - ...
- **Clustering:**
  - Hierarchical clustering [Dimitrova et al. 05]
  - K-means
  - Gaussian mixture models [Mörchen et al. 05b]
  - ...
- **Smoothing:**
  - Regression trees [Geurts 01]
  - Smoothing filters
    - Moving averaging
    - Savitzky-Golay filters [Mörchen et al. 05b]
  - ...

- **All-in-one methods:**
  - **SAX** [Lin et al. 07]
  - **Persist** [Mörchen et al. 05a]
  - Continuous hidden Markov models [Mörchen et al. 05a]
Persist [Mörchen et al. 05a]

- Assumption:
  Time series tries to stay at one of the discrete levels (= *states*) as long as possible

- Persist greedily chooses the breakpoints so that less state changes occur → a role of smoothing
Continuous hidden Markov models

- Two-step procedure
  - Train the HMM
  - Find the most probable state sequence by the Viterbi algorithm
  \[
  \text{State sequence} = \text{Discrete time series}
  \]

Positions and shapes of Gaussians are adjusted by EM.
Preliminary experiment [Mörchen et al. 05]

- Comparison on the predictive performance among the discretizers
- We used an artificial dataset called the “enduring-state” dataset

How well do the discretizers recover the answers?

- SAX
- Persist
- HMMs
- Equal width binning (EQW)
- Equal frequency binning (EQF)
- Gaussian mixture model (GMM)
Preliminary experiment (Cont’d)

• Error analysis: Persist
  – Levels are correctly identified
  – However many noises go across the boundaries
Preliminary experiment (Cont’d)

- Error analysis: HMMs
  - Some levels are misidentified
  - Small noises are correctly smoothed

5 levels
5 % outliers
Motivation

- From preliminary experiments, we can see:
  - **Persist**: robust in identifying the discrete levels (because its heuristic score captures the global behavior of the time series)
  - **HMMs**: good at local smoothing

**Our proposal:**
Hybridization of heterogeneous discretizers based on variational Bayes
**Variational Bayes**

- Efficient technique for Bayesian learning [Beal 03]
  - Empirically known as robust against outliers
  - Gives a principled way of determining # of discrete levels

- An HMM is modeled as: $p(x, z, \theta) = p(\theta) \ p(x, z \mid \theta)$
  - $x$: input time series
  - $z$: hidden state sequence (discretized time series)
  - $\theta$: parameters
  - $p(\theta)$: prior
  - $p(x, z \mid \theta)$: likelihood

- Prior of means and variances in HMMs:

\[
p(\mu_k, \sigma^2_k) = p(\mu_k, \lambda_k^{-1}) = \mathcal{N}(\mu_k \mid m_k, (\tau \lambda_k)^{-1}) \mathcal{G}(\lambda_k \mid a, b)
\]

Normal-Gamma distribution (conjugate prior)

hyperparameters
Variational Bayes (Cont’d)

- Variational Bayesian EM in *general* form:
  - We try to find $q = q^*$ that maximizes the variational free energy $F[q]$:
    \[
    F[q] = \sum_z \int_{\Theta} q(z, \theta) \log \frac{p(x, z, \theta)}{q(z, \theta)} d\theta
    \]
  - $F[q]$ is a lower bound of the marginal likelihood $L(x)$:
    \[
    L(x) \equiv \log p(x) = \log \sum_z \int_{\Theta} p(x, z, \theta) d\theta
    \]
    \[\Rightarrow\] $F[q^*]$ is a good approximation of $L(x)$
  - To get $q^*$, assuming $q(z, \theta) \approx q(z)q(\theta)$, we iterate the two steps alternately:
    - **VB - E step:** $q(z) \propto \exp(\int_{\Theta} q(\theta) \log p(x, z | \theta) d\theta)$
    - **VB - M step:** $q(\theta) \propto p(\theta) \exp\left(\sum_z q(z) \log p(x, z | \theta)\right)$
  - From $L(x) - F[q^*] = \text{KL}(q^*(z, \theta), p(z, \theta | x))$, $q^*$ is a good approximation of the posterior distribution and so used for discretization
Hybridization

- The means of Gaussians are updated by:

\[ m_k := \frac{\tau m_k + T_k \bar{x}_k}{\tau + T_k} \]

- We simply set \( m_k := (\beta_{k-1} + \beta_k)/2 \)
  where \( \beta_k \) are the breakpoints obtained by Persist

- In a similar way, we can also combine HMMs with SAX
Experiment 1: “Enduring-state” dataset

Weight \( \tau = 0.5, 1, 5, 10, 20, 50, 70, 100 \)

Accuracy (Hybridization, 2 levels)

NMI (Hybridization, 2 levels)

raw time series (input)

20/Nov/2010
Experiment 1: “Enduring-state” dataset

Weight $\tau = 0.5, 1, 5, 10, 20, 50, 70, 100$

Accuracy (Hybridization, 3 levels)

NMI (Hybridization, 3 levels)

raw time series (input)
Experiment 1: “Enduring-state” dataset

Weight $\tau = 0.5, 1, 5, 10, 20, 50, 70, 100$

Accuracy (Hybridization, 4 levels)  NMI (Hybridization, 4 levels)

raw time series (input)

20/Nov/2010  TAAI-2010
Experiment 1: “Enduring-state” dataset

Weight $\tau = 0.5, 1, 5, 10, 20, 50, 70, 100$
Experiment 1: “Enduring-state” dataset

Weight $\tau = 0.5, 1, 5, 10, 20, 50, 70, 100$

Accuracy (Hybridization, 6 levels)

NMI (Hybridization, 6 levels)

ratio of outliers

20/Nov/2010

TAAI-2010
Experiment 1: “Enduring-state” dataset

Weight $\tau = 0.5, 1, 5, 10, 20, 50, 70, 100$

Under accuracy
HMM+ Persist is significantly better than Persist except several cases with a large # of levels and many outliers

Under NMI
HMM+ Persist is significantly better than Persist for all cases according to Wilcoxon’s rank sum test ($p = 0.01$)
Experiment 2: Background

- Also based on [Mörchen et al. 05a]
- Data on muscle activation of a professional inline speed skater
  - Nearly 30,000 points recorded in log-scale
Experiment 2: Goal

- Estimating a plausible # of discrete levels *automatically* with variational Bayes
- An expert prefers to have 3 levels [Mörchen et al. 05a]

---

Last kick to the ground to move forward

Gliding phase (muscle is used to keep stability)
Experiment 2: Settings

• Having so many (30,000) data points, we need to:
  – Use large pseudo counts ($\geq 500$)
    \[ \overline{\mu}_k := \frac{\tau m_k + \overline{T}_k \overline{x}_k}{\tau + \overline{T}_k} \]
  – Use PAA (used in SAX) to compress the time series

(frame size = 50)
Experiment 2: Discretization by CHMMs (Cont’d)

- PAA disabled
- Savitzky-Golay filter enabled with half-window size = 100
- Pseudo counts = 1
Experiment 2: Discretization by CHMMs (Cont’d)

- PAA disabled
- Pseudo counts = 1000

![Graph showing discretization comparison]
Experiment 2: Discretization by CHMMs (Cont’d)

- PAA enabled with frame size = 10
- Pseudo counts = 1
Experiment 2: Discretization by CHMMs (Cont’d)

- **PAA enabled** with frame size = 20
- Pseudo counts = 1
Summary

• Unsupervised discretization of time series data
• Hybridizing heterogeneous discretizers via variational Bayes
  – Fast approximate Bayesian inference
  – Robust against noises
  – Automatic finding of the plausible number of discrete levels

Future work

• Histogram-based discretizer