Business Intelligence
SS 2017

Temporal Data Mining

W. Grossmann

Content

- Problem Formulation
- Classification Based on Time warping
- Classification Based on Response Features
- Clustering of Time Sequences

Problem Formulation

- Basic ingredients
  - A sequence of ordered time values, called observation times, $t_1 \leq t_2 \leq \ldots \leq t_T$
  - Attribute values at these times $x_1 \leq x_2 \leq \ldots \leq x_T$
- Time sequence (time stamped data):

$$x = < (t_1, x_1), (t_2, x_2) \ldots (t_T, x_T) >$$

- Two important goals in temporal data mining:
  - Classification of time sequences into classes
  - Finding clusters of time sequences

Problem Formulation

- A main issue is representation of temporal data for the analysis
  - Non-adaptive representation: transform time sequence into a feature space, e.g. Fourier transform
  - Adaptive representation: extract features of the time sequence, which can be used for analysis
  - Data clipping: transform time sequence into a bit string
  - Model based representation: time sequence as input for a model, e.g. a Markov chain
- We will focus on methods based on adaptive representation

Analysis Template

- Relevant Business and Data: Customer behavior represented as time sequence
- Analytical Goals:
  - Classification of a new time sequence into one of the possible classes
  - Segmentation of time sequences according to their structural similarity
- Modeling Task: Using visualization techniques for the time sequences of the process instances decide for a useful method:
  - Time warping for defining distances
  - Response features

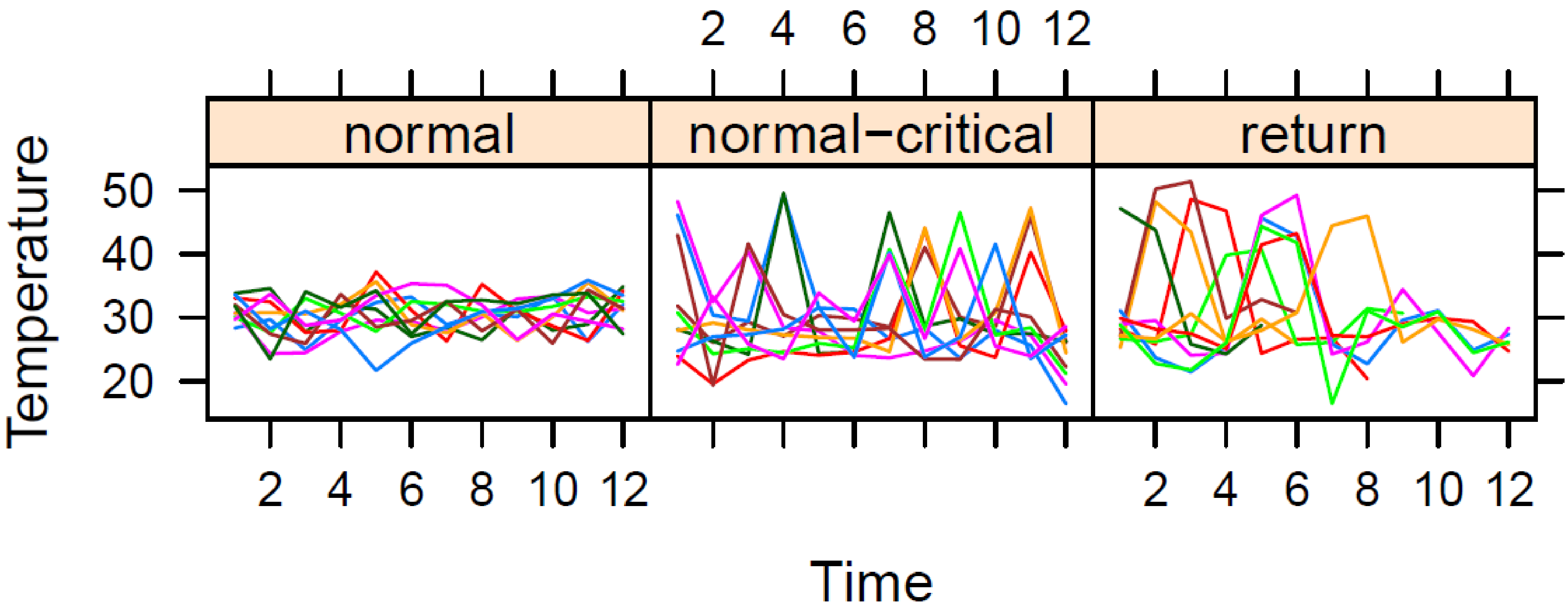
Analysis Template

- Analysis task:
  - Splitting Data: If possible split the data randomly in one set for training and one set for validation
  - Model Estimation: Estimate the warping path or the response features
  - Model Assessment: Assess quality of the model
  - Model Selection: Select a model
  - use the results of model estimation for segmentation or classification
- Evaluation and Reporting Task: Evaluate the results of segmentation or classification either with test data or by using cross validation

Classification based on time warping

- Example: Temperature monitoring of a container during transport
  - In case of “abnormal” temperature behavior the cargo is damaged and the transport has to be interrupted and returns to the origin
  - We distinguish three different scenarios as shown in the graphic

Classification based on time warping



Classification based on time warping

- General Problem formulation:
- Given are data of customer behavior represented as time sequences for process instances
- These data are classified into different groups
- Task: Find a classification rule which allows the assignment of a time sequence to one of the classes

Classification based on time warping

- Basic idea behind time warping:
  - Classes are defined by time series which show a similar “pattern”
  - The term similarity is understood in the sense of speech waves: different persons spell words differently but we can classify the waves to words

Classification based on time warping

- Problem which have to be taken into account:
  - Time sequences may have different length
  - Similarity may be blurred by some temporal transformations like stretching or squeezing some parts of the time sequence (see example)
- We have to define the similarity by matching the observed values of two time sequences in such a way that the above defined effects are compensated

Classification based on time warping

- Dynamic time warping allows the calculation of similarity
- Basic is the definition of a warping path:

Given two sequences $(x_1, x_2, \ldots, x_N)$ and $(y_1, y_2, \ldots, y_M)$

Define a sequence $(p_1, p_2, \ldots, p_L)$ of matching indices pairs (i_ℓ, j_ℓ) such that

$$p_1 = (1,1) \quad p_L = (N, M)$$

$$(i_1 \leq i_2 \leq \ldots \leq i_L) \text{ and } (j_1 \leq j_2 \leq \ldots \leq j_L)$$

$$p_{\ell+1} - p_\ell \in \{(1,0), (0,1), (1,1)\}$$

Classification based on time warping

  - The last condition means that we increase the matching index at least by one step ahead
- The costs of a warping path is defined by

$$D_P = \sum_{\ell=1}^{L} d(i_\ell, j_\ell) = \sum_{\ell=1}^{L} | x_{i_\ell} - y_{j_\ell} |$$

- The dynamic time warping algorithm finds a warping path for two sequences with minimal costs
  - The word “dynamic” indicates that the algorithm is based on dynamic programming

Classification based on time warping

- Application of the dynamic warping algorithm for all pairs of sequences defines a distance matrix for the observed time sequences
- We can apply now k-nearest neighbor classification for obtaining the classification rule

Classification Based on Response Features

- In that case we extract from the time sequence a number of time independent characteristic features
- Some examples of features:
  - Maximum and minimum of the time sequence
  - Temporal location of maximum and minimum
  - Breakpoints in the time sequence
  - Largest difference between two sequenced values
  - Length of the sequence
  - Area under the polygon defined by the sequence

Classification Based on Response Features

- More theoretically motivated features:
  - Transformation to frequencies and looking at the maximum frequency (Time sequence is sound or light)
  - Definition of a regression model for the time sequence
  - Definition of a representation language
- Based on these attributes one can apply methods of the classification of cross sectional data

Clustering of Time Sequences

- Clustering of time sequences can be done using the same principles as in the case of classification
- The definition of time warping defines a distance for the sequences which cam be used as input for cluster analysis (hierarchical or k-means)
- In the case of response features the distance between the time sequences is based on the response features