The Representation Race
Handling Time Phenomena

Katharina Morik

Univ. Dortmund, www-ai.cs.uni-dortmund.de

- MiningMart -- an approach to the representation race
- Time related learning tasks
- Case studies
  - shop
  - intensive care

The Problem: Method Selection

- Criteria for selecting a learning method for an application are missing -- no expert knowledge available! (MLT Consultant)
- Empirical studies do neither result in clear guidelines. (StatLog)
- Learning the rules that recommend a method for an application requires well-chosen descriptions of methods and tasks. (MetaL, CORA)

Observation

Experienced users can apply any learning system successfully to any application, since they prepare the data well...

- The representation L_E of examples determines the applicability of learning methods.
- A chain of data transformations (learning steps) leads to L_E of the method that delivers the desired result.

Experienced users remember prototypical successful transformation/learning chains

MINING MART

The Representation Race

application:

data | users | performance system

L_{E1}

L_{H1}

L_{E2}

L_{H2}

...

L_{En-1}

$L_{Hn-1} = L_{En} \longrightarrow L_{Hn} = L_{En+1}$

learning/data mining

L_{Hn+1}

...

L_{En+m}

L_{Hn+m}

The Consortium

- Katharina Morik Univ. Dortmund, D (Coordinator)
- Lorenza Saitta Univ. Piemonte del Avogadro, I
- Pieter Adriaans Syllogic, NL
- Dietrich Wettscherek Dialogis, D
- Jörg-Uwe Kietz SwissLife, CH
- Fabio Malabocchia CSELT, I

The MiningMart Approach

Best practice cases of transformation/learning chains exist

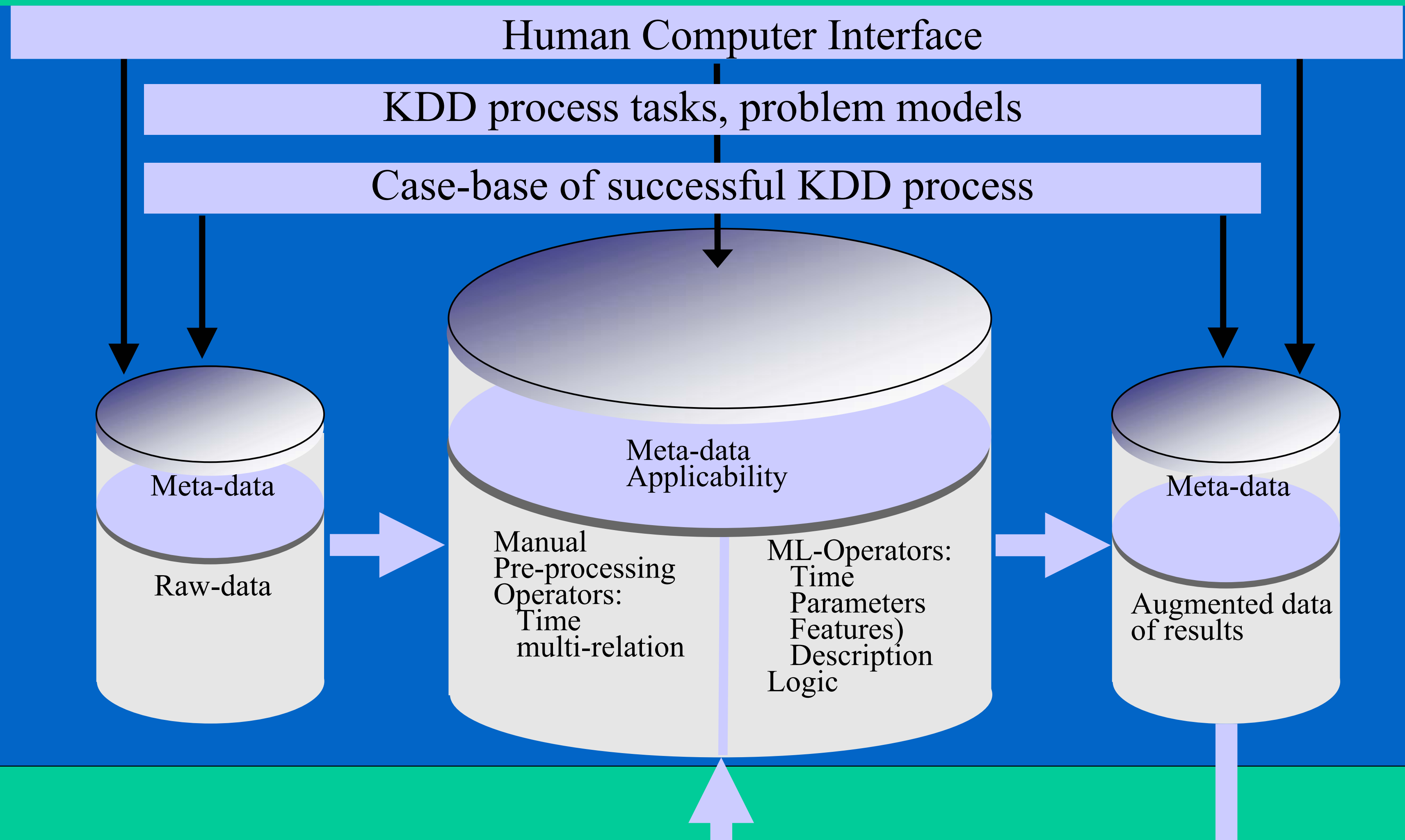
- Data, L_E and L_H are described on the meta level.
- The meta-level description is presented in application terms.
- MiningMart users choose a case and apply the corresponding transformation and learning chain to their application.

... and **more can be obtained!**

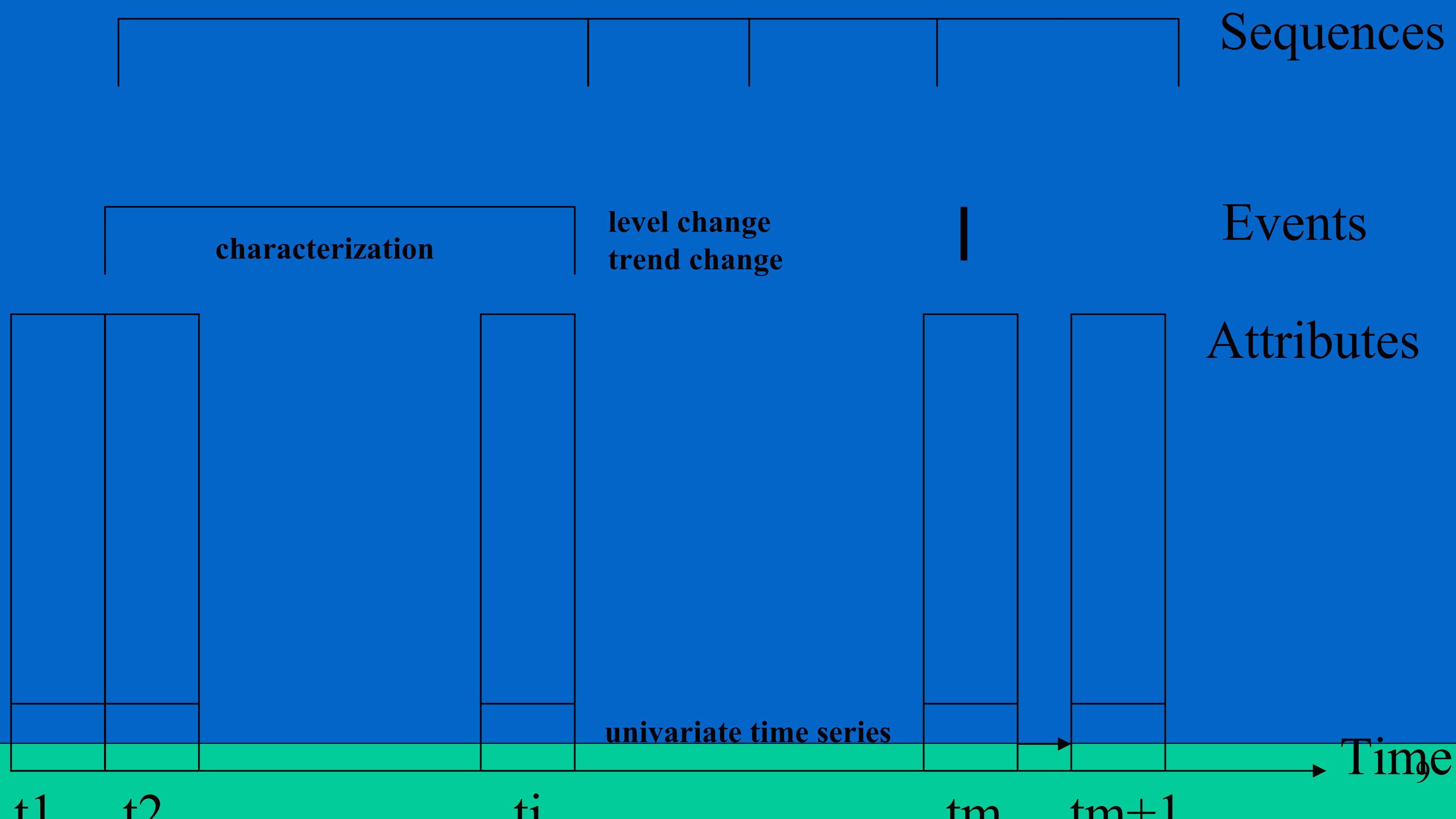
Call for Participation

- MiningMart is about to develop an operational meta-language for describing data and operators.
- MiningMart prepares the first cases of KDD.
- MiningMart will present the case-base in the WWW.
- You may contribute to the representation race!
  – Apply the meta-language to your application and deliver it as a positive example to the case-base; or
  – apply a case of MiningMart to your data.

The MiningMart System



Time Phenomena



Typical Time-Related Data

On-line measurements

- univariate time series
- multivariate time series

Database relations

- sales/contract data
- age/life situation

Granularity

- continuous measurements in day, hours, minutes, seconds
- time stamped events in years, half/quarter years, months, days

Learning Tasks -- Precedence

From a time series until t_m

univariate

- predict value at t_{m+n}
- find a common trend
- find cycles, seasons
- find level changes

Given sequences

- find clusters of similar subsequences

multivariate

- find co-occurrences
- find subsets of co-occurring attribute values (events)
- find time regions

Learning Tasks -- Dominance

Define sequences as

Frequent sequences:

precedence relation between sets of events (episodes)

Legal sequences:

proportions of time intervals (predicting actual time point)

Relations between time intervals:

overlap, inclusion, (direct) precedence

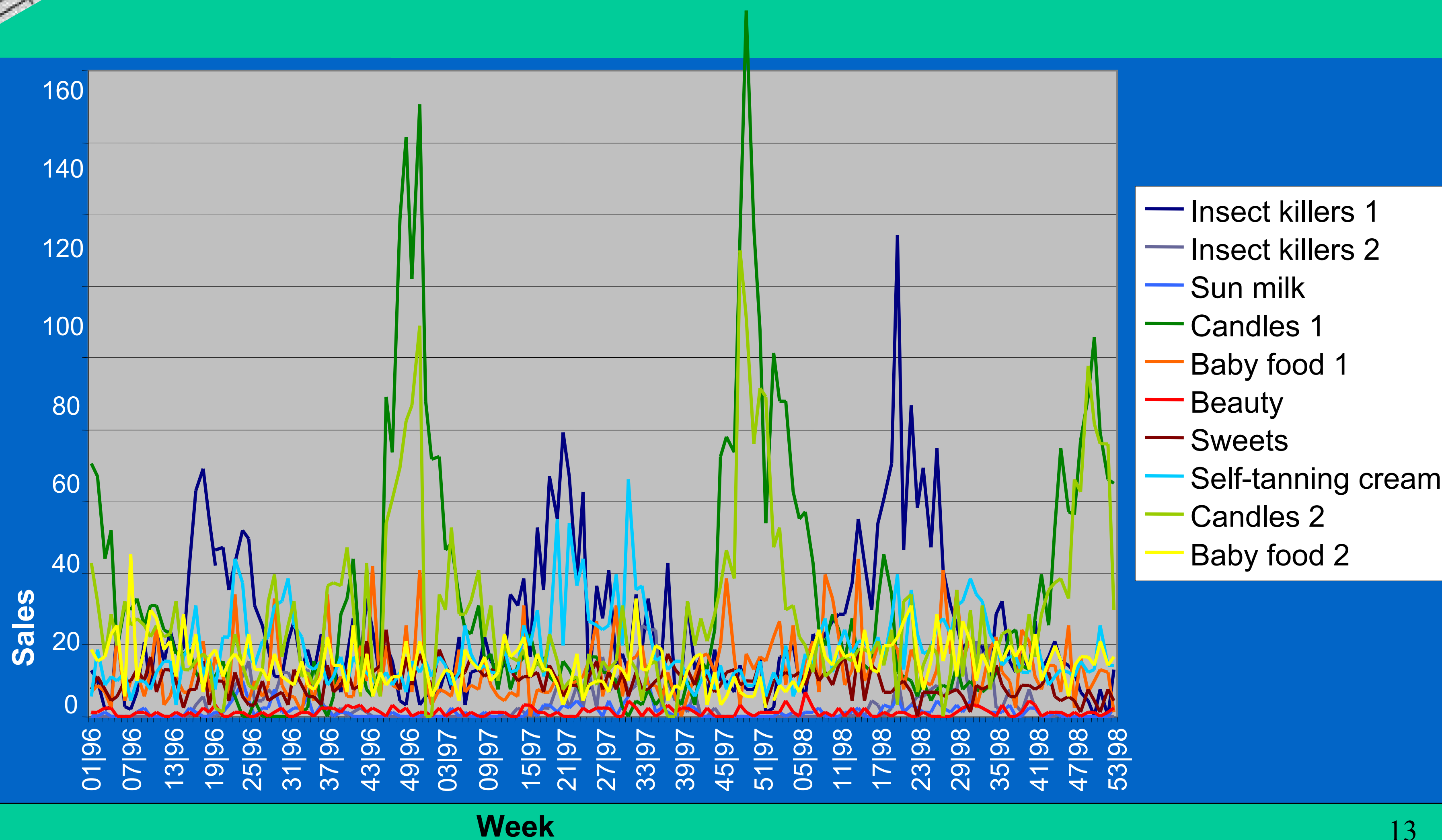
Higher-level categories:

a sequence of actions constitutes a category at the higher level

in terms of

- association rules
- first order logic
- prefix trees
- automata
- Hidden Markov Models

Sales of Items of a Drugstore

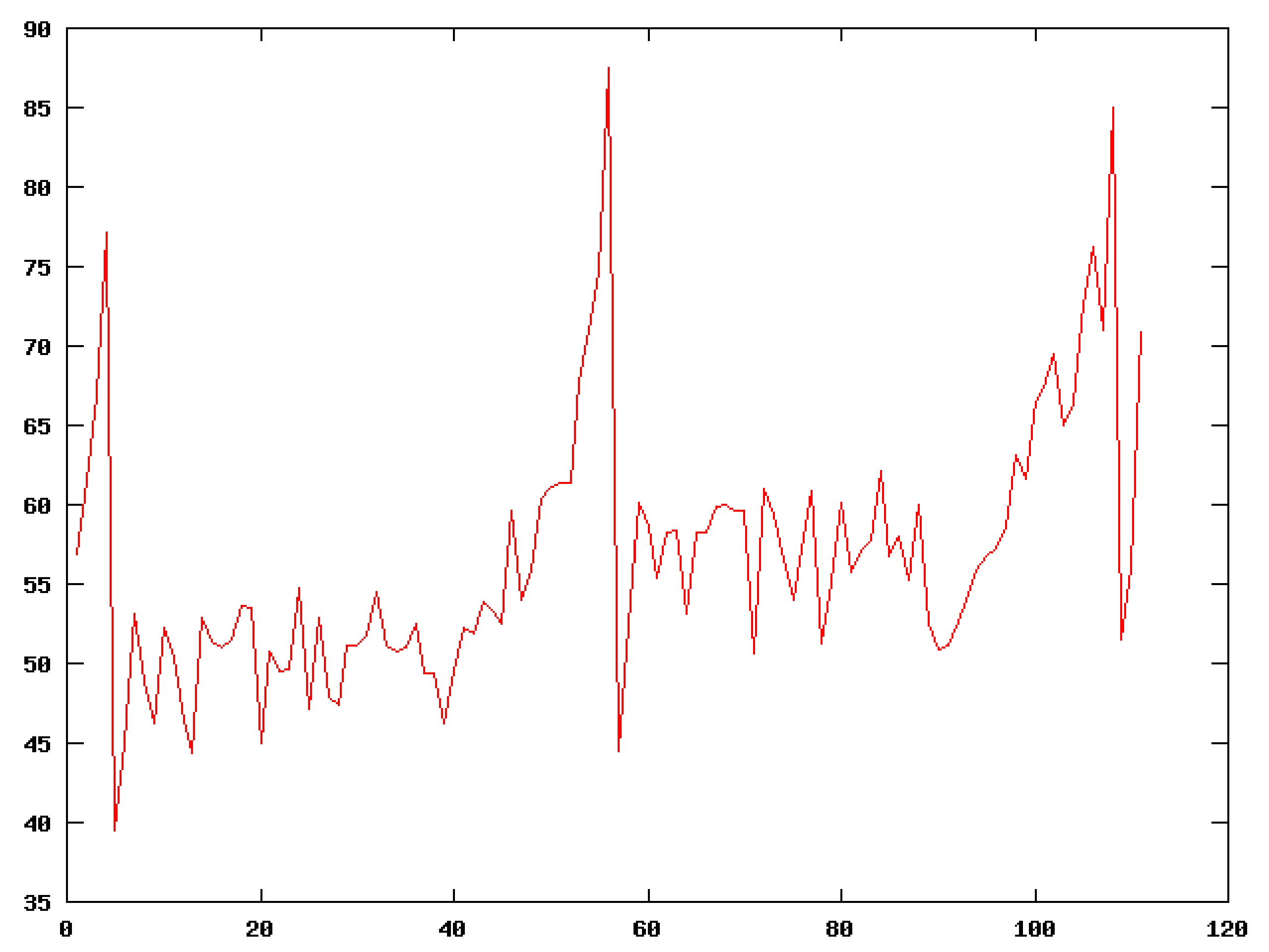


MINING
MART

Learn About All Sales

- Find seasons, cycles, trends in general
- Aggregate all items, all shops
- Define a standard function of sales in a year
- Inspect deviations of particular shops from the standard

Aggregation of All Items Over Time



Predict Sales of an Item

- Given drug store sales data of 50 items in 20 shops over 104 weeks
- Predict the sales of an item such that
  - the prediction never underestimates the sale,
  - the prediction overestimates less than the rule of thumb.
- Observation: 90% of the items are sold less than 10 times a week.
- Requirement: prediction horizon is more than 4 weeks ahead.

MINING MART

Shop Application -- Data

Shop	Week	Item1	...	Item50
Dm1	1	4	...	12
Dm1	...	...	...	...
Dm1	104	9	...	16
Dm2	1	3	...	19
...	...	...	...	...
Dm20	104	12	...	16

LE_{DB1}: I: T_1 A_1 ... A_{50}; set of multivariate time series

Transformations

- From shops to items: multivariate to univariate

$L_{E1'}$: i:t_1 a_1 … t_k a_k

For all shops for all items:
Create view Univariate as
Select shop, week, $item_i$
Where shop="dm_j"
From Source;

- Multiple learning

Dm1_Item1 ...	1	4 ...	104	9
Dm1_Item50	1	12...	104	16

....

Dm20_Item50	1	14...	104	16

Exponential Smoothing

- Univariate time series as input (LE_1),
- incremental method:
  current hypothesis h and new observation o yield next hypothesis by $h := h + \lambda \; o$,
  where λ is given by the user,
- predicts sales of n-next week by last h.

Transformations

- Obtaining many vectors from one series by sliding windows

L_{H5} i:t_1 a_1 ... t_w a_w
move window of size w by m steps

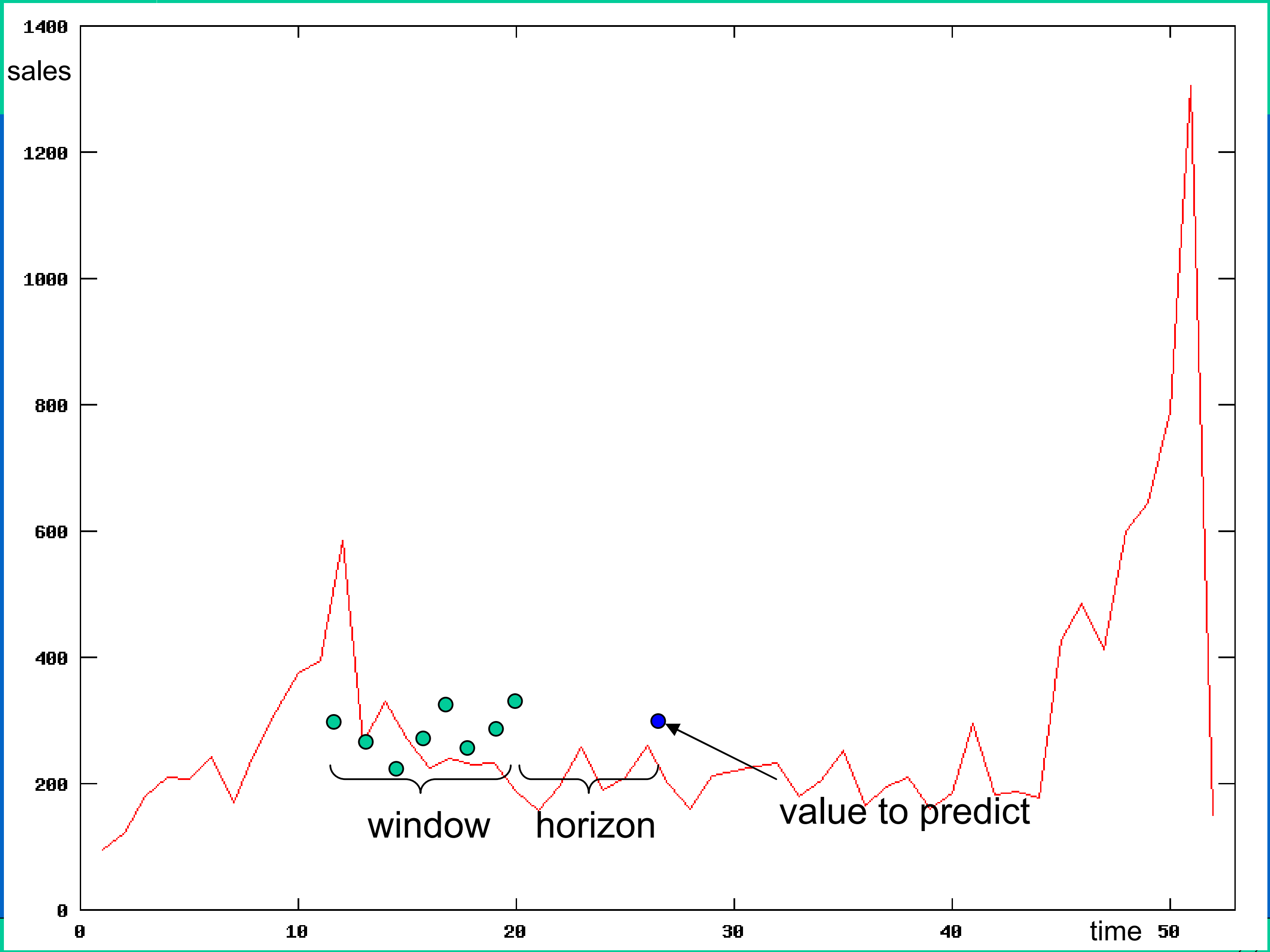
Dm1_Item1_1	1	4...	5	7
Dm1_Item1_2	2	4...	6	8
...				
Dm1_Item1_100	100	6...	104	9
...				
...				
Dm20_Item50_100	100	12...	104	16

SVM in the Regression Mode

- Multiple learning:
  for each shop and each item, the support vector machine learned a function which is then used for prediction.
- Asymmetric loss:
  - underestimation was multiplied by 20,
    i.e. 3 sales too few predicted -- 60 loss
  - overestimation was counted as it is,
    i.e. 3 sales too much predicted -- 3 loss

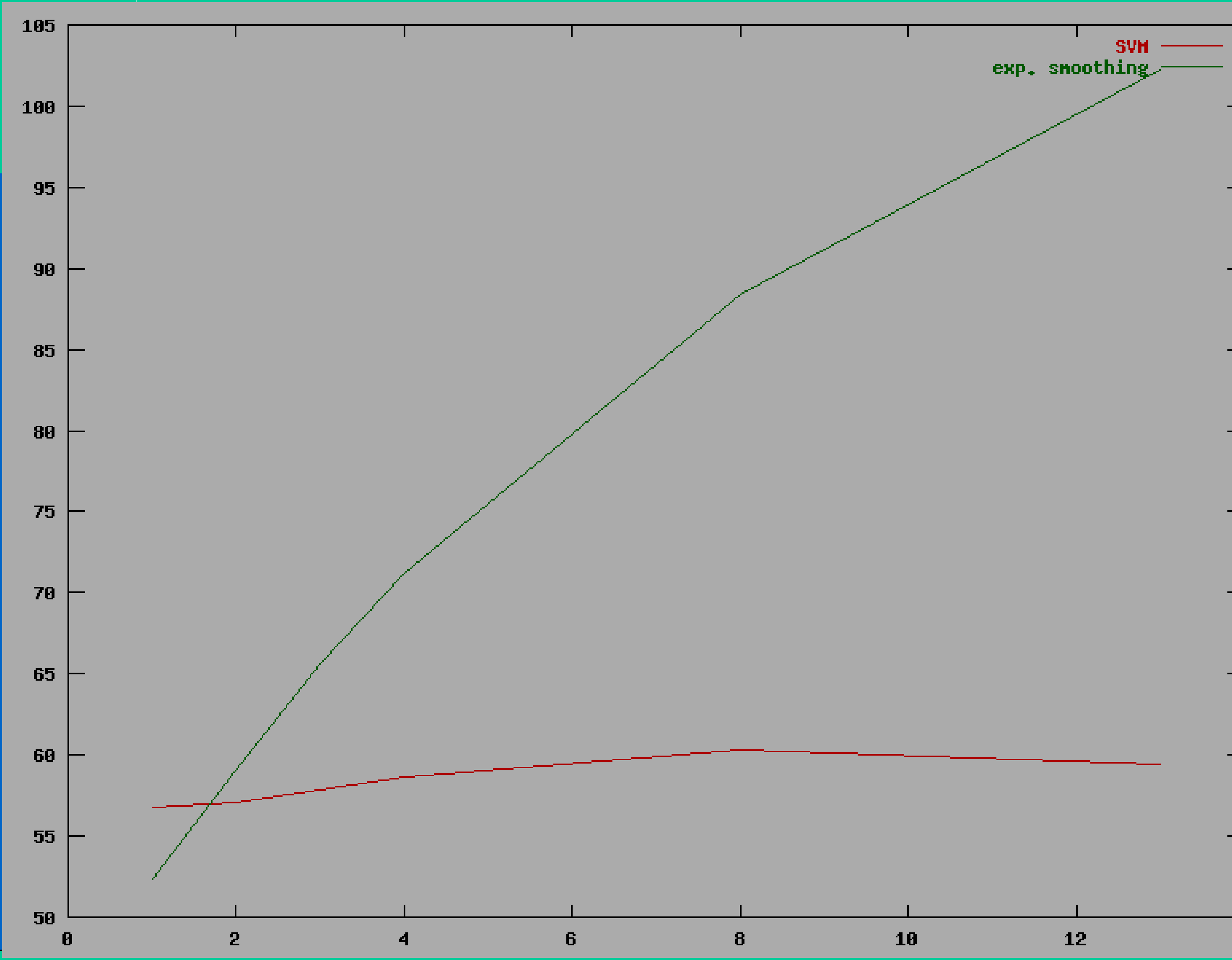
(Stefan Rüping 1999)

Article 766933 (bag?)



Comparison with Exponential Smoothing

horizon	SVM	exp. smoothing
1	56.764	52.40
2	57.044	59.04
3	57.855	65.62
4	58.670	71.21
8	60.286	88.44
13	59.475	102.24



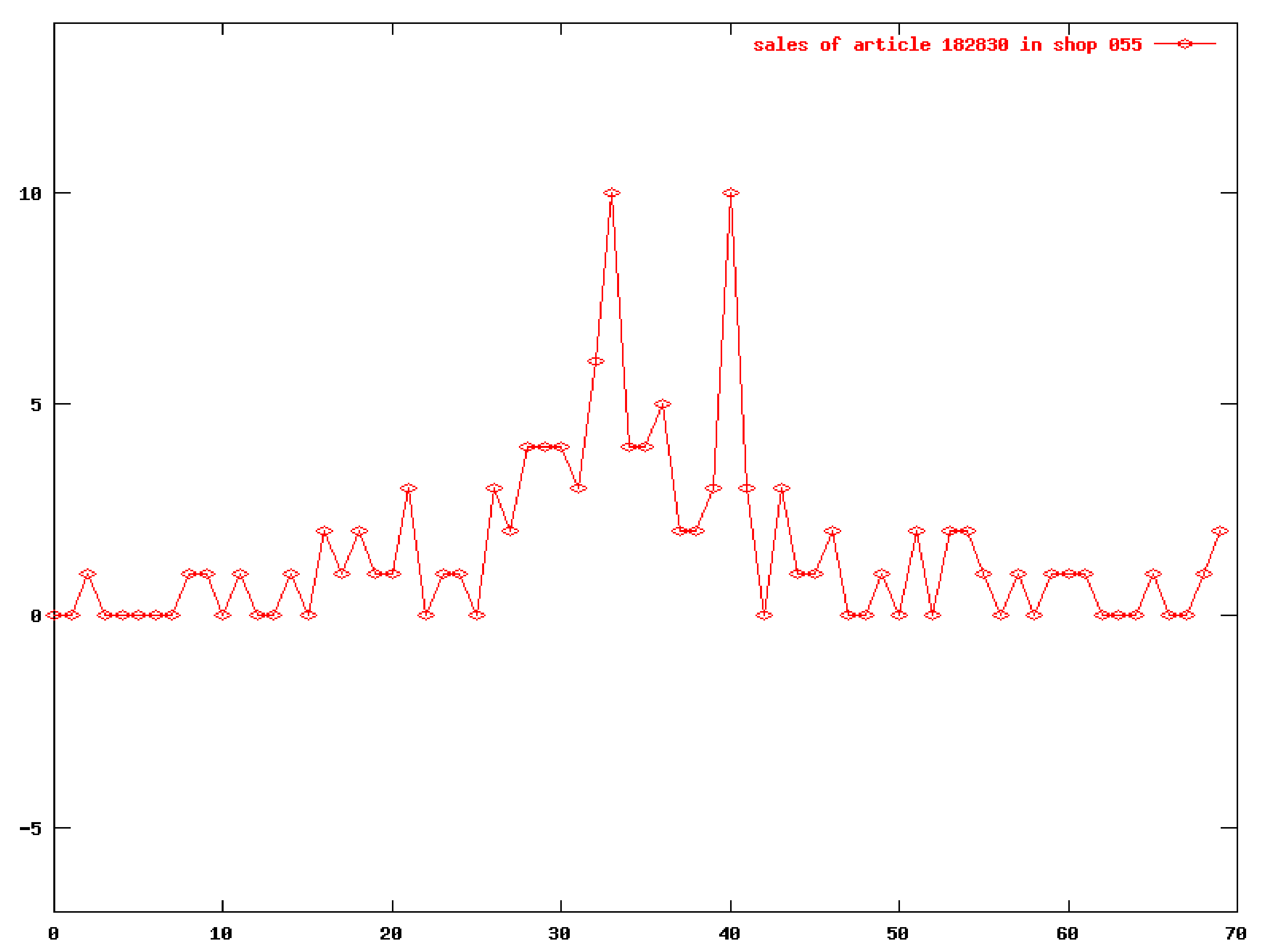
loss

horizon

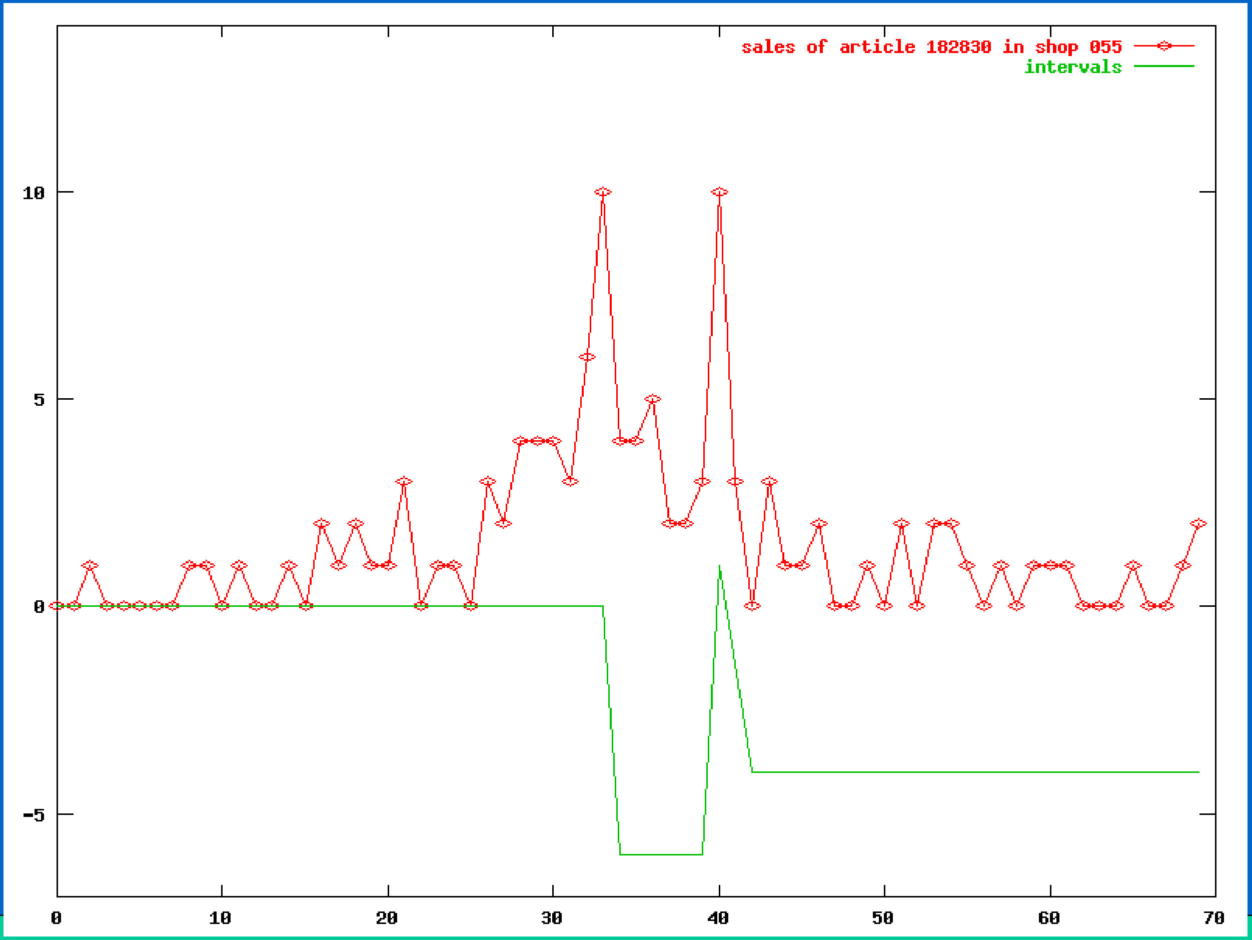
Learning Relations

- Are there typical sequences that are valid for <u>all</u> items? Prepocessing for rule learning about abstract episodes:
- Summarizing values within time intervals
  $L_{E1'}$: i:t_1 a_1 ... t_k a_k $\Rightarrow$
  L_{H6}: i: $[t_1, t_w]f(a_1,\ldots,a_w),\ldots, [t_m, t_{m+w}]\ g(a_1,\ldots,a_w)$
- Abstraction into classes of gradients valid for a time interval $\Rightarrow$
  L_{H2}:*Label* $_j$ $[t_1, t_w],\ldots,$*Label* $_l$ $[t_m, t_{m+w}]$

Sales of Item 182830 in Shop 55



Summarizing Sales



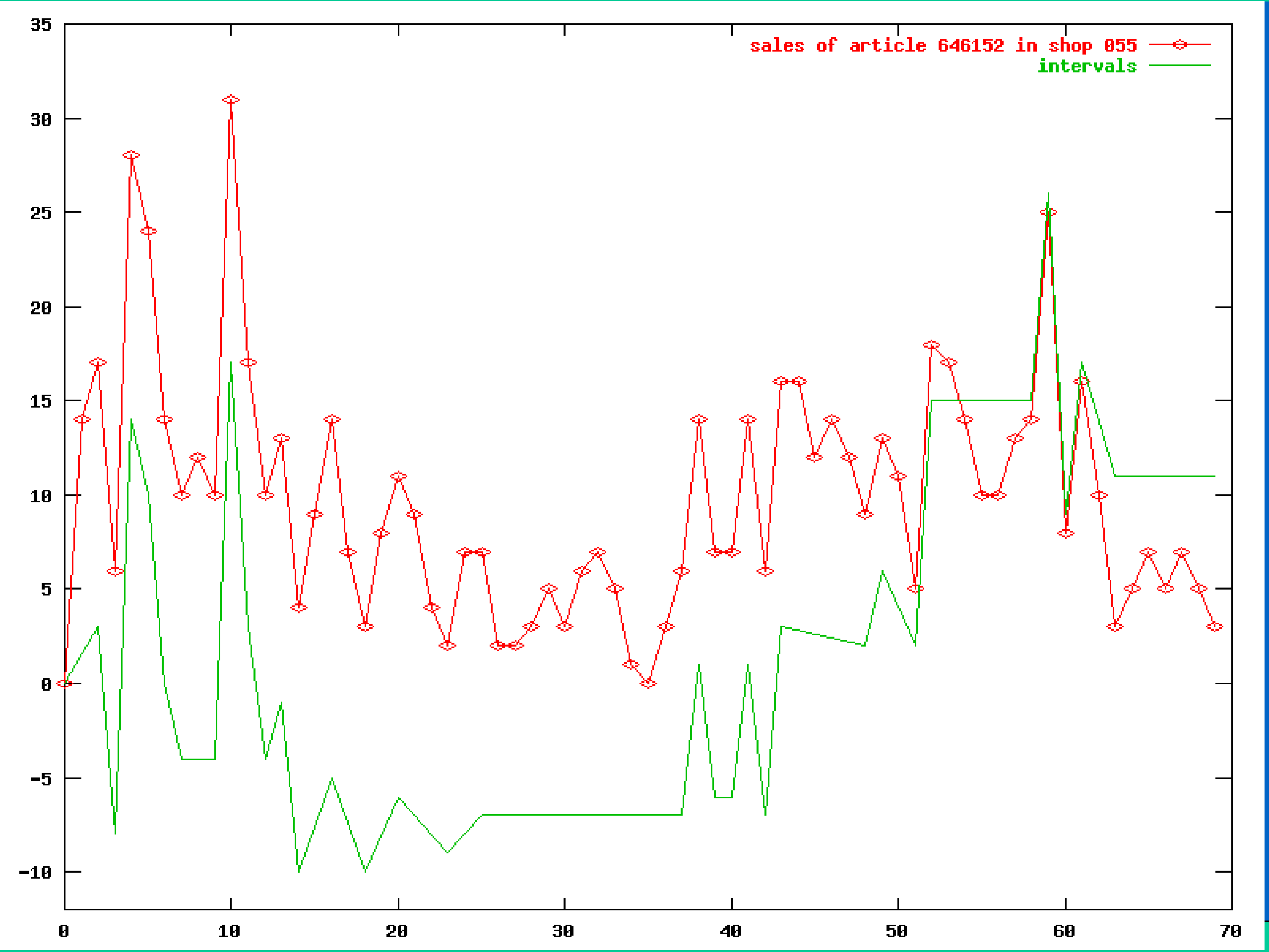
(Wessel, Morik 1999)

Transformation into Facts

L_{E4}:

stable(182830,1,33,0).
decreasing(182830, 33,34,-6).
stable(182830, 34, 39,0).
increasing(182830, 39, 40,7).
decreasing(182830, 40, 42,-5).
stable(182830, 42,108,0).

Summarizing Item 646152 in Shop 55



Corresponding Facts

increasing(646152,1,2,3).
decreasing(646152,2,3,-11).
increasingPeak(646152,3,4,22).
...
stable(646152, 25,37,0).
increasing(646152, 37, 38, 8).
decreasing(646152, 38, 39, -7).
stable(646152, 39,40, 0).
increasing(646152, 40, 41,7).
decreasing(646152, 41, 42,-8).
increasing(646152, 42, 43,10).
stable(646152, 43, 48,-1).

small time intervals

Rule Learning

- Transformations into facts:
  $L_{E\,4}$: $p(I, T_b, T_e, A_r, ..., A_s)$
- Rules about sequences:
  $p_1(I, T_b, T_e, A_r), p_2(I, T_e, T_{e2}, A_s) \rightarrow$
  $p_3(I, T_{e2}, T_{e3}, A_t)$
- results for sequences of sales trends:
  increasing $(Item, T_b, T_e) \rightarrow$ decreasing$(Item, T_e, T_{e2})$
  increasing $(Item, T_b, T_e)$, decreasing$(Item, T_e, T_{e2})$
  $\rightarrow$ stable$(Item, T_{e2}, T_{e3})$

Same Data -- Several Cases

- Find seasons or cycles in all sales

  aggregation of items and shops, description of the curve as a function
- Predict sales of a particular item in a particular shop

  multivariate to univariate, multiple exponential smoothing OR
  multivariate to univariate, sliding windows, multiple learning with SVM
- Find relations between trends that are valid for all sales in all shops

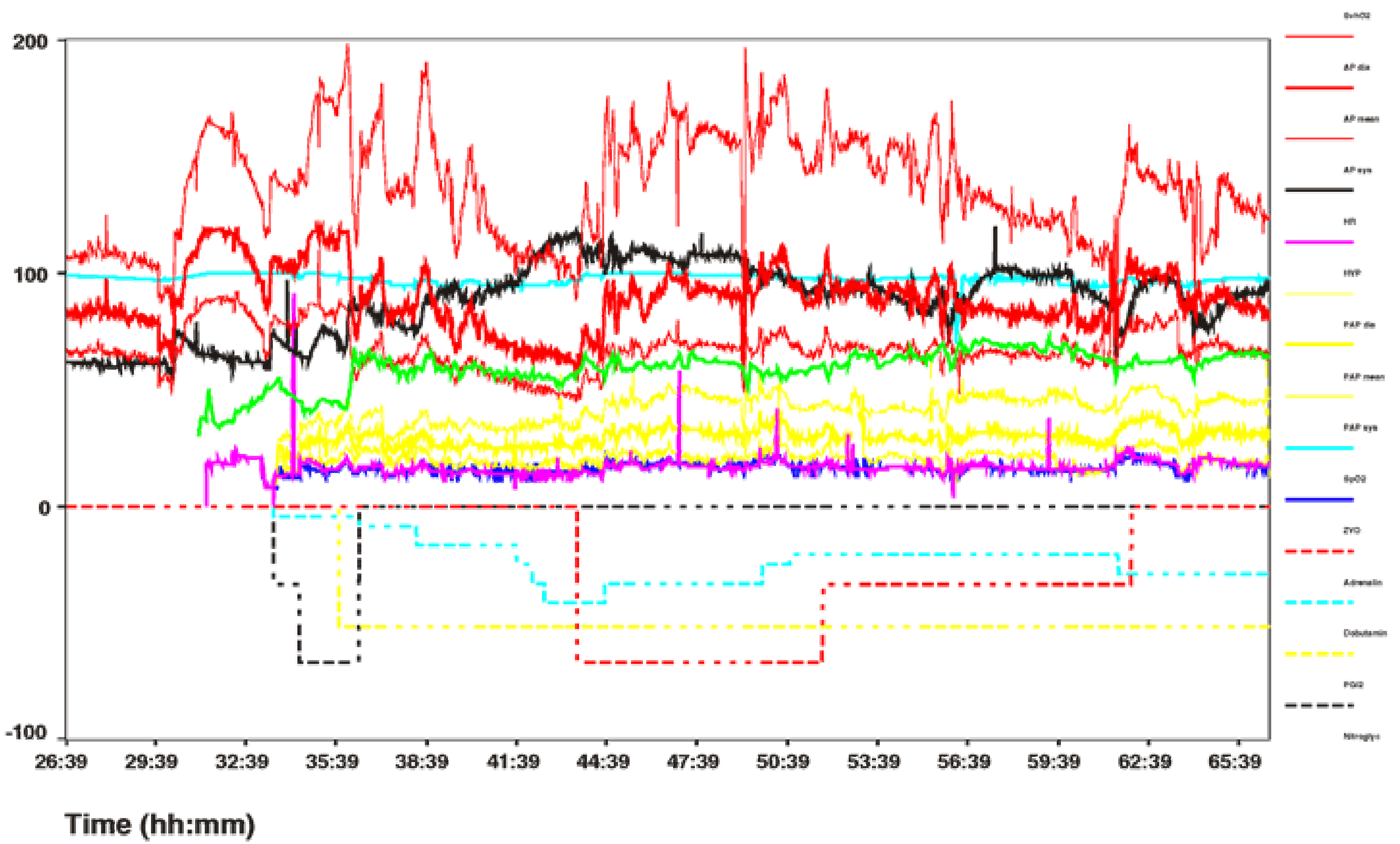
  summarizing, transformation into facts, rule learning

Applications in Intensive Care

- On-line monitoring of intensive care patients
- high-dimensional data about patient and medication
- measured every minute
- stored in the Emtec database of patient records ---
- learning when to intervene in which way.

Patient G.C., male, 60 years old

Hemihepatektomie right



The Data

LE_{DB2}

$i_1: t_1\ a_{1\,1} \ldots a_{1\,k}$
$i_1: t_2\ a_{2\,1} \ldots a_{2\,k}$
...
$i_2: t_1\ a_{1\,1} \ldots a_{1\,k}$
...

set of rows for each patient:
1 row for each minute

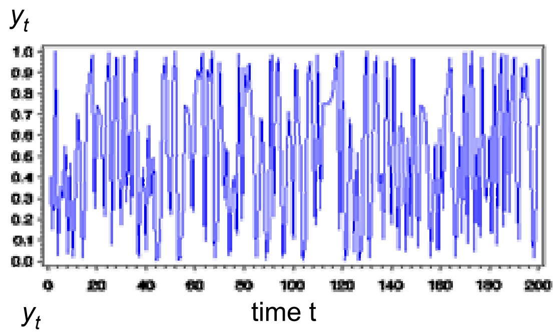
Transformations

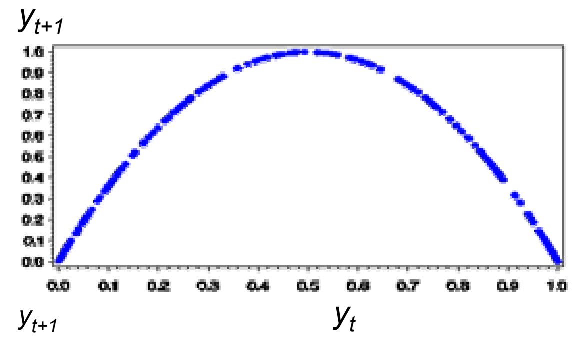
- Chaining database rows
  i_1: $t_1\ a_{1\,1}\ \ldots\ a_{1\,k}$, $t_2\ a_{2\,1}\ \ldots\ a_{2\,k}$, ...
- Multivariate to univariate
  i_1: $t_1\ a_1$, $t_2\ a_1\ \ldots\ t_m\ a_1$
  i_1: $t_1\ a_2$, $t_2\ a_2\ \ldots\ t_m\ a_2$
  ...
- Detecting level changes

Phase State Analysis

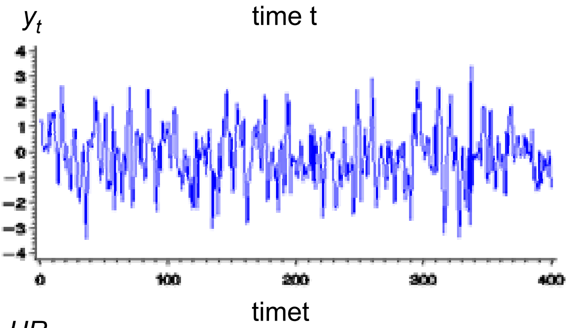
Time series $y_1, \ldots, y_N$ **Phase state** $\breve{y}_t = (y_t, y_{t+1})$

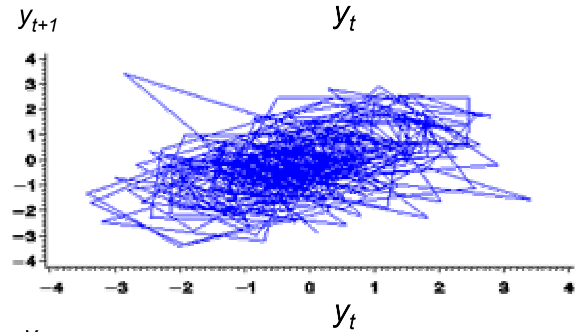
eter-
inistic
rocess



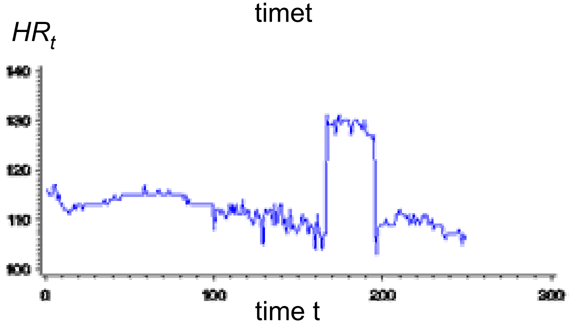


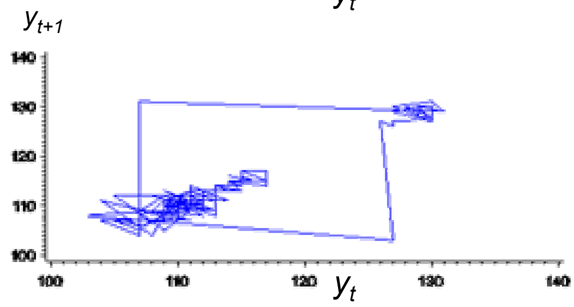
R(1)-process
ith outlier
AO)





eart rate





Level Change Detection

level_change(pat4999, 50, 112, hr, up)

level_change(pat4999, 112, 164, hr, down)

level_change(pat4999, 10, 74, art, constant)

level_change(pat4999, 74, 110, art, down)

Computed Feature

Comparing norm values for a vital sign and its mean in a time interval (± standard deviation):

deviation(pat4999, 10, 74, art, up)

MINING
MART

Learning Task

Are there valid rules

for all multivariate time series,

such that therapeutical interventions follow from a patient’s state?

Relational Learning

Given patient records in the form of facts:

- deviations -- time intervals
- therapeutical interventions -- time points
- types of vital signs (group1: hr, swi, co; group2: art, vr)

Learn rules about interventions:

group1(V), deviation(P, T1, T2, V, Dir)

→noradrenaline(P, T2, Dir)

The Chain of Preprocessing Steps

E $_{DB2}$:

$_1$: $t_1\ a_{1\,1}\ \ldots\ a_{1\,k}$

: $t_2\ a_{2\,1}\ \ldots\ a_{2\,k}$

.

: $t_1\ a_{1\,1}\ \ldots\ a_{1\,k}$

chaining db rows

i_1: $t_1\ a_{1\,1}\ \ldots\ a_{k1}\ t_2\ a_{1\,2}\ \ldots\ a_{k\,2}\ \ldots$

i_2: $t_1\ a_{1\,1}\ \ldots\ a_{k\,1}\ t_2\ a_{12}\ \ldots\ a_{k\,2}\ \ldots$

multi- to univariate

i_1: $t_1\ a_{1\,1}\ t_2\ a_{1\,2}$

i_1: $t_1\ a_{2\,1}\ t_2\ a_{2\,2}$

...

level changes

(i_1, t_i, t_j, A)

...

relational learning

$p_1(I, T_i, T_j, A, D)$, $p_2(I, T_j, T_k, A, D)$

$\rightarrow p_3(I, T_k, \mathrm{Dir})$

computed feature

$(i_1, t_i, t_j, A, D$

Disregarding Time

<u>Given</u> a patient's state at time t_i,

<u>learn</u> whether and how to intervene at t_{i+1}

Transformations:

- Selection of time points where an intervention was done
- Multiple to binary class

  for each drug, form the concepts drug_up, drug_down
- Multiple learning for each binary class resulting in classifiers for each drug and direction of dose change (SVM_light)

The Chain of Preprocessing Steps

E $_{DB2}$:
$_1$: $t_1\ a_{1\,1} \ldots a_{1\,k}$
: $t_2\ a_{2\,1} \ldots a_{2\,k}$
.
: $t_1\ a_{1\,1} \ldots a_{1\,k}$

Select time points with interventions
i_1: $t_i\ a_{1\,i} \ldots a_{ki}$
i_2: $t_j\ a_{1_j} \ldots a_{kj}$
...

Form binary classes
a_1_up$_+$: $a_2 \ldots a_k$
...
a_1_up$_-$: $a_2 \ldots a_k$
....
a_6_down$_+$: $a_2 \ldots a_{k..}$
a_6_down$_-$: $a_2 \ldots a_{k....}$

Learning classifiers using SVM_light
a_1_up$_+$: $w_2\ a_2 \ldots w_k\ a_k$
...
a_6_down$_+$: $w_2\ a_2 \ldots w_k\ a_k$

Same Data -- Several Cases

- Find time relations that express therapy protocols
  chaining db rows, multivariate to univariate, level changes, deviations, RDT
- Predict intervention for a particular drug
  select time points, multiple to binary class, SVM_light

Behind the Boxes

Db schema ndicating time ttribute(s), ranularity,...	Select statement in abstract form, instantiated by db schema	Creating views in abstract form, instantiated by db schema and learning task	Syntactic transformation for SVM Multiple learning contro

Calling SVM_light and writing results

MINING
ART

Summary of Cases Involving Time

Db schema
ndicating time
ttribute(s),their
ranularity,
niformity,
tarting point

Syntactic
transformations
L_{E1}
...
L_{E4}

Sliding windows
Summarizing windows
Level changes

Relational learning (RDT)
SVM_light for classification
SVM for regression
Exponential smoothing

MiningMart Approach to the Representation Race

- Manager -- end-user
  knows about the business case
- Database manager
  knows about the data
- Case designer -- power-user
  expert in KDD
- Developer
  supplies (learning) operators