Symbolization and pattern formation strategies for the assessment of cardiovascular control

Alberto Porta

Department of Biomedical Sciences for Health,
Galeazzi Orthopedic Institute,
University of Milan,
Milan, Italy

Introduction

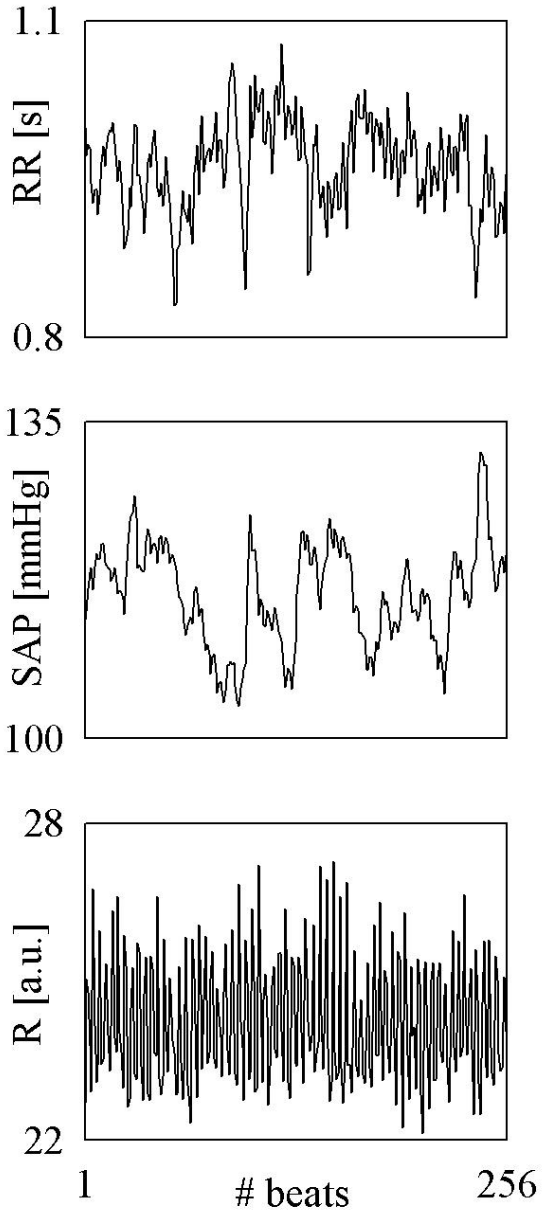
Cardiovascular variables exhibit, when observed on a beat-to-beat basis, rhythmical changes, referred to as spontaneous variability

Spontaneous variability reflects the action of cardiovascular control mechanisms operating to guarantee the functioning of our organism in every condition

In healthy condition modifications of the state of the cardiovascular control result in significant changes of spontaneous variability

Pathological conditions dramatically alter spontaneous variability

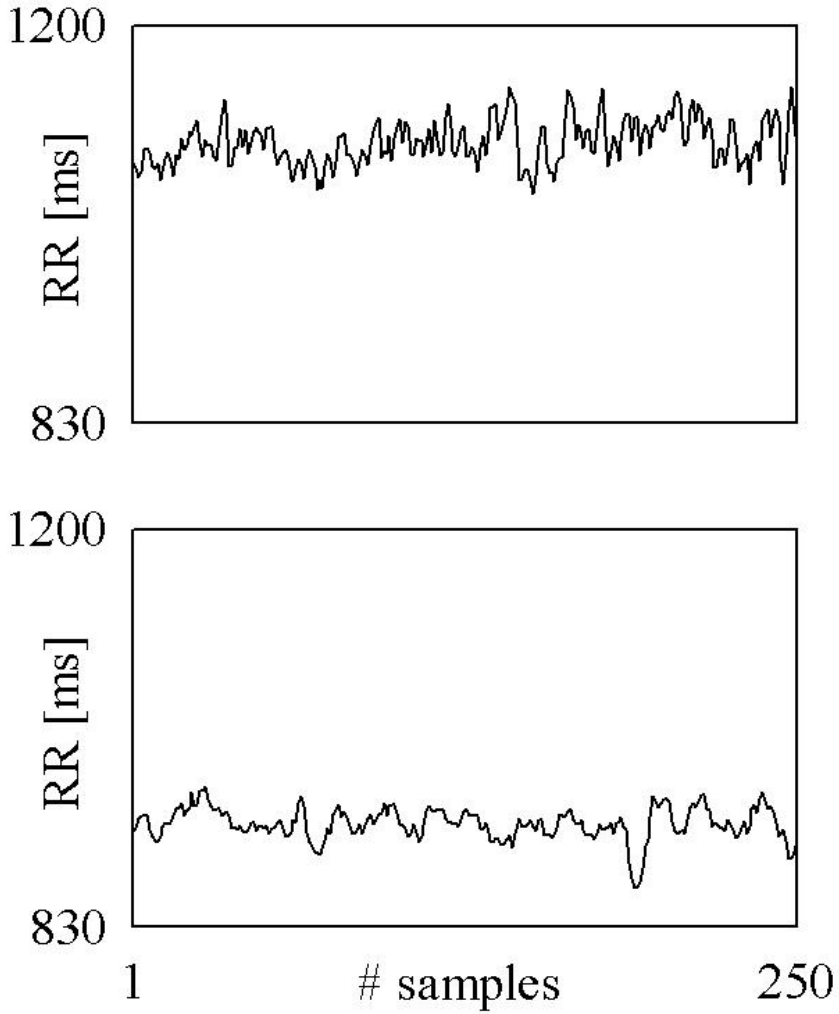
Spontaneous beat-to-beat variability



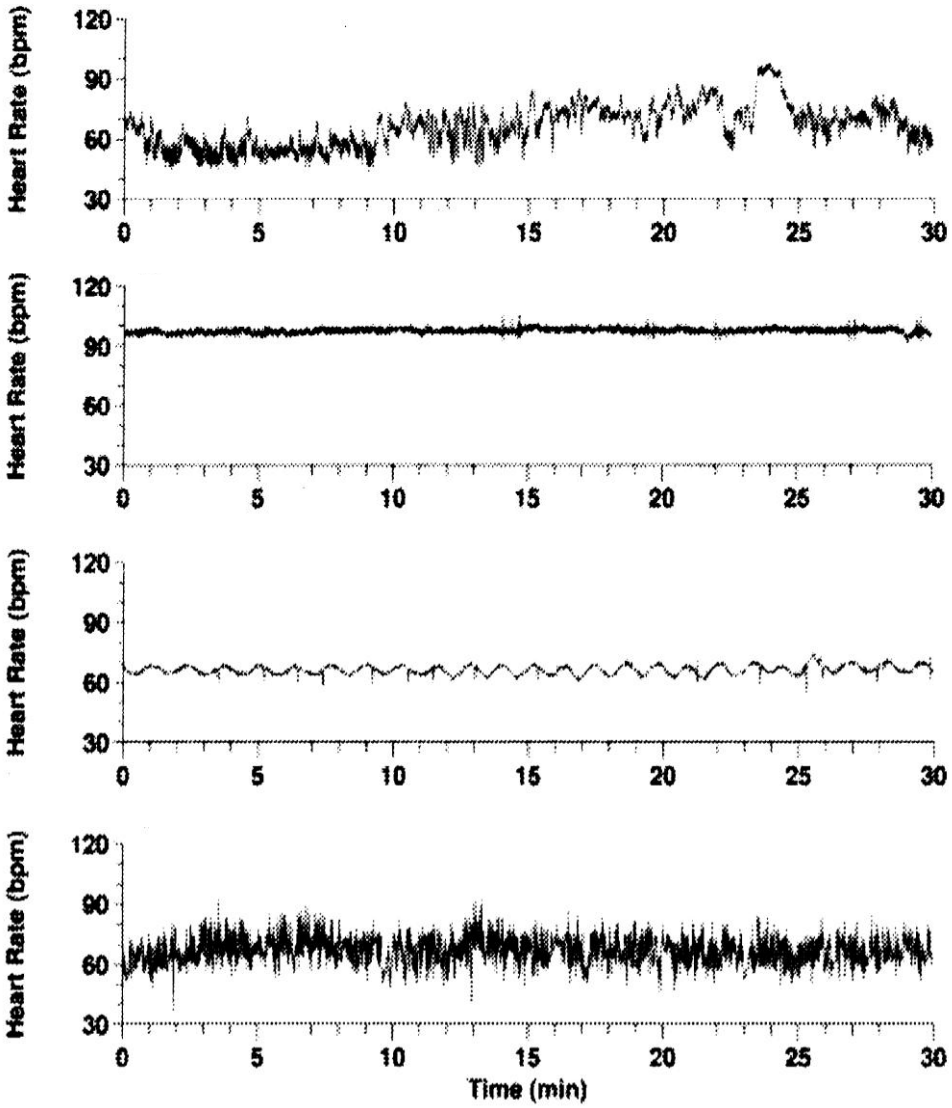
The state of the cardiovascular control influences spontaneous variability

Supine

Passive standing



Cardiovascular diseases influence spontaneous variability



Healthy subject

Congestive heart failure patient

Cheyne-Stokes breathing patient

Atrial fibrillation patient

A.L. Goldberger et al, Proc Natl Acad Sci, 99: 2466-2472, 2002

Introduction

Successful separation between different experimental conditions within the same population and between different populations within the same experimental condition depends on the significance of the features extracted from cardiovascular variability signals

Symbolization and pattern formation strategies, indissolubly linked to the concept of partition (and, more generally, to coarse graining), provide features and indexes helpful to distinguish experimental conditions and/or groups

Aim

To review symbolization and pattern formation strategies exploited to distinguish experimental conditions and/or groups in cardiovascular control studies

Outline

1) Symbolization strategies

2) Pattern construction via delay embedding procedure

3) Symbolization strategy and pattern construction technique impose a coarse graining of the multidimensional space

4) Techniques to optimize the pattern length

5) Approaches to reduce redundancy of patterns

6) Indexes derived from symbolization and pattern formation strategies

7) Application in clinics

Outline

1) Symbolization strategies

2) Pattern construction via delay embedding procedure

3) Symbolization strategy and pattern construction technique impose a coarse graining of the multidimensional space

4) Techniques to optimize the pattern length

5) Approaches to reduce redundancy of patterns

6) Indexes derived from symbolization and pattern formation strategies

7) Application in clinics

Symbolization

Symbolization is a procedure transforming samples of a series into symbols belonging to a finite alphabet

Code: all the samples belonging to the same bin are transformed into the same symbol

The most utilized binning procedures are:

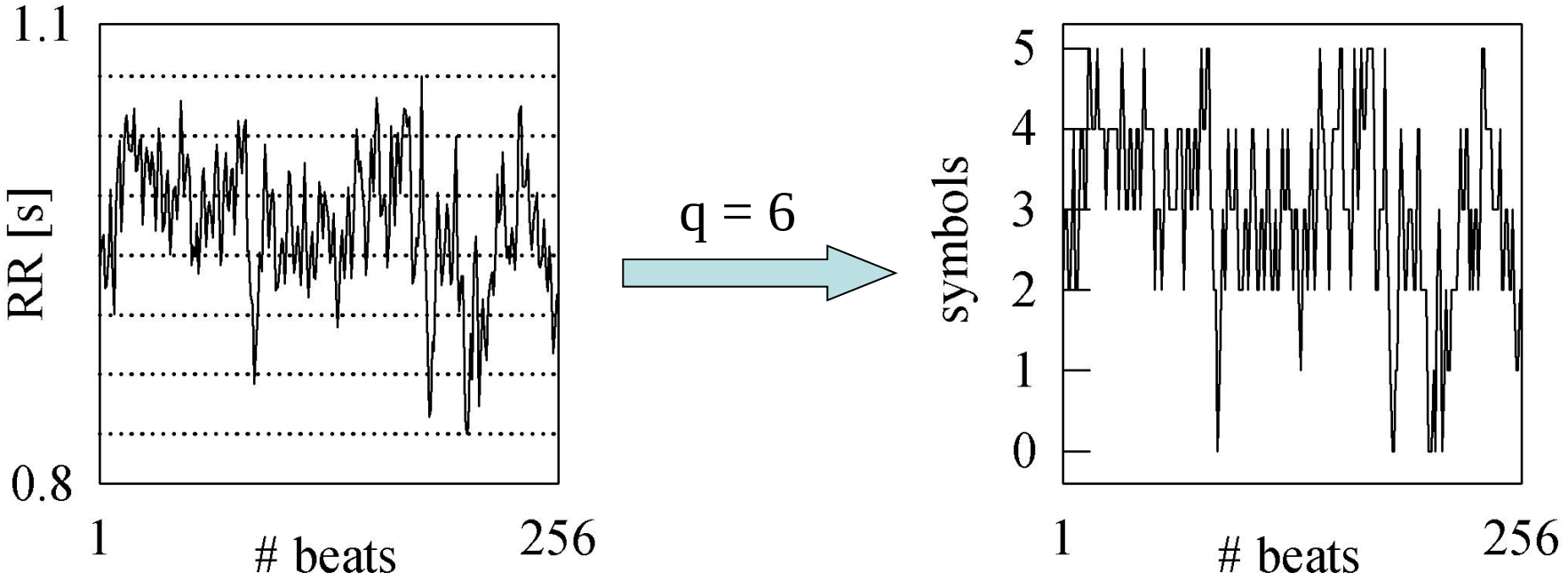
1) uniform quantization — A. Porta et al, Biol Cybern, 78:71-78, 1998

2) non uniform quantization — A. Porta et al, IEEE TBME, 54:94-106, 2007; N. Wessel et al, Phys Rev E, 61:733-739, 2000; D. Cysarz et al, AJP, 292:R368-R372, 2007

3) sample-centered uniform quantization — A. Porta et al, IEEE TBME, 54:94-106, 2007; S.M. Pincus, Chaos, 5:110-117, 1995; J.S. Richman et al, AJP, 278:H2039-H2049, 2000

4) sample-centered non uniform quantization — A. Porta et al, IEEE TBME, 54:94-106, 2007

Uniform quantization

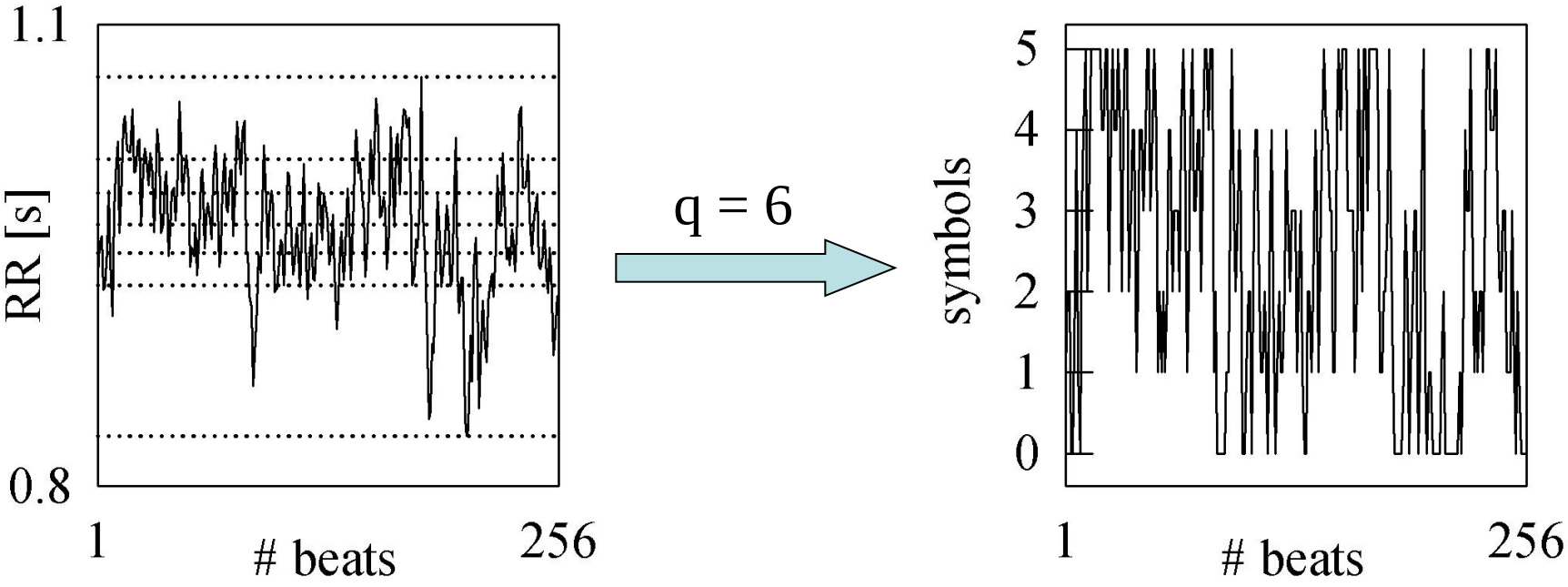


The full dynamical range (i.e. max-min) is divided into q bins of equal width ε

$$\varepsilon = \frac{\max(RR)-\min(RR)}{q}$$

Bins do not overlap and cover the full dynamical range → one sample is transformed into one symbol

Non uniform quantization

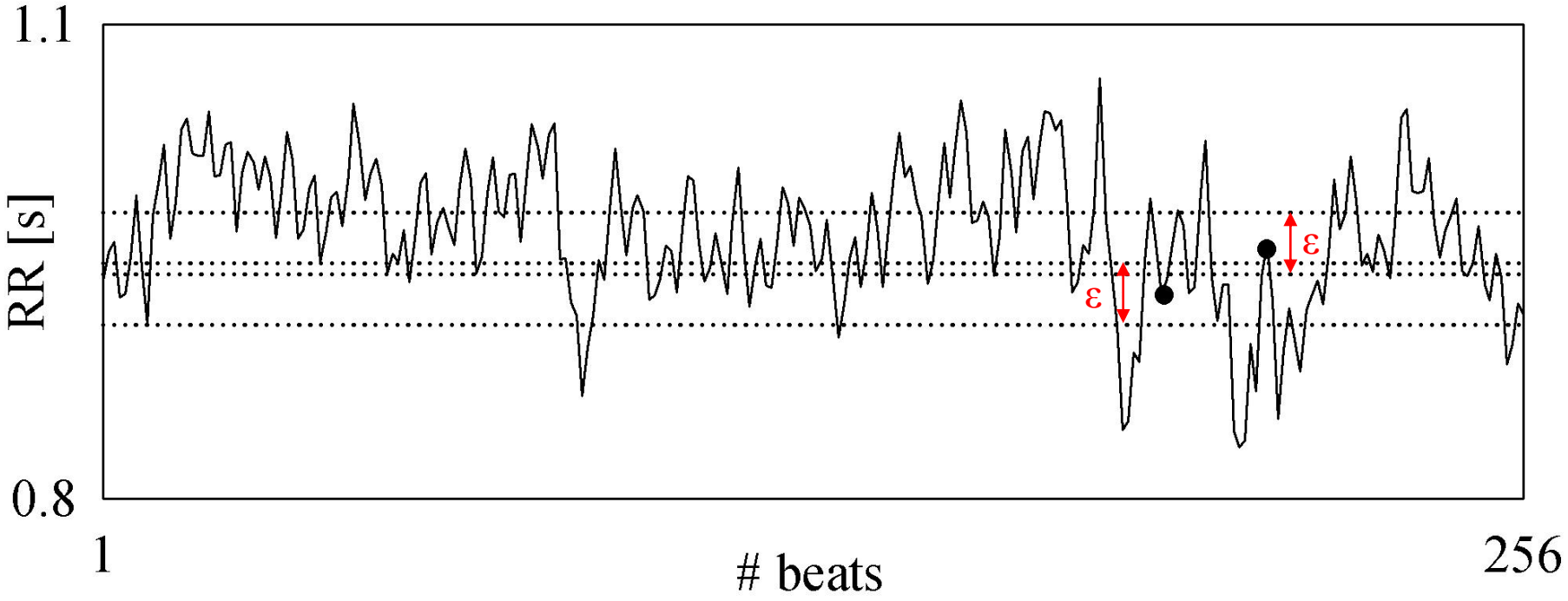


The full dynamical range (i.e. max-min) is divided into q bins of different width ε

Bins do not overlap and cover the full dynamical range

one sample is transformed into one symbol

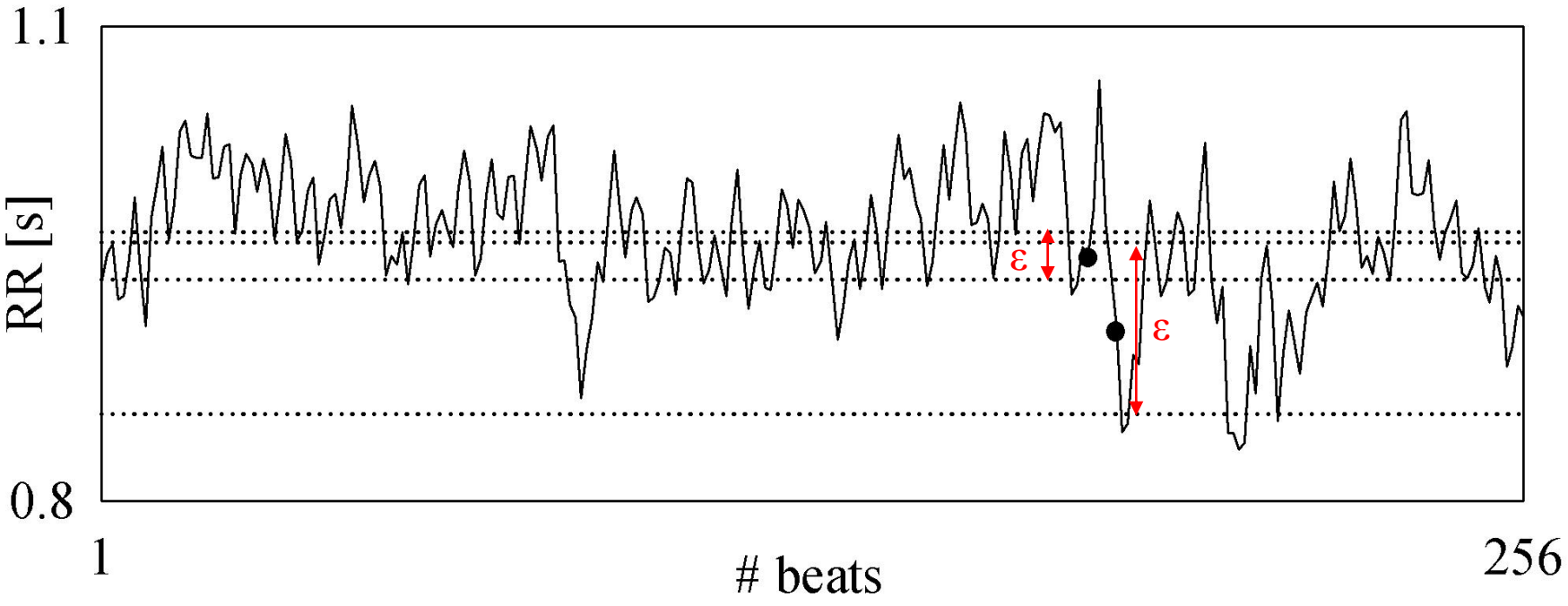
Sample-centered uniform quantization



Bins are constructed around each sample and have the same width ε

Bins overlap → several symbols can be associated to the same sample

Sample-centered non uniform quantization



Bins are constructed around each sample and have different width ε

Bins overlap ⇒ several symbols can be associated to the same sample

Outline

1) Symbolization strategies

2) Pattern construction via delay embedding procedure

3) Symbolization strategy and pattern construction technique impose a coarse graining of the multidimensional space

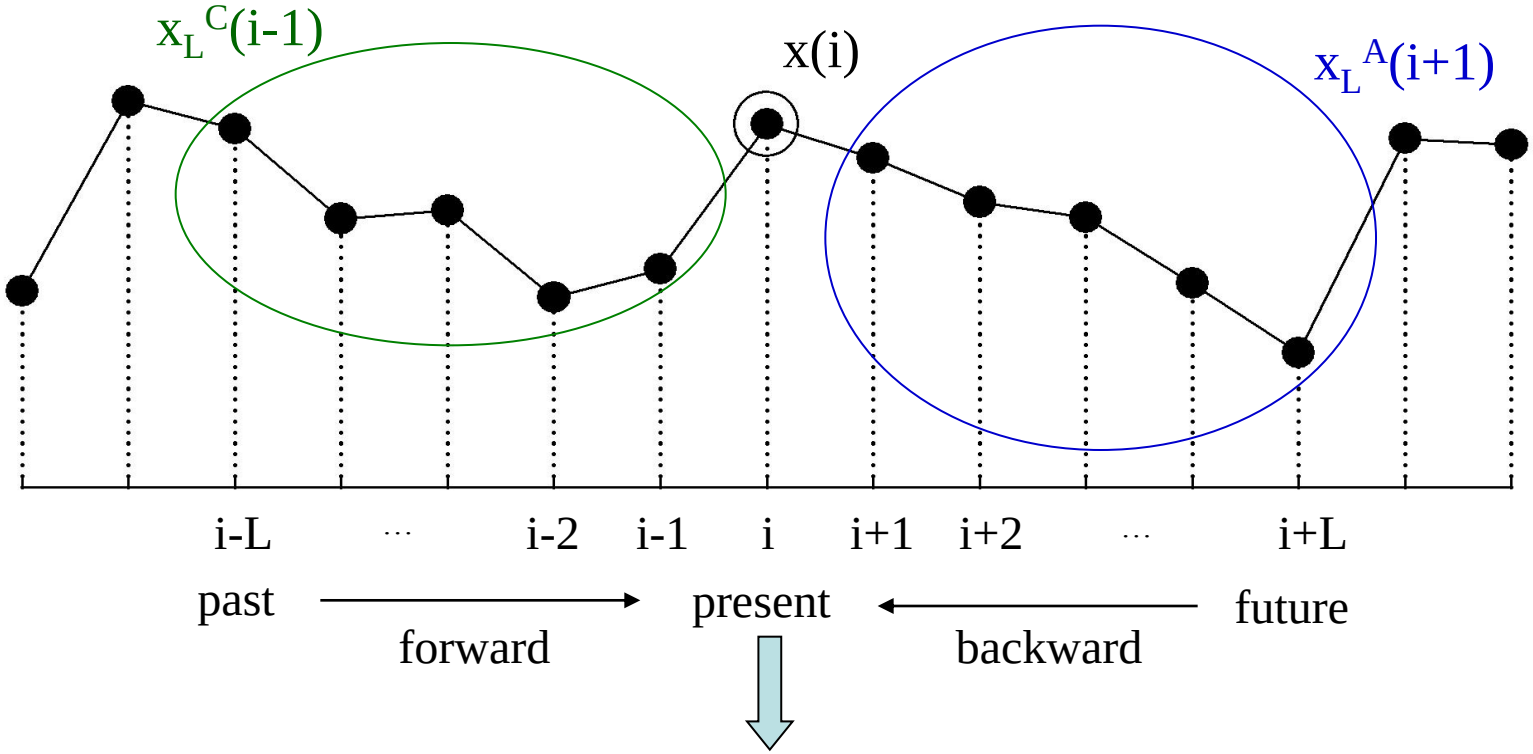
4) Techniques to optimize the pattern length

5) Approaches to reduce redundancy of patterns

6) Indexes derived from symbolization and pattern formation strategies

7) Application in clinics

Pattern formation via delay embedding reconstruction



$s_L^C(i-1) = (s(i-1), \ldots., s(i-L)) \Longrightarrow$ causal pattern of s(i)

$s_L^A(i+1) = (s(i+1), \ldots., s(i+L)) \Longrightarrow$ anti-causal pattern of s(i)

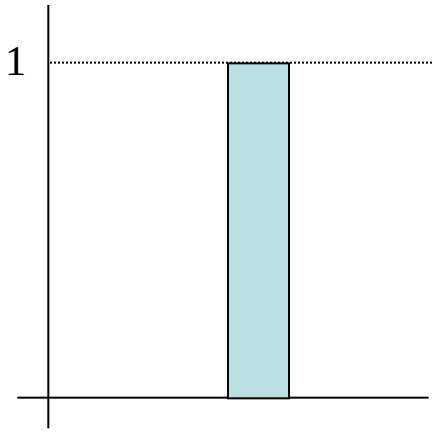
s(i) is the forward image of $s_L^C(i-1)$

s(i) is the backward image of $s_L^A(i+1)$

Causal patterns are linked to predictability

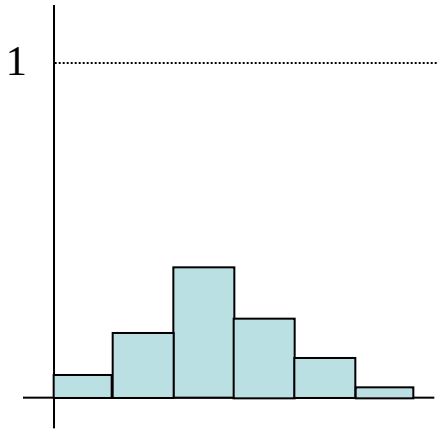
s(i) might be predicted using $s_L^C(i-1)$

Conditional distribution of s(i) given $s_L^C(i-1)$



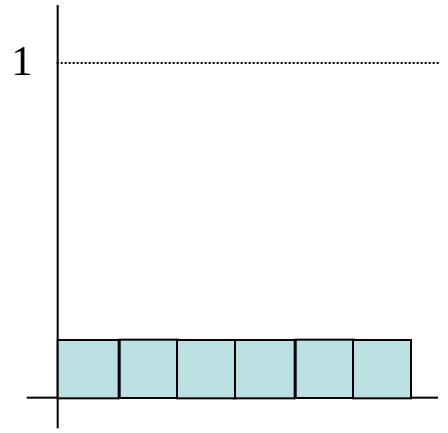
s(i) can be fully predicted given L previous symbols

Conditional distribution of s(i) given $s_L^C(i-1)$



s(i) can be largely predicted given L previous symbols

Conditional distribution of s(i) given $s_L^C(i-1)$

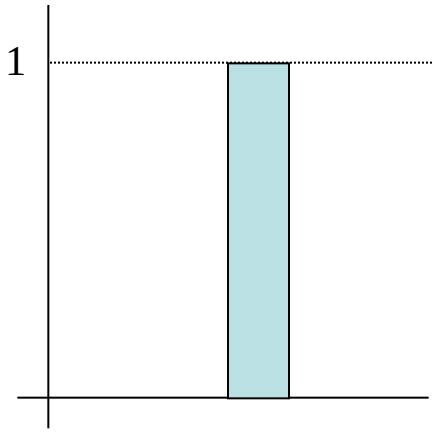


s(i) cannot be predicted given L previous symbols

Causal patterns are linked to conditional entropy

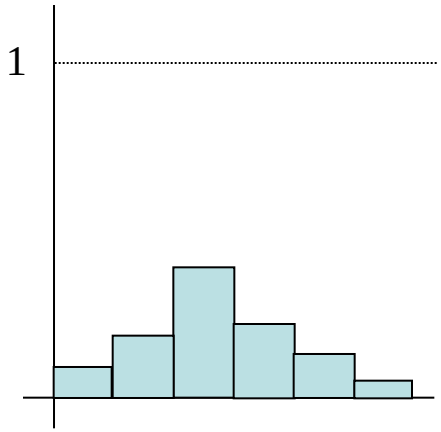
The information carried by s(i) might be reduced given $s_L^C(i-1)$

Conditional distribution of s(i) given $s_L^C(i-1)$



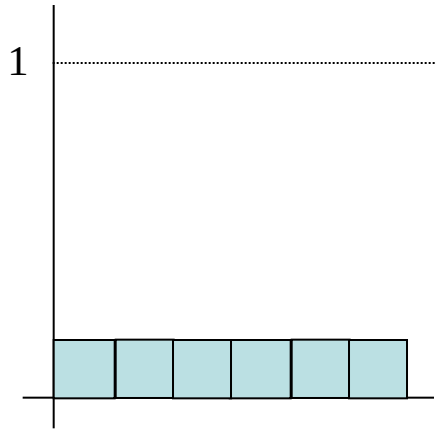
No information is carried by s(i) given L previous symbols

Conditional distribution of s(i) given $s_L^C(i-1)$



A certain amount of information is carried by s(i) given L previous symbols

Conditional distribution of s(i) given $s_L^C(i-1)$

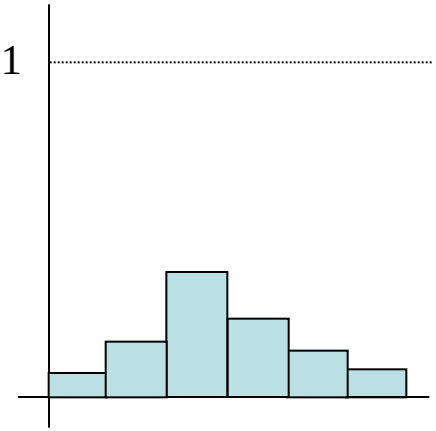


The information carried by s(i) cannot be reduced given L previous symbols

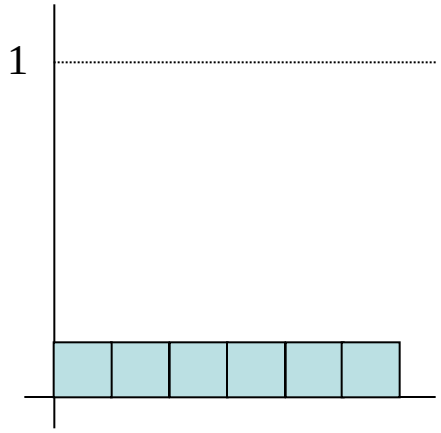
Anti-causal patterns are linked to reversibility

Since reversing time makes anti-causal patterns into causal ones (and vice versa), anti-causal patterns are helpful to test reversibility (i.e. the preservation of the statistical properties after time reversal)

Conditional distribution of s(i) given $s_L^C(i-1)$



Conditional distribution of s(i) given $s_L^A(i+1)$



Since the conditional distributions are different, statistical properties are not maintained after time reversal

Outline

1) Symbolization strategies

2) Pattern construction via delay embedding procedure

3) Symbolization strategy and pattern construction technique impose a coarse graining of the multidimensional space

4) Techniques to optimize the pattern length

5) Approaches to reduce redundancy of patterns

6) Indexes derived from symbolization and pattern formation strategies

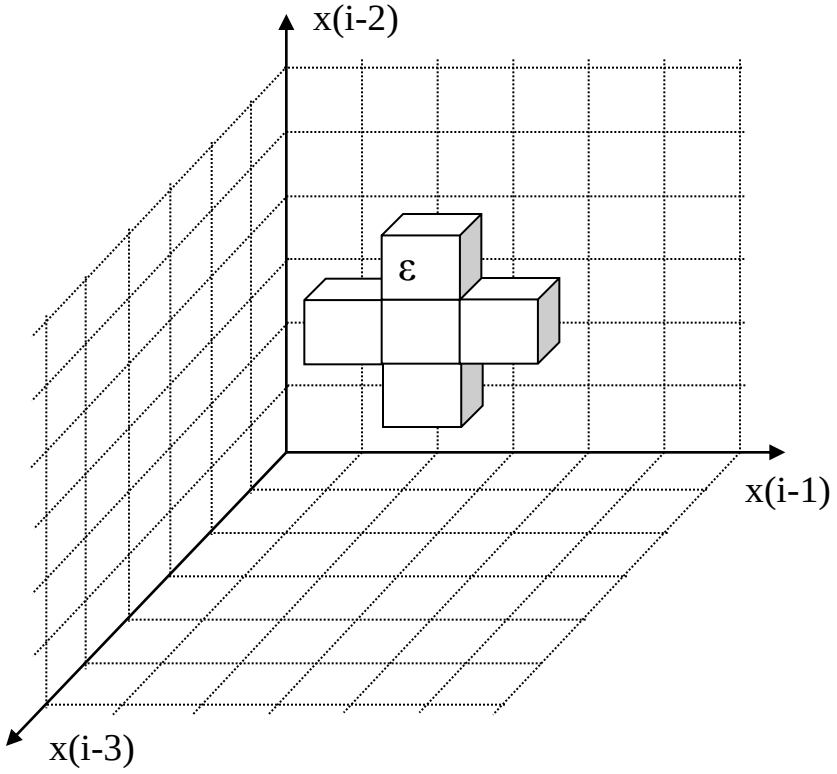
7) Application in clinics

Uniform partition

- uniform quantization
- delay embedding procedure

uniform partition

L=3, q=6



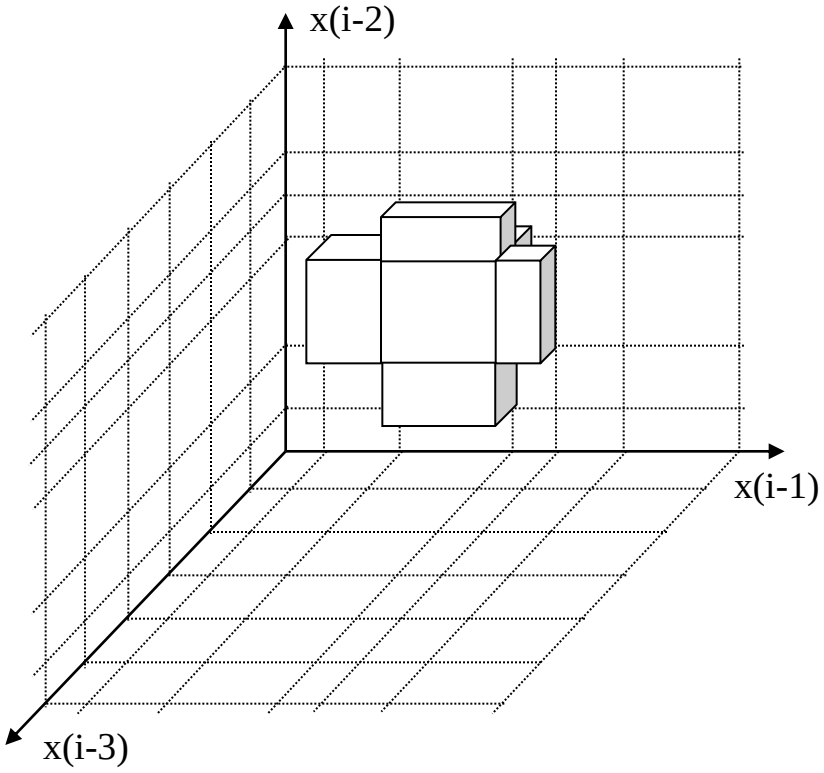
- Cells are hyper-cubes of side ε
- The number of cells are q^L
- Cells have the same size
- Cells cover the entire embedding space
- Cells do not intersect
- The number of patterns in each cell might be variable

Non uniform partition

- non uniform quantization
- delay embedding procedure

non uniform partition

L=3, q=6

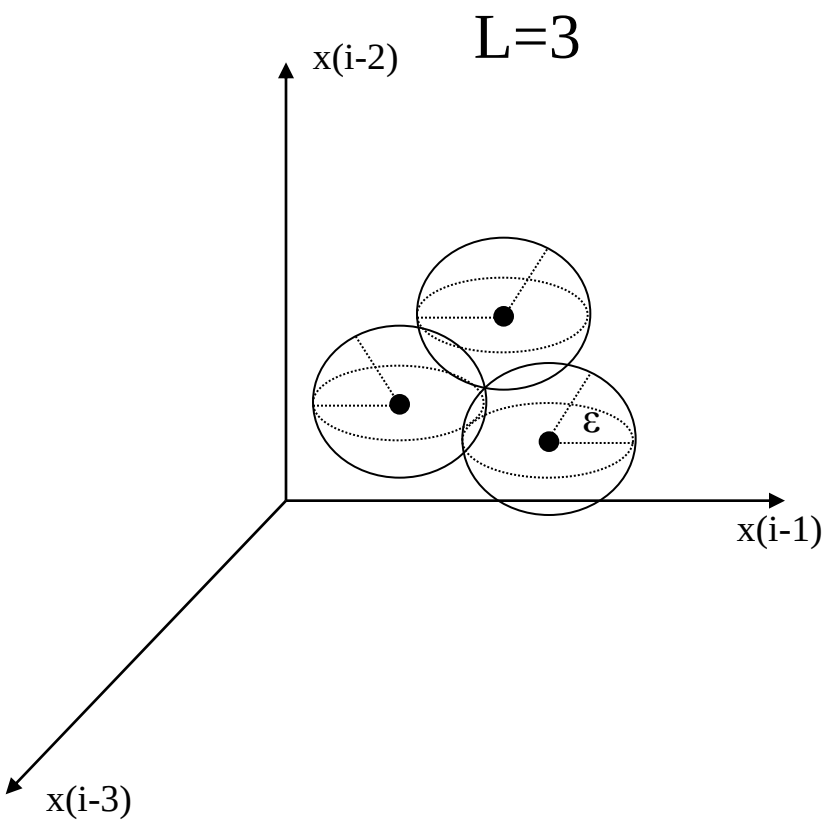


- Cells are hyper-boxes
- The number of cells are q^L
- Cells have different size
- Cells cover the entire embedding space
- Cells do not intersect
- The number of patterns in each cell might be made constant

Pattern-centered uniform coarse graining

- sample-centered uniform quantization
- delay embedding procedure

→ pattern-centered uniform coarse graining

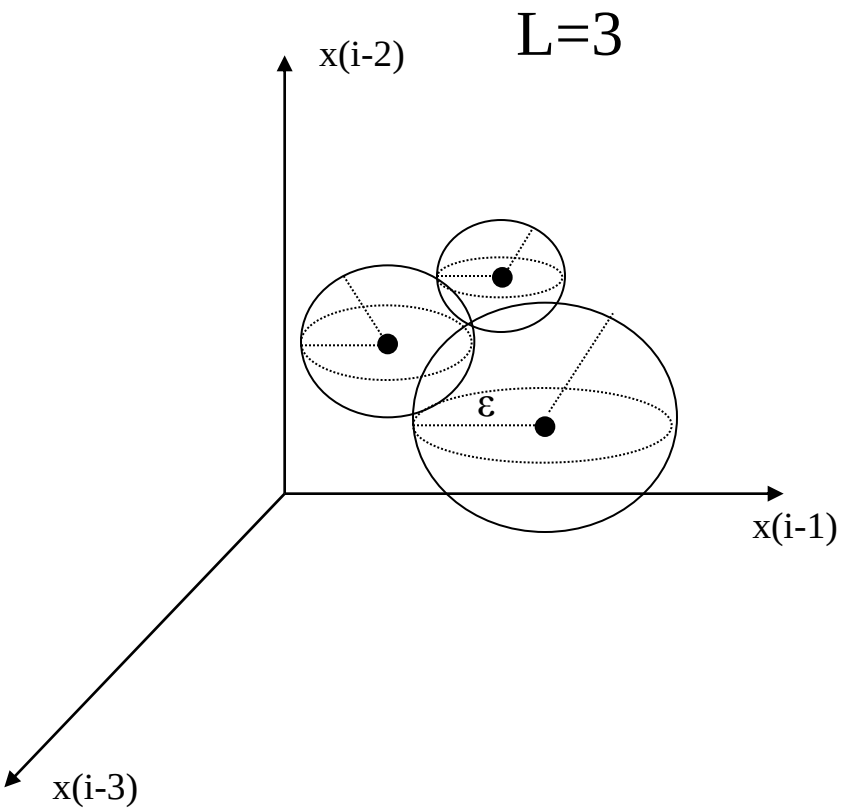


- Cells are hyper-spheres of radius ε
- The number of cells are N-L+1
- Cells are built around patterns
- Cells have equal size
- Cells cover the entire embedding space
- Cells intersect
- The number of patterns in each cell might be variable

Pattern-centered non uniform coarse graining

- sample-centered non uniform quantization
- delay embedding procedure

→ pattern-centered non uniform coarse graining



- Cells are hyper-spheres of radius ε
- The number of cells are N-L+1
- Cells are built around patterns
- Cells have different size
- Cells cover the entire embedding space
- Cells intersect
- The number of patterns in each cell might be made constant

Outline

1) Symbolization strategies

2) Pattern construction via delay embedding procedure

3) Symbolization strategy and pattern construction technique impose a coarse graining of the multidimensional space

4) Techniques to optimize the pattern length

5) Approaches to reduce redundancy of patterns

6) Indexes derived from symbolization and pattern formation strategies

7) Application in clinics

Techniques for the optimization of the pattern length

1) Fixing the pattern length, L, according to the series length, N, and the number of symbols, q

2) Using a strategy penalizing unreliable in-sample prediction or conditional entropy

3) Exploiting in-sample predictability loss or conditional entropy rise while increasing L using pattern-centered non uniform coarse graining

4) Taking advantage from out-of-sample prediction

Fixing the pattern length

L = pattern length
N = series length
q = number of symbols

$N >> q^L$ → Several patterns are present in each cell of partition

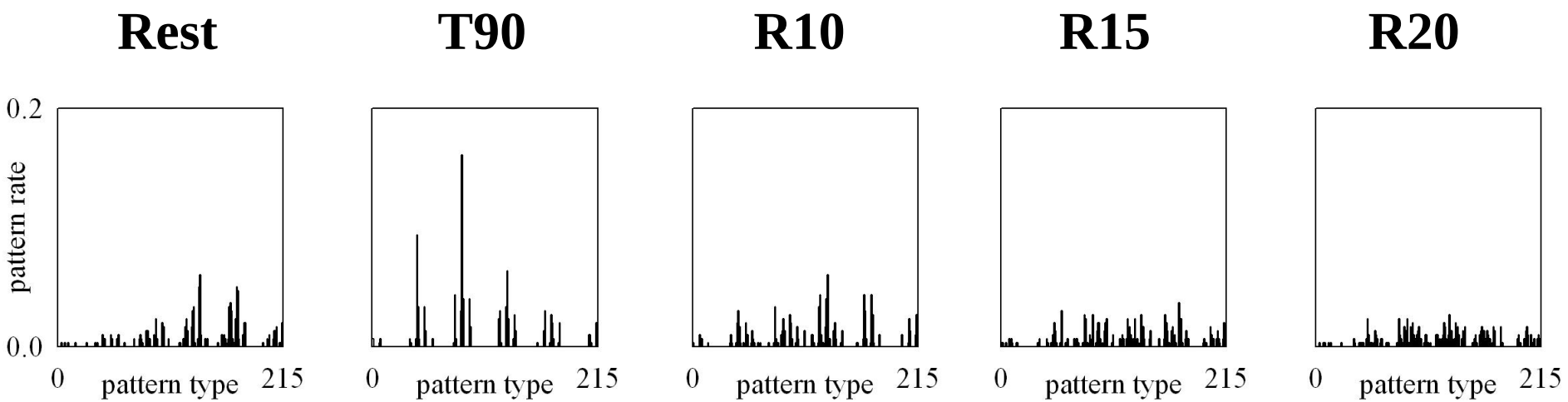
Example

In short-term cardiovascular variability analysis

q=6
N=300 → L = 3

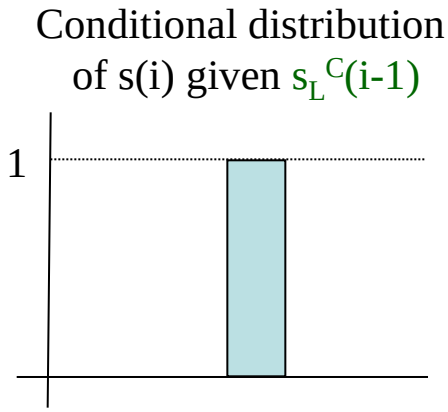
A. Porta et al, IEEE Trans Biomed Eng, 48:1282-1291, 2001

Length-three pattern distribution



A. Porta et al, IEEE Trans Biomed Eng, 48:1282-1291, 2001

Using a strategy penalizing unreliable in-sample prediction or conditional entropy



Perfect in-sample prediction or false certainty ?

If the number of $s_L^C(i-1)$ is large ⟹ Reliability of prediction is high

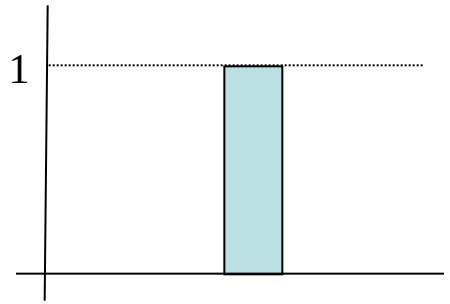
If the number of $s_L^C(i-1)$ is small ⟹ Reliability of prediction is low

If the number of $s_L^C(i-1)$ is 1 ⟹ Insufficient knowledge, likely to produce a false certainty

Using a strategy penalizing unreliable in-sample prediction or conditional entropy

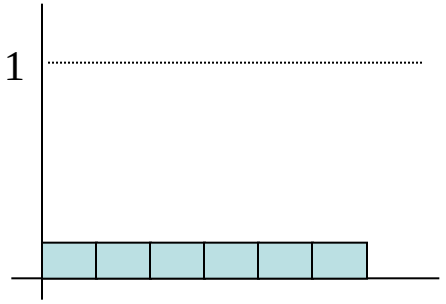
If the number of $s_L^C(i-1)$ is 1

Conditional distribution of s(i) given $s_L^C(i-1)$



Likely false certainty

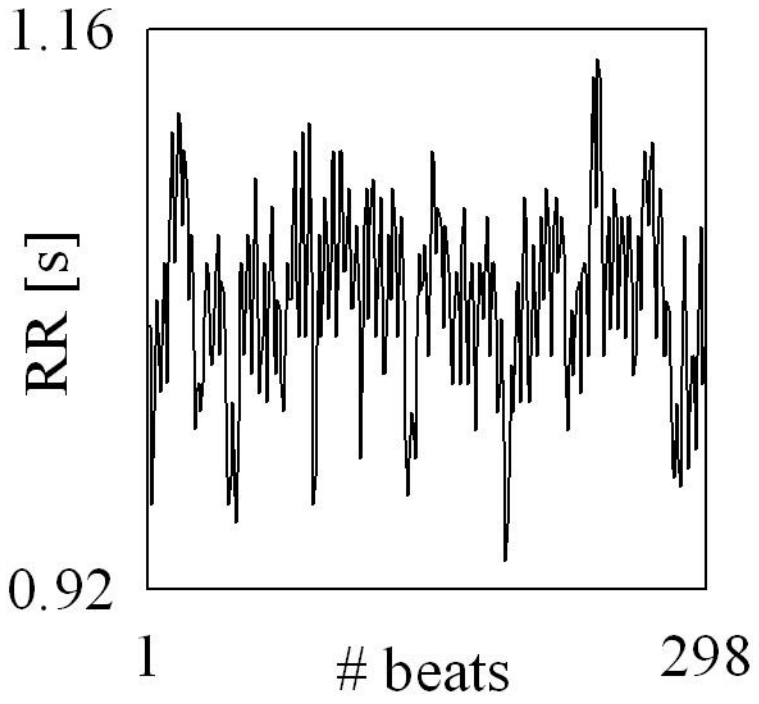
Conditional distribution of s(i) given $s_L^C(i-1)$

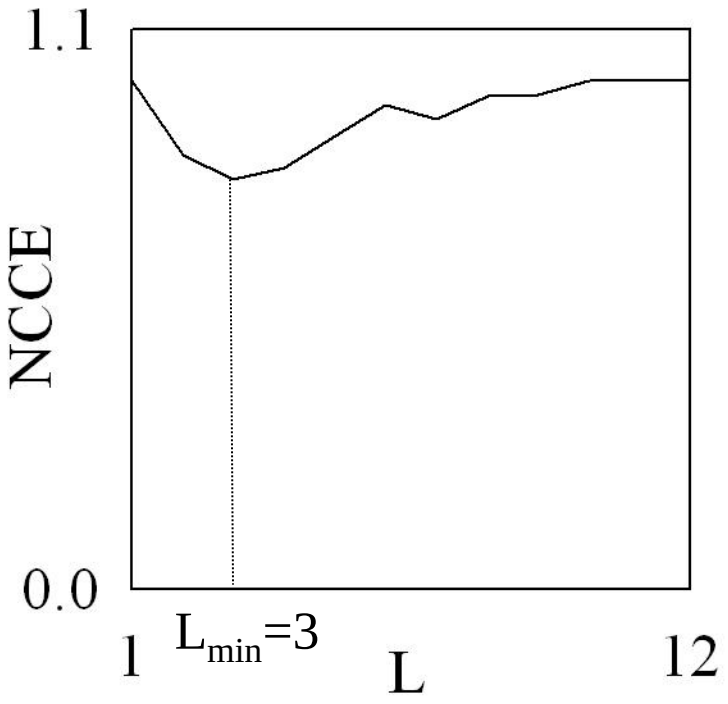


True uncertainty

Optimal L minimizes mean square prediction error or conditional entropy

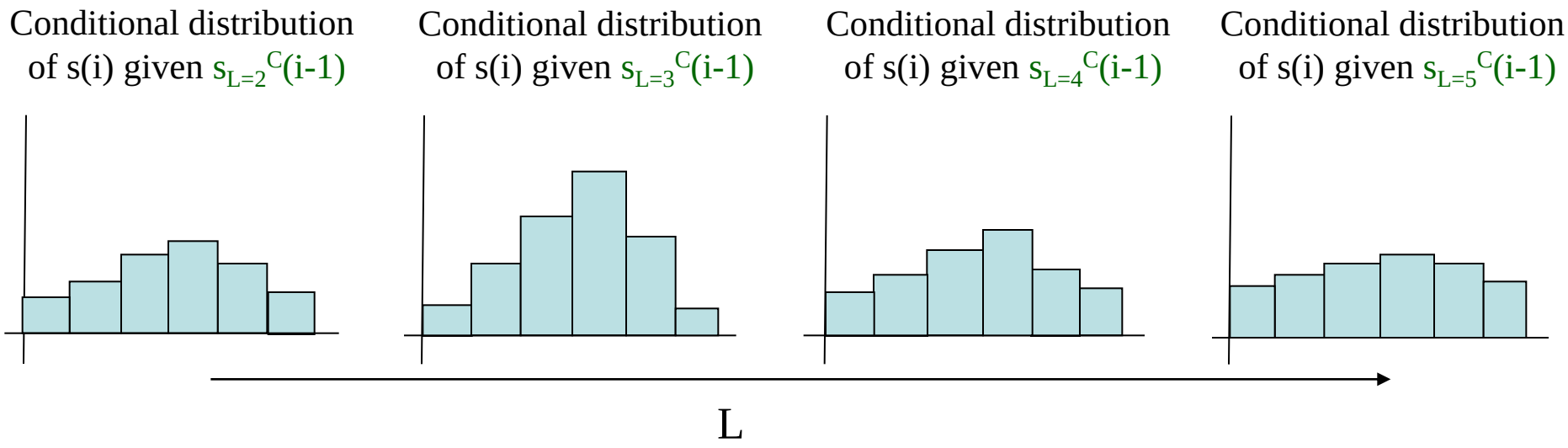
Normalized corrected conditional entropy





A. Porta et al, Biol Cybern, 78:71-78, 1998

Sample-centered non uniform coarse graining implies in-sample predictability loss or conditional entropy rise at large L

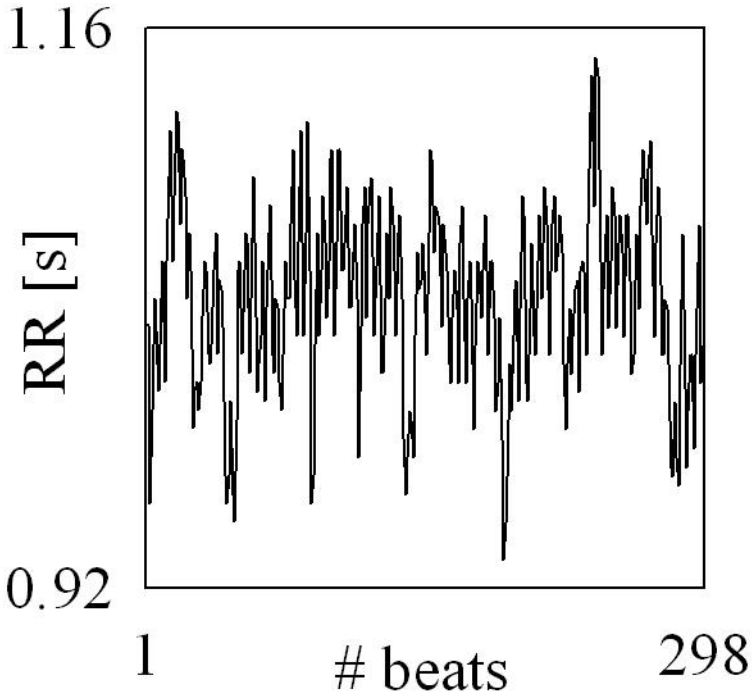


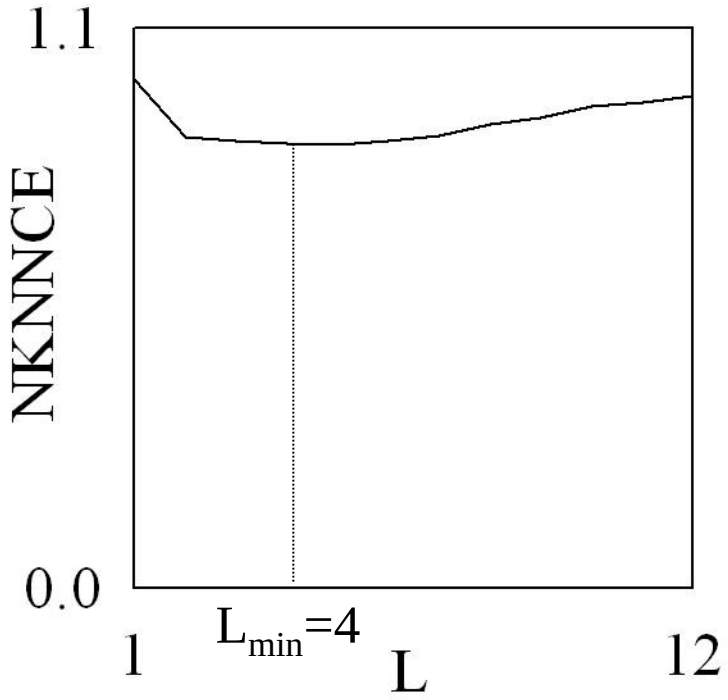
At low L, increasing L might be helpful to unfold dynamics, thus improving predictability and reducing conditional entropy

At high L, increasing L leads to the increase of cell size, thus decreasing predictability and increasing conditional entropy

Optimal L minimizes mean square prediction error or conditional entropy

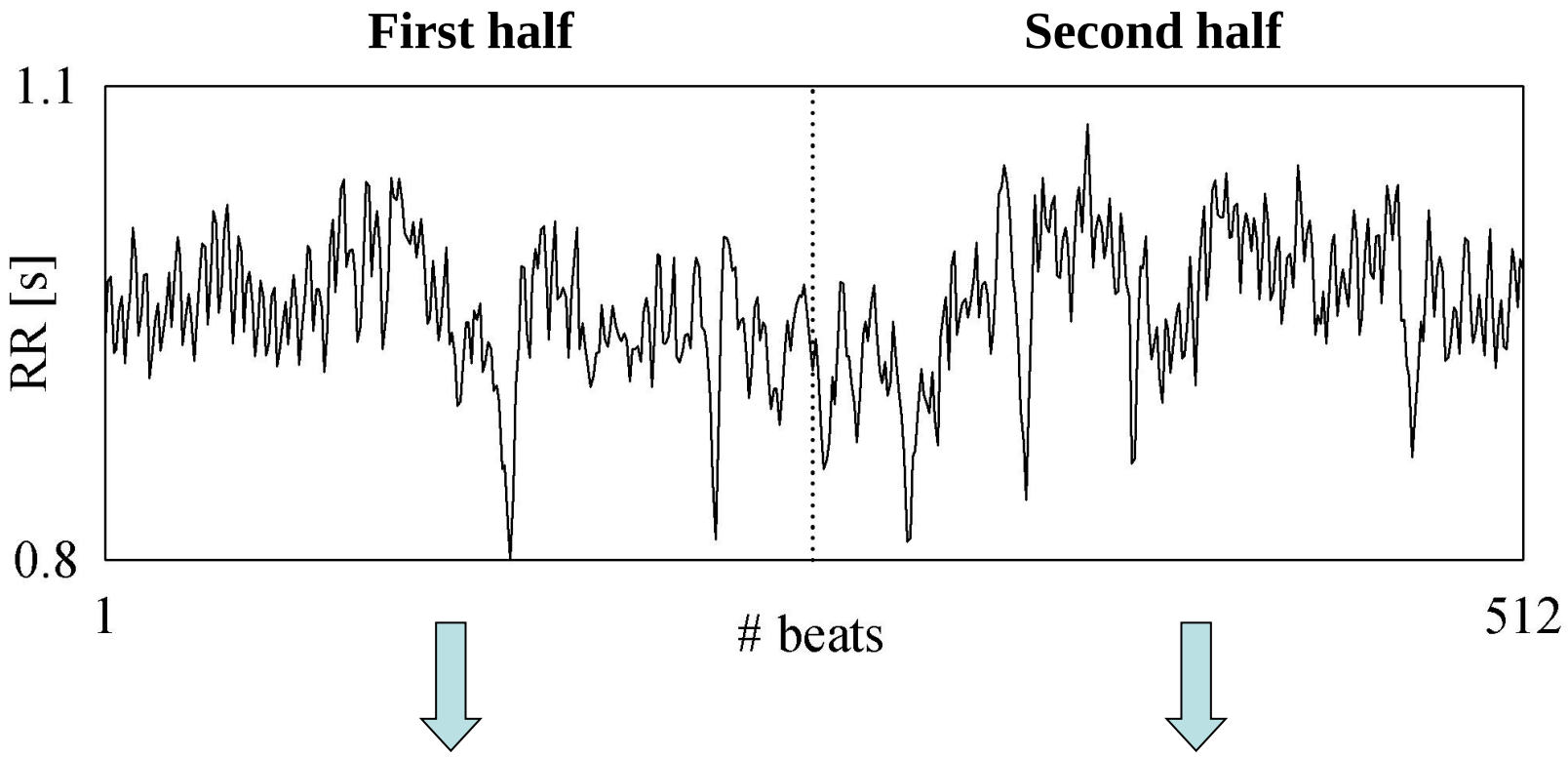
Normalized k-nearest-neighbor conditional entropy





Porta A et al, Physiol Meas, 34:17-33, 2013

Out-of-sample prediction



- Construction of the pattern library
- Assessment of the conditional distributions
- Definition of the predictor

- test for the predictor as a function of L

Optimal L minimizes mean square out-of-sample prediction error

Outline

1) Symbolization strategies

2) Pattern construction via delay embedding procedure

3) Symbolization strategy and pattern construction technique impose a coarse graining of the multidimensional space

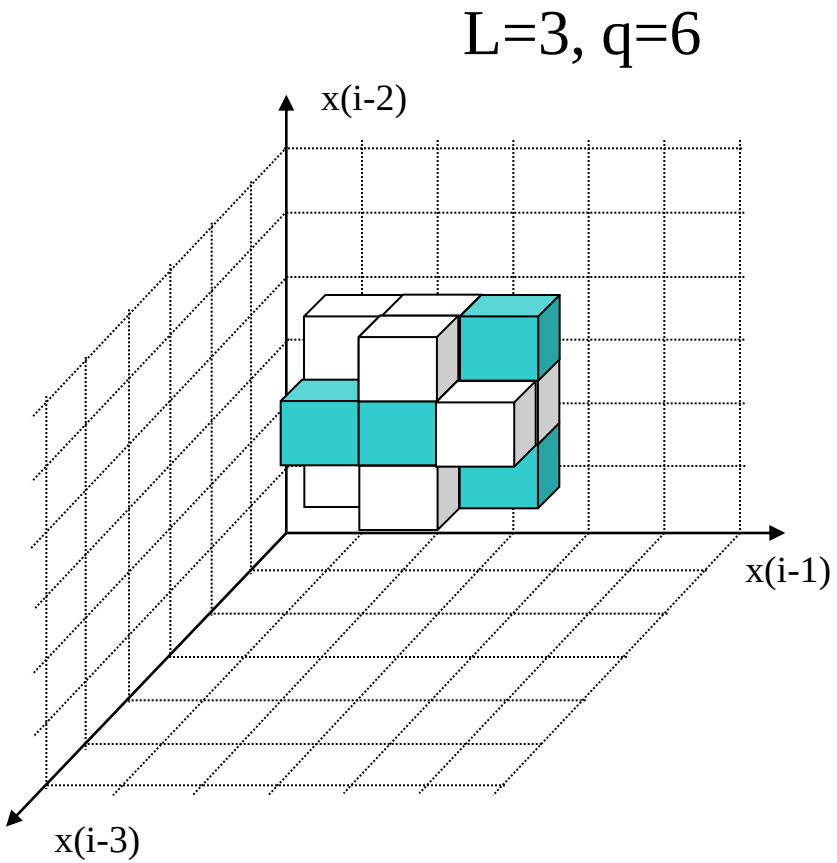
4) Techniques to optimize the pattern length

5) Approaches to reduce redundancy of patterns

6) Indexes derived from symbolization and pattern formation strategies

7) Application in clinics

Techniques to reduce the redundancy of patterns



Adjacent and non adjacent cells are merged according to some criteria

Redundancy reduction technique based on the frequency content of the patterns

$S=\{0,1\}$ and $s_L^C=\{s_L^C(i)$, with $i=1,\ldots,N-L+1\}$

Most stable patterns

$\{(0,0,\ldots,0,0), (1,1,\ldots,1,1)\}$

…

$\{\ldots, \ldots\}$

Most variable patterns

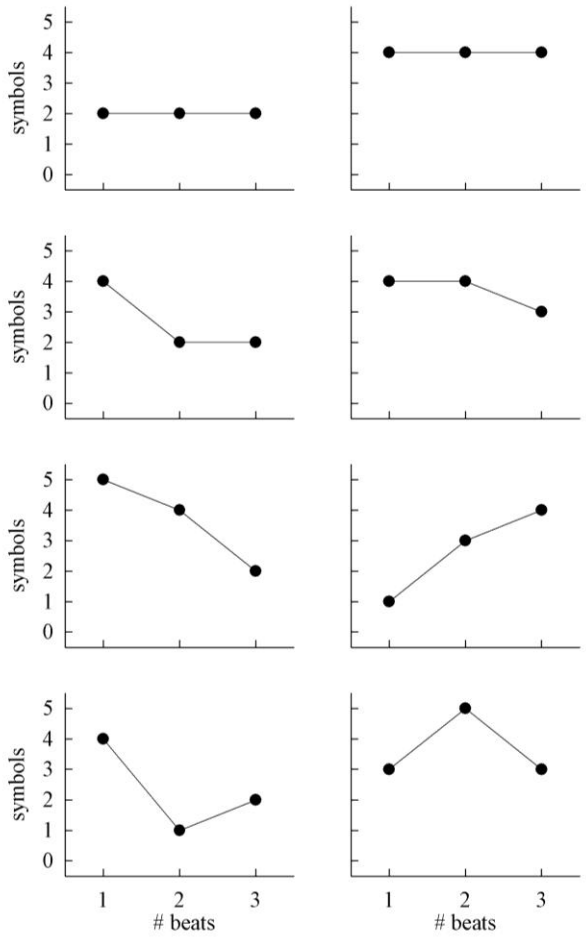
$\{(0,1,\ldots,0,1), (1,0,\ldots,1,0)\}$

Frequency of the dominant rhythm

Example of redundancy reduction technique based on the frequency content of the patterns

S={0,1,2,3,4,5} and s_L^C={s_L^C(i), with i=1,…,N-L+1} with L=3

Pattern number ⟹ 6^3=216



0V

1V

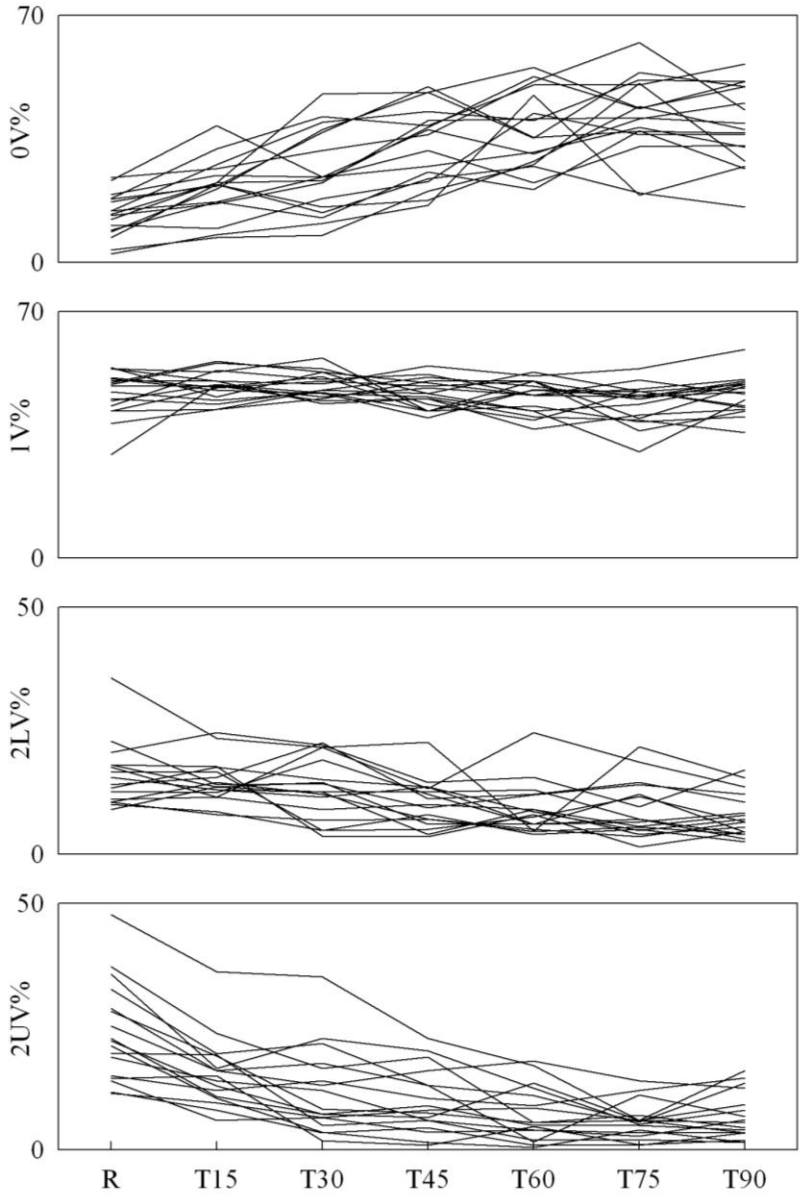
2LV

2UV

4 pattern classes

A. Porta et al, IEEE Trans Biomed Eng, 48:1282-1291, 2001

Symbolic dynamics and autonomic nervous system



A. Porta et al, Am J Physiol, 293:H702-H708, 2007

Redundancy reduction technique based of the Shannon entropy (SE) of the patterns

$S=\{0,1\}$ and $s_L^C=\{s_L^C(i)$, with $i=1,\ldots,N-L+1\}$

Least informative patterns

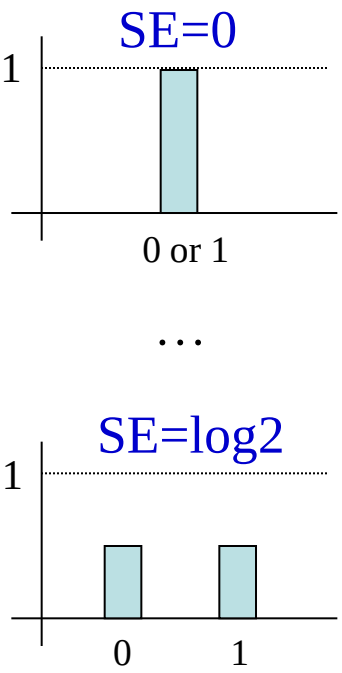
$\{(0,0,\ldots,0,0), (1,1,\ldots,1,1)\}$

…

$\{\ldots, \ldots\}$

Most informative patterns

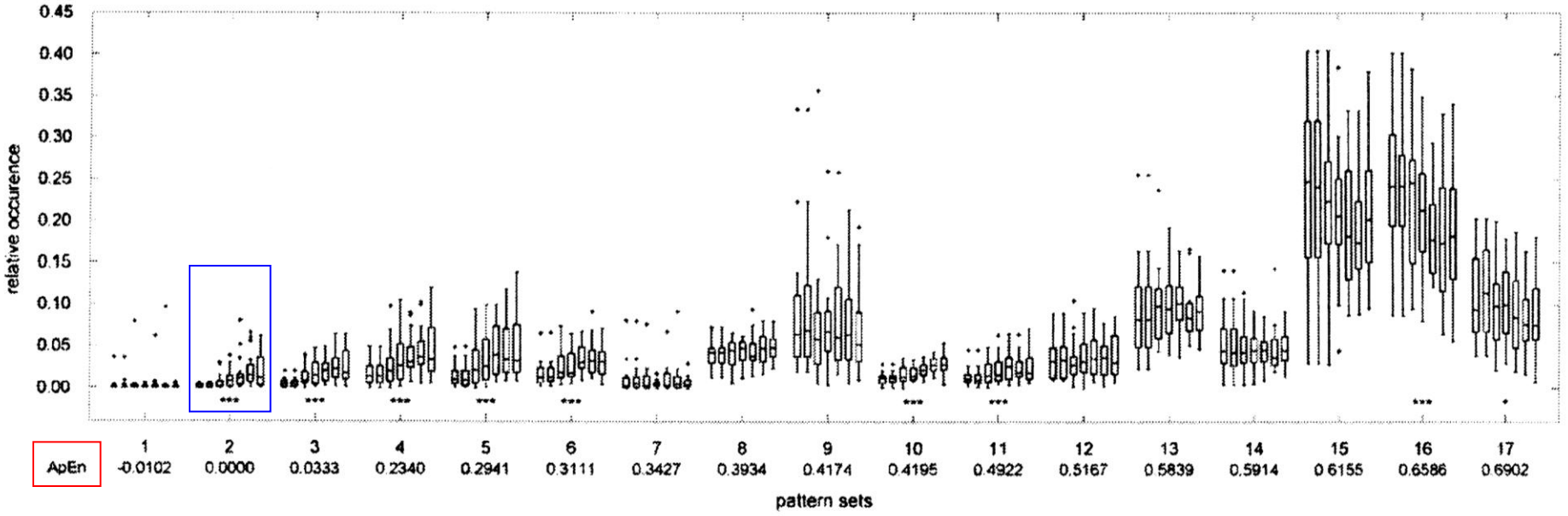
$\{(0,1,...,0,1), (0,\ldots,0,1,\ldots,1), \ldots\}$



Example of redundancy reduction technique based on the entropy of the patterns

S={0,1} and s_L^C={s_L^C(i), with i=1,…,N-L+1} with L=8

Pattern number ⟹ 2^8=256

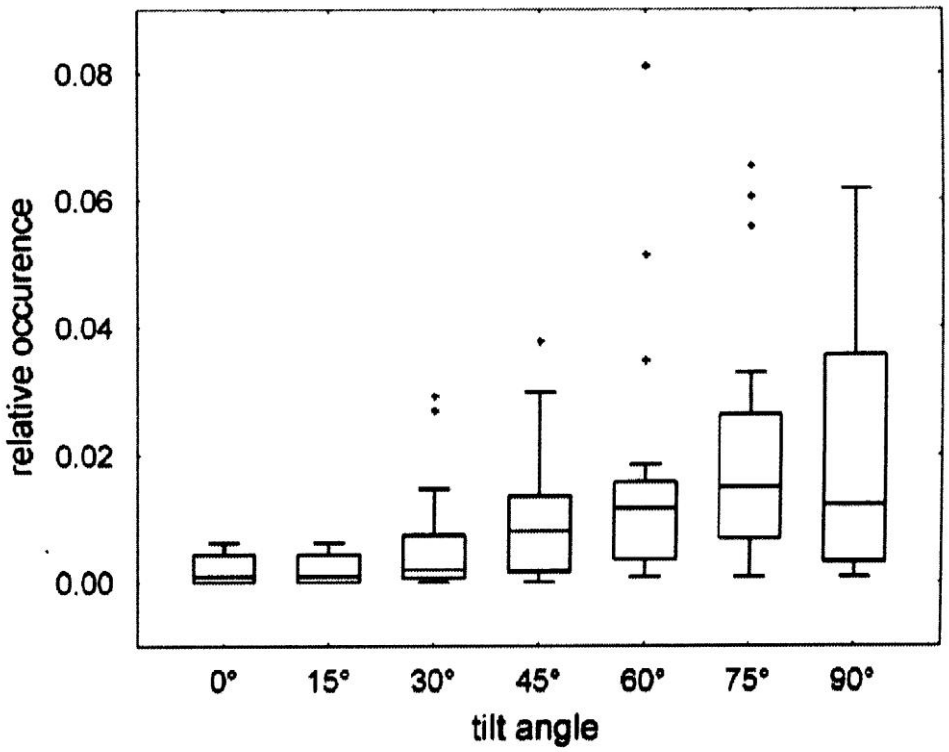


17 pattern classes

D. Cysarz et al, Comp Biol Med, 42:313-318, 2012

Symbolic dynamics and autonomic nervous system

Binary patterns of length L=8 with ApEn=0.0



D. Cysarz et al, Comp Biol Med, 42:313-318, 2012

Outline

1) Symbolization strategies

2) Pattern construction via delay embedding procedure

3) Symbolization strategy and pattern construction technique impose a coarse graining of the multidimensional space

4) Techniques to optimize the pattern length

5) Approaches to reduce redundancy of patterns

6) Indexes derived from symbolization and pattern formation strategies

7) Application in clinics

Time domain indexes derived from symbolization and pattern formation strategies

- Sample frequency of a specific pattern

N. Wessel et al, Phys Rev E, 61, 733-739, 2000

- Sample frequency of a specific pattern class

A. Porta et al, IEEE Trans Biomed Eng, 48, 1282-1291, 2001
D. Cysarz et al, Comp Biol Med, 42, 313-318, 2012

- Forward mean square prediction error between the current value and its best prediction based on causal patterns

A. Porta et al, IEEE Trans Biomed Eng, 47, 1555-1564, 2000
A. Porta et al, IEEE Trans Biomed Eng, 54, 94-106 , 2007
N. Wessel et al, Med Biol Eng Comput, 44, 321-330, 2006

- Backward mean square prediction error between the current value and its best prediction based on anti-causal patterns

A. Porta et al, Phil Trans R Soc A, 367:1359-1375, 2009

Information domain indexes derived from symbolization and pattern formation strategies

- Shannon entropy of the pattern distribution

N. Wessel et al, Phys Rev E, 61:733-739, 2000

A. Porta et al, IEEE Trans Biomed Eng, 48:1282-1291, 2001

- Entropy rate of the series

corrected approximate entropy (CApEn)

S.M. Pincus, Chaos, 5:110-117, 1995

A. Porta et al, J Appl Physiol, 103:1143-1149, 2007

sample entropy (SampEn)

J.S. Richman and J.R. Moorman, Am J Physiol, 278:H2039-H2049, 2000

corrected conditional entropy (CCE)

A. Porta et al, Biol Cybern, 78:71-78, 1998

- Mutual information

D. Hoyer et al, IEEE Trans Biomed Eng, 52:584-592, 2005

Outline

1) Symbolization strategies

2) Pattern construction via delay embedding procedure

3) Symbolization strategy and pattern construction technique impose a coarse graining of the multidimensional space

4) Techniques to optimize the pattern length

5) Approaches to reduce redundancy of patterns

6) Indexes derived from symbolization and pattern formation strategies

7) Application in clinics

Application on symbolic dynamics in clinics

Nonlinear Indices of Heart Rate Variability in Chronic Heart Failure Patients: Redundancy and Comparative Clinical Value

ROBERTO MAESTRI, M.S.,* GIAN DOMENICO PINNA, M.S.,* AGOSTINO ACCARDO, PH.D.,† PAOLO ALLEGRINI, PH.D.,‡ RITA BALOCCHI, PH.D.,¶ GIANNI D'ADDIO, M.S.,* MANUELA FERRARIO, M.S.,§ DANILO MENICUCCI, M.S.,¶ ALBERTO PORTA, PH.D.,** ROBERTO SASSI, PH.D.§,†† MARIA GABRIELLA SIGNORINI, PH.D.,§ MARIA TERESA LA ROVERE, M.D.,‡‡ and SERGIO CERUTTI, PH.D.§ [Correction added after online publication 7 March 2007: author listing has been amended to include more complete author names.]

From the *Dipartimento di Bioingegneria, Fondazione S. Maugeri, IRCCS, Montescano and Telese, Italy; †Dipartimento di Elettronica e Informatica, Università di Trieste, Italy; ‡Dipartimento di Fisica, Università di Pisa, Italy; ¶Istituto di Fisiologia Clinica, CNR, Pisa, Italy; §Dipartimento di Bioingegneria, Politecnico di Milano, Italy; **Dipartimento Scienze Precliniche, LITA Vialba, Università degli Studi di Milano, Italy; ††Dipartimento di Tecnologie dell'Informazione, Università degli Studi di Milano, Italy; and ‡‡Dipartimento di Cardiologia, Fondazione S. Maugeri, IRCCS, Montescano, Italy

Best multivariate clinical model for the prediction of the total cardiac death in heart failure population

TABLE 5

Cox Prognostic Model Based on Known Clinical and Functional Risk Factors

Variable	Regression Coefficient	Standard Error	P Value	Bootstrap Selection (%)
NYHA	0.718	0.245	0.003	90
LVEF	−0.053	0.019	0.006	91
Peak VO_2	−0.045	0.026	0.084	64
SAP	−0.015	0.009	0.123	61
Etiology ischemic	—	—	0.418	20
LVEDD	—	—	0.718	14
VPCs/hour	—	—	0.769	9
Sodium	—	—	0.842	15

Variable selection was carried out by backward elimination (significance level: 15%). The last column reports the frequency of selection of variables using the same procedure on 500 bootstrap samples.

For the definition of variables see Table 2. Variables with missing coefficient and standard error are those removed from the full model by the backward elimination procedure.

Note that at the selected 15% significance level only the top four variables are left in the final "clinical" model. As can be seen, these variables also performed well in the bootstrap validation.

R. Maestri et al, J Cardiovasc Electrophysiol 18:425-433, 2007

Linear and non linear indexes derived from heart rate variability

- Linear indexes in the time and frequency domains

TABLE 1

List of the 20 Nonlinear Indices of Heart Rate Variability Examined in the Study, with a Classification of the Families to Which They Belong

Variable	Description	Family
1VP	One variation pattern	Symbolic dynamics
2UVP	Two unlike variations pattern	Symbolic dynamics
BNI	Binary nonrandomness index	Symbolic dynamics
BLZC	Binary Lempel-Ziv complexity	Entropy
DELTA	Long-range memory in RR time series	Entropy
SampEn	Sample entropy	Entropy
DFA	Short-term detrended fluctuation analysis	Fractality-multifractality
HFD	Higuchi fractal dimension	Fractality-multifractality
1/f slope	Slope of the power-law regression line	Fractality-multifractality
SMFSr	Ratio between the width of the singularity multifractal spectrum and the same quantity after phase randomization	Fractality-multifractality
UPI	Non-normalized unpredictability index	Predictability
UPIn	Normalized unpredictability index	Predictability
IMAI1	Ratio between the power associated with the mode with frequency closest to 0.1 Hz (LF1) and the power of modes with frequencies higher than LF1	Empirical mode decomposition
IMAI2	Ratio between the power associated with the first mode with frequency <LF1 and the modes with frequencies higher than LF1 (see IMAI1)	Empirical mode decomposition
pLF2	Power associated with the first mode with frequency < LF1 (see IMAI1)	Empirical mode decomposition
LEN	Length of the bi-dimensional Poincaré plots	Poincaré plots
SD12	Ratio between the axes of the ellipse fitting bi-dimensional Poincaré plots	Poincaré plots
RAD_X	Radius of the semi-ellipse of inertia along the X axis of the 3-dimensional Poincaré plot	Poincaré plots
RAD_Y	Radius of the semi-ellipse of inertia along the Y axis of the 3-dimensional Poincaré plot	Poincaré plots
RAD_Z	Radius of the semi-ellipse of inertia along the Z axis of the 3-dimensional Poincaré plot	Poincaré plots

R. Maestri et al, J Cardiovasc Electrophysiol 18: 425-433, 2007

Additive predictive value to the best multivariate clinical model

TABLE 7

Additive Predictive Values of HRV Parameters to the Clinical Predictors

Variable	Family	P Value	Bootstrap $P \leq 0.05$ (%)
IMAI1	Empirical mode decomposition	0.166	27
1/f slope	Fractality-multifractality	0.571	8
RAD_Y	Poincaré plots	0.379	11
SMFSr	Fractality-multifractality	0.101	31
HFD	Fractality-multifractality	0.385	12
RAD_X	Poincaré plots	0.350	15
DELTA	Entropy	0.797	7
1VP	Symbolic dynamics	0.007	74
SampEn	Entropy	0.461	15
BNI	Symbolic dynamics	0.512	9
IMAI2	Empirical mode decomposition	0.023	67
SD12	Poincaré plots	0.123	35

The table reports the P values for selected variables in each cluster, after entering them into the prognostic model shown in Table 5. The last column reports the percentage of times the variable entered the model with a P value ≤ 0.05 in the 500 bootstrap samples.

R. Maestri et al, J Cardiovasc Electrophysiol 18: 425-433, 2007

Conclusions

Symbolization procedures and pattern formation techniques impose a coarse graining over the multidimensional space

Techniques to optimize the pattern length are necessary to make consistent indexes derived from symbolization and pattern formation strategies

Techniques to reduce redundancy of patterns are necessary to focus a limited amount of features

Symbolization and pattern formation strategies provided indexes improving the best multivariate model based on traditional clinical parameters for the prediction of total cardiac death in heart failure population