

PTC 2456 – Proc. Sinais Biomédicos
Análise, quantificação, classificação de sinais biol.

Análise e classificação: sinais biológicos

Prof. Sérgio S Furuie

EPUSP PTC/LEB - S.Furuie

Mobilização/motivação

- Subsídios objetivos para a tomada de decisão
- Auxílio para a interpretação médica
- Exemplos:
 - Área de QRS
 - Variabilidade de IRR
 - Máxima taxa de variação
 - Classe

EPUSP PTC/LEB - S.Furuie

Detecção do QRS

Time (s)

Der. 1^a

Der. 2^a

QRS

y1=[1a. difer.]

y2=[smooth y1]

y3=[der. y2]

y4=y2+y3

limiar em y4

EPUSP PTC/LEB - S.Furuie

Como classificar automaticamente ?

Símbolo	Descrição
.	Normal beat
V	Premature ventricular contraction
L	Left bundle branch block beat
R	Right bundle branch block beat
...	...

EPUSP PTC/LEB - S.Furuie

Can you tell health from disease?

- Pathology can affect physiologic recordings in unexpected and interesting ways. PhysioNet encourages researchers to find new ways to extract information hidden in data. For example, the figure above shows the instantaneous heart rates of four subjects. Can you tell which it is?

Heart Rate Dynamics in Health and Disease:
A Time Series Test

Heart Rate (bpm)

Time (min)

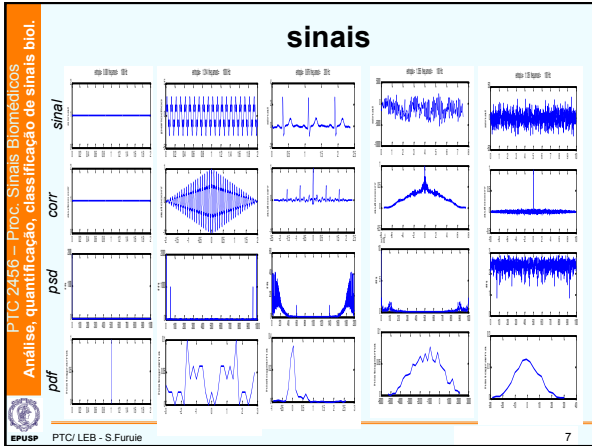
EPUSP PTC/LEB - S.Furuie

CONCEITOS IMPORTANTES

Formas para representar sinais:

- 1) Sinais no tempo
- 2) Sinais no domínio da frequência (projeção em bases ortonormais)
- 3) Autocorrelação
- 4) Densidade espectral de potência
- 5) Função distribuição de probabilidade
- 6) Por características (scatter plot)
- 7) ...

EPUSP PTC/LEB - S.Furuie



PTC 2456 – Proc. Sinais Biomédicos
Análise, quantificação, classificação de sinais biol.

Correlação entre sinais

Correlação [sinais $x(t)$ e $y(t)$]

- Função correlação cruzada: $corr_{xy}(\tau)$
- F. autocorrelação: $corr_{xx}(\tau)$
- Correlação: $corr(x,y)$

Coefficiente de correlação

- Normalizado entre [0,1]
- Extraídas as médias
- Normalizado pelo d. padrão

$$corr_{xy}(\tau) = \frac{1}{T} \int_0^T x(t).y(t+\tau) dt$$

$$corr_{xx}(\tau) = \frac{1}{T} \int_0^T x(t).x(t+\tau) dt$$

$$c_{xy} = corr_{xy}(0) = \frac{1}{T} \int_0^T x(t).y(t) dt$$

$$corr_{xx}(\tau) = \frac{1}{N} \sum_0^{N-1} x(n).y(n)$$

$$\rho_{xy} = \frac{\frac{1}{T} \int_0^T (x(t) - \mu_x).(y(t) - \mu_y) dt}{\sigma_x \sigma_y}$$

$$= \frac{\frac{1}{N} \sum_0^{N-1} (x(n) - \mu_x).(y(n) - \mu_y)}{\sigma_x \sigma_y}$$

σ_x, σ_y : desvio padrao

Matlab: `xcorr(x,y)`, `corr(x,y)`, `corrcoef(x,y)`

EPUSP PTC/LEB - S.Furule A rigor a definição envolve limite T-> infinito 8

PTC 2456 – Proc. Sinais Biomédicos
Análise, quantificação, classificação de sinais biol.

Correlação discreta

Correlação [sinais $x(n)$ e $y(n)$]

- Correlação: c_{xy}
- Função corr. cruzada: $corr_{xy}(k)$
- F. autocorrelação: $corr_{xx}(k)$

- Coef. de correlação
- F. Coeficiente de corr. Cruzada

$$c_{xy} = corr_{xy}(0) = \frac{1}{N} \sum_{n=0}^{N-1} x(n).y(n)$$

$$corr_{xy}(k) = \frac{1}{N} \sum_{n=0}^{N-1-k} x(n).y(n+k)$$

$$corr_{xx}(k) = \frac{1}{N} \sum_{n=0}^{N-1-k} x(n).x(n+k)$$

$$\rho_{xy} = \frac{\frac{1}{N} \sum_0^{N-1} (x(n) - \mu_x).(y(n) - \mu_y)}{\sigma_x \sigma_y}$$

$$\rho_{xy}(k) = \frac{\frac{1}{N} \sum_0^{N-1-k} (x(n) - \mu_x).(y(n+k) - \mu_y)}{\sigma_x \sigma_y}$$

μ_x, μ_y : médias
 σ_x, σ_y : desvio padrao

Matlab: `xcorr(x,y)`, `corr(x,y)`, `corrcoef(x,y)`

EPUSP PTC/LEB - S.Furule 9

PTC 2456 – Proc. Sinais Biomédicos
Análise, quantificação, classificação de sinais biol.

Correlação discreta (estocástica)

Correlação [V.As $X(n)$ e $Y(n)$]

- Correlação: c_{xy}
- Função corr. cruzada: $corr_{xy}(k)$
- F. autocorrelação: $corr_{xx}(k)$

- Coef. de correlação
- F. Coeficiente de corr. Cruzada

$$c_{xy} = corr_{xy}(0) = E[X(n).Y(n)]$$

$$corr_{xy}(k) = E[X(n).Y(n+k)]$$

$$corr_{xx}(k) = E[X(n).X(n+k)]$$

$$\rho_{xy} = \frac{E[(X - \mu_x)(Y - \mu_y)]}{\sigma_x \sigma_y}$$

$$\rho_{xy}(k) = \frac{E[(X(n) - \mu_x)(Y(n+k) - \mu_y)]}{\sigma_x \sigma_y}$$

μ_x, μ_y : médias
 σ_x, σ_y : desvio padrao

Matlab: `xcorr(x,y)`, `corr(x,y)`, `corrcoef(x,y)`

EPUSP PTC/LEB - S.Furule 10

PTC 2456 – Proc. Sinais Biomédicos
Análise, quantificação, classificação de sinais biol.

Extração de características

- Atributos relevantes e discriminativos
- Exemplos:
 - Área do QRS
 - Duração do QRS
 - Intervalo RR (IRR)
 - Correlação com uma referência
 - ...

EPUSP PTC/LEB - S.Furule 11

PTC 2456 – Proc. Sinais Biomédicos
Análise, quantificação, classificação de sinais biol.

Detecção do QRS

$y1 = |1a. difer. |$
 $y2 = |smooth y1 |$
 $y3 = |der. y2 |$
 $y4 = y2 + y3$
limiar em $y4$

EPUSP PTC/LEB - S.Furule 12

PTC 2456 – Proc. Sinais Biomédicos
Análise, quantificação, classificação de sinais biol.

Classificação não-supervisionada

- Clusters
 - K-means
 - Fuzzy c-means
 - SOM (Kohonen)

EPUSP PTC/LEB - S.Furule 13

PTC 2456 – Proc. Sinais Biomédicos
Análise, quantificação, classificação de sinais biol.

Classificação supervisionada

- Existem amostras de cada classe (C) tornando possível caracterizar parcialmente cada classe
 - Classificação direta
 - NN-Nearest-neighbor
 - K-NN: classificação pela maioria
 - Definição de distâncias:
 - Euclidiana
 - estatística (Mahalanobis)
 - Treinamento de algoritmos
 - Classificadores otimizados
 - Linear de Fischer
 - Bayesiano
 - SVM
 - Redes neurais (backpropagation)
 - Avaliação de algoritmos

EPUSP PTC/LEB - S.Furule 14

PTC 2456 – Proc. Sinais Biomédicos
Análise, quantificação, classificação de sinais biol.

Classificação de padrões: 1 atributo

Ideal, desejada

0 d_{QRS}

Realidade:

0 ? d_{QRS}

EPUSP PTC/LEB - S.Furule 15

PTC 2456 – Proc. Sinais Biomédicos
Análise, quantificação, classificação de sinais biol.

Classificação de padrões: 2 atributos

0 d_{QRS}

EPUSP PTC/LEB - S.Furule 16

PTC 2456 – Proc. Sinais Biomédicos
Análise, quantificação, classificação de sinais biol.

P atributos, C classes

0 d_{QRS}

IRR

Clusters (nuvens)

EPUSP PTC/LEB - S.Furule 17

PTC 2456 – Proc. Sinais Biomédicos
Análise, quantificação, classificação de sinais biol.

Classificação de padrões: 1 atributo

Ideal, desejada

0 d_{QRS}

Realidade:

0 ? d_{QRS}

EPUSP PTC/LEB - S.Furule 18

PTC 2456 – Proc. Sinais Biomédicos
Análise, quantificação, classificação de sinais biol.

Separando pêras e maçãs usando apenas uma balança ...

EPUSP PTC/LEB - S.Furule 19

PTC 2456 – Proc. Sinais Biomédicos
Análise, quantificação, classificação de sinais biol.

Obtendo o ponto de decisão: ML

T: valor de y com igual possibilidade de ser classe A ou B

$$P(A|y) = P(B|y)$$

$$P(A|y) = \frac{p(y|A) \cdot P(A)}{p(y)}$$

$$P(B|y) = \frac{p(y|B) \cdot P(B)}{p(y)}$$

$$\therefore p(T|A) \cdot P(A) = p(T|B) \cdot P(B)$$

se Gaussianas com mesma variância,

$$p(y|A) = \frac{1}{\sqrt{2\pi}\sigma_A} \exp\left(-\frac{(y-m'_A)^2}{2\sigma_A^2}\right)$$

$$P(A) \cdot \exp\left(-\frac{(T-m'_A)^2}{2\sigma^2}\right) = P(B) \cdot \exp\left(-\frac{(T-m'_B)^2}{2\sigma^2}\right)$$

$$T = \frac{m'_A + m'_B}{2} + \frac{\sigma^2}{m'_A - m'_B} \ln\left(\frac{P(B)}{P(A)}\right)$$

se $P(A) = P(B) \Rightarrow T = \frac{m'_A + m'_B}{2}$

EPUSP PTC/LEB - S.Furule 20

PTC 2456 – Proc. Sinais Biomédicos
Análise, quantificação, classificação de sinais biol.

E SE TIVERMOS MAIS DE 1 ATRIBUTO?

EPUSP PTC/LEB - S.Furule 21

PTC 2456 – Proc. Sinais Biomédicos
Análise, quantificação, classificação de sinais biol.

Classificador Linear de Fischer

Dadas 2 classes A e B com alta dimensionalidade, como obter um classificador linear otimizado?

d: atributos

- Obter a combinação linear (pesos w) que maximiza a separabilidade entre as 2 classes
- Obter o ponto de decisão (threshold)

$$\sigma_{A,B} = \frac{|m'_A - m'_B|^2}{S'_A + S'_B}$$

$$m'_A = \frac{1}{N_A} \sum_{i=1}^{N_A} x^i$$

$$m'_B = \frac{1}{N_B} \sum_{i=1}^{N_B} x^i$$

$$S'_A = \sum_{i=1}^{N_A} (x^i - m'_A)^2$$

$$S'_B = \sum_{i=1}^{N_B} (x^i - m'_B)^2$$

EPUSP PTC/LEB - S.Furule 22

PTC 2456 – Proc. Sinais Biomédicos
Análise, quantificação, classificação de sinais biol.

$$\sigma_{A,B} = \frac{|m'_A - m'_B|^2}{S'_A + S'_B}$$

$$m'_A = \frac{1}{N_A} \sum_{i=1}^{N_A} x^i$$

$$m'_B = \frac{1}{N_B} \sum_{i=1}^{N_B} x^i$$

$$S'_A = \sum_{i=1}^{N_A} (x^i - m'_A)^2$$

$$S'_B = \sum_{i=1}^{N_B} (x^i - m'_B)^2$$

Seja:

$$x \in \mathbb{R}^d$$

$$m_A = \frac{1}{N_A} \sum_{i=1}^{N_A} x$$

$$m_B = \frac{1}{N_B} \sum_{i=1}^{N_B} x$$

$$x^i = w^T \cdot x = [w_1 \ w_2 \ \dots \ w_d] \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_d \end{bmatrix}$$

$$x^i = \sum_{j=1}^d w_j \cdot x_j$$

$$\sigma_{A,B}(w) = \frac{w^T S_E w}{w^T S_I w}$$

$$\max_w \sigma_{A,B}(w) \Rightarrow S_I w = m_A - m_B$$

$$w = S_I^{-1} (m_A - m_B)$$

S_E : espalhamento entre-classes
 S_I : espalhamento intra-classes

EPUSP PTC/LEB - S.Furule 23

PTC 2456 – Proc. Sinais Biomédicos
Análise, quantificação, classificação de sinais biol.

$$x^i = w^T \cdot x$$

$$m'_A = \frac{1}{N_A} \sum_{i=1}^{N_A} w^T \cdot x = w^T \cdot m_A$$

$$m'_B = w^T \cdot m_B$$

$$S'_A = \sum_{i=1}^{N_A} (w^T \cdot x - w^T \cdot m_A)^2 = \sum_{i=1}^{N_A} (x - m_A) \cdot (x - m_A)^T \cdot w$$

$$= w^T \cdot \left[\sum_{i=1}^{N_A} (x - m_A) \cdot (x - m_A)^T \right] \cdot w$$

$$S'_B = w^T \cdot \left[\sum_{i=1}^{N_B} (x - m_B) \cdot (x - m_B)^T \right] \cdot w$$

$$|m'_A - m'_B|^2 = |w^T (m_A - m_B)|^2 = w^T (m_A - m_B) \cdot (m_A - m_B) \cdot w = w^T \cdot S_E \cdot w$$

$$S'_A + S'_B = w^T \cdot \left[\sum_{i=1}^{N_A} (x - m_A) \cdot (x - m_A)^T + \sum_{i=1}^{N_B} (x - m_B) \cdot (x - m_B)^T \right] \cdot w = w^T \cdot S_I \cdot w$$

$$S_E = (m_A - m_B) \cdot (m_A - m_B)^T$$

singular $\Rightarrow S_E \cdot w = K \cdot (m_A - m_B)$

$$S_I = \left[\sum_{i=1}^{N_A} (x - m_A) \cdot (x - m_A)^T + \sum_{i=1}^{N_B} (x - m_B) \cdot (x - m_B)^T \right]$$

$$\sigma_{A,B}(w) = \frac{w^T S_E w}{w^T S_I w}$$

EPUSP PTC/LEB - S.Furule 24

Classificação NN-Nearest Neighbor

1. Seja $S = \{s_i, i=1, N\}$ as amostras e $c(s)$ a classe do elemento s , $c: S \rightarrow \{1, 2, \dots, C\}$
2. Determinar a distância do vetor x a todos os elementos em S
3. $c(x) \leq c(s_{NN})$

distância euclideana

$$\vec{s}_i = [s_{i1} \quad s_{i2} \quad \dots \quad s_{ip}]^T$$

$$\vec{x} = [x_1 \quad x_2 \quad \dots \quad x_p]^T$$

$$d^2(\vec{s}_i, \vec{x}) = \sum_{k=1, p} (s_{ik} - x_k)^2$$

distância estatística

$$d^2(\vec{s}_i, \vec{x}) = (\vec{s}_i - \vec{x})^T \cdot \text{Cov}^{-1} \cdot (\vec{s}_i - \vec{x})$$

Cov: matriz de covariância de interesse

Classificação k-NN

1. Definido: k
2. Obter os k vizinhos mais próximos
3. Classe de x será aquela com maioria entre os k vizinhos

Exemplo de aplicação: arritmias

- Tendo-se amostras dos diversos tipos de arritmias:
 - Supra-ventricular
 - Bigeminada
 - ...
- Obter conjunto de atributos para caracterizar os clusters de cada classe
- Aplicar classificador, p. ex. k-NN p/ decisão

Exemplo: arritmias em tempo real

- Usar trecho inicial, sob supervisão humana, para “aprender” padrão de QRS normal ($k=1$)
- Para cada novo QRS, comparar (correlação) com o normal e as demais classes existentes.
 - Caso difira substancialmente de todas, criar nova classe (arritmia tipo $k+1$)
 - Se pertencer a uma certa classe, atualizar o padrão de forma adaptativa.

Avaliação

- Poucos casos classificados
 - Leave-one-out
 - Treinamento com $N-1$ casos
 - Avaliação com o elemento restante
- Amostras em grande número
 - Conjunto para treinamento
 - Conjunto independente para avaliação

Bibliografia

- Biomedical Signal Analysis. R.M. Rangayyan. Wiley Interscience, 2002