



THE DEVELOPER'S CONFERENCE





Trilha Machine Learning - Mineração de dados em séries temporais para sistemas biométricos

Henrique Passos



THE
DEVELOPER'S
CONFERENCE



-  Mestre em Sistemas de Informação (EACH-USP)
-  Lead consult (Systems Architect, Machine Learning Engineer)
-  Automobilismo (Kart, Virtual)
-  <https://www.linkedin.com/in/henrique-s-passos/>



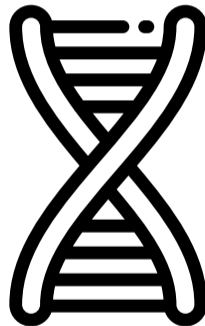
THE
DEVELOPER'S
CONFERENCE

Biometria

Biometria



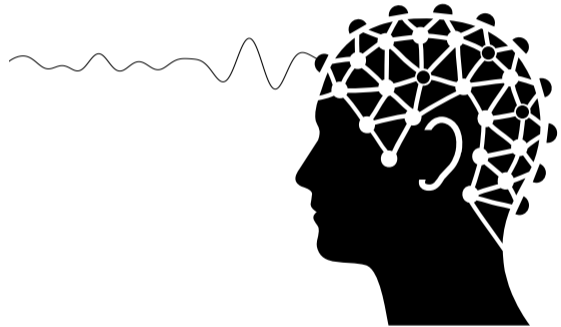
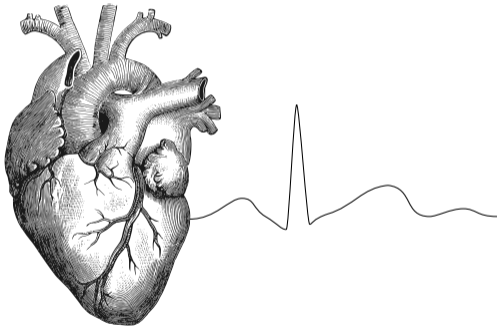
THE
DEVELOPER'S
CONFERENCE

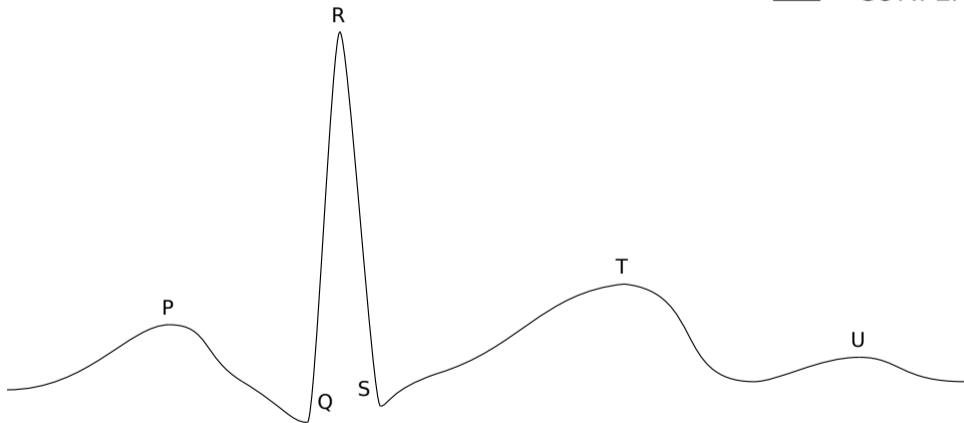


Biometria



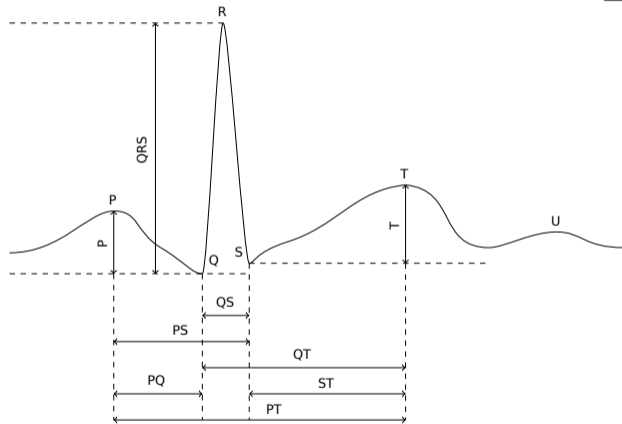
THE
DEVELOPER'S
CONFERENCE





An ecg classifier designed using modified decision based neural networks ¹

¹[Simon and Eswaran, 1997]










Ecg analysis: a new approach in human identification ²

²[Biel et al., 2001]










Sistema biométrico baseado em sinais de ECG:

-  Pré-processamento de dados brutos
-  Extração de características
 -  Características fiduciais
 -  Características não fiduciais
 -  Características híbridas
-  Classificação
-  Identificação










Sistema biométrico baseado em sinais de ECG:

-  Pré-processamento de dados brutos (Pan Tompkins)
-  Extração de características
 -  Características fiduciais
 -  Características não fiduciais
 -  Características híbridas
-  Classificação
-  Identificação



Sistema biométrico baseado em sinais de ECG:

-  Pré-processamento de dados brutos (Pan Tompkins)
-  Extração de características
 -  Características fiduciais
 -  Características não fiduciais
 -  Características híbridas
-  Classificação (KNN - distância euclidiana)
-  Identificação



Sistema biométrico baseado em sinais de ECG:

- 🎨 Pré-processamento de dados brutos (Pan Tompkins)
- 🎨 Extração de características
 - 🎨 Características fiduciais
 - 🎨 Características não fiduciais
 - 🎨 Características híbridas
- 🎨 Classificação (KNN - distância euclidiana)
- 🎨 Identificação (FAR, FRR)



Sistema biométrico baseado em sinais de ECG:

- 🎨 Pré-processamento de dados brutos (Pan Tompkins)
- 🎨 Extração de características ←
 - 🎨 Características fiduciais
 - 🎨 Características não fiduciais
 - 🎨 Características híbridas
- 🎨 Classificação (KNN - distância euclidiana)
- 🎨 Identificação (FAR, FRR)



Sistema biométrico baseado em sinais de ECG:

- 🎨 Pré-processamento de dados brutos (Pan Tompkins)
- 🎨 Extração de características ←
 - 🎨 Características fiduciais
 - 🎨 Características não fiduciais ←
 - 🎨 Características híbridas
- 🎨 Classificação (KNN - distância euclidiana)
- 🎨 Identificação (FAR, FRR)

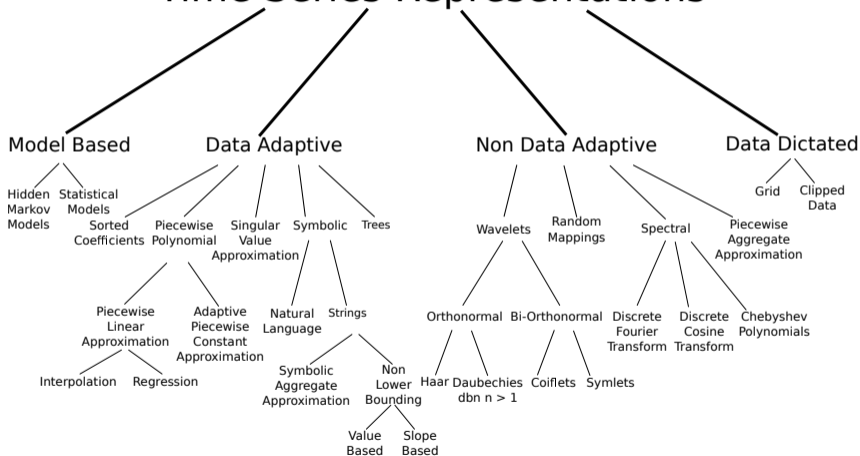


THE
DEVELOPER'S
CONFERENCE

Representações de séries temporais



Time Series Representations



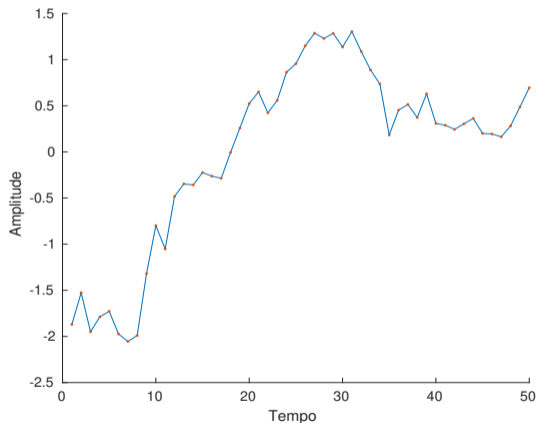
Fonte: [Lin et al., 2007]

Representações de séries temporais

A review on time series data mining³



THE
DEVELOPER'S
CONFERENCE



Mineração de dados em séries temporais

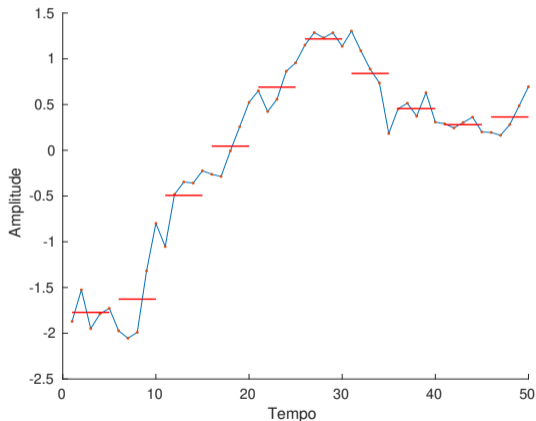
³[Fu, 2011]

Representações de séries temporais

Aproximação Agregada por Parte (PAA) ⁴



THE
DEVELOPER'S
CONFERENCE



Parâmetros: tamanho do segmento = 5

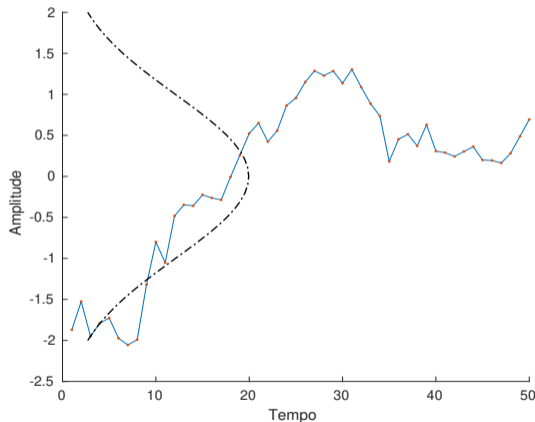
⁴[Keogh et al., 2001]

Representações de séries temporais

Aproximação Agregada Simbólica (SAX) ⁵



THE
DEVELOPER'S
CONFERENCE



Parâmetros: tamanho do segmento = 5, alfabeto = 4

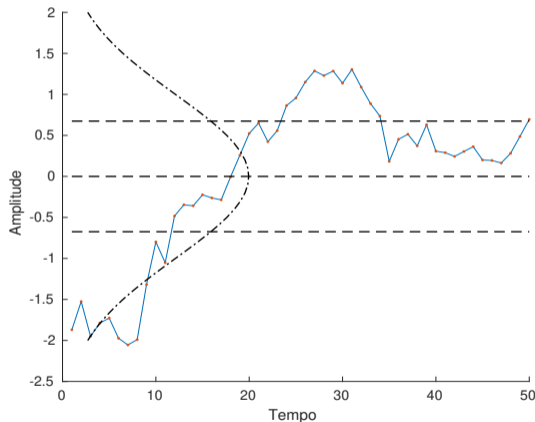
⁵[Lin et al., 2003]

Representações de séries temporais

Aproximação Agregada Simbólica (SAX) ⁵



THE
DEVELOPER'S
CONFERENCE



Parâmetros: tamanho do segmento = 5, alfabeto = 4

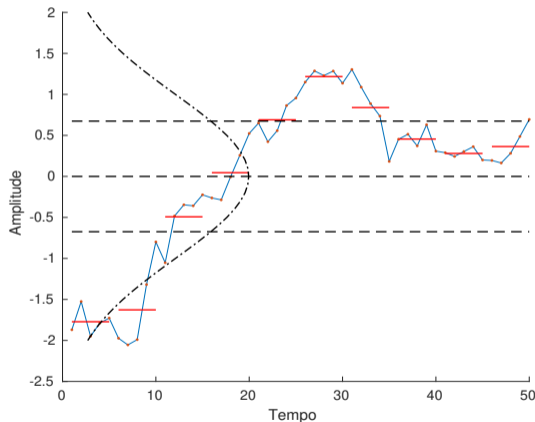
⁵[Lin et al., 2003]

Representações de séries temporais

Aproximação Agregada Simbólica (SAX) ⁵



THE
DEVELOPER'S
CONFERENCE



Parâmetros: tamanho do segmento = 5, alfabeto = 4

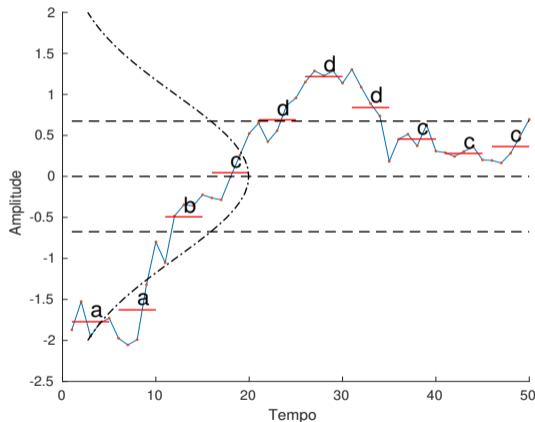
⁵[Lin et al., 2003]

Representações de séries temporais

Aproximação Agregada Simbólica (SAX) ⁵



THE
DEVELOPER'S
CONFERENCE



Parâmetros: tamanho do segmento = 5, alfabeto = 4

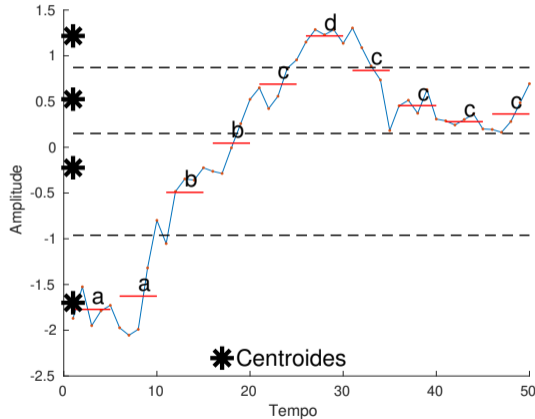
⁵[Lin et al., 2003]

Representações de séries temporais



THE
DEVELOPER'S
CONFERENCE

Aproximação Agregada Simbólica Adaptativa (ASAX) ⁶



Parâmetros: tamanho do segmento = 5, alfabeto = 4

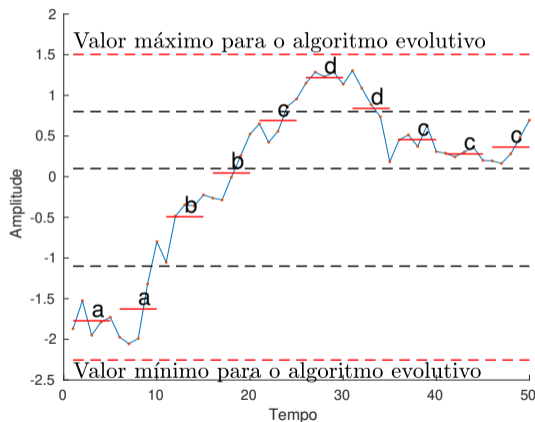
⁶[Pham et al., 2010]

Representações de séries temporais

Aproximação Agregada Simbólica baseado em GA (GASAX)⁷



THE
DEVELOPER'S
CONFERENCE



Parâmetros: tamanho do segmento = 5, alfabeto = 4

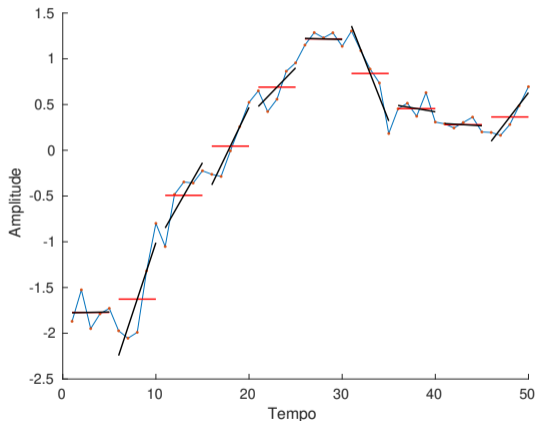
⁷[Fuad, 2012]

Representações de séries temporais



THE
DEVELOPER'S
CONFERENCE

1D-Aproximação Agregada Simbólica (1DSAX) ⁸



Parâmetros: tamanho do segmento = 5, alfabeto = 4

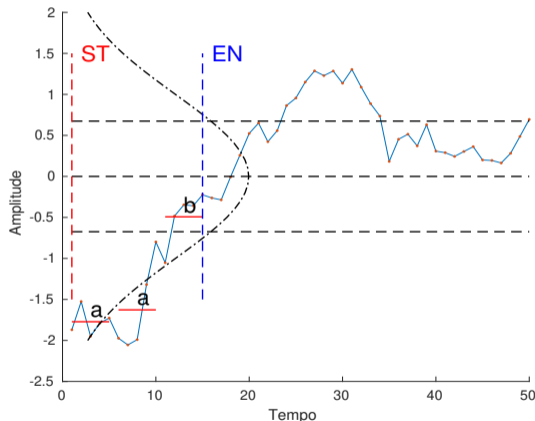
⁸[Simon et al., 2013]

Representações de séries temporais

Aproximação Agregada Simbólica baseado em Variância (VWSAX) ⁹



THE
DEVELOPER'S
CONFERENCE



Parâmetros: tamanho do segmento = 5, alfabeto = 4, limiar = 1.2

⁹[Sun et al., 2012]



THE
DEVELOPER'S
CONFERENCE

Sistemas Biométricos



Symbolic representations of time series applied to biometric recognition based on ECG signals

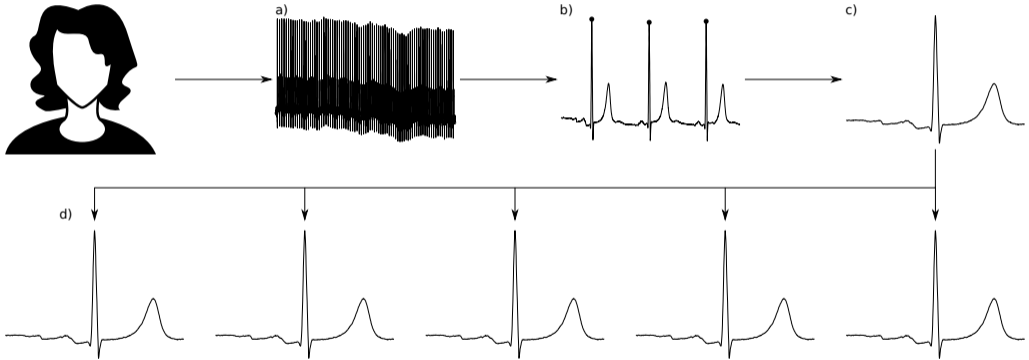
Henrique dos Santos Passos^{*}, Felipe Gustavo Silva Teodoro[†], Bruno Matarazzo Duru[‡],
Edenilton Lima de Oliveira[§], Sarajane M. Peres[¶] and Clodoaldo A. M. Lima^κ

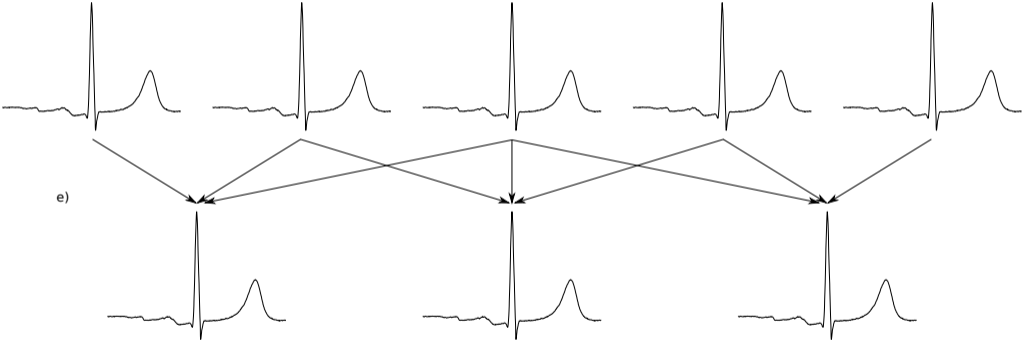
Information Systems Graduate Program
School of Arts, Science and Humanities
University of São Paulo
São Paulo, SP, Brazil

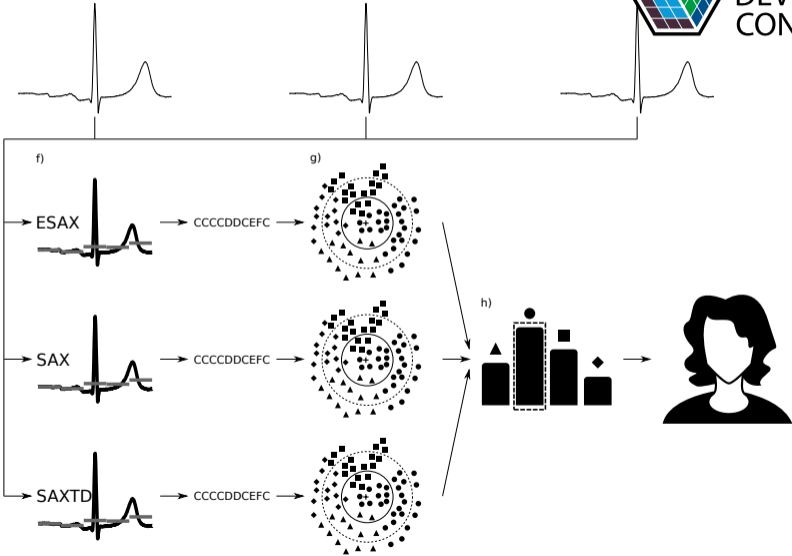
Email: ^{*}henrique.passos@usp.br, [†]fteodoro@usp.br, [‡]bruno.duru@usp.br,
[§]edenilton@usp.br, [¶]sarajane@usp.br, ^κc.lima@usp.br

Abstract—One reason for researching new biometric modalities is to improve the capabilities of security systems against threats. Biometric modalities based on biomedical signals, in particular the electrocardiogram signal (ECG), have been widely adopted. These can be represented by time series. However, in this

is intentionally removed to prevent a person from being identified. With the increased use of Biometric Systems, these kinds of attacks are becoming more frequent and some serious questions are beginning to be raised about this technology.









Resultado obtido pelo Ensemble com BCPSO usando a base de dados PTB

Resultados do Ensemble por otimização global

Repr.	Tx. treino	Tx. teste	EER	Wilcoxon	Parâmetro
EFD	99.1687	99.0839	69.1219	V - 0.0002	$\lambda = 110; \mu = 4; \alpha = 29; \nu = 4; \kappa = 2$
EFVD	98.0843	97.6312	77.9495	V - 0.0002	$\lambda = 103; \mu = 1; \alpha = 62; \nu = 8; \kappa = 2$
ESAX	96.3916	92.8657	87.3971	V - 0.0002	$\lambda = 50; \mu = 4; \alpha = 46; \nu = 9; \kappa = 21$
EWD	99.8753	99.7502	93.8373	F - 0.2247	$\lambda = 114; \mu = 4; \alpha = 36; \nu = 1; \kappa = 1$
SAXTD	99.0176	99.3060	88.7203	V - 0.0003	$\lambda = 125; \mu = 4; \alpha = 47; \nu = 2; \kappa = 4$
Ensemble	99.8866	99.8334	96.1969	-	Representações acima





Resultados do Ensemble por otimização individual

Repr.	Tx. treino	Tx. teste	EER	Wilcoxon	Parâmetro
EFD	99.7355	99.5558	44.0178	V - 0.0004	$\lambda = 104; \mu = 4; \alpha = 64; \nu = 5; \kappa = 2$
EFVD	99.7280	99.7779	79.0321	V - 0.0237	$\lambda = 122; \mu = 4; \alpha = 54; \nu = 5; \kappa = 2$
ESAX	99.5126	99.6761	82.4558	V - 0.0006	$\lambda = 123; \mu = 4; \alpha = 62; \nu = 5; \kappa = 1$
EWD	99.9622	99.9260	95.4659	F - 1.0000	$\lambda = 125; \mu = 4; \alpha = 4; \nu = 1; \kappa = 1$
SAXTD	99.5126	99.6854	85.0097	V - 0.0020	$\lambda = 116; \mu = 4; \alpha = 57; \nu = 1; \kappa = 1$
Ensemble	99.9358	99.9260	95.7065	-	Representações acima

Conclusões



THE
DEVELOPER'S
CONFERENCE

-  Identificação (classificação)
-  FAR/FRR (biométrica)
-  Ensemble x Representações
-  Aplicações







THE DEVELOPER'S CONFERENCE



Referências







THE
DEVELOPER'S
CONFERENCE


-  Biel, L., Pettersson, O., Philipson, L., and Wide, P. (2001).
Ecg analysis: a new approach in human identification.
Instrumentation and Measurement, IEEE Transactions on, 50(3):808–812.
-  Fu, T.-c. (2011).
A review on time series data mining.
Engineering Applications of Artificial Intelligence, 24(1):164–181.
-  Fuad, M. M. M. (2012).
Genetic algorithms-based symbolic aggregate approximation.
Springer.
-  Keogh, E., Chakrabarti, K., Pazzani, M., and Mehrotra, S. (2001).
Dimensionality reduction for fast similarity search in large time series databases.
Knowledge and information Systems, 3(3):263–286.

Referências



-  Lin, J., Keogh, E., Lonardi, S., and Chiu, B. (2003).
A symbolic representation of time series, with implications for streaming algorithms.
In Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery, pages 2–11. ACM.
-  Lin, J., Keogh, E., Wei, L., and Lonardi, S. (2007).
Experiencing sax: a novel symbolic representation of time series.
Data Mining and knowledge discovery, 15(2):107–144.
-  Pham, N. D., Le, Q. L., and Dang, T. K. (2010).
Two novel adaptive symbolic representations for similarity search in time series databases.
In Web Conference (APWEB), 2010 12th International Asia-Pacific, pages 181–187. IEEE.
-  Simon, B. P. and Eswaran, C. (1997).
An ecg classifier designed using modified decision based neural networks.
Computers and Biomedical Research, 30(4):257–272.



-  Simon, M. et al. (2013).
1d-sax; a novel symbolic representation for time series.
IDA'13.
-  Sun, C., Stirling, D., Ritz, C., and Sammut, C. (2012).
Variance-wise segmentation for a temporal-adaptive sax.
In Proceedings of the Tenth Australasian Data Mining Conference-Volume 134, pages
71–77. Australian Computer Society, Inc.