

Extração de Atributos para Classificação de Morfologias em Sinais ECG

Wesley Lioba Caldas*, João Paulo do Vale Madeiro**, João Paulo Pordeus Gomes*

*Universidade Federal do Ceará, Fortaleza, Brasil

**Universidade da Integração Internacional da Lusofonia Afro-Brasileira, Redenção, Brasil

e-mail: wesleylec@lia.ufc.br

Abstract - *Computer diagnostic using electrocardiograms (ECGs) is widely used to aid in the identification of various heart diseases. By the way, the classification of ECG signal morphologies can be an important step in the detection of various cardiac diseases, such as malignant arrhythmias, which can lead to death if there is no adequate treatment. In this work, it will be presented the extraction of 11 features based on mathematical functions (Rayleigh, Gaussian and Mexican Hat), for the classification of 20 morphologies of synthetic ECG signals. Under this methodology, was obtained an accuracy rate of 99% for the classification of artificial ECG signals, using an algorithm based on linear regression.*

Keywords: *Classification, ECG, Features.*

Introdução

O diagnóstico por computador tem se popularizado bastante como ferramenta de auxílio para detecção de doenças de origem cardíaca, relacionadas a distúrbios de condução e arritmias. Dentre os diversos campos aplicados, o estudo do eletrocardiograma (ECG) merece destaque por ter apresentado resultados promissores em auxiliar no diagnóstico de doenças, como arritmias cardíacas, morte súbita, acidente vascular cerebral isquêmico e hemorrágico [1].

O ciclo cardíaco padrão presente no sinal ECG é composto por ondas características que ocorrem devidas as alterações elétricas provocadas pela despolarização e repolarização que ocorrem no coração. As despolarizações resultam nas contrações dos átrios, gerando as ondas a onda P, e nas contrações dos ventrículos, gerando o complexo QRS. Já a polarização resulta no retorno da massa ventricular ao estado de relaxamento, gerando a onda T. A junção de todas essas ondas formam o ciclo PQRST.

O laudo descritivo de um exame ECG deve apresentar detalhes sobre a duração, o eixo elétrico de P, QRS e T, bem como a morfologia do complexo QRS, dentre outras características [2]. Em especial a morfologia do complexo QRS tem atraído

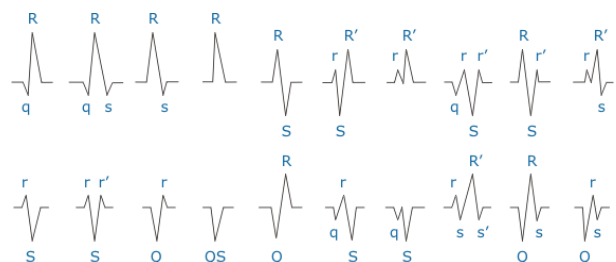
a atenção da comunidade acadêmica por indicar diversos tipos de doenças como a taquicardia ventricular, que em seu estado mais grave pode causar morte súbita [3]. Uma vez que a análise visual do ECG continua como uma das principais metodologias dos cardiologistas para fins de diagnóstico, formas de automatizar este procedimento são de grande interesse da comunidade acadêmica.

Vale ressaltar que, devido a grande variabilidade dos batimentos cardíacos, encontrar atributos que ajudem a identificar todas as possíveis morfologias do complexo QRS é uma tarefa árdua. Neste trabalho propomos usar 5 modelos matemáticos baseados em 3 funções bem conhecidas (Rayleigh, Gaussiana e Chapéu Mexicano) para construir atributos a fim de identificar as morfologias presentes nos batimentos cardíacos de um indivíduo qualquer, em que estes modelos matemáticos estão relacionados somente com a forma e proporção de cada batimento e não com a amplitude e duração de suas ondas.

Infelizmente, para que seja possível realizar uma tarefa como a classificação supervisionada, se faz necessário uma base inicial de dados. Devido a grande dificuldade em se conseguir sinais rotulados (i.e em que cada batimento tem sua morfologia anotada), foi adotado o uso de um simulador, capaz de gerar sinais sintéticos ECG realísticos.

Materiais e métodos

Figura 1 – Representação das possíveis morfologias do complexo QRS, conforme [4] e [5].



O complexo QRS apresenta uma grande variabilidade uma vez que tanto a forma quanto a amplitude das ondas são regidas por múltiplos fatores

individuais, em particular pela forma e posição do coração, e a presença e natureza de patologias, entre outras fatores [4] [5]. Como resultado, o complexo QRS pode ser categorizado dependendo da forma de suas ondas, neste estudo utilizamos as 20 variações mais conhecidas presentes na Figura 1.

A duração e amplitudes normais das ondas podem ser vistas na Tabela 1.

Tabela 1 – Amplitude e duração das ondas conforme [4] e [5]

Onda	Amplitude(mV)	Duração(Segundos)
P	0,00-0,25	0,08 - 0,10
Q	0,00-0,30	0,06 - 0,10
R	0,60-2,10	0,06 - 0,10
S	0,00-0,60	0,06 - 0,10
T	0,00-0,50	0,10 - 0,25

0.1 Simulador

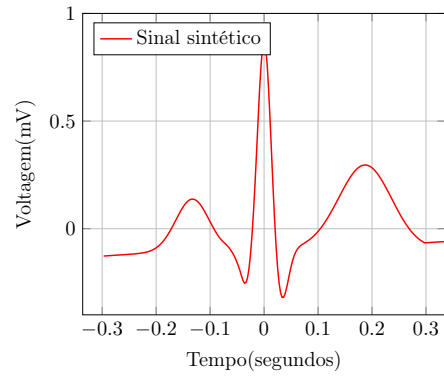
Construir um modelo de reconhecimento de padrões baseado nas diferentes morfologias exige sinais ECG específicos de cada morfologia a fim de compôr uma base de conhecimento inicial. Infelizmente, rotular os sinais ECG exige tempo e o uso de um especialista, além da permissão do paciente. Como forma de mitigar esse problema, foi desenvolvido um simulador de batimentos cardíacos capaz de aproximar com grande exatidão um batimento real [6].

O simulador é composto por um sistema dinâmico baseado em três equações diferenciais capazes de gerar sinais sintéticos ECG realísticos. É possível especificar a frequência e variação dos batimentos cardíacos, o espectro de potência do tacograma RR, a morfologia do complexo QRS, ondas de banda de Baixa Frequência (BF) e Alta Frequência (AF) juntamente com a proporção BF/AF. Por fim, as amplitudes e modulações do ciclo PQRST também podem ser especificadas permitindo uma grande variabilidade de sinais ECG.

0.2 Extração de atributos para classificação

A qualidade da classificação de batimentos cardíacos depende bastante da etapa de pré-processamento dos dados. Em especial, a detecção automática do complexo QRS é de fundamental importância, tendo diversos algoritmos implementados baseados em derivadas de sinais, filtros digitais, redes neurais dentre outros [7]. Uma vez que

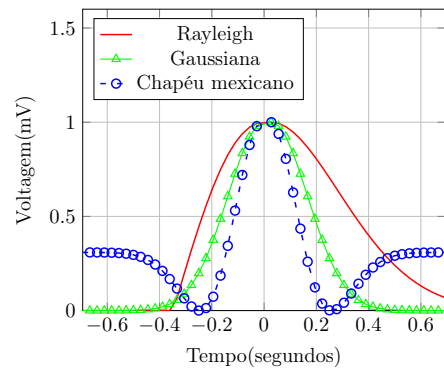
Figura 2 – Sinal do ciclo PQRST sintético produzido pelo simulador.



a segmentação não é objeto de estudo desse artigo, utilizaremos uma técnica simples de passagem-baixa e passagem-alta, em conjunto com limiares para a detecção do complexo QRS [8].

Devida a grande variabilidade dos batimentos presentes no sinal ECG, tornam-se necessárias ferramentas que identifiquem a forma de cada batimento independente da amplitude e duração dos sinais. Nesse trabalho, foram adotadas 3 funções capazes de identificar a morfologia de cada batimento, levando em conta somente a estrutura do batimento, sendo estas as funções Chapéu Mexicano, Gaussiana e Rayleigh, cuja as formas podem ser vistas na Figura 3.

Figura 3 – Funções Rayleigh, Gaussiana e Chapéu Mexicano.



A composição de cada modelo foi feita misturando duas, podendo haver repetidas, das três equações abaixo:

$$\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} \quad (1)$$

$$R(x|\lambda) = \frac{x}{\lambda^2} e^{-x^2/(2\lambda^2)} \quad (2)$$

$$C(x|\lambda) = \frac{2}{\sqrt{3\lambda\pi^{1/4}}} \left(1 - \left(\frac{x}{\lambda}\right)^2\right) e^{-\frac{x^2}{2\lambda^2}} \quad (3)$$

Sendo as Equações, 1, 2 e 3, respectivamente as funções Gaussiana, Rayleigh e Chapéu Mexicano, em que μ é a média, σ^2 a variância, e λ a escala das funções. A frequência de amostragem dos modelos deve ser condizente com a do sinal analisado, sendo por padrão é adotado o número de 360 Hz. Ao todo, foram compostos 5 modelos matemáticos descritos pela Tabela 2.

Tabela 2 – Composição de cada modelo

Modelo	Função A.	Função B.
1	Gaussiana	Gaussiana
2	Ch. mexicano	Ch. mexicano
3	Rayleigh	-Rayleigh
4	-Rayleigh	Rayleigh
5	Rayleigh	Rayleigh

Nas Figuras 4, 5, 6 e 7, podem ser vistas as formas de cada composição.

Figura 4 – Modelo matemático 1: composição entre duas Gaussianas.

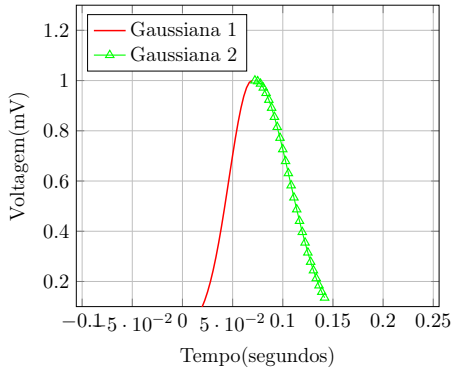
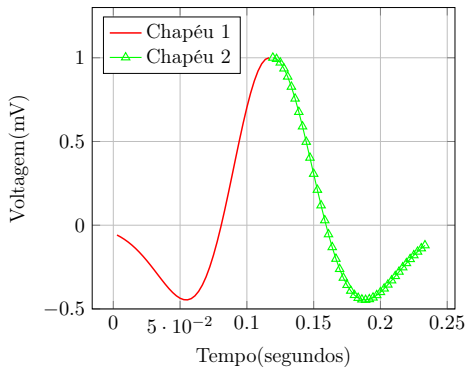


Figura 5 – Modelo matemático 2: composição entre dois Chapéus mexicanos.



Os Modelos 1 e 2, são delimitadas a esquerda e a direita do pico de interesse (pico com maior valor absoluto do sinal), existindo uma sobreposição dos picos dos sinais que o compõem. Já os Modelos 3, 4 e 5 permanecem com os picos de ambas as

Figura 6 – Modelo matemático 3: composição entre duas funções Rayleigh. O modelo matemático 4 é idêntico ao modelo 3, porém seus picos estão invertidos.

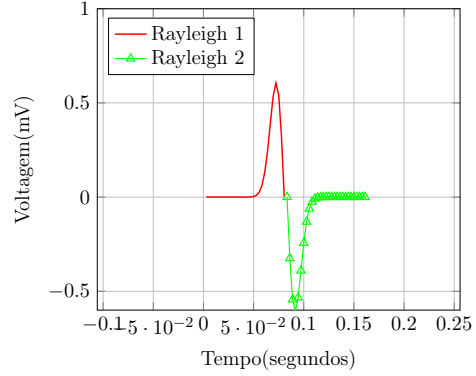
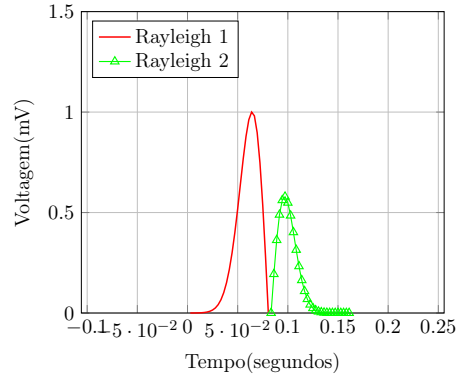


Figura 7 – Modelo matemático 5: composição entre duas funções Rayleigh. Note que o modelo 5 difere do 3 pelo sinal das funções Rayleigh.



funções de composição, sendo o ponto de interseção o primeiro ponto com valor zero entre as duas funções. As variáveis μ , σ e λ são otimizadas por força bruta e a amplitude do sinal final é normalizada de forma que a menor e a maior amplitude de cada modelo se equiparem a menor e maior amplitude do batimento analisado.

Para a seleção das variáveis de cada modelo, adotou-se a escolha do menor Erro Quadrático Médio (EQM). Para tal, ajusta-se o pico de interesse do sinal final de cada modelo com o do batimento analisado, e então, calcula-se o EQM a esquerda e direita deste pico. O cálculo EQM pode ser visto nas equações abaixo:

$$EQM_i - Esquerda = \sqrt{\frac{1}{n-j+1} \sum_{t=j}^n e_{it}^2} \quad (4)$$

Em que e_{it} representa a diferença elétrica entre o modelo matemático i (variando de 1 a 5) no tempo t somente para a parte esquerda ao pico de interesse, iniciando no instante de tempo j e finalizando

no instante de tempo n .

$$EQM_i - Direita = \sqrt{\frac{1}{k-n} \sum_{t=n+1}^k e_{it}^2} \quad (5)$$

A Equação 5 é análoga à Equação 4, porém iniciando no instante de tempo $n+1$ até o final do batimento cardíaco, localizado no instante de tempo k . Os índices j , n e k delimitam um batimento qualquer em questão, não importando a unidade de tempo utilizada.

São totalizados então, 10 atributos referentes aos erros à esquerda e à direita de cada modelo matemático. Por fim, um último atributo representando o sinal do pico de interesse é adicionado, em que o valor 1 indica um pico positivo e 0 caso contrário.

Resultados

Utilizando o simulador, foram criados 10 sinais sintéticos para cada uma das 20 morfologias, totalizando 200 sinais, com variação nas amplitudes e durações dos sinais, conforme a Tabela 1. A frequência dos sinais foi estabelecida em 360 Hz, com a taxa de batimentos cardíacos variando entre 60 e 100 batimentos por minuto e proporção BF/AF de 0.5 e com aproximadamente 2 horas de duração.

Foi utilizado como método preditivo em nossos experimentos, a versão linear do algoritmo de aprendizagem de máquina, Máquina de Vetores de Suporte [9]. Além disso, um procedimento de validação cruzada usando 70% dos sinais para treino e 30% para teste, repetindo o experimento 10 vezes e obtendo uma taxa de acurácia de $99,00\% \pm 00,12$ em relação as 20 classes de um batimento cardíaco.

Discussão e Conclusões

A classificação utilizando os 11 atributos propostos para a identificação da morfologia em sinais ECG, obteve uma excelente taxa de acurácia, demonstrando a eficácia dos atributos em identificar as morfologias apresentadas. Além disso, a morfologia de um batimento cardíaco pode ajudar no diagnóstico computadorizado de diversos tipos de doenças cardiovasculares, em especial a taquicardia ventricular que pode em casos mais graves ocasionar a morte súbita. Em outras palavras, a classificação das morfologias pode ajudar em detectar tais doenças.

Em trabalhos futuros, o mesmo procedimento será realizado utilizando sinais ECC reais. Em

adicional, novos modelos matemáticos deverão ser propostos utilizando outras funções, além de testes para identificação de arritmias e outras doenças cardiovasculares.

Referências

- [1] R. J. Martis, U. R. Acharya, and H. Adeli, "Current methods in electrocardiogram characterization," *Computers in Biology and Medicine*, vol. 48, no. Supplement C, pp. 133–149, 2014.
- [2] C. Pastore, J. Pinho, C. Pinho, N. Samesima, H. Pereira Filho, J. Kruse, A. Paixão, A. Pérez-Riera, A. Ribeiro, C. Oliveira, *et al.*, "Tii diretrizes da sociedade brasileira de cardiologia sobre análise e emissão de laudos eletrocardiográficos," *Arquivos Brasileiros de Cardiologia*, vol. 106, no. 4, pp. 1–23, 2016.
- [3] S. Chatterjee and N. Changawala, "Framed qrs complex: a novel marker of cardiovascular disease," *Clinical cardiology*, vol. 33, no. 2, pp. 68–71, 2010.
- [4] T. S. Lugovaya, "Biometric human identification based on ecg," 2005.
- [5] A. Nemirko and T. Lugovaya, "Biometric human identification based on electrocardiogram," in *Proceedings of XII-th Russian Conference on Mathematical Methods of Pattern Recognition*. MAKS Press, Moscow, pp. 387–390, 2005.
- [6] P. E. McSharry, G. D. Clifford, L. Tarassenko, and L. A. Smith, "A dynamical model for generating synthetic electrocardiogram signals," *IEEE transactions on biomedical engineering*, vol. 50, no. 3, pp. 289–294, 2003.
- [7] R. J. Martis, U. R. Acharya, and H. Adeli, "Current methods in electrocardiogram characterization," *Computers in biology and medicine*, vol. 48, pp. 133–149, 2014.
- [8] J. Pan and W. J. Tompkins, "A real-time qrs detection algorithm," *IEEE transactions on biomedical engineering*, no. 3, pp. 230–236, 1985.
- [9] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and their applications*, vol. 13, no. 4, pp. 18–28, 1998.