

# Análise da Complexidade de Tempo e Espaço em Algoritmos para Classificação de Padrões

Eduardo Luz, David Menotti  
PPGCC - Programa de Pós-Graduação em Ciência da Computação  
UFOP - Universidade Federal de Ouro Preto  
Ouro Preto, Minas Gerais, Brasil  
email: eduluz@gmail.com, menottid@gmail.com

**Resumo**—O reconhecimento de padrões pode ser automatizado com o auxílio de algoritmos computacionais, e assim, ações podem ser tomadas, como a classificação dos dados em diversas categorias. Neste trabalho, analisaremos alguns algoritmos clássicos destinados ao reconhecimento e classificação de padrões de arritmias em sinais de eletrocardiograma. Propomos fazer uma análise da complexidade de tempo e espaço com objetivo de analisar a viabilidade de tais algoritmos frente à quantidade de dados disponíveis no banco de dados de arritmia do MIT-BIH.

**Keywords**-Arritmia, ECG, Complexidade, Classificadores, Reconhecimento de Padrões.

## I. INTRODUÇÃO

O problema de reconhecer padrões em grandes quantidades de dados é fundamental [1]. O reconhecimento de padrões pode ser automatizado com o auxílio de algoritmos computacionais, e assim, ações podem ser tomadas, como a classificação dos dados em diversas categorias.

Neste trabalho, analisaremos alguns algoritmos clássicos destinados ao reconhecimento e classificação de padrões, com foco em abordagens que utilizam aprendizado supervisionado paramétrico, *i.e.*, métodos que utilizam um conjunto de dados pré-rotulados, chamados de conjunto de treino, em que este conjunto de treino é utilizado para ajustar os parâmetros de algum modelo adaptativo. Uma vez que o modelo é treinado, ele pode classificar dados desconhecidos (conjunto de teste).

O trabalho consiste no uso destes modelos objetivando a classificação de arritmias no sinal de Eletrocardiograma (ECG) [2], [3], [4]. Propomos fazer uma análise teórica da complexidade de tempo e espaço dos algoritmos relacionados ao treino dos modelos, ou seja, encontrar a função e ordem de complexidade dos algoritmos e também realizar experimentos, a fim de comparar o tempo gasto e quantidade de memória necessários. Deste modo, analisar a viabilidade dos mesmos.

O artigo está organizado da seguinte forma. A Seção II apresenta as justificativas do estudo proposto. A Seção III descreve os objetivos do trabalho. A Seção IV apresenta a metodologia utilizada na pesquisa e por fim, os resultados esperados na Seção V.

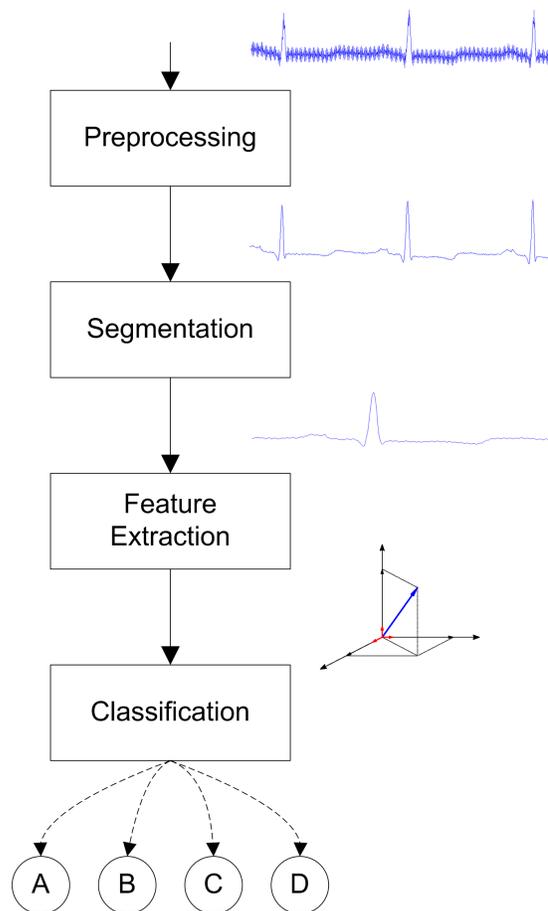


Figura 1. Um diagrama de um sistema de classificação de arritmias

## II. JUSTIFICATIVAS

Alguns algoritmos utilizados para reconhecimento de padrões podem ser altamente impraticáveis, dependendo da quantidade de dados para o treino, dimensionalidade das características e da quantidade de categorias (classes) em que os dados estão divididos [5]. Para grandes conjuntos de dados, o tempo de treino pode ser proibitivo, ou a quantidade de memória requerida ser fisicamente ou economicamente

impossível.

Portanto, é interessante conhecer a complexidade computacional de cada algoritmo a fim de se determinar a viabilidade do seu uso, especialmente para grandes quantidades de dados. Neste trabalho, três métodos muito utilizados na literatura serão analisados : *Multilayer Neural Networks* (MLP), *Probabilistic Neural Networks* (PNN) e *Support Vector Machine* (SVM).

### III. OBJETIVOS

No problema da detecção de arritmias no sinal de ECG, temos 4 etapas fundamentais : Pré-processamento, Segmentação, Extração de características e Classificação (ver Figura 1). Este trabalho manterá o foco apenas na etapa de classificação.

Para se classificar arritmias no sinal de ECG, pode-se utilizar vários modelos computacionais. Muitos modelos podem ser vistos na literatura, mas neste artigo vamos nos ater apenas à três métodos que utilizam modelos com abordagem de aprendizado supervisionado [2], [3], [4].

O objetivo deste trabalho é determinar a função/ordem de complexidade (tempo e espaço) dos algoritmos utilizados para se construir os modelos computacionais para classificação de arritmias dos três métodos selecionados, bem como realizar experimentos. Estes métodos foram escolhidos por utilizarem modelos computacionais amplamente empregados em métodos visando a detecção de arritmias no sinal de ECG.

### IV. METODOLOGIA

Para se determinar a complexidade dos algoritmos utilizados pelos três métodos propostos para estudo, usaremos uma abordagem descrita em [6]. Uma vez que o algoritmo é implementado em alguma linguagem, um custo é associado a cada expressão. Para se determinar a complexidade da função, soma-se o custo de todas as expressões pelo número de vezes que elas ocorrem.

A fim de simplificar a análise proposta por este trabalho, vamos desprezar expressões de baixo custo computacional. Utilizando-se a mesma medida de custo para todos os algoritmos, podemos efetuar comparações entre os mesmos e determinar os mais eficientes. Assim, consideramos apenas dois tipos de expressões para fazer a análise : atribuições e comparações.

Em nossa análise, concentramos em encontrar apenas o pior caso do tempo de execução, *i.e.*, o maior tempo gasto para se executar um algoritmo com entrada de tamanho  $n$ , ou ainda, o limite superior de tempo de execução.

Como o código dos métodos apresentados nos artigos analisados não foram disponibilizados pelos autores, a análise dos métodos que fazem uso de redes neuronais artificiais (MLP e PNN) será feita com base na implementação encontrada em *Neural Network Toolbox* do Matlab e para o SVM, será utilizado o código da biblioteca LIBSVM [7].

Nos três métodos selecionados para análise, utilizou-se o banco de dados de arritmia do MIT-BIH [8]. O banco de dados de arritmia do MIT-BIH contém 48 registros de 30 minutos, amostrados a 360Hz, e dezoito tipos de batimentos pré-classificados e rotulados, ou seja, dezoito categorias (classes). O banco de dados de arritmia do MIT-BIH é público e amplamente utilizado como *benchmark* na literatura.

Para este trabalho, também utilizaremos os banco de dados de arritmia do MIT-BIH e faremos uma implementação dos métodos em Matlab<sup>1</sup> o mais fiel possível.

### V. RESULTADOS ESPERADOS

Espera-se, com esta pesquisa, encontrar teoricamente a ordem de complexidade dos algoritmos de treino dos três métodos estudados assim como o tempo de execução, por meio de experimentos. Com base nos resultados, poderemos determinar a viabilidade dos algoritmos diante dos dados disponíveis no banco de dados de arritmia do MIT-BIH, que é o banco de dados mais utilizado para teste e validação de métodos destinados à detecção de arritmias no sinal de ECG. Também poderá ser feita uma comparação entre os métodos, o que permitirá apontar o mais eficiente dentre os três.

### REFERÊNCIAS

- [1] C. M. Bishop, *Pattern Recognition and Machine Learning*, 1st ed. Springer, 2006.
- [2] S. Yu and Y. Chen, "Electrocardiogram beat classification based on wavelet transformation and probabilistic neural network," *Pattern Recognition Letters*, vol. 28, no. 10, pp. 1142–1150, 2007.
- [3] I. Güler and E. D. Übeyli, "ECG beat classifier designed by combined neural network model," *Pattern Recognition*, vol. 38, no. 2, pp. 199–208, 2005.
- [4] C. Ye, M. T. Coimbra, and B. V. K. V. Kumar, "Arrhythmia detection and classification using morphological and dynamic features of ECG signals," in *IEEE International Conference on Engineering in Medicine and Biology Society (EMBC)*, 2010, pp. 1918–1921.
- [5] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed., 2001.
- [6] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 3rd ed. The MIT Press, 2009.
- [7] C.-C. Chang and C.-J. Lin, *LIBSVM: a library for support vector machines*, 2001, software disponível em <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [8] "MIT-BIH ECG database," disponível em <http://ecg.mit.edu/>.

<sup>1</sup>The source code of implementations are available at <http://code.google.com/p/embc-ecg-paper/>.