

Aplicação do algoritmo de seleção negativa na detecção de arritmias cardíacas

Daniel Henrique Petrikicz, Bruno Rodrigues de Oliveira, Caio Cesar Enside de Abreu

Resumo—Atualmente a maior causa de mortes no mundo está relacionada às doenças cardíacas. O desenvolvimento de algoritmos para detecção automática de arritmias em eletrocardiogramas (ECGs) é de grande importância, pois, podem auxiliar os especialistas na obtenção de diagnósticos mais precisos e em menor tempo. Este artigo investiga a aplicação de um sistema imunológico artificial, o algoritmo de seleção negativa (ASN), para a detecção automática de arritmias. Simulações realizadas utilizando o banco de dados *MIT-BIH arrhythmia database* mostraram resultados satisfatórios: sensibilidade de 81,96% e especificidade de 62,41%, indicando que o ASN é uma ferramenta com potencial a ser explorado pelos pesquisadores da área.

Palavras-Chave—Algoritmo de Seleção Negativa, detecção de arritmia, eletrocardiograma.

Abstract— *Currently the leading cause of death in the world is related to heart disease. Development of algorithms for automatic ECG arrhythmia detection is of high importance to assist experts in faster and more accurate diagnostic tests. This article investigates the use of an artificial immune system, the negative selection algorithm (NSA), for automatic arrhythmia detection. Simulations performed using the MIT-BIH arrhythmia database showed satisfactory results: sensitivity of 81.96% and specificity of 62.41%, indicating that the NSA is a powerful tool to be exploited by researchers.*

Keywords—*Negative Selection Algorithm, arrhythmia detection, electrocardiogram.*

I. Introdução

Atualmente a maioria dos óbitos está relacionada às doenças cardíacas [4], destacando-se entre elas as arritmias. Arritmias são alterações que ocorrem nos batimentos cardíacos, podendo ser consequência da ingestão de substâncias que provocam alterações do ritmo cardíaco ou devida a doenças que afetam o funcionamento do nódulo sinusal, responsável pelo ritmo cardíaco normal [4]. Dentre os diversos métodos para diagnosticar cardiopatias, a análise do ECG ainda é aquele preferencial. O ECG é obtido pelo eletrocardiograma que registra os fenômenos elétricos do coração decorrente das atividades realizadas pelo músculo cardíaco, através de sensores dispostos em pontos específicos do corpo humano. Analisando a morfologia das suas ondas características P, Q, R, S e T, e o intervalo entre batimentos consecutivos, os especialistas conseguem diagnosticar as arritmias cardíacas [3]. O desenvolvimento de algoritmos de identificação automática de anormalidades no ritmo cardíaco é de grande importância, pois podem auxiliar os especialistas na obtenção de diagnósticos mais precisos e em menor tempo [3].

Os sistemas imunológicos artificiais (SIAs) são considerados metodologias inteligentes que se aplicam na resolução de problemas em diversas áreas [1]. Exemplos são aplicações envolvendo segurança computacional, detecção de anomalias e reconhecimento de padrões [5]. A área de pesquisa sobre SIAs é extensa e a escolha da metodologia depende das características do problema que será estudado. O Algoritmo de Seleção Negativa foi proposto por [1], onde o problema de segurança computacional foi abordado. Apesar de ser uma ferramenta com potencial reconhecido pelos pesquisadores da área, o ASN ainda é pouco explorado pela comunidade científica. Além disso, existem poucos trabalhos na literatura que utilizam de alguma forma SIAs para análise de ECGs. Como exemplo, cita-se [5].

O objetivo deste trabalho é investigar as principais características do ASN, que são úteis ao problema da identificação de arritmias em ECGs, e apresentar alguns resultados que reforcem esta premissa.

II. Algoritmo de Seleção Negativa

O ASN foi desenvolvido baseado no processo de seleção negativa das células T que ocorre dentro do timo. No timo existem apenas moléculas do próprio corpo, denominadas de próprio. Dessa forma, o processo de seleção negativa é caracterizado pela eliminação de qualquer célula portadora de anticorpos capazes de reconhecer antígenos próprios. Após este processo, as células T maturadas são liberadas e atuarão na identificação de antígenos agressores, denominados não-próprios [1]. Com base na discriminação das células em próprio ou não-próprio, o ASN é dividido em duas etapas [1]:

1. **Censoriamento:** a) É definido o conjunto de cadeias próprias **P** que se deseja proteger; b) São geradas cadeias aleatoriamente e a afinidade (*Match*) entre cada uma delas e as cadeias próprias é avaliada. Caso a afinidade seja superior a um limiar estipulado, a cadeia é rejeitada. Caso contrário, armazena-se em um conjunto de detectores **R**.

2. **Monitoramento:** a) Dado o conjunto de cadeias que se deseja proteger (cadeias protegidas), avalia-se a afinidade entre cada uma delas e o conjunto de detectores. Se a afinidade for superior a um limiar preestabelecido, um elemento não-próprio é identificado.

A etapa de censoriamento é realizada em modo *off-line* e visa apenas a construção de um conjunto de detectores de falhas ou anomalias. Já a etapa de monitoramento é executada em tempo real, de maneira que a cadeia identificada por um detector seja classificada como uma anormalidade. Por analogia aos conceitos biológicos, os detectores são

Daniel Henrique Petrikicz¹, Bruno Rodrigues de Oliveira² e Caio Cesar Enside de Abreu³, Departamento de Computação^{1,3}, Universidade do Estado de Mato Grosso-UNEMAT, Alto Araguaia - MT, Brasil, Distribuição/Contadoria², Tribunal de Justiça de Mato Grosso do Sul, Chapadão do Sul, Brasil, E-mails: dpetrikicz@gmail.com¹, bruno.rodrigues@tjms.jus.br², caioenside@unemat.br³

comparados às células T maturadas que podem reconhecer agentes patogênicos [1].

III. SIMULAÇÕES

A base de dados utilizada nesse estudo é denominada MIT-BIH *arrhythmia database* [2]. Ela foi obtida pelo laboratório de arritmia do *Boston's Israel Beth Hospital* em parceria com o MIT (*Massachusetts Institute of Technology*) e está disponível para investigações da comunidade científica possibilitando comparações entre diferentes algoritmos [2]. O banco de dados contém 48 registros de ECG com duração de 30 minutos cada. Todos os 47 pacientes observados possuem características físicas distintas um dos outros [2].

Para a detecção das arritmias foram utilizados os tacogramas, conforme [3]. O tacograma consiste de uma série temporal contendo as distâncias entre cada pico R de um ECG. Para avaliar o método, metade dos registros foi utilizada na etapa de sensoriamento do ASN (treinamento) e o restante na etapa de monitoramento (teste).

A análise foi realizada em três fases. Na primeira fase foi realizado um janelamento retangular dos sinais de ECG (representados por seus tacogramas), segmentando-os em amostras contendo 32 intervalos RR e as seguintes características foram extraídas: Curtose; Expoente de Hurst (estimador para a auto-similaridade de uma série temporal); Média e Variância. Essas medidas foram selecionadas por apresentarem os melhores resultados nos testes preliminares. Cada segmento do tacograma foi caracterizado como normal ou arritmico utilizando as anotações de cada batimento fornecidas pelo MIT-BIH *arrhythmia database*. Assim como em [3], um segmento contendo 32 intervalos RR foi considerado normal se ele contivesse mais do que 30 batidas normais, caso contrário ele foi considerado arritmico.

Na segunda fase um conjunto de cadeias próprias **P** foi construído contendo 1082 vetores de características extraídos de segmentos normais dos tacogramas. O conjunto **P** foi construído de forma aleatória, considerando todos os registros do conjunto de treinamento. Definido o conjunto **P**, a etapa de sensoriamento do ASN foi executada e um conjunto de detectores de arritmias **R** foi construído.

Na terceira fase a etapa de monitoramento do ASN foi executada sobre o conjunto de teste. O objetivo foi classificar cada segmento dos tacogramas em normal ou arritmico, tendo como base o conjunto de detectores de arritmia **R** e uma medida de afinidade. Quando a afinidade entre um segmento RR do ECG em análise e algum detector $r_i \in \mathbf{R}$ foi superior a um limiar λ , então o segmento foi classificado como arritmico.

A medida de afinidade utilizada para ambas as etapas do ASN foi a distância de Canberra:

$$d(p, q) = \sum_{i=1}^n \frac{|p_i - q_i|}{|p_i| + |q_i|}. \quad (1)$$

Tanto a medida de distância quanto o limiar foram definidos empiricamente, de acordo com o desempenho do sistema. Dessa forma, definiram-se dois limiares: $\lambda_1=2,6$ para a etapa de sensoriamento e $\lambda_2=1$ para a etapa de monitoramento. Vale ressaltar que quanto menor a distância entre dois padrões, maior a afinidade entre eles.

A Tabela I apresenta os resultados obtidos pelo ASN para 1498 segmentos RR classificados. Foram consideradas 13 tipos

de arritmias conforme [3]. Para efeito de comparação, apresentam-se também os resultados obtidos em [3], para características extraídas no domínio do tempo, onde os autores utilizaram o mesmo banco de dados e a mesma segmentação dos tacogramas utilizados no presente estudo. Além disso, em [3] um conjunto bem maior de características e várias redes neurais foram utilizadas para a classificação dos segmentos RR.

TABELA I. RESULTADOS OBTIDOS

Método	Sensibilidade (%)	Especificidade (%)
Proposto em [3]	87,50	89,50
ASN	81,96	62,41

Com base na Tabela I, nota-se que os resultados obtidos pelo ASN não estão muito distantes daqueles obtidos pela comunidade científica. O ASN possui uma característica intrínseca de identificação de anomalias. Para isso, uma comparação direta entre padrões desconhecidos e padrões considerados anormais é realizada com base em uma medida de afinidade. O maior problema identificado durante as simulações foi resultado das similaridades entre segmentos normais e anormais, culminando com a classificação de diversos segmentos normais como arritmicos. Isto se deve as características dos sinais de ECG, que tem morfologia bastante variável de paciente para paciente. Por outro lado, a medida de sensibilidade indica que o algoritmo mostrou-se eficiente na tarefa de identificação de arritmias, apresentando resultados satisfatórios.

IV. CONSIDERAÇÕES FINAIS

A proposta deste trabalho foi apresentar o ASN como ferramenta para a análise de ECGs, visando a identificação de registros com arritmias cardíacas. O ASN mostrou-se eficiente devido a sua característica intrínseca de identificação e classificação. Os resultados obtidos foram: sensibilidade de 81,96% e especificidade de 62,41%.

Em trabalhos futuros pretende-se melhorar o desempenho do ASN propondo a extração de novas características que possam fornecer uma classificação mais adequada frente a padrões muito semelhantes.

AGRADECIMENTOS

Os autores agradecem a Universidade do Estado de Mato Grosso - UNEMAT, pelo apoio concedido.

REFERÊNCIAS

- [1] S. FORREST et al. "Self-Nonsel Self Discrimination in a Computer", *IEEE Symposium on Research in Security and Privacy*, pp. 202-212, 1994.
- [2] A. L. GOLDBERGER et al. *PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals*. *Circulation* 101(23):e215-e220.
- [3] G. T. MARKOS and I. F. DIMITRIOS "Automatic arrhythmia detection based on time and time-frequency analysis of heart rate variability", *Computer Methods and Programs in Biomedicine*, Volume 74, Issue 2, May 2004, Pages 95-108, ISSN 0169-2607
- [4] W. H. ORGANIZATION et al. *Health statistics and health information systems*. World Health Organization, 2012.
- [5] K. POLAT et al. "A new method to medical diagnosis: Artificial immune recognition system (AIRS) with fuzzy weighted pre-processing and application to ECG arrhythmia", *Expert Systems with Applications* 31. vol. 2, pp. 264-269, 2006.