CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DO PARANÁ -
CEFET-PR
COORDENAÇÃO DE PÓS-GRADUAÇÃO EM ENGENHARIA
ELÉTRICA E INFORMÁTICA INDUSTRIAL - CPGEI
LABORATÓRIO DE BIOINFORMÁTICA - BIOINFO

Autores *Authors*	Bruno Marchesi, Álvaro Luiz Stelle, Heitor Silvério Lopes
Título *Title*	Programação Genética na Detecção de Eventos Epilépticos: Resultados Preliminares
Publicação *Publication*	L. Caloba, J.M. Barreto (editores), III Congresso Brasileiro de Redes Neurais, p. 374-378
Detalhes *Details*	Florianópolis (SC), 21-25/julho/1997 ISBN: 85-900382-1-1

CEFET-PR / CPGEI / BIOINFO
Av. 7 de setembro, 3165
80230-901 - Curitiba (PR) - Brasil
Tel: +55-41-322-4544 ext. 663/671
Fax: +55-41-224-5170
http://bioinfo.cpgei.cefetpt.br

Programação Genética na Detecção de Eventos Epilépticos: Resultados Preliminares

Bruno Marchesi, Álvaro Luiz Stelle, Heitor Silvério Lopes

Centro Federal de Educação Tecnológica do Paraná - CEFET-PR
Curso de Pós-graduação em Engenharia e Informática - CPGEI
Av. 7 de Setembro, 3165 - 80230-901 - Curitiba - PR
e-mail: {*marchesi, stelle, hslopes}@cpgei.cefetpr.br*

Resumo

Este artigo apresenta um método para detecção automática de complexos ponta-e-onda de 3 Hz em sinais de eletroencefalografia (EEG) empregando técnicas de programação genética. Parâmetros de treinamento foram calculados a partir de quadros de EEG de 1s de duração, escolhidos aleatoriamente de arquivos pré-gravados. Os quadros foram classificados visualmente como complexo ponta-e-onda (*spike-and-slow-wave complex* - SASWC) ou ausência de complexo ponta-e-onda (*non-spike-and-slow-wave* complex - NSASWC). Técnicas de programação genética são então aplicadas a estes dados a fim de se obter um programa capaz de reconhecer tais complexos.

Palavras-chave: programação genética, processamento de sinais, EEG.

Abstract

This paper presents a method using genetic programming for automatic detection of 3 Hz spike-and-slow-wave complexes in EEG signals. Training features are extracted from 1s EEG frames, randomly chosen from pre-recorded files. The frames are visually classified as spike-and-slow-wave complex (SASWC) or non-spike-and-slow-wave complex (NSASWC). Genetic programming techniques are then applied to these data to build a program capable of recognizing such complexes.

Keywords: genetic programming, signal processing, EEG

I. Introdução

A. Epilepsia

Não há uma doença chamada “epilepsia”. Ao invés disso, existem crises epilépticas, que são reações anormais do cérebro, causadas por um grande número de doenças. O tipo de crise depende da extensão da área cerebral afetada e todo o cérebro, ou parte dele, pode estar envolvido. Em geral, a patologia é localizada no cérebro, entretanto disfunções em órgãos importantes, associados com trocas tóxico-metabólicas, podem causar encefalopatias secundárias, levando a crises epilépticas. Existe também a possibilidade de predisposição genética à epilepsia. Estima-se que 0,5% da população mundial tenha crises epilépticas de algum tipo [1].

O EEG escalpelar é um dos recursos mais importantes no diagnóstico e diferenciação dos vários tipos de epilepsia, em especial aquelas associadas a lesões corticais que são freqüentemente detectadas. Em geral, o EEG em pacientes epilépticos é anormal durante e entre as crises. Os padrões característicos são as pontas, isoladas ou em associações com outras ondas [2].

O período ictal da crise de ausência é caracterizado por descargas síncronas generalizadas de pontas-e-ondas de 3 Hz. Não é possível a ocorrência de tal crise se não houver estas descargas. Todavia, sua existência não implica a ocorrência de uma crise de ausência [1].

A descarga de ponta-e-onda é máxima na linha centro-fontral da cabeça e pode começar numa taxa de 4 Hz, reduzindo para 3 a 3,5 Hz e, na fase final do ataque, caindo para 2,5 Hz. Tanto o início quanto o fim da crise são abruptos e podem ser precedidos e seguidos de atividade EEG normal. A detecção automática destes complexos é essencial na análise de sinais de longa duração [3].

B. Programação Genética

O paradigma de programação genética é descrito a seguir. Este texto usa como referência Koza [4]. Por conveniência, o termo *fitness* (adequabilidade) será usado em inglês.

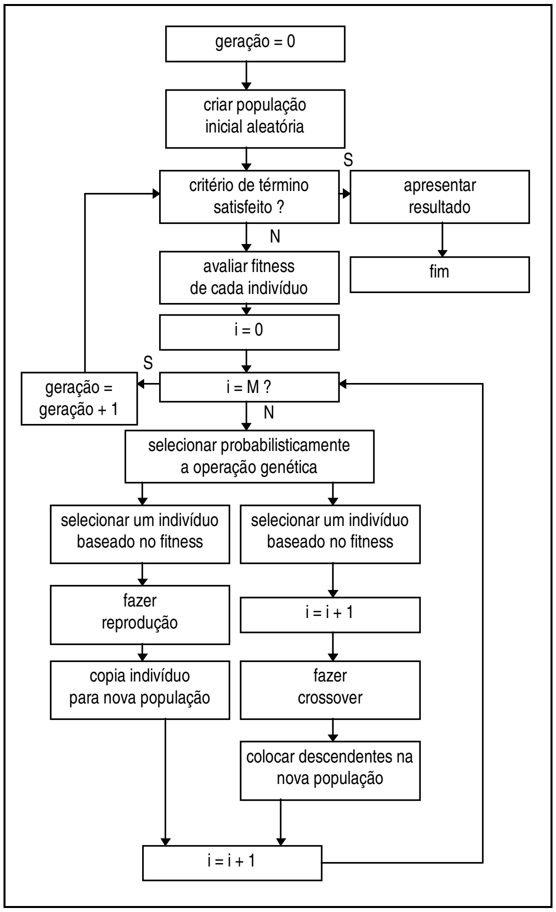

Figura 1: Fluxograma para o paradigma de programação genética. 'M' representa o tamanho da população e o índice 'i' refere-se a um indivíduo na população.

Muitos problemas aparentemente diferentes em inteligência artificial podem ser vistos como a necessidade de se encontrar um programa de computador que produza uma saída desejada segundo entradas particulares. O processo de solução desses problemas pode ser reformulado como uma busca de um programa de *fitness* elevado num universo de programas possíveis [4].

Em programação genética, populações de centenas ou milhares de programas são criados geneticamente, usando-se o princípio Darwiniano de reprodução e sobrevivência. A operação de recombinação genética (*crossover*) é usada para recombinar programas.

A programação genética começa com uma população inicial de programas compostos por funções e terminais apropriados para o problema. As funções podem ser operações aritméticas, funções matemáticas e programas ou funções lógicas. Cada indivíduo da população de programas é medido em termos de quão bem ele se comporta no ambiente do problema. Esta medida é chamada *fitness measure* (medida de adequabilidade). A natureza da medida de *fitness* varia com o problema.

Tipicamente, cada programa da população é executado sobre certo número de diferentes casos de *fitness* (*fitness cases*), de modo que o seu *fitness* é medido como uma soma ou uma média sobre várias situações. Tais casos representam, por exemplo, amostras de uma variável independente ou diferentes condições iniciais de um sistema.

A menos que o problema seja simples, de maneira que a busca aleatória possa resolvê-lo, o programa indivíduo da geração inicial terá *fitness* bastante pobre. Apesar disso, alguns indivíduos da população terão *fitness* mais alto e essas diferenças de desempenho dos programas são então exploradas através do princípio de reprodução e sobrevivência do melhor adaptado e da recombinação (*crossover*).

O processo de reprodução envolve a seleção, baseada no *fitness*, de um programa da população atual de programas, permitindo que ele sobreviva copiando-o para a nova população.

A reprodução entre dois programas é usada para criar uma descendência baseada no *fitness*. Tipicamente, os programas-pais são de formas e tamanhos diferentes, bem como os seus descendentes.

Intuitivamente, se dois programas são de alguma forma eficientes para resolver um problema, então cada um deles têm seus méritos. Recombinando-se partes aleatoriamente escolhidas destes programas, pode-se criar, eventualmente, um novo programa de *fitness* melhor do que os anteriores.

Depois que as operações de reprodução e *crossover* são executadas, a população de descendentes substitui a população antiga e o processo é repetido ao longo de muitas gerações.

Tipicamente, o melhor indivíduo que apareceu ao longo de todas as gerações é designado como o resultado da programação genética. A figura 1 apresenta um fluxograma para o paradigma da programação genética. O índice *i* refere-se a um indivíduo da população de tamanho *M*. A variável *gen* é o número da geração atual. Os passos a seguir resumem o paradigma da programação genética:

1. Criar uma população inicial, composta por uma combinação aleatória das funções e terminais relativos ao problema (programas).
2. Executar iterativamente as ações a seguir, até que um critério de parada, a ser definido, seja atingido:
   a. Executar cada programa na população e associar um valor de *fitness* de acordo com sua adequação na solução do problema proposto.
   b. Criar uma nova população de programas, pela aplicação das operações a seguir. Tais operações são probabilisticamente aplicadas aos programas segundo seus valores de *fitness*.
      i. Copiar programas exitentes para uma nova população.
      ii. Criar novos programas pela recombinação genética de duas partes aleatoriamente escolhidas dos programas existentes.
3. O melhor indivíduo de todas as gerações é considerado o resultado produzido pela programação genética.

C. Métodos aplicados ao reconhecimento de padrões em sinais de EEG

Värri et al. [5] aplicaram técnicas de redes neurais na redução de sinais de EEG de longa duração, através da seleção de períodos significativos do sinal, isto é, que contêm ondas alfa, padrões ponta-e-onda, ou artefatos de EOG (eletro-oculograma). Obtiveram uma detecção de 92 % dos eventos e um acerto de 13% na discriminação desses eventos.

Mylonas e Comley [6] propuseram um método para detecção automática de pontas num sinal de EEG aplicando filtros adaptativos, com filtragem e predição linear adaptativa. O método foi testado unicamente em um sinal de EEG de 10min de duração, e os resultados obtidos foram: 4 alarmes falsos, 62 detecções e nenhuma ponta perdida ou não detectada.

Kalayci e Özdamar [7] empregaram *wavelets* para pré-processar o sinal de EEG, antes de executar o reconhecimento automático dos padrões definidos como SSW (*spike and spike and wave* - ponta e ponta-e-onda) e *non*-SSW (qualquer outro padrão), obtendo os seguintes resultados: em média, a exatidão de reconhecimento durante os testes foi 2,1% menor que a exatidão de treinamento das redes neurais; o reconhecimento de ponta-e-onda com positivo-verdadeiro foi de 50,6% para o pior caso e 93,9% no melhor caso.

Schiff et al. [8] preocuparam-se com a obtenção de um algoritmo para processamento de sinais de EEG em tempo real, apresentando o CSDDWT (*critically sampled discrete dyadic wavelet transform)*, baseado na multi-resolução das *wavelets*, e na técnica de sobre-amostragem do sinal de EEG. Obtiveram uma capacidade de detecção de 94% de pontas num sinal de EEG típico, havendo a possiblidade de implementação em tempo real.

II. MÉTODOS

Serão abordadas a seguir as considerações sobre a terminologia usada neste trabalho. Posteriormente, serão apresentados os passos para a construção dos bancos de dados de treinamento e teste.

A. Terminologia

O banco de dados de teste é composto por um conjunto de parâmetros, os quais são valores numéricos calculados a partir de um vetor de dados, neste caso um sinal de EEG de 1segundo de duração (aqui denominado "quadro"). Foram utilizados sete parâmetros. Estes parâmetros são descritos por Hudgins et al [9]: *mean absolute value* (MAV), *number of zero crossings* (ZC), *number of slope sign changes* (SC) e *waveform length* (LEN). Outros parâmetros, descritos por Fernandez et al [10] foram também utilizados: *average value* (AVG), *up slope* (UP) e *down slope* (DOWN). Além disso, utilizou-se o seguinte critério (baseado em Niedermeyer [1]): se o quadro sob análise contém três pontas, é considerado como característico de descargas do tipo pequeno mal, e classificado como SASWC; caso contrário, é definido como NSASWC. Esta informação é passada para o programa genético, associando-se um número ao quadro: 1.0 se SASWC e -1.0 se NSASWC. Este número é tratado como um parâmetro também, a despeito de sua natureza diversa dos sete parâmetros previamente mencionados.

B. Procedimento de Treinamento

O procedimento para a obtenção da base de dados, usada para as etapas de treinamento e teste, é apresentada a seguir.

1. Foi utilizada uma base de dados confiável de sinais de EEG, publicamente disponível na Internet, na Universidade de Tampere, Finlândia*.
2. Dois arquivos contendo descargas do tipo pequeno mal foram selecionados da base de dados.
3. Cada arquivo é composto do registro de quatro canais do eletroencefalógrafo. Os dois primeiros canais de cada arquivo foram usados para a criação do banco de dados de treinamento, e os dois últimos foram usados para o banco de testes.
4. Cada canal tem, em média, 3 minutos de gravação a uma freqüência de amostragem de 200Hz. Assim, cada canal tem em torno de 36.000 amostras.
5. Quadros de 1 segundo (200 amostras) são

aleatoriamente escolhidos de cada um dos quatro canais usados até que dez quadros, visualmente classificados como NSASWC, sejam obtidos.

6. Cinco quadros SASWC foram extraídos de cada canal.
7. Finalmente, foram criados 40 quadros NSASWC e 40 SASWC. Para cada um destes 80 quadros, foram calculados os oito parâmetros anteriormente mencionados. Os 640 parâmetros resultantes constituem a base de dados de treinamento.

A base de dados de treinamento constitui a informação de entrada para o programa genético. Inicialmente, um programa indivíduo, isto é, uma expressão matemática de até sete variáveis, é aleatoriamente gerada. A expressão é avaliada, e o resultado é comparado com o oitavo parâmetro. A figura 2 ilustra as faixas de valores esperados.

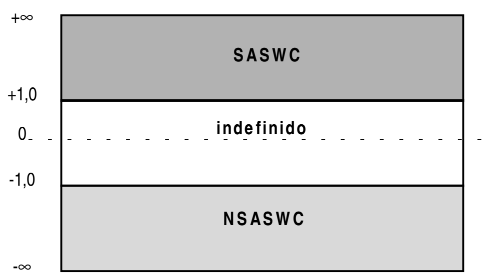

Figura 2: Faixas de valores esperados para um programa indivíduo. O símbolo ∞ representa um número arbitrariamente grande.

Por exemplo, se o programa genético fornece 1,5 e o número SASWC/NSASWC é 1,0, o número de *hits* é incrementado. O critério de *fitness* foi definido como:

$$\mathit{fitness} = \mathit{fitness\ cases} - \mathit{hits} \qquad (1)$$

onde *fitness cases* é o número de casos de treinamento e *hits* é o número de acertos de cada programa indivíduo.

Dois critérios de parada para o programa genético foram definidos, sendo que a execução é terminada se algum dos dois é satisfeito:

(a) Um programa indivíduo tem o número de *hits* igual ao número de casos de treinamento.

(b) O número máximo de gerações permitidas foi atingido.

Depois que um dos critérios de parada é atingido, o desempenho do programa de melhor *fitness* é verificado com os quadros de teste. O procedimento de criação do banco de dados de teste é o mesmo do banco de treinamento.

III. Resultados

O programa (expressão algébrica) a seguir foi o melhor programa indivíduo encontrado:

$$\begin{aligned} &2\cdot sc\cdot[\alpha - sc - \cos(sc)] + \\ &e^{avg+\alpha+e^{\alpha-\cos(len)}} + \\ &R\log(2\cdot sc) + R\log(2\cdot\alpha - mav) + \\ &R\log[2\cdot sc + [mav - avg - sc + (avg + 2\cdot sc)\cdot(mav + up)]\cdot\cos(len)] \end{aligned} \qquad (2)$$

onde:

$$\alpha = mav - sc + up \qquad (3)$$

$$R\log(x) = \begin{cases} 0, & \text{se } x = 0 \\ \log(\text{abs}(x)), & \text{se } x \neq 0 \end{cases} \qquad (4)$$

A expressão mostrada em (2), aplicada ao banco de teste composto por 40 quadros SASWC e 40 NSASWC, forneceu os seguintes resultados:

Tabela 1: Resultados do teste de detecção.

verdadeiro positivo	29	72,5%
verdadeiro negativo	39	97,5%
falso positivo	1	2,5%
falso negativo	11	27,5%
total de acertos	68	85%

O segundo teste foi feito com o todo sinal de EEG de um dos canais escolhidos para o banco de dados de teste. A expressão 2 foi calculada para cada amostra do sinal, e o resultado é apresentado na figura 3. A curva em negrito é o resultado da expressão, e a outra é o sinal de EEG original. A figura 4 mostra o valor máximo da expressão calculada, que coincide com a ocorrência da crise epiléptica que se pretendia detectar.

IV. Discussão e Conclusão

O propósito deste trabalho foi investigar a possibilidade de reconhecimento de SASWCs de 3Hz, característicos de crises epilépticas do tipo pequeno mal, em sinais de EEG de curta duração, empregando-se técnicas de programação genética. Este método foi anteriormente utilizado por Fernandez et al. [9], que usou o paradigma da programação genética no reconhecimento de padrões aplicado a sinais mioelétricos.

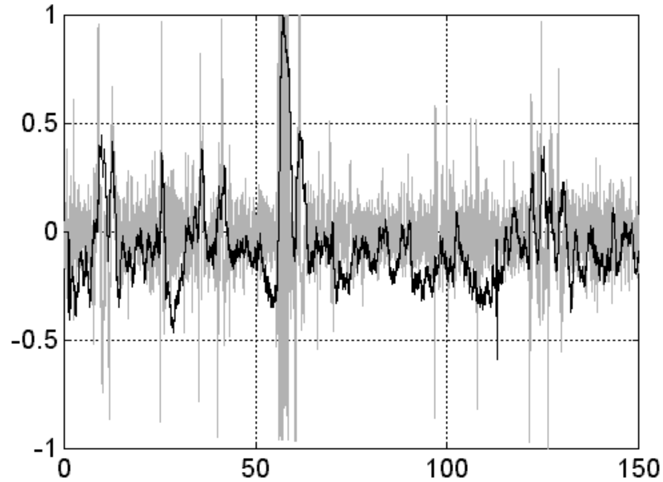

Figura 3: Expressão calculada para um canal de EEG.

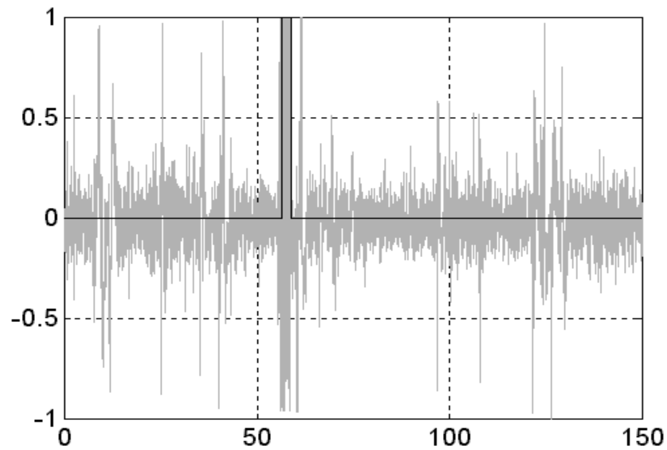

Figura 4 : O sinal de EEG e o valor máximo da equação 2, indicando a posição onde ocorrem os complexos ponta-e-onda.

Resultados satisfatórios foram obtidos nos testes experimentais, o que sugere a possibilidade de emprego da programação genética como um método de otimização de programas de reconhecimento de padrões, no caso específico de sinais de EEG de curta duração com complexos ponta-e-onda de 3 Hz. A generalização das conclusões obtidas deve ser feita com cuidado, porque os resultados apresentados não apontam no sentido de uma generalização; são específicos para a base de dados de EEG usada. Todavia, os resultados encorajam a continuação dos experimentos e estudos envolvendo esta metodologia.

AGRADECIMENTO

O autor Bruno Marchesi agradece à CAPES pelo fornecimento da bolsa de mestrado, que possibilitou a execução deste trabalho.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] NIEDERMEYER, E. Epileptic seizure disorders. In: _____.; SILVA, F. H. L. da. **Electroencephalography: basic principles, clinical applications, and related fields**. 1. ed. Baltimore : Urban & Schwarzenberg, 1987. p. 405-510.

[2] GEDDES, L. A.; BAKER, L. E. Electroencephalography. In: _____. **Principles of applied biomedical instrumentation**. New York : J. Wiley & Sons, 1989. p. 717-743.

[3] STELLE, A. L. **Real-time computer aided analysis of the electroencephalogram: a two-dimensional approach**. PhD Thesis, The City Univertity, London, 1991.

[4] KOZA, J. R. **Genetic Programming**. Massachessetts: MIT Press, 1994.

[5] VÄRRI, A. et al. **Pattern recognition of the EEG in epilepsy with neural networks.** In: 2nd International Hans-Berger Congress. "Quantitative and Topological EEG and MEG Analysis". Jena, Germany, 8-11 sep., 1993.

[6] MYLONAS, S. A.; COMLEY, R. A. Detection of epileptic spikes in the EEG using adaptive filters. **Anais do I Fórum Nacional de Ciência e Tecnologia em Saúde**, Campos do Jordão, SP, 1992, p. 268-271.

[7] KALAYCI, T.; ÖZDAMAR, Ö. Wavelet preprocessing for automated neural network detection of EEG spikes. **IEEE Engineering in Medicine and Biology**, v.14, n.2, p. 160-166, 1995.

[8] SCHIFF, S. J.; ALDROUBI, A.; UNSER, M. et al. Fast wavelet transformation of EEG. **Electroencephalography and Clinical Neurophysiology**, v. 91, n. 5, p. 442-455, 1994

[9] HUDGINS, B.; PARKER, P. e SCOTT, R. N. A New strategy for multifunction myoelectric control. **IEEE Transactions on Biomedical Engineering**, v. 40, p. 82-94, 1993.

[10] FERNANDEZ, J. J.; FARRY, K. e CHEATHAM, J. B. Waveform recognition using genetic programming: the myoelectric signal recognition problem. **Proceedings of Genetic Programming Conference,** Stanford, USA, 1995.

*Os endereços da Universidade de Tampere são:
ftp://sigftp.cs.tut.fi/pub/eeg-data/
http://www.tut.fi/