



CARACTERIZAÇÃO DE PADRÕES DE DISPAROS EM SÉRIES NEURONAIS

Rosângela Follmann¹, Elbert E. N. Macau², Rafael dos Santos³

¹Pós-Graduação em Computação Aplicada-CAP-INPE, São José dos Campos, Brasil, rosangela@lac.inpe.br

²Laboratório Associado de Computação e Matemática Aplicada, São José dos Campos, Brasil, elbert@lac.inpe.br

³Laboratório Associado de Computação e Matemática Aplicada-LAC-INPE, São José dos Campos, Brasil, rafael@lac.inpe.br

Resumo: Neste trabalho analisa-se o comportamento de *bursting* em sinais neuronais. Os sinais neuronais são provenientes da solução do modelo FitzHugh-Nagumo, o qual descreve a excitação e a propagação das propriedades eletroquímicas que percorrem a membrana do axônio. Para identificar os padrões de *burstings*, são usadas duas metodologias. A primeira, explora as propriedades estatísticas da série temporal para caracterizar padrões pré-definidos, através do *Casamento de Padrões*. A segunda metodologia busca por padrões desconhecidos ao longo da série. Essa técnica aplica uma “janela móvel” ao longo da série e são extraídas subsequências, que são convertidas em símbolos. A partir das subsequências de símbolos, constrói-se uma matriz de distâncias. Executando a análise desta matriz de distâncias, são encontrados os padrões que se repetem ao longo da série. Ambas as metodologias apresentam resultados interessantes.

Palavras Chaves: Modelo FitzHugh-Nagumo, Reconhecimento de Padrões, Casamento de Padrões.

1. INTRODUÇÃO

Muitos fenômenos naturais podem ser descritos com base em mecanismos não lineares de reação-difusão (RD), entre os quais se destacam o disparo dos neurônios no cérebro, o ritmo cardíaco e a organização celular [1–3]. Nessa classe de sistemas de reação-difusão se encaixa o modelo FitzHugh-Nagumo [4, 5], o qual captura características-chaves de muitos processos biológicos. Em particular, este modelo simula o comportamento do potencial de ação ou *spike* que é essencial para a vida animal, pois este comportamento está associado ao rápido transporte das informações entre e dentro dos tecidos. Basicamente, um potencial de ação é uma onda de descarga eletroquímica que percorre a membrana de uma célula, que é o mecanismo básico usado pelo sistema nervoso para comunicação entre os neurônios.

No sistema nervoso existe um tipo comum de desordem conhecida por epilepsia. A epilepsia é caracterizada por crises recorrentes, que são desencadeadas por distúrbios temporários das funções do cérebro, e podem durar de segundos a minutos. As crises são resultantes de súbitas descar-

gas paroxísticas, excessivas e em padrões rítmicos organizados (sincrônicos) de uma população de neurônios no córtex cerebral [6]. Estes padrões geralmente iniciam e terminam espontaneamente, sem intervenção externa. Esta inicialização e terminação aparentemente abrupta das crises epiléticas reflete em propriedades intrínsecas do cérebro epilético, uma vez que as atividades de *bursting* estão diretamente relacionadas às crises epiléticas. *Bursting* é uma forma rápida de disparos que ocorre nos neurônios. Tais disparos podem estar agrupados em dois ou mais potenciais de ação, ver Figura 1. Em alguns pacientes, os fatores desencadeantes das crises podem ser luzes piscantes, certos tipos de ruídos, leitura prolongada, privação de sono, fadiga, uso de álcool, hipoglicemia, etc.

Neste trabalho é realizado um estudo sobre a influência de um estímulo periódico sobre um neurônio, principalmente a caracterização desta excitação sobre a dinâmica dos *burstings* no modelo de FitzHugh-Nagumo. O estímulo periódico simula uma luz piscante, que pode ser o fator que desencadeia uma crise. Para a caracterização destas atividades serão necessárias técnicas de reconhecimento de padrões que possam identificar os *burstings*.

As técnicas de Reconhecimento de Padrões possuem aplicações na análise exploratória de base de dados, monitoramento e diagnóstico de sistemas críticos, classificação de séries temporais, e também na descoberta não supervisionada de padrões recorrentes [7–9]. Para resolver o problema de encontrar padrões em séries temporais, pode-se essencialmente optar por duas abordagens. Na primeira, os métodos usam padrões pré-definidos como moldes e fazem uma busca por tais padrões ao longo da série, como é o caso dos métodos: Casamento de Padrões (*Template Matching*) [10–12], que será usado neste trabalho, Alinhamento Temporal Dinâmico (*Dynamic Time Warping*) [13], Modelos Escondidos de Markov (*Hidden Markov Models*) [10], wavelets [14] e Redes Neurais Artificiais [15]. Na segunda abordagem, não é necessária a identificação prévia de um molde, pois os métodos buscam por similaridades que ocorrem ao longo da série temporal. Um exemplo de aplicação é a descoberta de certos padrões em sequências biológicas que assumem um papel importante na identificação de doenças genéticas,

e em decifrar mecanismos biológicos [16, 17]. A investigação deste tipo de abordagem vem gerando diversos trabalhos sobre o assunto, dos quais pode-se citar [18, 19]. Entre os métodos existentes, destacam-se os Modelos de Mistura [7], Mapas Auto-Organizáveis (*self-organizing map*, SOM) [15] e o algoritmo EMMA (Enumeration of Motifs through Matrix Approximation) [18].

2. MODELO NEURONAL

Desde os primórdios, o ser humano busca o entendimento do funcionamento do sistema nervoso, tendo descoberto a natureza elétrica dos sinais nervosos e que o cérebro é constituído por bilhões de células, os neurônios, que se comunicam através das sinapses.

Neurônio é a célula do sistema nervoso responsável pela condução do impulso nervoso. Também pode ser considerado a unidade básica da estrutura do cérebro e do sistema nervoso. Os modelos de neurônio procuram caracterizar o neurônio através de sistemas de equações diferenciais e parâmetros que representam as propriedades biológicas presentes no neurônio real, como por exemplo, a concentração de íons, difusão de neurotransmissores, etc. [20]. Os pioneiros foram Hodgkin e Huxley, que propuseram um modelo fenomenológico composto por quatro equações diferenciais para explicar os eventos observados durante a ocorrência de um potencial de ação no axônio gigante da lula [21].

Em 1961, FitzHugh e Nagumo propuseram um modelo simplificado, que reproduz os principais resultados das equações de Hodgkin-Huxley. Este modelo isola conceitualmente as propriedades matemáticas essenciais de excitação e propagação das propriedades eletroquímicas do fluxo de íons de potássio e sódio na membrana do axônio [4, 5]. O sistema de equações é dado por:

$$\dot{V} = k(4V - V^3 - W); \quad (1)$$

$$\dot{W} = (V - b(t)W - c)/k; \quad (2)$$

onde $a, c, k = 2.4$ são parâmetros, V representa o potencial da membrana e W a variável de restauração. A fim de investigar a influência de fotoestímulos sobre os neurônios, neste trabalho será considerado uma perturbação periódica, $b(t) = a \text{sen}(\omega t)$, sobre o modelo neuronal, que se refletirá na dinâmica rápida e lenta do modelo [22].

O comportamento temporal do potencial de ação é ilustrado na Figura 1, onde podem ser vistos o comportamento de *bursting* que é uma transição recorrente entre um estado tranquilo e um estado de repetida excitação. O espaço de fase do modelo FitzHugh-Nagumo na Figura 2 que descreve a trajetória no plano x-y para uma determinada condição inicial, e em vermelho são mostradas as curvas quando $\dot{V} = 0$ e $\dot{W} = 0$, onde o sinal de \dot{V} e \dot{W} passa por zero.

3. METODOLOGIAS

A identificação dos diferentes padrões nas séries provenientes da solução do modelo de FitzHugh-Nagumo será realizada por meio de duas metodologias: O Casamento de

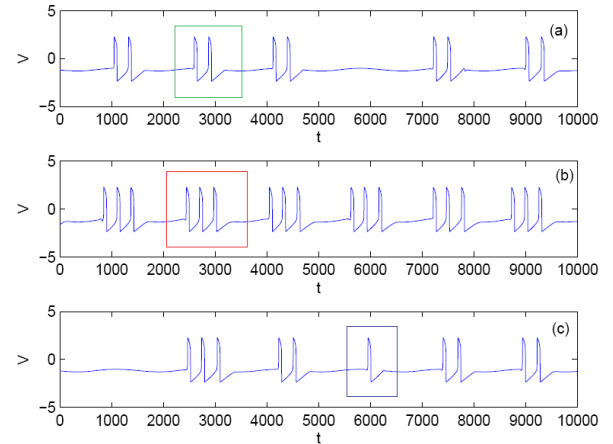


Figure 1 – Comportamento do modelo FitzHugh-Nagumo ao longo do tempo. (a) *bursting* com dois pulsos, (b) *bursting* com 3 pulsos, e (c) *bursting* com 3, 2 e 1 pulsos.

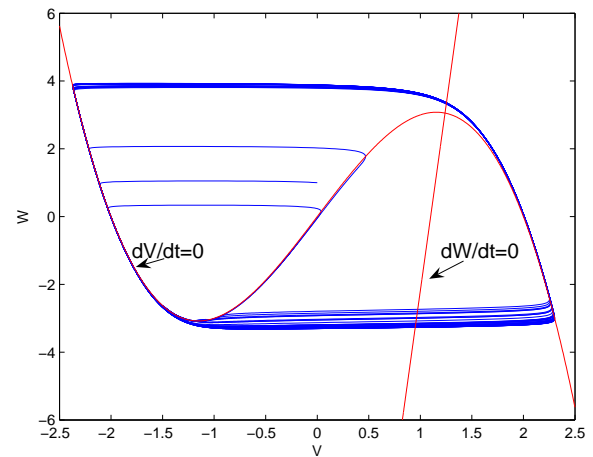


Figure 2 – Espaço de fase do modelo FitzHugh-Nagumo.

Padrões implementado de forma supervisionada e uma Técnica não Supervisionada Baseada em Representação Simbólica para encontrar o padrões de interesse.

3.1. Casamento de Padrões

A abordagem por casamento de padrões ou *template matching* é geralmente usada em casos bi-dimensionais de reconhecimento de padrões. Basicamente, esta técnica usa uma máscara de convolução feita para uma característica específica da imagem a ser analisada [23]. Em séries temporais, a aplicação é semelhante: neste caso usa-se a técnica para encontrar subsequências da série que sejam similares a um padrão (*template*) de interesse.

Para verificar similaridade entre um padrão de referência e um padrão de teste, é necessário uma medida de similaridade, que neste estudo foi usada a distância Euclidiana.

Seja $r(k), k = 1, \dots, W$ um segmento de um sinal $y(t)$. E seja $s(k), k = 1, \dots, W$ um padrão de referência, então a

distância, $D(\mathbf{r}, \mathbf{s})$, entre $r(k)$ e $s(k)$ é dada por:

$$D(\mathbf{r}, \mathbf{s}) = \left(\sum_{i=1}^W |r_i - s_i|^2 \right)^{\frac{1}{2}}, \quad (3)$$

sendo a distância da série estimada, segmento a segmento. W é o tamanho do padrão de referência e τ , o passo de deslocamento da janela, ver Figura 3. Por exemplo, se $W = 100$ e $\tau = 5$, então o segmento de sinal a ser analisado será $y(1), \dots, y(100)$, o próximo segmento será $y(6), \dots, y(106)$. De uma forma geral, cada segmento de sinal será dado por: $y(n \cdot \tau + k)$, com $n = 0, \dots, N$ e $k = 1, \dots, W$, onde N é o tamanho do sinal e W é o tamanho da janela. Note que o tamanho da janela será igual ao tamanho do padrão de referência. O desempenho do casamento é baseado na distância mínima. Se a distância mínima é menor que um limiar (ξ), então a série temporal possui o padrão representado pelo padrão de referência. O limiar é escolhido heurísticamente, neste caso o limiar será $\xi = 20$. O deslocamento da janela foi de passo unitário, ou seja $\tau = 1$, com esta escolha garantimos maior a representatividade do sinal.

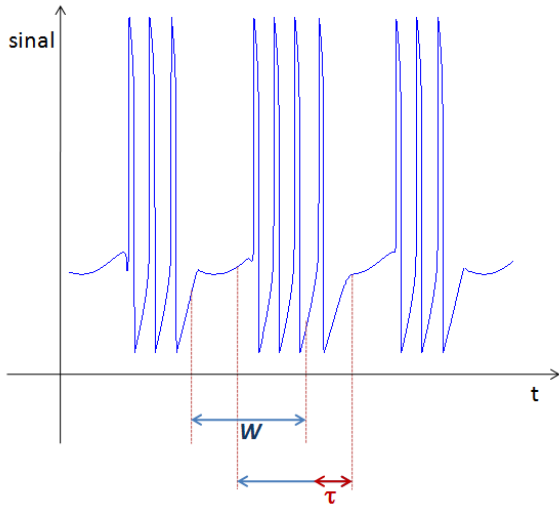


Figure 3 – Ilustração de um segmento de sinal, sendo W é o tamanho da janela e τ o passo de deslocamento da janela.

3.2. Técnica Baseada em Representação Simbólica

Muitas das pesquisas experimentais tem por base a observação de eventos ao longo do tempo, que podem representar fenômenos naturais de interesse. Isso pode gerar séries muito longas, muitas vezes com bilhões de observações. Porém, nem sempre há interesse em propriedades globais das séries temporais, mas sim em subseções destas séries, as quais são chamadas de subseqüência [24].

A técnica que será apresentada, faz uso de representação simbólica para tais séries temporais, permitindo a redução na dimensionalidade, de modo a encontrar padrões que se repetem ao longo da mesma.

A representação simbólica é obtida a partir do SAX (*Symbolic Aggregate Approximation*) apresentada por [25]. O

SAX permite que uma série de tamanho arbitrário n seja convertida em uma seqüência de símbolos de tamanho m , onde $m < n$, e o tamanho do alfabeto é um inteiro α , onde $\alpha > 2$. Esta conversão se dá em duas etapas: A primeira etapa é reduzir a dimensionalidade da série original e a segunda é a discretização, que produz símbolos equiprováveis.

Os dados são transformados por uma aproximação que agrega segmentos, conhecida por PAA (*Piecewise Aggregate Approximation*), reduzindo a dimensionalidade da série temporal [26]. Seja uma série temporal $C = c_1, \dots, c_n$, que pode ser representada em um espaço m -dimensional por um vetor $\bar{C} = \bar{c}_1, \dots, \bar{c}_m$.

O i -ésimo elemento de \bar{C} é calculado pela seguinte equação:

$$\bar{c}_i = \frac{m}{n} \sum_{j=\frac{n}{m}(i-1)+1}^{\frac{n}{m}i} c_j \quad (4)$$

De forma simples, para reduzir séries temporais de n para m dimensões, cada série é dividida em m partes de tamanhos iguais. O valor médio de cada ponto da série que cai dentro de uma destas partes é calculado, e o vetor destes valores torna-se a representação da série reduzida. Essa representação pode ser visualizada como uma tentativa de se aproximar a série original em uma combinação linear de “funções caixa”, como mostrado na Figura 4. Neste caso, uma seqüência C de tamanho 128 foi reduzida para uma seqüência $\bar{C} = \bar{c}_1, \bar{c}_2, \dots, \bar{c}_8$, de 8 dimensões.

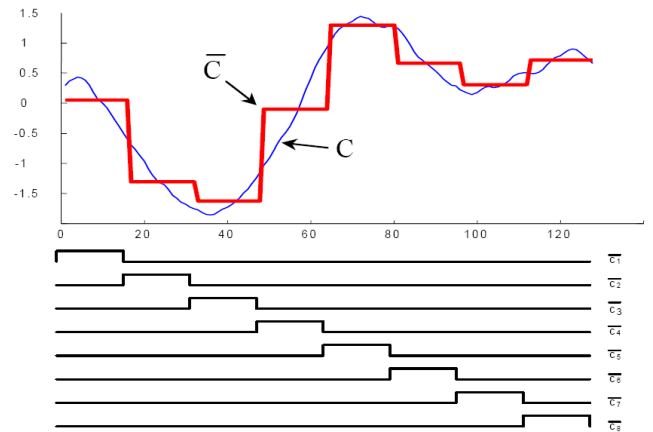


Figure 4 – Exemplo de redução de dimensão. Neste caso, uma seqüência (em azul) de tamanho 128 foi reduzida para 8 dimensões.

FONTE: Adaptada de [25, p.3].

Após o processo da redução de dimensão, aplica-se uma transformação, com o objetivo de se obter uma representação discreta da série reduzida. Admitindo que os dados normalizados possuem distribuição gaussiana, pode-se determinar os “pontos de quebra” que geram α áreas iguais abaixo da curva gaussiana.

Os pontos de quebra são uma lista ordenada de números $B = \beta_1, \dots, \beta_{\alpha-1}$ tal que a área abaixo da curva gaussiana com $N(0, 1)$ de β_i à β_{i+1} é igual a $\frac{1}{\alpha}$. Estes pontos de quebra podem ser determinados com base em uma tabela estatística.

Na Figura 5, são ilustrados os valores dos pontos de quebra que dividem uma distribuição gaussiana em um número arbitrário, α , de regiões equiprováveis, onde α pode assumir valores de 3 à 10.

α	3	4	5	6	7	8	9	10
β_1	-0.43	-0.67	-0.84	-0.97	-1.07	-1.15	-1.22	-1.28
β_2	0.43	0	-0.25	-0.43	-0.57	-0.67	-0.76	-0.84
β_3		0.67	0.25	0	-0.18	-0.32	-0.43	-0.52
β_4			0.84	0.43	0.18	0	-0.14	-0.25
β_5				0.97	0.57	0.32	0.14	0
β_6					1.07	0.67	0.43	0.25
β_7						1.15	0.76	0.52
β_8							1.22	0.84
β_9								1.28

Figure 5 – Tabela contendo os pontos de quebra para α de 3 a 10.
FONTE: Adaptada de [25, p.4].

A partir dos pontos de quebra, a série reduzida é discretizada da seguinte forma: Todos os coeficientes de PAA que estão abaixo do menor ponto de quebra são mapeados para o símbolo “a”; todos os coeficientes maiores ou iguais do que o menor ponto de quebra e menores do que o segundo menor ponto de quebra são mapeados pelo símbolo “b” e assim por diante. A Figura 6 ilustra este procedimento. De acordo com esta descrição, é então introduzido a representação simbólica SAX.

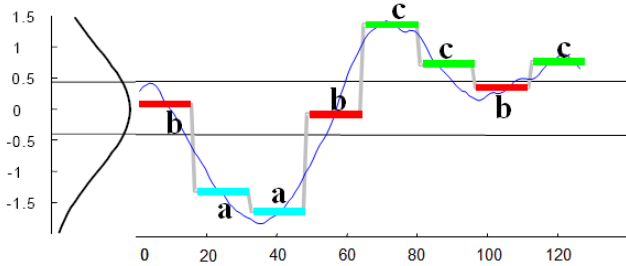


Figure 6 – Usando os pontos de quebra para mapear os coeficientes PAA em símbolos. Neste exemplo, $n = 128$, $m = 8$ e $\alpha = 3$, a série foi mapeada para **baabcbcb.**
FONTE: Adaptada de [25, p.4].

Obtida a representação simbólica da série é necessário definir-se uma medida de distância entre símbolos. Sejam duas palavras, $\hat{Q} = \hat{q}_1, \dots, \hat{q}_m$ e $\hat{C} = \hat{c}_1, \dots, \hat{c}_m$, define-se uma função distância D_s como:

$$D_s(\hat{Q}, \hat{C}) \equiv \frac{1}{n} \sqrt{\sum_{i=1}^m (dist(\hat{q}_i, \hat{c}_i))}, \quad (5)$$

onde a função $dist$ retorna a distância obtida a partir de uma tabela de distâncias entre os símbolos do alfabeto. Cada célula, $cel_{r,c}$, da tabela de distâncias pode ser calculado pela

seguinte expressão:

$$cel_{r,c} = \begin{cases} 0, & \text{se } r = c \\ |r - c|, & \text{caso contrario} \end{cases} \quad (6)$$

Para um determinado alfabeto α , o cálculo da tabela se faz uma única vez. A Figura 7 mostra um exemplo de tabela de distâncias para $\alpha = 3$, onde a distância entre dois símbolos pode ser obtida examinando a correspondente linha e coluna, por exemplo, $dist(\mathbf{a}, \mathbf{a}) = 0$ e $dist(\mathbf{a}, \mathbf{c}) = 2$.

	a	b	c	d
a	0	1	2	3
b		0	1	2
c			0	1
d				0

Figure 7 – Tabela usada pela função D_s .

De forma geral, a técnica utiliza uma janela móvel para extrair subsequências e converter estas em seqüências de símbolos, que é realizado com o auxílio do SAX. Logo após é calculada a distância entre todas as subsequências obtendo-se uma matriz de distância. A partir das distâncias mínimas, são identificados os padrões que se repetem com maior frequência ao longo da série.

4. RESULTADOS

As duas metodologias foram aplicadas a três séries com diferentes características, as quais podem ser vistas na Figura 8. Estas foram obtidas da solução do modelo FitzHugh-Nagumo, e contêm padrões bem definidos, bons para testar as metodologias. Na Figura 8(a) tem-se a variação do potencial de ação ao longo do tempo com padrões repetidos de um pulso. Em (b) são observados seqüências de *bursting* com quatro pulsos e em (c) mostra-se uma série mista contendo *bursting* com um, dois e três pulsos.

No casamento de padrões são necessários padrões de referência (*templates*), que foram extraídos da solução da Equação 1, dos quais foram escolhidos padrões de *bursting* com 01 pulso, 02 pulsos, 03 pulsos e 04 pulsos, ilustrados na Figura 9.

O casamento de padrões é realizado num sinal $y(t)$ para os quatro padrões de referência, gerando quatro vetores de distância (\mathbf{D}) referentes a cada molde apresentado. O mínimo entre estes vetores indica os padrões similares. Por exemplo, se o mínimo for proveniente do padrão s_1 , isto significa que este padrão (s_1) foi encontrado ao longo da série analisada, e assim por diante.

Na Figura 10(a) podemos observar a ocorrência de 11 padrões s_1 ao longo da série. Em (b) são ilustrados os quatro vetores de distância referentes a cada molde apresentado, em azul a \mathbf{D} para o padrão s_1 , em verde a \mathbf{D} para o padrão s_2 , em vermelho a \mathbf{D} para o padrão s_3 e em azul claro a \mathbf{D} para o

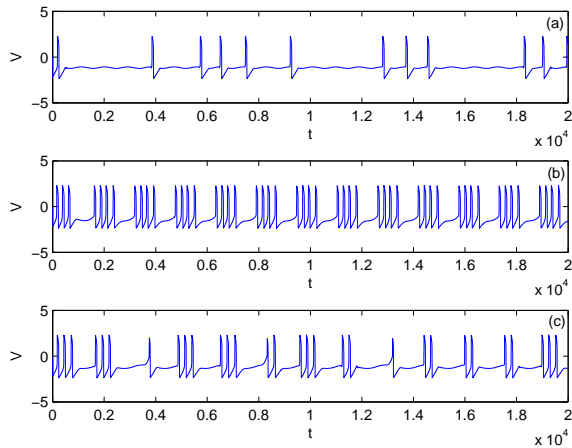


Figure 8 – Sinais escolhidos para testar a metodologia. (a) sinal com 11 padrões de 1 pulso, (b) sinal com 12 padrões de 4 pulsos, (c) sinal misto, contendo 2 padrões de 1 pulso, 5 padrões de 2 pulsos e 6 padrões de 3 pulsos.

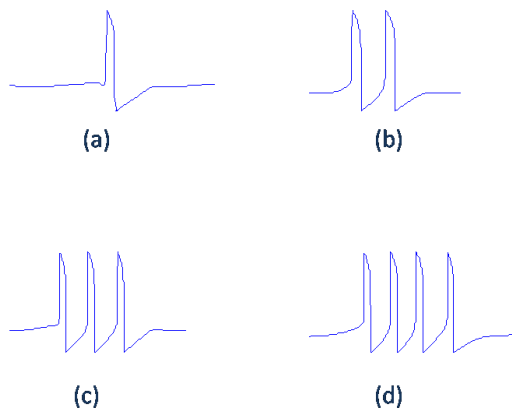


Figure 9 – Padrões usados para o casamento. (a) um pulso (s1), (b) dois pulsos (s2), (c) três pulsos (s3), (d) quatro pulsos (s4).

padrão s_4 . Observa-se que os mínimos entre as quatro séries são referentes ao padrão s_1 (em azul), portanto traçando-se o limiar $\epsilon = 20$ (linha preta em (c)) foram identificados 10 padrões s_1 , mostrando que a técnica encontrou quase todos os padrões s_1 contidos na série analisada, indicando um acerto em torno de 90%.

Na Figura 11 uma série contendo 12 padrões com seqüências de *bursting*, com quatro pulsos, é analisada e o mínimo obtido direciona ao padrão de referência s_4 , sendo que identificou-se 11 padrões s_4 indicando também um acerto de $\sim 90\%$.

Uma série mista, contendo diferentes padrões foi analisada. Esta série contém 2 padrões s_1 , 5 padrões s_2 e 6 padrões s_3 . O resultado pode ser visto na Figura 12, onde foi possível identificar a maior parte dos padrões, exceto o padrão s_3 destacado pelo retângulo vermelho, dos quais foram encontrados apenas 3 dos 6 padrões s_3 existentes na série. Os padrões s_1 (em azul) e s_2 (em verde) foram iden-

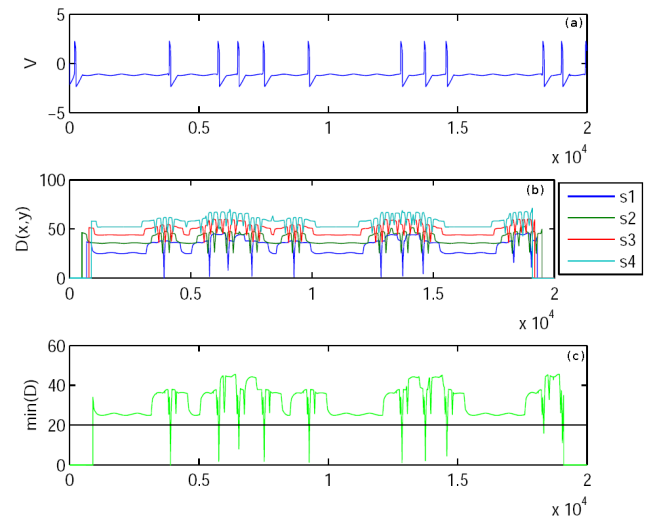


Figure 10 – (a) série contendo 1 spike, (b) vetor de distâncias do casamento entre os 4 padrões de referência, (c) mínimo das distâncias de todos os templates.

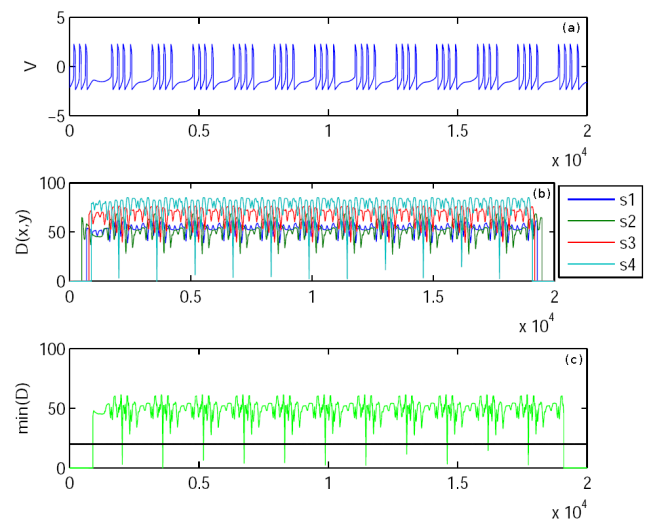


Figure 11 – (a) série contendo 4 spikes, (b) vetor de distâncias do casamento entre os 4 padrões de referência (c) mínimo das distâncias de todos os moldes, a linha preta é o limiar.

tificados completamente. Devido ao maior erro na detecção do padrão s_3 , obteve-se um acerto de aproximadamente 77% em toda a série.

Com base nestes testes, a metodologia mostrou-se eficaz no reconhecimento dos padrões, e observa-se que os padrões localizados nas extremidades (porção inicial e final de uma série) nem sempre são identificados pela técnica. Assim, o bom desempenho da técnica está relacionado com padrões bem definidos nas bordas da série.

Com base nestes resultados preliminares, a caracterização de padrões via casamento de padrões pode ser aplicada a uma grande quantidade de sinais. Portanto, com o propósito de investigar a resposta do modelo FitzHugh-Nagumo (Equação

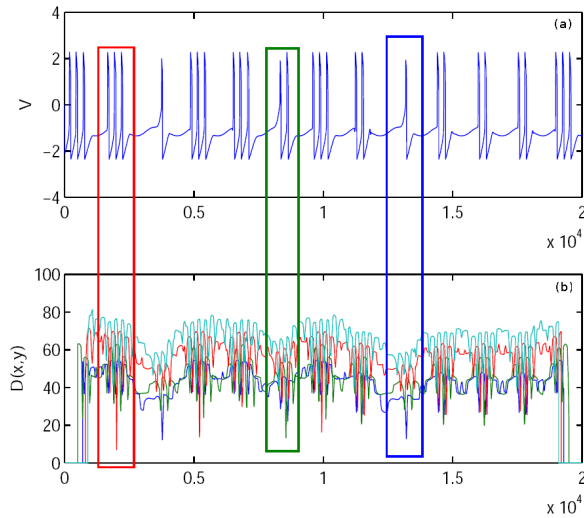


Figure 12 – (a) série mista contendo 1, 2 e 3 pulsos, (b) casamento de padrões para os 4 padrões.

1) sob diferentes estímulos, foram gerados diferentes sinais através da variação dos parâmetros a , c e ω deste modelo da seguinte forma:

- 100 amostras variando $\omega \in [0.001, 0.5]$ e mantendo fixos $a = 0.2, c = 2/\sqrt{3}$.
- 100 amostras variando $a \in [0.01, 0.5]$ e mantendo fixos $\omega = 0.04, c = 2/\sqrt{3}$.
- 100 amostras $c \in [0.95, 1.6]$ e mantendo fixos $\omega = 0.04, a = 0.2$.

Sendo assim, obteve-se 300 amostras com 20000 pontos cada. Para a análise, aplicou-se o método casamento de padrões em todas as amostras, e então quantificou-se a ocorrência dos padrões de referência nesses sinais.

Na Figura 13 podem ser observados os padrões encontrados a partir das 100 amostras geradas com a variação do parâmetro ω . Neste caso, foram identificados 32 padrões, dos quais 3 são referentes ao padrão $s1$, 14 ao padrão $s2$, 15 ao padrão $s3$ e nenhum padrão $s4$. Nas 100 amostras geradas com a variação de a foram identificados 303 padrões, dos quais 24 são referentes ao padrão $s1$, 81 ao padrão $s2$, 100 ao padrão $s3$ e 98 ao padrão $s4$, ver Figura 14. E na Figura 15 foram identificados um total de 280 padrões, dos quais 82 são referentes ao padrão $s1$, 120 ao padrão $s2$, 70 ao padrão $s3$ e 3 ao padrão $s4$, das 100 amostras obtidas da variação de c .

Para a melhor compreensão da influência destes parâmetros, usou-se coordenadas paralelas para visualizar estes resultados. A visualização de Coordenadas Paralelas, proposta por Inselberg [27], consiste em mapear um espaço n -dimensional em uma estrutura bidimensional que utiliza n eixos equidistantes denominados coordenadas. Os eixos verticais representam as dimensões dos dados e cada linha representa um item, permitindo a visão de padrões.

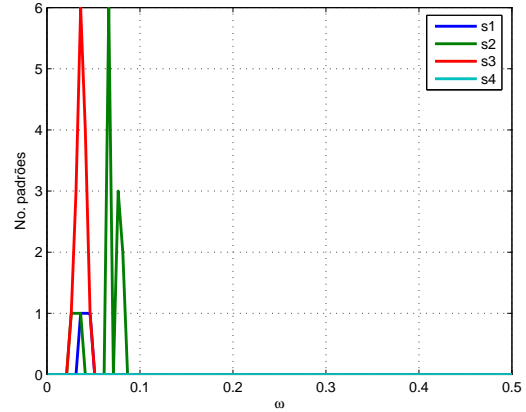


Figure 13 – Padrões encontrados no conjunto de amostras, para $a = 0.2$ e $c = 2/\sqrt{3}$

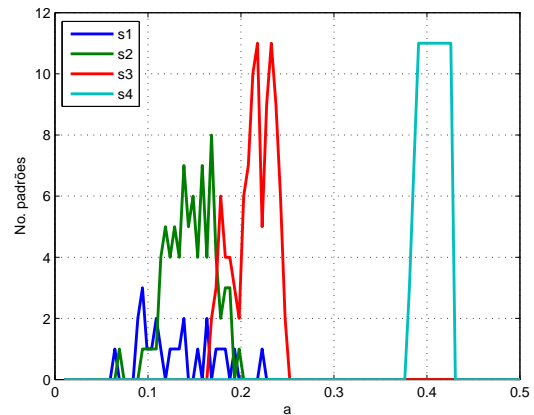


Figure 14 – Padrões encontrados no conjunto de amostras, para $\omega = 0.04$ e $c = 2/\sqrt{3}$

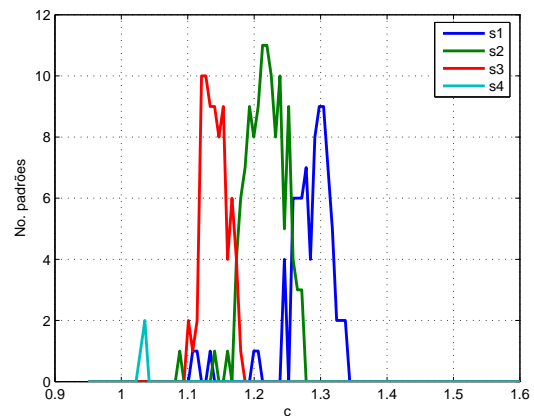


Figure 15 – Padrões encontrados no conjunto de amostras, para $\omega = 0.04$ e $a = 0.2$

A Figura 16 mostra a visualização, em coordenadas paralelas, para o parâmetro ω e as ocorrências de pulsos encontrados. As linhas destacadas em vermelho ilustram os val-

ores que geram comportamento excitatório no modelo. Neste caso, o padrão de referência $s1$ está associado à ocorrência de 1 *spike*; $s2$ de 2 *spikes*, e assim por diante. Observa-se que apenas uma pequena parte do vetor de parâmetros gera tais padrões de referência, o que também ocorre para os parâmetros a e c , como pode ser visualizado nas Figuras 17 e 18. Comparando estes resultados, observa-se que os parâmetros a e c têm mais influência sobre o modelo, gerando mais atividades de *bursting* do que quando o parâmetro ω é variado. Note que o parâmetro ω está associado a frequência do estímulo periódico, portanto isto indica que a alteração na frequência do piscar de uma lampada, por exemplo, não influencia tanto quanto sua amplitude (relacionado ao parâmetro a) para desencadear comportamentos de *bursting* no neurônio.

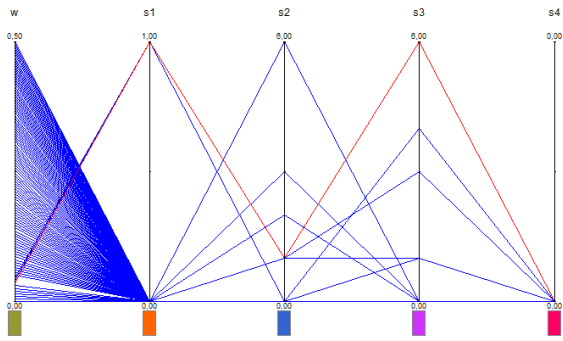


Figure 16 – Coordenadas paralelas para o parâmetro ω , em vermelho $\omega = 0.0363$, $a = 0.2$, $c = 2/\sqrt{3}$.

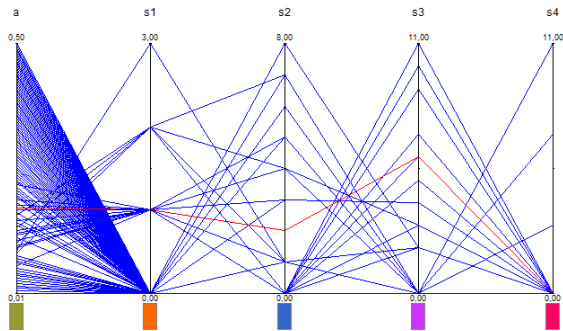


Figure 17 – Coordenadas paralelas para o parâmetro a , em vermelho $\omega = 0.04$, $a = 0.1733$, $c = 2/\sqrt{3}$.

Estes resultados são interessantes, pois podem auxiliar na escolha dos parâmetros para o modelo. Se há interesse em determinado comportamento do modelo, a escolha pode ser baseada nesse tipo de análise.

A técnica baseada em representação simbólica encontra apenas os padrões, porém não os classifica. Dessa forma, não será possível utilizar esta técnica para a análise mostrada anteriormente, com o casamento de padrões. A seguir são mostrados os resultados usando a técnica baseada em representação simbólica descrita na seção anterior. Primeiramente, usa-se o SAX para obter a representação simbólica. O SAX pode ser usado para converter toda a série em uma seqüência de símbolos, ou pode-se usar uma janela móvel

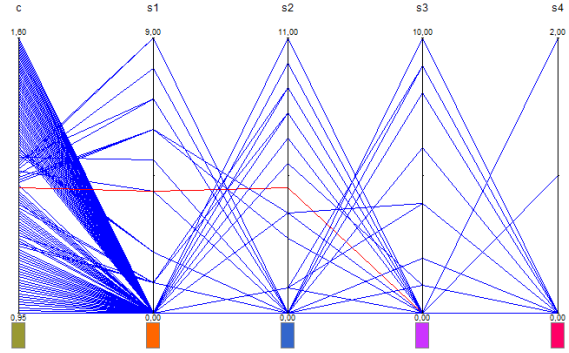


Figure 18 – Coordenadas paralelas para o parâmetro c , em vermelho $\omega = 0.04$, $a = 0.2$, $c = 1.2455$.

para extrair subseqüências e converter estas em seqüências de símbolos. Para os resultados que aqui serão mostrados, usou-se uma janela móvel e gerou-se uma matriz de dados simbólicos, \mathbf{S} , contendo todas as subseqüências. A distância entre todas as subseqüências é calculada e obtém-se uma matriz de distância. A partir das distâncias mínimas, são identificados os padrões que se repetem na série.

Para aplicar o SAX é necessário se determinar um alfabeto adequado. Foram testados diferentes valores de alfabeto, e escolheu-se $\alpha = 6$ para todos os casos analisados. A escolha de N é baseada em termos dos padrões de referência utilizados na metodologia anterior. A dimensão a que a subseqüência será reduzida foi escolhida de forma intuitiva, tendo em vista que esta escolha não pode afetar a representatividade (comportamento) desta seqüência ao longo do tempo. Para todos os casos escolheu-se $n = 100$.

Seja a série contendo padrões repetidos de 1 pulso, ilustrada na Figura 8(a). Para aplicar a técnica descrita, é necessário indicar o tamanho da janela, o tamanho da dimensão a que será reduzida e o alfabeto. Aqui o SAX é aplicado com $N=1400$, $m=100$ e $\alpha = 6$. O valor de N foi escolhido de modo a ser o mesmo aplicado na técnica de casamento de padrões.

Na Figura 19 é ilustrado a redução de dimensão da primeira subseqüência através do PAA. Então, obtém-se a matriz de dados simbólicos \mathbf{S} e, em seguida, calcula-se a distância entre todas as seqüências de símbolos contidas em \mathbf{S} , umas com as outras. Na Figura 20, pode ser visualizada a matriz de distâncias, D_s , cujo cálculo foi efetuado entre todas as subseqüências de símbolos. A cor azul escura representa os menores valores, o que significa que nestas regiões a distância entre duas subseqüências é pequena, indicando similaridade entre elas. A seguir, escolhem-se as distâncias menores que um determinado limiar, ilustradas na Figura 21, e estas indicam os padrões similares. Com esta análise são encontrados diversos padrões, que ocorrem duas ou mais vezes ao longo da série. O padrão com maior ocorrência é o objetivo.

Os padrões que ocorrem com mais frequência identificados pela técnica são mostrados em verde na Figura 22. Observa-se que não foram encontrados padrões com pulsos, como os que encontramos na técnica casamento de padrões.

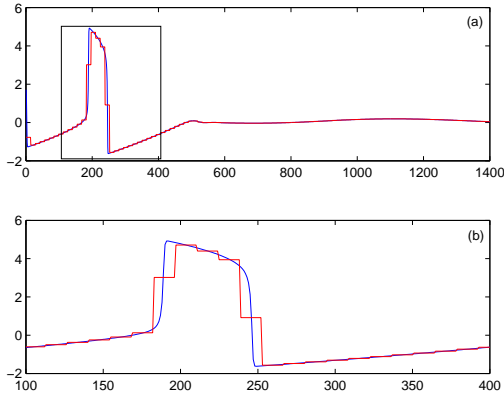


Figure 19 – (a) Ilustra a redução de dimensão da primeira subsequência obtida da série em análise, através do PAA; (b) Ampliação da parte destacada pelo retângulo;

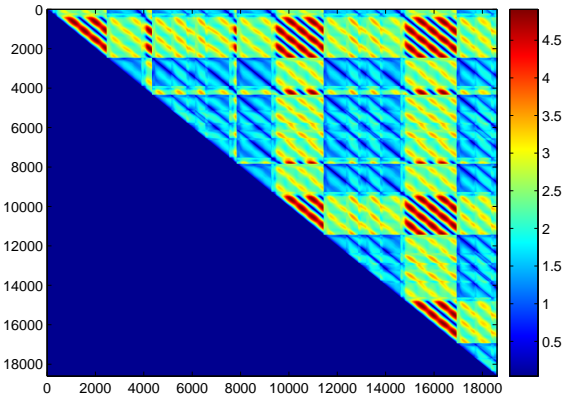


Figure 20 – Matriz de distâncias D_s entre todas as subsequências de símbolos.

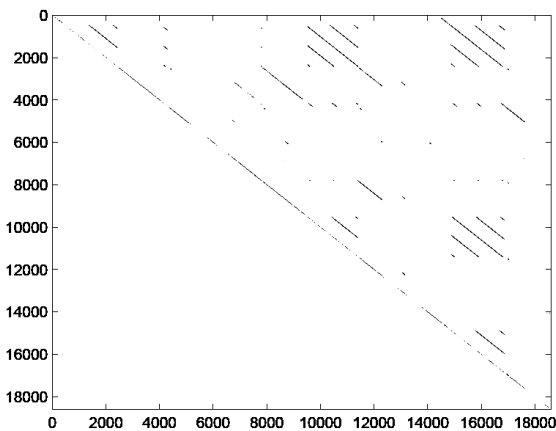


Figure 21 – Distâncias menores que 0.05.

Como estamos interessados em encontrar os *bursting* dos sinais neuronais, foram realizados alguns testes eliminando-se os pontos iniciais da série, e também alterou-se o tamanho

da janela, obtendo dessa forma melhores resultados. Na Figura 23 são destacados em vermelho os novos padrões identificados pela técnica na série, para este caso $N = 1200$ e com o descarte dos primeiros 2500 pontos. Onde observam-se 4 padrões similares de pulsos.

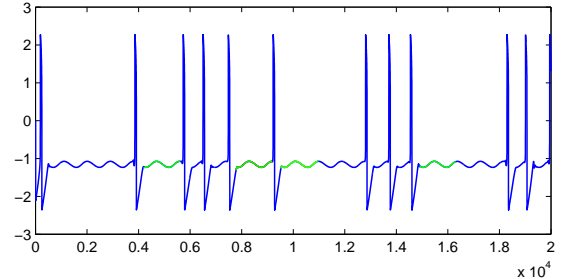


Figure 22 – Padrões identificados pela técnica baseada em representação simbólica.

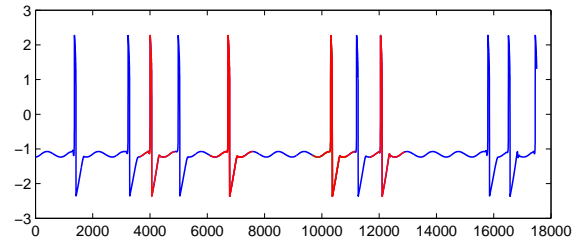


Figure 23 – Padrões encontrados após descartados os primeiros 2500 pontos da série e $N = 1200$.

Na Figura 24 são apresentados os diferentes padrões encontrados ao longo da série mostrada na Figura 10(b), utilizando $N = 1800$, $n = 100$ e $\alpha = 6$. O padrão encontrado no gráfico (b) da Figura 24 não é de interesse, pois o que se busca são padrões como os definidos anteriormente, ver Figura 9(d). Porém, os padrões mostrados em (a) e (c) são semelhanças, mas não o suficiente para a técnica encontrar como um mesmo padrão.

Com a finalidade de melhorar os resultados, novamente foram feitos alguns ajustes: O tamanho da janela foi reduzido para $N = 1600$; e foram descartados os primeiros 1200 pontos da série. Os resultados com essas modificações são apresentados na Figura 25, onde pode-se observar que a técnica identificou 7 ocorrências de um mesmo padrão, mostrados em verde e o retângulo identifica o tipo de padrão encontrado. Neste caso foram encontrados *bursting* de 4 pulsos.

A mesma metodologia foi aplicada para uma série mista, mostrada na Figura 10(c) com $N = 1000$, $n = 100$ e $\alpha = 6$. Os resultados são mostrados na Figura 26, onde em (a) e (b) foram identificados dois padrões similares e em (c) foram encontrados três. Observa-se que não foi possível identificar os *burstings* com 3 pulsos ou com 2 pulsos existentes ao longo da série.

Novamente, foram feitos ajustes para melhorar a identificação dos padrões, nesse caso, aumentou-se o tamanho da janela para $N = 1300$ e foram descartados os primeiros 1000 pontos. Os resultados são apresentados na Figura 27, onde

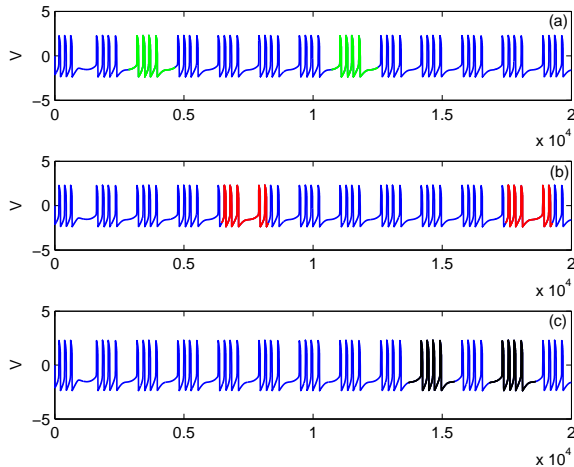


Figure 24 – Diferentes padrões identificados ao longo da série, para $N = 1800$.

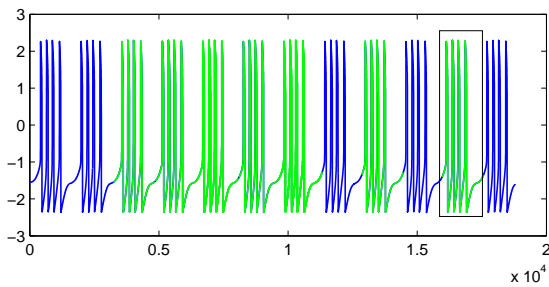


Figure 25 – Diferentes padrões identificados para $N = 1600$, e descartando os primeiros pontos da série.

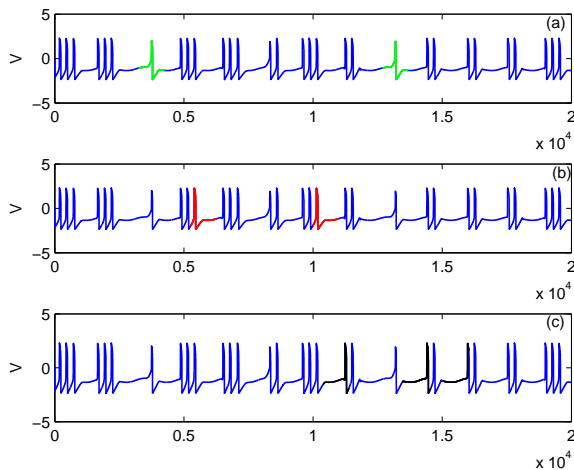


Figure 26 – Diferentes padrões identificados ao longo da série, para $N = 1000$.

observa-se a identificação de 4 ocorrências de *burstings* com 2 pulsos.

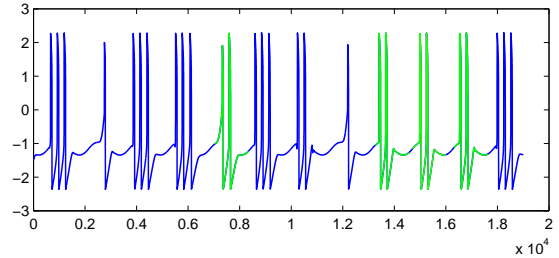


Figure 27 – Diferentes padrões identificados ao longo da série, para $N = 1300$ e com o descarte dos primeiros 1000 pontos.

5. CONSIDERAÇÕES FINAIS

Neste estudo apresentou-se duas metodologias para encontrar padrões em séries temporais. O casamento de padrões, aplicado de forma supervisionada mostrou-se eficaz na identificação e classificação dos padrões. Tendo em vista que, os padrões bem definidos nas séries, provenientes do modelo de FitzHugh-Nagumo, podem ter auxiliado no bom resultado desta metodologia. E a partir da análise das 300 amostras de sinais neuronais pôde-se identificar, por exemplo, quais os melhores parâmetros que se adequam ao comportamento de repetidos *burstings* com 3 pulsos. Pode-se concluir também que a alteração na frequência do piscar de uma lampada, por exemplo, não influencia tanto quanto sua amplitude para desencadear comportamentos de *bursting* no neurônio. A técnica não supervisionada baseada em representação simbólica encontrou boa parte dos padrões similares nas séries analisadas. Esta metodologia destaca-se na aplicação em séries que não apresentam a primeira vista padrões claros e repetitivos, e também nos casos em que são necessárias análises de grandes bancos de dados, com séries muito longas, tendo em vista que a redução de dimensão reduz o custo computacional. Pode-se observar que melhores resultados foram obtidos com uma escolha adequada da janela de busca (N).

ACKNOWLEDGMENTS

Ao Instituto Nacional de Pesquisa Espacial - INPE pelo suporte técnico e a Fundação de Amparo à Pesquisa do Estado de São Paulo - Fapesp pelo apoio financeiro.

REFERÊNCIAS

- [1] A. Roxin, A. Ledberg, Neurobiological models of two-choice decision making can be reduced to a one dimensional nonlinear diffusion equation, *Computing Biology* 4 (3).
- [2] Y. Belhamadia, A time-dependent adaptive remeshing for electrical waves of the heart, *IEEE Transactions on Biomedical Engineering* 55 (2).
- [3] D. Lebiedz, U. Brandt-Pollmann, Manipulation of self-aggregation patterns and waves in a reaction-diffusion system by optimal boundary control strategies, *Phys. Rev. Lett.* 91 (20) (2003) 208301.

- [4] R. FitzHugh, Impulses and physiological states in theoretical models of nerve membrane, *Biophysical Journal*.
- [5] J. Nagumo, S. Arimoto, S. Yoshizawa, An active pulse transmission line simulating nerve axon, *Proceedings of the Ire*.
- [6] J. Engel, T. A. Pedley, *Epilepsy: a comprehensive textbook*, 2nd Edition, Copyright, Philadelphia, 2008.
- [7] T. L. Bailey, C. Elkan, Fitting a mixture model by expectation maximization to discover motifs in biopolymers, in: *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, PubMed, San Diego, 1994, pp. 28–36.
- [8] T. Oates, M. D. Schmill, P. R. Cohen, A method for clustering the experiences of a mobile robot that accords with human judgements, in: *American Association for Artificial Intelligence*, American Association for Artificial Intelligence, Amherst, 2000, pp. 846–851.
- [9] D. Minnen, T. Starner, I. Essa, C. Isbell, Improving activity discovery with automatic neighborhood estimation, in: *International Joint Conference on Artificial Intelligence*, International Joint Conference on Artificial Intelligence, Hyderabad, India, 2007, pp. 6–12.
- [10] X. Ge, P. Smyth, Deformable markov model templates for time-series pattern matching, in: *Technical Report UCI-ICS*, AAAI Press, New York, 2000, pp. 81–90.
- [11] Z. Chi, W. Wu, H. Zach, N. G. Hatsopoulos, D. Margoliash, Template-based spike pattern identification with linear convolution and dynamic time warping, *Journal of neurophysiology* 97 (2).
- [12] T. Fu, F. Chung, R. Luk, C. Ng, Stock time series pattern matching: Template-based vs. rule-based approaches, *Eng. Appl. Artif. Intell.* 20 (3) (2007) 347–364.
- [13] D. J. Berndt, J. Clifford, Using dynamic time warping to find patterns in time series, in: *AAAI Technical Report*, copyright, 1994, pp. 359–370.
- [14] K.-P. Chan, A. W.-C. Fu, Efficient time series matching by wavelets, in: *15th International Conference on Data Engineering*, 1999, pp. 126–133.
- [15] S. Haykin, *Redes Neurais Artificiais: Princípios e Prática*, Bookman, Porto Alegre, 2001.
- [16] S. S. Abeysinghe, N. Chuzhanova, M. Krawczak, E. V. Ball, D. N. Cooper, Translocation and gross deletion breakpoints in human inherited disease and cancer i: Nucleotide composition and recombination-associated motifs, *Human Mutation* 22 (3).
- [17] T. Li, H. He, Y. Wang, H. Zheng, G. Skogerbo, R. Chen, In vivo analysis of *caenorhabditis elegans* noncoding rna promoter motifs, *BMC Molecular Biology* 9 (1) (2008) 71.
- [18] J. Lin, E. Keogh, S. Lonardi, P. Patel, Finding motifs in time series, *Proceedings of the 2nd Workshop on Temporal Data Mining*, at the 8th ACM SIGKDD Int Conference on Knowledge Discovery and Data Mining.
- [19] B. Chiu, E. Keogh, S. Lonardi, Probabilistic discovery of time series motifs, *Sigkdd* 1 (2003) 24–27.
- [20] E. R. Kandel, J. H. Schwartz, T. M. Jessell, *Principles of Neural Science*, 4th Edition, Copyright, New York, 2000.
- [21] A. L. Hodgkin, A. F. Huxley, A quantitative description of membrane current and its application to conduction and excitation in nerve, *The Journal of Physiology*.
- [22] C. Doss-Bachelet, J. P. Francoise, Bursting oscillations in two coupled fitzhugh-nagumo systems, *Complexus*.
- [23] S. Theodoridis, *Pattern Recognition*, 2nd Edition, Academic Press, USA, 2003.
- [24] E. Keogh, K. Chakrabarti, S. Mehrotra, M. Pazzani, Locally adaptive dimensionality reduction for indexing large time series databases, in: *Proceedings of ACM SIGMOD Conference on Management of Data*, ACM SIGMOD, Santa Barbara, 2001, pp. 151–162.
- [25] J. Lin, E. Keogh, S. Lonardi, B. Chiu, 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*, ACM Press, Riverside, 2003, pp. 2–11.
- [26] E. Keogh, K. Chakrabarti, M. Pazzani, S. Mehrotra, Dimensionality reduction for fast similarity search in large time series databases, *Journal of Knowledge and Information Systems* 3 (3) (2001) 263–286.
- [27] A. Inselberg, B. Dimsdale, Parallel coordinates: a tool for visualizing multi-dimensional geometry, in: *VIS '90: Proceedings of the 1st conference on Visualization*, IEEE Computer Society Press, Los Alamitos, 1990, pp. 361–378.