

Aplicação de Métodos de Séries Temporais para a Identificação de Seções em Exames de Manometria Anorretal *

Carlos Andres Ferrero^{1,2}, Huei Diana Lee¹,
Maria Carolina Monard², Wu Feng Chung^{1,3}, Cláudio Sady Rodrigues Coy³,
João José Fagundes³, Juvenal Ricardo Navarro Góes³

¹Centro de Engenharias e Ciências Exatas – Universidade Estadual do Oeste do Paraná
Laboratório de Bioinformática – LABI
Parque Tecnológico Itaipu – PTI
Caixa Postal 39, 85856-970 – Foz do Iguaçu, PR, Brasil

²Instituto de Ciências Matemáticas e de Computação – Universidade de São Paulo
Laboratório de Inteligência Computacional – LABIC
Caixa Postal 668, 13560-970 – São Carlos, SP, Brasil

³Faculdade de Ciências Médicas – Universidade Estadual de Campinas
Serviço de Coloproctologia
Caixa Postal 6111, 13083-970 – Campinas, SP, Brasil

{anfer86, hueidianalee}@gmail.com

Abstract. *Time-series databases store sequences of values that change with time, such as data collected in the area of medicine regarding patients' exams. In this case, data mining techniques can be used to find the characteristics and trend of changes for the evolution of timestamps of such exams. This work presents a methodology to identify sections of anorectal manometry exams data, which is provided by several time series. The methodology uses time series analysis and artificial intelligence methods in order to detect the most relevant sections of this kind of exams. We consider that the results of such analysis could facilitate the decision making process of the domain specialist.*

Resumo. *O avanço da tecnologia tem permitido o armazenamento de informações ao longo do tempo em de bases de dados temporais, como na área médica em que são registrados exames de pacientes. Nesse sentido, técnicas de mineração de dados podem ser aplicadas para analisar características e tendências desses dados. Neste trabalho, é apresentada uma metodologia para identificar seções em exames de manometria anorretal, a partir de dados temporais coletados durante esse exame. A metodologia utiliza métodos de análise de séries temporais e de inteligência artificial e permitirá selecionar os dados interessantes dos exames para, posteriormente, aplicar métodos para extração de padrões, o que poderia auxiliar especialistas da área na tomada de decisões.*

1. Introdução

Sistemas computacionais para gerenciamento de dados permitem cada vez mais o armazenamento de informações de diversas área de conhecimento. Esse acúmulo de informações

*Trabalho desenvolvido com o apoio da Fundação Parque Tecnológico Itaipu – FPTI

faz com que haja necessidade de utilizar métodos computacionais que permitam organizá-los e analisá-los com o objetivo de extrair informações adicionais que permitam, por exemplo, auxiliar especialistas em processos de tomada de decisões. Na área médica, essa análise é de grande importância, pois os exames registrados nas diversas especialidades descrevem informações a respeito do estado de saúde dos pacientes.

A incontinência fecal é a perda da habilidade e da capacidade de controlar a passagem de fezes ou gases em tempo e lugar adequados e socialmente aceitáveis. Essa doença, que leva com o paciente uma problemática social muito preocupante, tem uma incidência de 2% na população em geral, sendo 66% desses casos mulheres, e encontra-se presente em 17% das pessoas com idade superior aos setenta anos [Saad et al. 2002]. O exame de Manometria Anorretal — MA — é imprescindível para o diagnóstico desse tipo de enfermidade, sendo um dos testes fisiológicos mais utilizados e estudados. A MA, assim como outros exames médicos, como o eletrocardiograma, tem a particularidade de armazenar dados ao longo do tempo, isto é, a característica mensurada pelo exame sobre o estado de saúde do paciente é medida no decorrer do tempo do exame. Nesse sentido, esses exames podem ser entendidos computacionalmente como Séries Temporais — ST —, as quais consistem em qualquer amostra de dados (acontecimentos) que ocorram no tempo. Desse modo, métodos matemáticos e computacionais podem ser aplicados sobre essas séries para extrair informações importantes, que permitam auxiliar especialistas, por exemplo, no diagnóstico de doenças.

Neste trabalho em andamento é apresentada uma metodologia para a identificação das seções de exames de MA, utilizando métodos de análise de séries temporais, com o intuito de, posteriormente, estudar detalhadamente cada uma das seções do exame. Este trabalho constitui parte do Projeto de Análise Inteligente de Dados, desenvolvido em uma parceria entre o Laboratório de Bioinformática — LABI — da Universidade Estadual do Oeste do Paraná — UNIOESTE — juntamente com o Laboratório de Inteligência Computacional — LABIC — da Universidade de São Paulo — USP / São Carlos e o Serviço de Coloproctologia da Faculdade de Ciências Médicas da Universidade Estadual de Campinas — UNICAMP.

Paralelamente, este estudo relacionado a ST poderá contribuir no desenvolvimento do Projeto de Análise Inteligente de Dados do Centro de Estudos Avançados em Segurança de Barragens — CEASB —, desenvolvido em uma parceria entre a Hidrelétrica Itaipu Binacional e o LABI, juntamente com o LABIC, auxiliando na aplicação e desenvolvimento de técnicas para a análise dos dados temporais referentes à segurança de barragens.

Este trabalho está organizado do seguinte modo. Na Seção 2 é descrita a metodologia proposta; na Seção 3 são apresentadas as discussões do trabalho e na Seção 4 as conclusões e os trabalhos futuros.

2. Metodologia Proposta

O exame de MA consiste em um cateter axial, o qual é posicionado no esfíncter anal externo do paciente com o intuito de medir a pressão anal de contração do paciente. Geralmente, esse cateter é composto por oito canais dispostos circularmente. Esses sensores permitem quantificar a pressão do esfíncter anal externo, a uma frequência de oito vezes por segundo. Nesse sentido, os exames de MA podem ser representados computacional-

mente por um conjunto S de cardinalidade oito, $S = \{C_1, C_2, \dots, C_8\}$, em que cada C_i representa a série temporal coletada por cada canal que compõe o cateter axial. Como os dados são coletados pelos oito canais simultaneamente, as séries temporais de S têm o mesmo tamanho. Cada ST de S é definida como $C_i = \{x_1, x_2, \dots, x_d\}$, em que x_j , é valor numérico que representa a pressão do canal anal no instante de tempo j e, d , o número de amostras coletadas pelo sensor. Um exemplo de exame de MA de paciente normal é ilustrado na Figura 1, a qual apresenta um gráfico para a visualização da série temporal correspondente a um dos oito canais de coleta de dados.

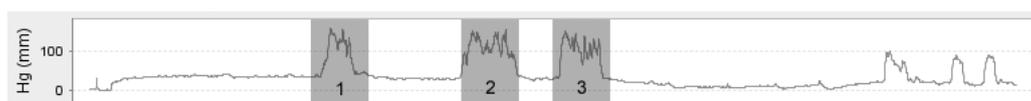


Figura 1. Exemplo de exame de MA de paciente normal.

Na Figura 2, é apresentado um exemplo de exame de paciente com incontinência do tipo grau III.

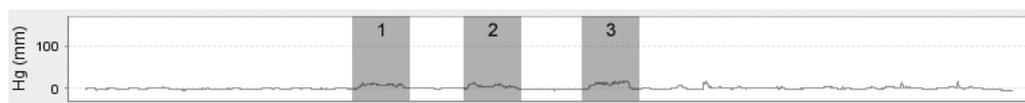


Figura 2. Exemplo de exame de MA de paciente com incontinência do tipo grau III.

Em ambas figuras foram indicadas, juntamente com especialistas da área, cada seção do exame de MA. Pode ser evidenciada a existência de três seções do exame, nas quais o especialista solicita ao paciente contrair voluntariamente o esfíncter anal para, desse modo, quantificar a capacidade do paciente para reter fezes e/ou gases [Saad 2002]. É importante ressaltar que, conforme apresentado nas Figuras 1 e 2, o início e o fim de cada seção encontram-se melhor definidos em exames de pacientes normais do que em pacientes com algum tipo de enfermidade.

O objetivo deste trabalho consiste em identificar computacionalmente o início e o fim de cada uma dessas seções, com o intuito de automatizar a seleção de intervalos de tempo interessantes para o estudo detalhado dessas seções. A metodologia proposta tem como objetivo a aplicação de métodos de análise de séries temporais para a identificação de seções contidas nos exames de MA. A metodologia está constituída de três fases: definição de características, extração sequencial de características no tempo e análise e seleção de intervalos de interesse. Essas etapas são descritas a seguir.

2.1. Fase 1: Definição de Características

No contexto de análise de séries temporais, existem diversos métodos que podem ser aplicados para transformar a ST em outras formas de representação, que permitam evidenciar mais facilmente informações presentes nas séries que poderiam ser interessantes no contexto do problema. A seguir, são apresentadas duas abordagens desses métodos: modelos matemáticos e discretização de séries temporais.

2.1.1. Modelos Matemáticos

Os modelos matemáticos permitem reduzir a dimensionalidade das séries para que sejam representadas por um conjunto reduzido de parâmetros ou pontos, de modo a auxiliar no entendimento dos fenômenos presentes nas ST, tais como a transformada de Fourier, de *wavelets* e fractais, descritas brevemente a seguir.

Transformada de Fourier

A análise de Fourier é uma das maneiras mais tradicionais de modelagem matemática de ST, e consiste na representação através de uma combinação linear de senos e cossenos, que por sua vez podem ser representados pelos seus coeficientes. Assim, dada uma ST, ela pode ser descrita por meio desse conjunto de coeficientes. Na Figura 3 são apresentadas, graficamente, as componentes que constituem a ST $f_1(t)$, em que t representa um instante de tempo. Os gráficos rotulados por *a*, *b* e *c* representam as componentes que conformam $f_1(t)$ e, o gráfico rotulado por *d*, representa a combinação linear dessas três componentes, ou seja, a própria função $f_1(t)$.

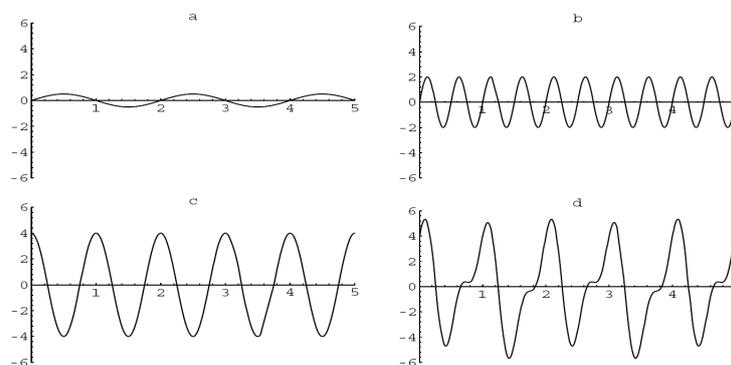


Figura 3. Representação da função $f_1(t)$ (*d*) e de suas componentes 1(*a*), 2(*b*) e 3(*c*) [Shatkay 1995].

Como mencionado, cada uma das componentes é representada por uma função seno ou cosseno, considerando uma frequência e uma amplitude para cada uma dessas. A transformada de Fourier permite obter informações a respeito das frequências presentes na série e, assim, obter informações que não podem ser lidas no domínio do tempo.

Transformada de Wavelets

A transformada de *Wavelets* ou de ondaletas, tem se apresentado como uma eficiente alternativa de substituição da transformada de Fourier em várias aplicações da computação gráfica e de processamento de imagens, de fala e de sinais. Essa transformada permite realizar uma análise mais completa da ST que está sendo estudada, considerando os atributos tempo e frequência simultaneamente, sendo possível obter informações como quais frequências ocorrem em um determinado intervalo de tempo [Chan and Fu 1999]. Na Figura 4 é ilustrado um exemplo de uma transformada de ondaleta de um sinal.

Fractais

Os Fractais constituem uma das formas de representação mais utilizadas para detecção de fenômenos que transcendem no tempo, e consistem na quantificação da propriedade de auto-similaridade de um conjunto de pontos, isto é, a similaridade, parcial ou

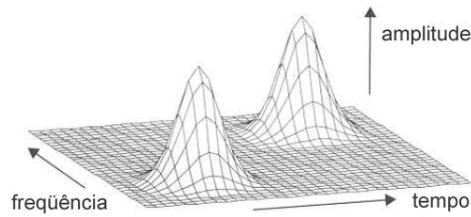


Figura 4. Exemplo de transformada *wavelets* de um sinal.

integralmente, das características para diferentes escalas em que esse conjunto de pontos pode ser analisado. Um dos métodos utilizados para medir a dimensão fractal a partir de um conjunto de pontos é o método de Katz [Esteller et al. 1999]. A fórmula correspondente ao cálculo da dimensão fractal, segundo esse método, é definida pela Equação 1.

$$D = \frac{\log_{10}(M)}{\log_{10}(e)} \quad (1)$$

onde M representa a soma das distâncias entre os pontos sucessivos e e representa a distância entre o primeiro ponto e o ponto mais distante. Esse método compara o número de unidades que compõem a curva com o número mínimo necessário para representar o padrão em um espaço de comprimento igual.

2.1.2. Discretização de Séries Temporais

A discretização de ST permite utilizar um conjunto de símbolos para descrever os eventos através de um conjunto limitado de valores. Esse processo é realizado por meio do mapeamento dos valores expressos em números reais da ST original, para um conjunto de valores discretos. A seguir são apresentadas duas etapas para discretização: redução da dimensionalidade de séries temporais e discretização dos valores que compõem a série temporal.

Redução da Dimensionalidade de Séries Temporais

O método de *Piecewise Aggregate Approximation* — PAA — permite representar uma série temporal em um conjunto reduzido de dimensões. Essa redução de dimensionalidade consiste na separação da ST em pequenas partes. Em PAA, todas as partes tem o mesmo tamanho, sendo que para cada uma dessas partes é atribuído um valor, que pode corresponder, por exemplo, à média dos valores que compõem a parte [Keogh et al. 2001].

Considere uma série temporal $X = \{x_1, x_2, \dots, x_{N_X}\}$, em que x_j corresponde ao valor da série no instante j e N_X ao número de pontos da série. Define-se como $X' = \{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_{N_{X'}}\}$ a série temporal construída a partir de X , utilizando a técnica PAA, em que \bar{x}_i representa o valor da i -ésima parte, e $N_{X'}$ o número de partes para a qual a série temporal é reduzida. A Equação 2 apresenta o cálculo dos valores da nova representação da ST.

$$\bar{x}_i = \frac{N_X}{N_{X'}} \sum_{j=\frac{N_{X'}}{N_X}(i-1)+1}^{\frac{N_{X'}}{N_X} * i} x_j \quad (2)$$

Em [Keogh et al. 2001] é proposto um método alternativo denominado *Adaptive Piecewise Constant Approximation* — APCA. A redução de dimensionalidade através de APCA é realizada de modo que o tamanho das partes se ajuste de acordo com mudanças acentuadas de comportamento. Na Figura 5 são ilustrados os dois métodos apresentados anteriormente para a redução da dimensionalidade de séries temporais.



Figura 5. Comparação visual dos métodos PAA e APCA [Keogh et al. 2001].

Discretização de Valores em Séries Temporais

Na maioria dos casos, o tipo de valores utilizados para descrever as ST são números reais. A discretização de ST permite utilizar um alfabeto, Σ , para descrever os eventos através de um conjunto limitado de símbolos, por meio do mapeamento dos valores originais da série para um conjunto de valores discretos. A seguir são apresentados os métodos de discretização por particionamento uniforme, entropia máxima e *Symbolic Aggregate Approximation* — SAX.

Particionamento Uniforme

Uma das abordagens para discretização, denominada particionamento uniforme de valores, consiste em, a partir do intervalo entre o limite inferior e o superior da ST, dividir esse intervalo em um conjunto de regiões do mesmo tamanho [Freitas 2002, Rajagopalan and Ray 2006]. Define-se $\hat{X}_u = \{\hat{x}_1, \hat{x}_2, \dots, \hat{x}_{N_{X'}}\}$, como sendo o conjunto criado a partir de X' em que os valores reais de cada dimensão são transformados em valores discretos, isto é, o espaço é discretizado. Por exemplo, um espaço discretizado, de tamanho $|\Sigma_u| = 4$ poderia ser descrito pelo alfabeto $\Sigma_u = \{\alpha, \beta, \gamma, \delta\}$. Na Figura 6(a) é apresentado um exemplo a divisão do intervalo em quatro regiões. Nesse exemplo, a ST original pode ser descrita pela palavra $\hat{X}_u = \{\gamma, \alpha, \alpha, \gamma, \delta, \delta, \gamma, \delta\}$.

Particionamento por Entropia Máxima

Uma outra técnica para a discretização de ST é o particionamento por entropia máxima, que se baseia no conceito de que certos espaços são mais ricos em informação do que outros. O cálculo de entropia máxima é realizado da seguinte maneira: inicialmente todos os pontos de, 1 até N , são organizados em ordem crescente. Desse modo, $\frac{N}{|\Sigma_{me}|}$, em que Σ_{me} representa o alfabeto utilizado para discretizar a ST, indica o número de pontos em cada partição, garantindo que a quantidade de informação em cada partição seja a mesma. Desse modo, define-se $\hat{X}_{me} = \{\hat{x}_1, \hat{x}_2, \dots, \hat{x}_{N_{X'}}\}$, como sendo o conjunto criado a partir de X' em que os valores reais de cada dimensão são transformados em um espaço discretizado de acordo com o método de particionamento por entropia máxima.

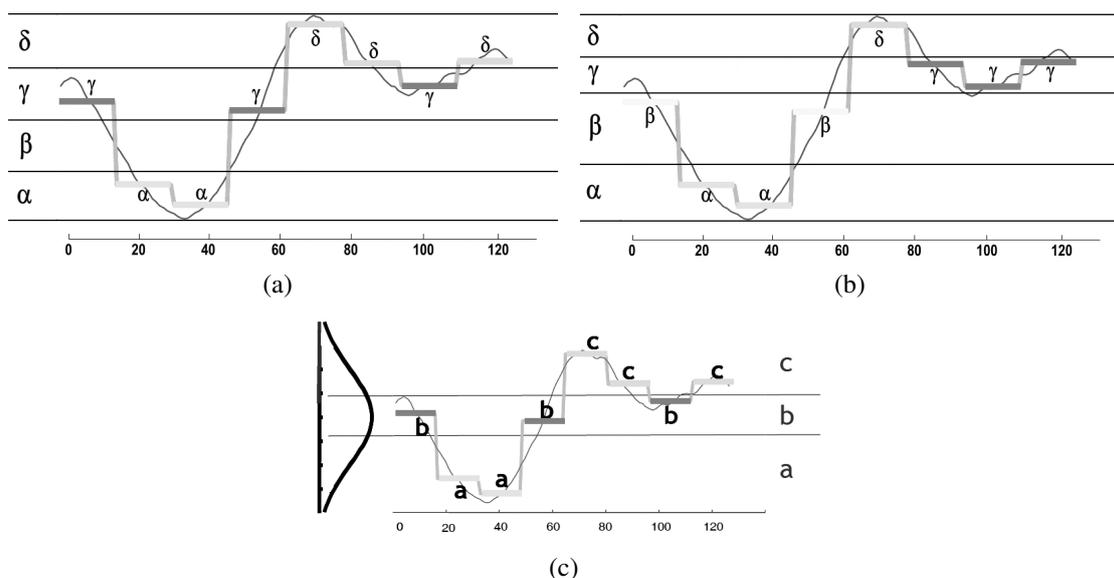


Figura 6. Exemplos de particionamento uniforme (a), de entropia máxima (b), SAX (c) [Rajagopalan and Ray 2006].

Na Figura 6(b) é ilustrado um exemplo da divisão do intervalo em quatro regiões considerando o método apresentado. Nesse exemplo, a ST original pode ser descrita pela palavra $\hat{X}_u = \{\beta, \alpha, \alpha, \beta, \delta, \gamma, \gamma, \gamma\}$.

Symbolic Aggregate approXimation

Uma terceira abordagem para a discretização, denominada *Symbolic Aggregate approXimation* — SAX —, foi proposta para a representação de ST também por meio da utilização de um alfabeto [Lin et al. 2007]. Define-se $\hat{X}_{sax} = \{\hat{x}_1, \hat{x}_2, \dots, \hat{x}_{N_{X'}}\}$ como o conjunto construído a partir de X' em que o espaço $N_{X'}$ -dimensional é discretizado pelo alfabeto $\Sigma_{sax} = \{a, b, c\}$ através da análise por distribuição Gaussiana dos dados. Na Figura 6(c) é apresentada a discretização do espaço de valores utilizando a técnica SAX, onde $\hat{X}_{sax} = \{b, a, a, b, c, c, b, c\}$.

2.2. Fase 2: Extração Seqüencial de Características no Tempo

A segunda fase consiste na extração das características, as quais foram definidas na fase 2, ao longo do tempo. Essa tarefa é realizada estabelecendo uma janela de tamanho L que percorre a ST original extraindo características que permitam a análise mas completa dos fenômenos que estão presentes na ST. Na Figura 7 é apresentada uma ST indicando a janela de tamanho L que percorre a série da esquerda para a direita, isto é, na ordem em que os acontecimentos ocorreram.

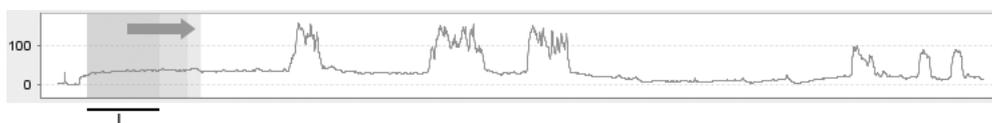


Figura 7. Exemplo de extração da seqüencial de características.

2.3. Fase 3: Análise e Seleção de Intervalos de Interesse

Com base na fase anterior é realizada a análise das características que foram extraídas no tempo para, posteriormente, realizar a seleção dos intervalos de interesse correspondentes às seções de exames de MA. Essa análise possibilitará a identificação dos intervalos de interesse em exames de MA que não contenham seções do exame bem definidas, o que é um problema comum em exames de pacientes que padeçam algum tipo de enfermidade. Entre os métodos de Inteligência Artificial que podem ser aplicados nesse tipo de problemas, os de classificação são comumente utilizados [Wei and Keogh 2006]. Dentre os métodos para classificação mais utilizados têm-se os baseados no método do vizinho mais próximo e baseados em redes neurais [Last et al. 2001, Wei and Keogh 2006].

3. Discussão

A manometria anorretal é um dos exames fisiológicos mais utilizados para o diagnóstico de incontinência fecal. Como mencionado, os dados provenientes desse exame são coletados através do tempo, os quais podem ser representados por séries temporais. A aplicação de métodos para a análise de séries temporais pode auxiliar numa melhor compreensão dos dados, podendo, posteriormente, complementar e auxiliar no processo de tomada de decisões associado a esse tipo de enfermidades.

Como evidenciado pelas Figuras 1 e 2, as seções em exames de pacientes com enfermidade não são tão bem definidas como as seções de paciente normais. Com isso, a utilização de métodos para a análise de séries temporais pode auxiliar a identificar esses intervalos de tempo de interesse. Para isso, neste trabalho, foi proposta uma metodologia, constituída de três fases, para a identificação desses intervalos. Inicialmente, são definidas as características que serão extraídas da série temporal. Posteriormente, é realizada a extração dessas características no tempo e, por fim, é realizada a análise desses dados extraídos e identificados os intervalos de interesse.

A importância da definição das características está relacionada às maneiras de representar uma série temporal, de modo que seja possível obter informações adicionais sobre o início e o fim das seções, a partir dos dados coletados. Desse modo, foram apresentadas duas abordagens de representação: modelos matemáticos e discretização de ST.

A análise de Fourier é uma das técnicas mais utilizadas na representação e tratamento de sinais e permite que informações que não podem ser lidas no domínio do tempo, sejam obtidas no domínio das frequências. As ST estacionárias, ou seja, as que repetem o comportamento em um determinado período de tempo, bem como as não estacionárias podem ser representadas no domínio das frequências. Porém, é desejável que essas últimas sejam estudadas pelo domínio do tempo e das frequências, simultaneamente. Isso permite relacionar em quais intervalos de tempo estão presentes determinadas frequências. Nesse caso, a transformada de Fourier não é a adequada, sendo necessária a aplicação de outra técnica, como a transformada *wavelets*.

A teoria de *wavelets* auxilia na localização e identificação de acumulações de pequenas ondas, provendo suporte para o entendimento desses fenômenos. Diferentemente da análise de Fourier e da teoria espectral, são baseadas na representação de frequências locais e não nas globais. Entre as vantagens da utilização da transformada de *wavelets* destaca-se a extração de coeficientes com alta representatividade, devido ao alto

grau de correlação com o padrão a ser reconhecido. Existem diversas funções *wavelets* que tem sido propostas, descritas e amplamente difundidas na literatura, as quais diferem substancialmente em diversos aspectos, como a ordem dos filtros de decomposição, os momentos de decaimento, o grau de oscilação, a simetria em relação ao eixo y , a ortogonalidade, a possibilidade de implementação por um algoritmo rápido, entre outras [Argoud et al. 2004]. Conseqüentemente, cada transformada será mais ou menos adequada para um determinado propósito. Em [Argoud et al. 2004] foram utilizadas 47 funções *wavelets* em eletroencefalogramas para a diferenciação de eventos epileptogênicos (relacionados à epilepsia) e eventos de fundo (normais). Esse trabalho constitui um passo para a análise de eletroencefalogramas através de *wavelets*, realizando uma avaliação da aplicabilidade, ou não, dessas funções para esse domínio. Outra abordagem está relacionada à utilização da dimensão fractal das curvas descritas pelos pontos, os quais consistem basicamente em quantificar a propriedade de auto-similaridade de um conjunto de pontos. Essa abordagem tem sido aplicada para a detecção de anomalias em ST [Esteller et al. 1999].

Em relação aos métodos de discretização foram apresentadas duas etapas para realizar a discretização de séries temporais. A primeira etapa consiste na divisão da curva em pequenas partes, com o intuito de diminuir o número de pontos que representa a curva, o que pode facilitar a identificação de padrões comportamentais. Para isso, duas técnicas para a transformação da ST foram apresentadas: PAA e APCA. A primeira consiste em dividir a curva em partes iguais, sem levar em consideração o comportamento local de cada fragmento de curva e, a segunda divide a curva de acordo com as variações que ocorrem no transcurso da série temporal. A segunda etapa, consiste em discretizar o espaço de valores dos pontos da série temporal, isto é, ao invés de utilizar números reais para representar os pontos da ST, o que promove uma grande quantidade de valores possíveis, o espaço de valores é reduzido para um conjunto menor de símbolos, utilizando algum método de particionamento desse espaço. Entre os diversos métodos para particionamento do espaço foram apresentados três métodos: uniforme, por entropia máxima e *Symbolic Aggregate approXimation*. O particionamento uniforme consiste em dividir o espaço de valores de modo uniforme. Entretanto, comumente as informações não se encontram dispostas de maneira uniforme. Para isso, os métodos de particionamento por entropia máxima e SAX permitem dividir o espaço de valores através das análises de entropia e de distribuição Gaussiana, respectivamente.

4. Conclusão

O desenvolvimento de sistemas computacionais que permitam o armazenamento e a organização de informações tem promovido um aumento das bases de dados em todas as áreas do conhecimento. Nesse sentido, diversas áreas de pesquisa têm se interessado em compreender fenômenos que transcendem no tempo. Uma dessas áreas é a medicina, que procura o avanço no entendimento dos fenômenos biológicos, tal como a incontinência fecal, foco deste trabalho. A incontinência fecal é uma doença com grande incidência dentro da população, principalmente em idosos. A manometria anorretal é um exame imprescindível para o diagnóstico desse tipo de enfermidades e os registros desse exame correspondem a dados de pressão anal do paciente, que transcendem no tempo. Desse modo, aplicação de conceitos relacionados à área de séries temporais e inteligência artificial podem auxiliar numa análise mais aprofundada desse tipo de dados.

Neste trabalho foi apresentado um estudo para a análise de exames de manometria anorretal, que tem como primeiro objetivo a identificação dos intervalos de tempo de interesse nos exames. Como trabalhos futuros têm-se a construção de um aplicativo para dar suporte à metodologia proposta neste trabalho e o estudo de novos métodos e técnicas presentes na literatura para a representação e análise de séries temporais.

Referências

- Argoud, F. I. M., de Azevedo, F. M., and Neto, J. M. (2004). Estudo comparativo entre funções wavelets e suas diferentes aplicabilidades em reconhecimento de padrões epileptogênicos em eletroencefalograma. *Revista Brasileira de Engenharia Biomédica*, 20(2-3):49–59.
- Chan, K.-P. and Fu, A. W.-C. (1999). Efficient time series matching by wavelets. In *Data Engineering, 1999. Proceedings 15th International Conference on*, pages 126–133.
- Esteller, R., Vachtsevanos, G., Echauz, J., and Lilt, B. (1999). A comparison of fractal dimension algorithms using synthetic and experimental data. In *Proceedings of the 1999 IEEE International Symposium on Circuits and Systems*, volume 3, pages 199–202.
- Freitas, A. A. (2002). *Data Mining and Knowledge Discovery with Evolutionary Algorithms*. Springer, New York — NY, USA.
- Keogh, E. J., Chakrabarti, K., Mehrotra, S., and Pazzani, M. J. (2001). Locally adaptive dimensionality reduction for indexing large time series databases. In *SIGMOD Conference*.
- Last, M., Klein, Y., and Kandel, A. (2001). Knowledge discovery in time series databases. *Systems, Man and Cybernetics, Part B, IEEE Transactions on*, 31(1):160–169.
- Lin, J., Keogh, E., Li, W., and Lonardi, S. (2007). Experiencing sax: A novel symbolic representation of time series. *Data Mining and Knowledge Discovery Journal*, pages 1–31. A ser publicado.
- Rajagopalan, V. and Ray, A. (2006). Symbolic time series analysis via wavelet-based partitioning. *Signal Process*, 86(11):3309–3320.
- Saad, L. H. C. (2002). *Quantificação da Função Esfincteriana pela Medida da Capacidade de Sustentação da Pressão de Contração Voluntária do Canal Anal*. Tese de Doutorado, Faculdade de Ciências Médicas, Universidade Estadual de Campinas — UNICAMP.
- Saad, L. H. C., Coy, C. S. R., Fagundes, J. a. J., Ariyazono, M. d. L., Shoji, N., and Góes, J. R. N. (2002). Quantificação da função esfincteriana pela medida da capacidade de sustentação da pressão de contração voluntária do canal anal. *Arquivos de Gastroenterologia*, 39:233–239.
- Shatkay, H. (1995). The Fourier transform — a primer. Technical report, Department of Computer Science, Brown University, Providence — RI, USA.
- Wei, L. and Keogh, E. (2006). Semi-supervised time series classification. In *KDD '06: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 748–753, New York, NY, USA. ACM Press.