

# Aplicação de Clustering de Séries Temporais em Exames de Manometria Ano-retal no Estudo da Incontinência Fecal

Carlos Andres Ferrero<sup>1</sup>, André Gustavo Maletzke<sup>1</sup>, Hwei Diana Lee<sup>1</sup>,  
Newton Spolaôr<sup>1,3</sup>, Cláudio Saddy Rodrigues Coy<sup>2</sup>,  
João José Fagundes<sup>2</sup>, Feng Chung Wu<sup>1,2</sup>

<sup>1</sup>*Centro de Engenharias e Ciências Exatas – Universidade Estadual do Oeste do Paraná, Laboratório de Bioinformática – LABI, Parque Tecnológico Itaipu – PTI, Foz do Iguaçu, PR, Brasil*

<sup>2</sup>*Faculdade de Ciências Médicas – Universidade Estadual de Campinas, Serviço de Coloproctologia, Campinas, SP, Brasil*

<sup>3</sup>*Centro de Matemática, Computação e Cognição – Universidade Federal do ABC, Santo André, SP, Brasil*

**Resumo** — Com o crescimento das bases de dados, principalmente devido ao avanço tecnológico, a aplicação de métodos computacionais para a aquisição de conhecimento tornou-se uma tarefa de crescente interesse em diversas áreas do conhecimento, inclusive a área de saúde. Usualmente, exames médicos são armazenados em hospitais e clínicas médicas com o objetivo de complementar o diagnóstico realizado pelos especialistas. No contexto da incontinência fecal, o exame médico de Manometria Ano-retal é fundamental para o diagnóstico desse distúrbio. Nesse exame, são descritos os comportamentos relacionados à função esfinteriana do paciente quando realizada uma contração voluntária ao longo do tempo. Neste trabalho é apresentado um método para a análise desses exames por meio da construção de agrupamentos (clusters). O método foi implementado em uma ferramenta computacional, a qual permite a seleção de diferentes medidas de similaridade, bem como diferentes algoritmos de construção de clusters, entre outros, e foi aplicado a 20 exames de manometria ano-retal. Os agrupamentos construídos foram analisados por especialistas do domínio e por meio de medidas objetivas, apresentando resultados promissores na identificação de padrões nos dados.

**Palavras chave** — inteligência computacional, agrupamento de dados, séries temporais, manometria ano-retal, incontinência fecal.

## 1. Introdução

O desenvolvimento de tecnologias tem outorgado ao ser humano a possibilidade de capturar dados das mais variadas fontes e armazená-los em distintos formatos. Nesse sentido, a coleta de dados de fenômenos que variam ao longo do tempo é uma tarefa que tem despertado interesse em distintas áreas, mas que pode gerar grandes conjuntos de dados, dificultando a análise das informações por meio de processos manuais. Uma estratégia amplamente aplicada é a representação desses dados como Séries Temporais – ST. O estudo de séries temporais provê um conjunto de métodos que podem auxiliar na análise desses dados por meio de tarefas como classificação, previsão, agrupamento, modelagem e descrição [8].

Nesse contexto, cada vez mais, métodos computacionais têm sido propostos e aplicados na análise de grandes conjuntos de dados. Dentre esses processos, o de Mineração de Dados – MD –, apoiado pela área de Aprendizado de Máquina – AM –, tem sido utilizado na tarefa de análise de dados. Esse processo tem como objetivo a extração de padrões contidos

nos dados, de modo que esses padrões constituam uma fonte de informação interessante e relevante para especialistas de diversos domínios [11].

Em hospitais e clínicas médicas, é relevante a quantidade de informações produzidas ao longo do tratamento e acompanhamento de pacientes. Essas informações apresentam-se em diversos formatos, inclusive no domínio temporal, como é o caso de exames de eletrocardiograma, eletroencefalograma e manometria ano-retal, entre outros. Assim, a análise mais completa dessas informações pode auxiliar especialistas em processos de tomada de decisão, por exemplo, no diagnóstico de doenças.

Nesse cenário, as tarefas de MD podem ser classificadas em preditivas e descritivas [14]. As tarefas preditivas consistem em, por exemplo, dado um conjunto de informações referente à sintomatologia de pacientes, ser capaz de prever o estado de saúde desse paciente, isto é, o rótulo ou atributo classe desse paciente. Entretanto, existem casos em que não é possível realizar a tarefa de predição devido ao fato de existir restrições nos dados, como a ausência da variável que se deseja prever. Nesse caso, podem ser aplicadas tarefas descritivas que buscam identificar comportamentos intrínsecos presentes nesses dados. Existem diversas tarefas descritivas, em [20] são apresentadas algumas das principais, dentre as quais a tarefa de agrupamento (*clustering*) que constitui o foco deste trabalho. O agrupamento de dados tem como finalidade construir grupos (*clusters*) a partir de conjuntos de exemplos, de acordo com algum critério, por exemplo, a similaridade entre os exemplos.

Este trabalho tem como objetivo apresentar a aplicação de métodos de agrupamento, implementados em uma ferramenta computacional, para a análise de dados de exames de manometria ano-retal, no intuito de encontrar padrões nos dados, que permitam auxiliar no diagnóstico do distúrbio de incontinência fecal, por meio desse exame médico. O trabalho constitui parte do projeto Análise Inteligente de Dados, mais especificamente de Séries Temporais, que está sendo desenvolvido em uma parceria entre o Laboratório de Bioinformática – LABI – da Universidade Estadual do Oeste do Paraná – UNIOESTE / Foz do Iguaçu –, 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 – e colaboração do Centro de Matemática, Computação e Cognição – CMCC – da Universidade Federal do ABC – UFABC.

O trabalho está organizado da seguinte forma: na Seção 2, são discutidos os materiais utilizados e o método proposto para a construção dos agrupamentos; na Seção 3, são apresentados os resultados do trabalho e realizada a discussão desses resultados; e, na Seção 4, são apresentadas as conclusões e os trabalhos futuros.

## 2. Materiais e Métodos

Como mencionado, a MD é um campo de pesquisa que está sendo amplamente aplicado em distintas áreas, dentre as quais a área de medicina tem recebido destaque. O processo de MD é caracterizado como iterativo e interativo e pode ser dividido em três fases: pré-processamento, construção de padrões ou modelos e pós-processamento. A primeira fase é a que, geralmente, demanda maior custo de tempo e tem como objetivo entender o domínio ao qual o processo será aplicado e preparar os dados para que sejam submetidos à próxima fase. Na segunda fase são selecionados e aplicados um ou mais algoritmos considerando o tipo de tarefa de interesse, isto é, preditiva ou descritiva. Após, os resultados devem ser validados e avaliados na última fase desse processo tanto por medidas objetivas quanto por

reuniões com especialistas do domínio. É importante ressaltar a participação do especialista em cada fase e, principalmente, na primeira e terceira fases, nas quais, respectivamente, o conhecimento de domínio é incorporado ao processo e os resultados são analisados.

Nesse contexto, este trabalho tem como foco a realização de tarefas descritivas, especificamente, tarefas de agrupamento de dados, que têm como objetivo construir agrupamentos de modo que exemplos pertencentes a um mesmo grupo sejam similares entre si, de acordo com algum critério de similaridade, enquanto que os exemplos pertencentes a grupos distintos apresentem menor similaridade.

Diversas definições de agrupamento são encontradas na literatura, dentre essas, a proposta em [4]: “Um agrupamento dos pontos no espaço de teste é definido de modo que a distância entre quaisquer dois pontos em um mesmo grupo é menor que a distância entre qualquer ponto desse grupo e outro ponto não pertencente a esse grupo”.

Os algoritmos de agrupamento podem ser classificados considerando vários aspectos, entre eles o método utilizado para definir os grupos. Nesse sentido, duas classificações amplamente utilizadas são: o agrupamento particional e o agrupamento hierárquico. O agrupamento particional consiste, basicamente, em dividir o conjunto de exemplos em  $k$  partições, de modo iterativo. Nessa abordagem, todos os exemplos são utilizados para determinar os grupos, implicando no cálculo de distância para cada exemplo, tornando a tarefa custosa em relação ao esforço computacional. Esse problema é resolvido elegendo-se um representante do agrupamento, por exemplo, o elemento central do grupo, para realizar o cálculo de similaridade. Por outro lado, o agrupamento hierárquico possui como característica a organização dos grupos por meio de uma estrutura hierárquica. Dentre essas estruturas, a mais frequentemente utilizada é denominada de dendograma, que consiste em um tipo especial de estrutura de árvore [2], a qual descreve agrupamentos diferentes a cada nível da hierarquia determinados por pontos de corte.

Uma das maiores vantagens na utilização de algoritmos de agrupamento hierárquico está relacionada à não necessidade em se determinar, a priori, o número de agrupamentos a ser construído. Isso permite flexibilidade durante a análise de agrupamentos, com maior ou menor granularidade, dependendo do ponto de corte determinado. Neste trabalho, devido a essa vantagem, em conjunto com as características do domínio do problema, foi utilizada a abordagem hierárquica para a qual os principais algoritmos de agrupamento hierárquico presentes na literatura são [4][6]:

- **Single-link:** é o método mais utilizado e possui como principal característica a determinação da distância inter-*clusters* a partir do par de exemplos de maior similaridade;
- **Complete-link:** considera o par de exemplos de menor similaridade para definir a distância inter- *clusters*. Os grupos gerados por esse método, geralmente, tendem a ser mais compactos;
- **Average-link:** esse algoritmo define a similaridade entre os grupos por meio da média de todas as distâncias entre os pares de exemplos. Esse algoritmo pode amenizar a presença de *outliers*, os quais podem afetar negativamente os demais algoritmos apresentados.

Uma vez selecionado o algoritmo de agrupamento a ser aplicado, outro aspecto a ser considerado refere-se à seleção de uma medida de similaridade, isto é, de um critério que

possibilite determinar de maneira objetiva a semelhança entre os exemplos de um conjunto. A medida de similaridade utilizada apresenta forte influência no resultado da aplicação do algoritmo de agrupamento. Em [4] podem ser encontradas algumas das medidas de similaridade propostas na literatura, as quais devem ser selecionadas de acordo com as características dos dados aos quais se pretende aplicar a tarefa de agrupamento.

O agrupamento de dados representados por meio de séries temporais é uma tarefa de interesse e, conforme [5][7], pode ser classificado em duas abordagens:

- **Whole clustering:** nessa abordagem a tarefa de agrupamento é realizada considerando a ST em sua totalidade, isto é, todas as observações da série;
- **Subsequence clustering:** nessa abordagem, são extraídas subsequências de cada ST, de modo que somente as subsequências extraídas sejam fornecidas ao algoritmo de agrupamento.

Desse modo, os resultados da aplicação da tarefa de agrupamento sobre séries temporais são fortemente dependentes da escolha do método de agrupamento, medida de similaridade e da abordagem selecionada para aplicação em dados de séries temporais.

### *2.1 Descrição dos Dados*

A incontinência fecal consiste em um distúrbio associado à incapacidade de controle de fezes ou gases em locais socialmente aceitáveis que atinge uma incidência na população mundial de 0,5% a 5% e apresenta, de acordo com alguns estudos, uma ocorrência maior em idosos e em mulheres submetidas ao parto normal [10][15][16]. Uma classificação proposta na literatura identifica três níveis de intensidade do distúrbio [19], descritos a seguir:

- Grau I: incontinência somente para gases;
- Grau II: incontinência para gases e fezes líquidas;
- Grau III: incontinência para fezes sólidas.

Nesse contexto, a Manometria Ano-retal – MA – constitui um dos exames complementares mais aplicados para a avaliação das funções fisiológicas de continência e evacuação, bem como para o diagnóstico e o acompanhamento dessa problemática [9]. Nesse exame, são descritas informações referentes à história clínica do paciente e à sua capacidade de contração voluntária. Um cateter axial é posicionado no esfíncter anal externo do paciente com o intuito de medir a Pressão de Contração Voluntária – PCV [15]. Geralmente, esse cateter é composto por oito canais dispostos circularmente, cujos sensores permitem quantificar a pressão do esfíncter anal externo, em milímetros de mercúrio, exercida sobre o cateter, em uma determinada frequência.

Em [15] foi proposto um protocolo para a coleta de medições de PCV com a finalidade de reduzir as variáveis envolvidas no processo de coleta das informações, as quais podem influenciar negativamente na precisão do exame de manometria ano-retal na detecção de pacientes com incontinência. Esse protocolo consiste na coleta de informações durante três períodos de contração voluntária ano-retal, sendo cada período de aproximadamente 40 segundos, intercalando com períodos de repouso. Os períodos que apresentam elevados valores de PCV podem ser significativos para a análise da fisiologia ano-retal.

O conjunto de dados utilizado neste trabalho consiste em 20 exames de MA, elaborados entre Maio de 1995 e Novembro de 1996 pelo Serviço de Coloproctologia da UNICAMP, a partir dos quais foram extraídos os valores de PCV, coletados a uma frequência de oito observações por segundo, pelos oito sensores de um cateter associado ao equipamento pneumo-hidráulico de perfusão capilar *Arndorfer Medical Specialties* e com o auxílio do software *Polygram Power GI* versão 6.4. A história clínica de cada exame foi analisada com o auxílio de especialistas do domínio para identificar os pacientes em condições normais e com graus de incontinência. Para avaliar a aplicação de técnicas computacionais no domínio, neste trabalho inicial, foram considerados somente exames referentes à pacientes com incontinência em Grau III (oito pacientes) e normais (doze pacientes). A consideração desses extremos pode auxiliar a identificar a viabilidade de aplicação técnicas.

## 2.2 Método para Construção de Agrupamentos

O método proposto neste trabalho para a construção de agrupamentos é composto de três etapas, ilustradas na Figura 1 e descritas a seguir:

1. Preparação dos dados;
2. Seleção da medida de similaridade;
3. Aplicação do algoritmo de agrupamento.

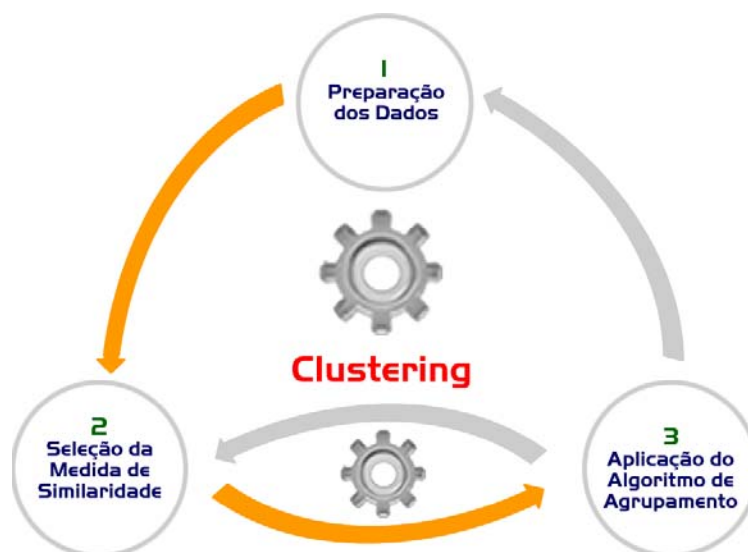


Figura 1. Etapas do método para construção de agrupamentos

Na primeira etapa, os dados temporais são preparados e transformados em um formato adequado para a utilização de métodos de agrupamento. Esse formato deve ser o atributo-valor, no qual, as linhas representam os exemplos<sup>1</sup> e as colunas representam as observações da série temporal. Como mencionado, os exames de manometria ano-retal consistem na coleta de valores de pressão por oito sensores ao longo do tempo. Assim, cada exame é definido por  $S = (S_1, S_2, \dots, S_8)$ , em que  $S_i$ , corresponde à série temporal gerada pelo

<sup>1</sup> Neste trabalho, uma série temporal é considerada um exemplo.

$i$ -ésimo sensor  $S_i = (s_{i,1}, s_{i,2}, \dots, s_{i,n})$ , em que  $s_{i,j}$ , corresponde ao valor de pressão coletado pelo sensor  $i$  no instante  $j$ . Neste trabalho, é considerada uma resultante dos valores de pressão captados pelos oito sensores a cada instante de coleta, a qual consiste na soma dos valores coletados pelos sensores. Assim, cada exame é representado por uma única série temporal, resultante da combinação das séries  $S_1, S_2, \dots, S_8$ , definida por  $X = (x_1, x_2, \dots, x_n)$ , em que  $x_j$ , corresponde à soma dos valores coletados pelos oito sensores no instante  $j$ , ou seja,  $x_j = s_{1,j} + s_{2,j} + \dots + s_{8,j}$ .

Ainda nessa etapa, é construída a tabela atributo-valor que representa os dados a serem usados no restante do método no formato adequado para a análise. Na Tabela 1 é apresentada a representação atributo-valor, em que cada linha refere-se à série temporal resultante da soma das oito séries temporais de um exame, conforme definido anteriormente, e as colunas representam os valores de pressão de cada série temporal resultante.

Tabela 1. Representação atributo-valor dos dados a serem analisados<sup>2</sup>

$X_i$	$X_{i,1}$	$X_{i,2}$	...	$X_{i,nc}$
$X_1$	$X_{1,1}$	$X_{1,2}$	...	$X_{1,nc}$
$X_2$	$X_{2,1}$	$X_{2,2}$	...	$X_{2,nc}$
...	...	...	...	...
$X_{nl}$	$X_{nl,1}$	$X_{nl,2}$	...	$X_{nl,nc}$

A segunda etapa consiste na seleção da medida de similaridade para quantificar a semelhança entre os exemplos contidos no conjunto de dados. Nessa etapa é construída uma matriz de distância  $M$  de ordem  $nl$ , em que  $nl$ , corresponde ao número de exemplos do conjunto de dados. A matriz descreve a distância entre os exemplos do conjunto dois a dois de acordo com a medida selecionada. Dentre as medidas de similaridade mais comumente utilizadas têm-se as baseadas na medida de Minkowsky ou normas  $L_p$ , que incluem a distância Manhattan, para  $p=1$ , e a distância Euclidiana, para  $p=2$ . Uma discussão mais aprofundada dessas medidas pode ser encontrada em [1]. Outra medida que tem sido utilizada para a comparação de dados sequenciais consiste na medida *Dynamic Time Warping – DTW*, a qual consiste em calcular a distância para vários alinhamentos, por meio de alguma medida convencional de similaridade, com a finalidade de encontrar o melhor alinhamento entre as sequências [3].

Na terceira etapa, são aplicados os métodos de agrupamento a partir das matrizes de distância geradas na etapa anterior. Os agrupamentos construídos podem ser avaliados por medidas objetivas, por exemplo, pelo Coeficiente de Qualidade Geral – CQG [13][21], que avalia a qualidade de agrupamentos a partir das relações entre as distâncias inter-*clusters* (entre os agrupamentos) e intra-*clusters* (entre os exemplos dentro de cada agrupamento). O valor desse coeficiente é estimado a partir da diferença entre a menor distância inter-*cluster* e a maior distância intra-*cluster*. Assim, agrupamentos com alto valor de distância inter-*clusters* e baixo valor de distância intra-*clusters*, são descritos por valores baixos de CQG. É importante ressaltar que a avaliação juntamente com o especialista é fundamental para o entendimento, mais completo, dos padrões representados por esses agrupamentos.

<sup>2</sup> Nessa tabela,  $nl$  e  $nc$ , definem o número de exemplos e o tamanho das séries temporais consideradas, respectivamente.

As relações entre as etapas – Figura 1 – ilustram a possibilidade de retornar para as etapas anteriores, no intuito de utilizar configurações experimentais que permitam obter melhores resultados.

### 2.3 Aplicativo Computacional TimeSSys

O método para a construção de agrupamentos apresentado anteriormente foi implementado por meio um módulo computacional, denominado *CluST*, que permite a definição dos parâmetros associados à medida de similaridade, ao algoritmo de agrupamento e à medida de avaliação objetiva desses agrupamentos. Esse módulo está inserido dentro de um aplicativo maior denominado *Time Series System – TimeSSys*. O objetivo desse sistema é disponibilizar, em um mesmo *framework*, conjuntos de ferramentas que auxiliem nas diversas etapas do processo de análise de séries temporais. A ideia desse sistema é favorecer a integração de ferramentas de código livre implementadas, bem como de ferramentas desenvolvidas pelos pesquisadores parceiros, em um mesmo *framework*, disponibilizando funcionalidades para agrupamento, visualização, pré-processamento, previsão, recuperação de conteúdo, entre outras, de series temporais.

Como mencionado, o desenvolvimento do sistema *TimeSSys* é baseado em tecnologias e ferramentas livres. A camada de negócio está sendo implementada na linguagem *R* [12], a qual disponibiliza uma grande diversidade de funcionalidades relacionadas à análise estatística, à análise de séries temporais e à visualização. Para compartilhar as ferramentas integradas no *TimeSSys*, é utilizada a tecnologia *Rpad*<sup>3</sup>, baseada no paradigma cliente-servidor, a qual disponibiliza um ambiente simples e flexível para o desenvolvimento de interfaces WEB, além de estar diretamente associada com a linguagem *R*, o que facilita a integração das ferramentas desenvolvidas e a realização e o compartilhamento de experimentos científicos.

Todos os experimentos realizados neste trabalho foram executados utilizando o módulo implementado. Na Figura 2 é apresentada a interface gráfica do módulo, na qual é possível observar alguns dos parâmetros envolvidos no processo de construção de agrupamentos.

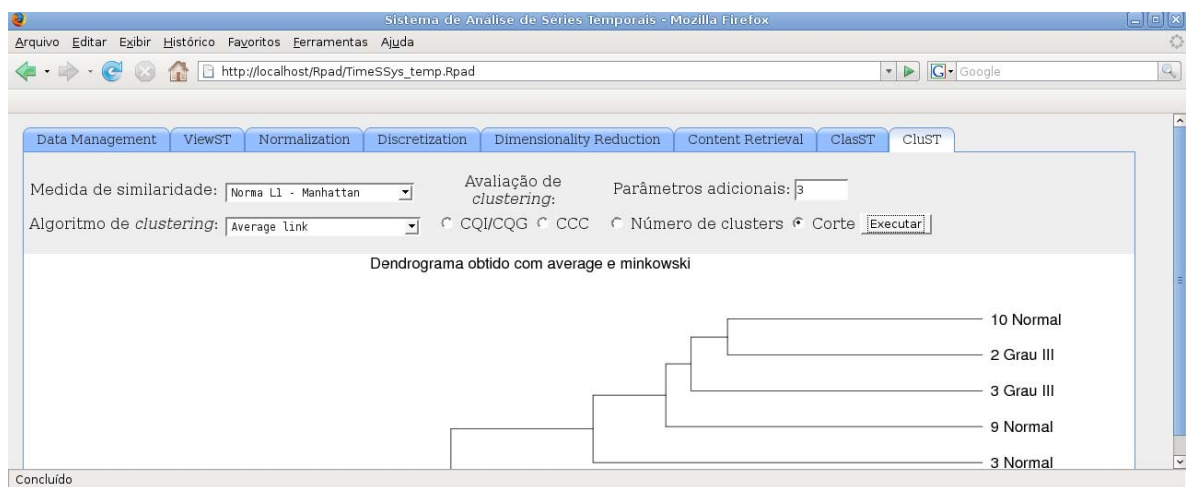


Figura 2. Interface gráfica do módulo *CluST*

<sup>3</sup> <http://www.rpad.org/Rpad/>.

### 3. Resultados e Discussão

Como mencionado, o método proposto e apresentado anteriormente foi aplicado ao conjunto de 20 exames de manometria ano-retal. Os experimentos foram realizados considerando quatro medidas de similaridade e três algoritmos de agrupamento, totalizando doze configurações experimentais. Os agrupamentos construídos foram avaliados objetivamente e de acordo com o parecer dos especialistas. Neste trabalho, é apresentada uma parte desses resultados e a descrição completa pode ser encontrada em [18].

#### 3.1 Preparação dos dados

Inicialmente, em cada exame foram delimitadas, juntamente com especialistas, três seções com valores elevados de PCV, cada uma com o intervalo de tempo de 40 segundos. Na Figura 3 é ilustrado do procedimento utilizado para a identificação das seções do exame.

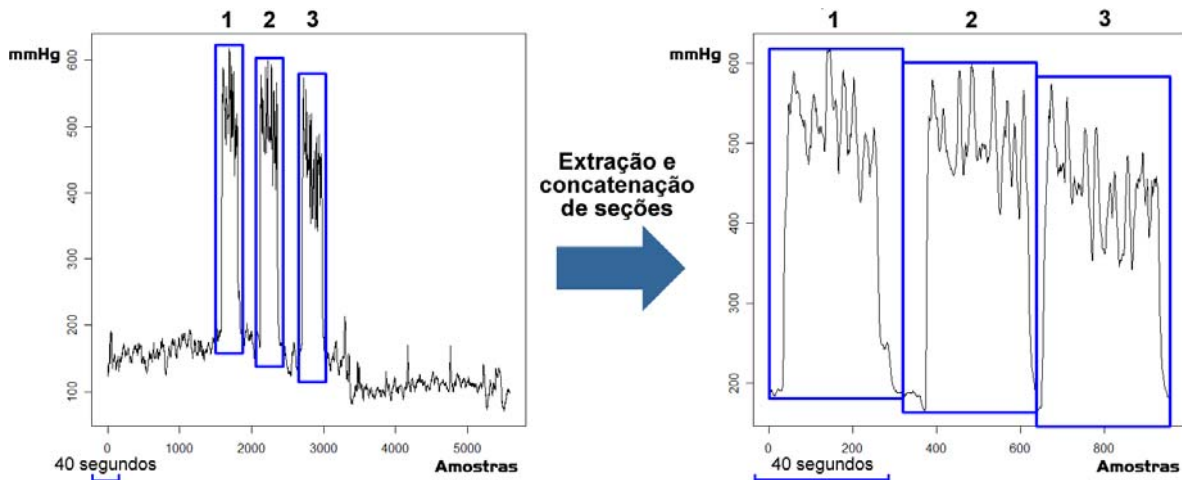


Figura 3. Extração e concatenação de seções das seções do exame [17]

Considerando a frequência de coleta do instrumento de oito observações por segundo e a utilização de três seções do exame de 40 segundos cada, as séries temporais  $X_1, X_2, \dots, X_{20}$ , que representam os exames, são compostas por 960 observações cada uma.

#### 3.2 Seleção da medida de similaridade

Nessa etapa, foram utilizadas quatro medidas de similaridade para a construção das matrizes de distância entre os exemplos. Três dessas medidas correspondem às baseadas nas normas  $L_p$ , para  $p = 1, 2$  e  $3$ . Em [1] é discutida a adequabilidade das medidas de Minkowsky para conjuntos de dados com grandes dimensões, observando que valores de  $p$  menores podem apresentar melhores resultados. A quarta medida corresponde à medida DTW, a qual tem se apresentado promissora para a análise de séries temporais. Essa medida apresenta maior robustez em relação à comparação de séries temporais defasadas no eixo temporal ou que apresentam tamanhos distintos [3].

#### 3.3 Aplicação do algoritmo de agrupamento

Na terceira etapa foram aplicados os métodos hierárquicos *Average-link*, *Complete-link* e *Single-link* para a construção dos agrupamentos a partir das matrizes de distância geradas



na etapa anterior. A aplicação desses três algoritmos associados a cada uma das quatro medidas de similaridade permitiu a construção de doze dendogramas.

Os agrupamentos foram analisados inicialmente de modo objetivo, usando o coeficiente de qualidade geral. A configuração experimental que apresentou menor valor de CQG foi referente à medida de similaridade DTW usando o algoritmo de agrupamento *Complete-link*. Desse modo, os agrupamentos referentes a essa configuração foram analisados juntamente com o especialista do domínio.

Deve ser observado que os agrupamentos foram construídos sem considerar a classe (em condições normais e Grau III) de cada exame. Entretanto, no intuito de encontrar correlações entre os agrupamentos gerados computacionalmente a partir dos exames e o diagnóstico, a classe a que cada exame corresponde foi utilizada na avaliação com os especialistas. Na Figura 4 é ilustrado o dendograma gerado pelo aplicativo *TimeSSys*. Nessa figura é possível observar o agrupamento das séries temporais, em quatro grupos 1, 2, 3 e 4 representados pelas cores azul, verde, amarelo e cinza, correspondentes aos oito pacientes com incontinência em Grau III e aos doze pacientes em condições normais, à medida que varia o valor de similaridade.

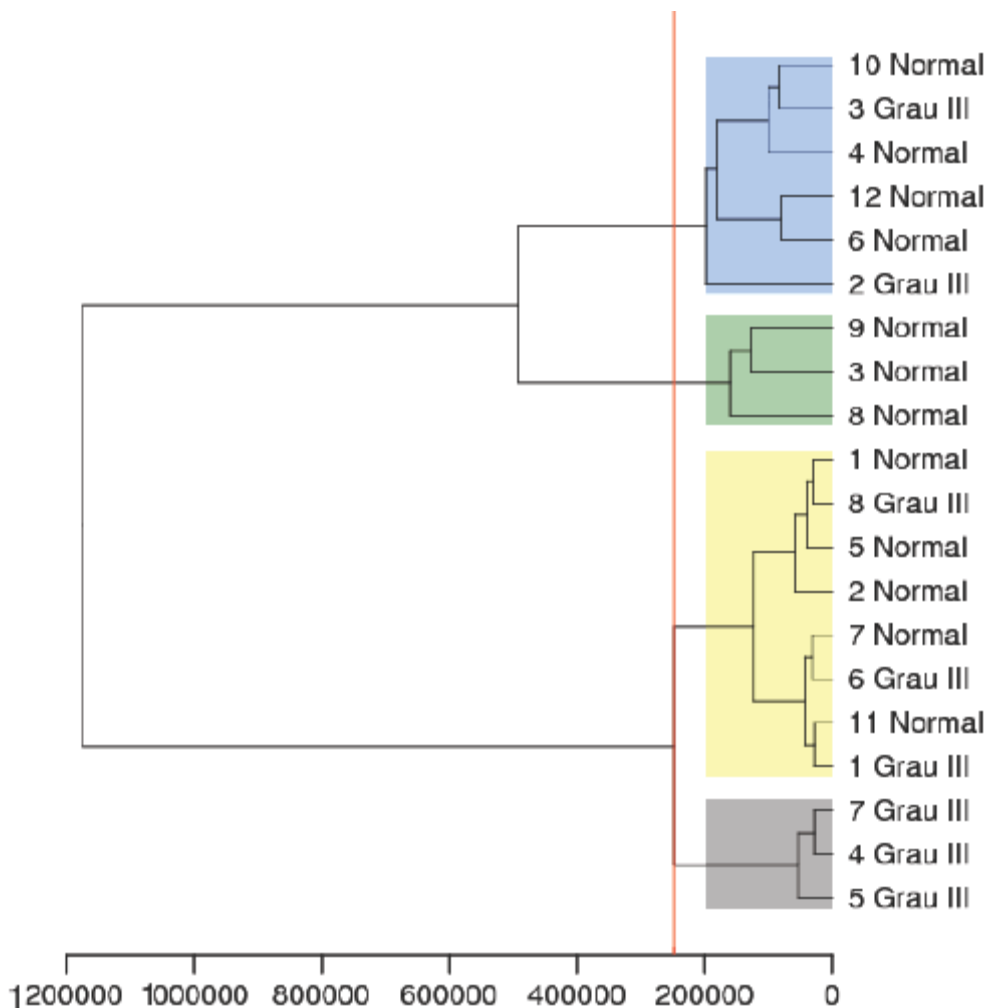


Figura 4. Dendograma correspondente ao Complete-link associado ao DTW

Os grupos 2 e 4 apresentaram séries temporais de exames de uma única classe. Especificamente o grupo 2, apresenta exemplos da classe de pacientes em condições normais e, o grupo 4, da classe Grau III. Nos grupos 1 e 3, embora existam séries de ambas as classes, se observa a predominância de pacientes em condições normais em relação a pacientes de incontinência em Grau III. De acordo com o especialista, os grupos apresentam características que podem auxiliar na identificação de informações relacionadas às classes consideradas. O especialista sugeriu a realização de uma análise mais detalhada dos exames, de modo que sejam verificados o valor de média, desvio-padrão e valores máximos e mínimos de pressão. Outra questão a ser considerada consiste na análise conjunta dos grupos encontrados em relação à história clínica desses pacientes associada aos exames.

A partir desta análise será possível explorar mais detalhadamente os motivos que produziram os agrupamentos relacionando pacientes de incontinência em Grau III e pacientes em condições normais. De acordo com os especialistas, uma justificativa para esse comportamento, que deve ser comprovada ou descartada por uma análise futura, consiste na presença, em pacientes com diagnóstico de incontinência em Grau III, de propriedades clínicas que caracterizariam exames mais próximos aos dos pacientes em condições normais.

#### **4. Conclusão**

No contexto atual, a tecnologia vem se incorporando ao cotidiano da humanidade de forma cada vez mais intensa. A evolução de hardwares e softwares propicia à computação um crescente poder de processamento que pode ser utilizado, por exemplo, para prover suporte à área da saúde. No domínio de manometria ano-retal observam-se anormalidades e doenças, cujos sintomas provocam no paciente desde situações de desconforto social até graves quadros clínicos.

Neste trabalho, foi apresentado um método para a construção de agrupamentos a partir de exames de manometria ano-retal. Esse método tem como objetivo a extração de padrões nos dados por meio de técnicas de agrupamento, no intuito de auxiliar no estudo e no diagnóstico do distúrbio de incontinência fecal. Esse método foi aplicado a um conjunto de 20 exames, dentre os quais, oito referem-se a pacientes com incontinência em Grau III e doze referem-se a pacientes em condições normais. Desse modo, foram gerados agrupamentos considerando diferentes configurações de medidas de similaridade e técnicas de *clustering*.

Os agrupamentos construídos foram avaliados objetivamente e em conjunto com os especialistas. De acordo com essas avaliações, o método proposto e as técnicas aplicadas apresentaram-se promissoras para a construção de agrupamentos que possam dar apoio no estudo do distúrbio de incontinência fecal, além de auxiliar na definição das características dos pacientes em condições normais.

Trabalhos futuros incluem: a seleção de outros grupos a partir do dendograma selecionado e dos dendogramas referentes às outras configurações experimentais; a análise de história clínica dos exames envolvidos para auxiliar no entendimento mais completo das relações entre as séries temporais associadas a cada grupo; e a utilização de medidas objetivas de avaliação, baseadas em outros critérios, que permitam ressaltar características interessantes para a área de domínio.

## Agradecimentos

Ao Programa de Desenvolvimento Tecnológico Avançado – PDTA/FPTI-BR – pelo auxílio por meio da linha de financiamento de bolsas.

## Referências

- [1] Aggarwal, C. C., Hinneburg, A., e Keim, D. A. (2001). On the surprising behavior of distance metrics in high dimensional space. In *Lecture Notes in Computer Science*, Springer, pages 420–434.
- [2] Alpaydin, E. (2004). *Introduction to Machine Learning*. MIT Press, Cambridge, Inglaterra.
- [3] Chu, S., Keogh, E. J., Hart, D., e Pazzani, M. J. (2002). Iterative deepening dynamic time warping for time series. In *Second SIAM International Conference on Data Mining*, pages 1–18, Virginia, USA.
- [4] Everitt, B. S. (1993). *Cluster Analysis*. Edward Arnold, Londres, Inglaterra, 3 edição.
- [5] Han, J. and Kamber, M. (2006). *Data Mining: concepts and techniques*. Morgan Kaufmann, California, USA, 2 edição.
- [6] Jain, A. K. and Dubes, R. C. (1988). *Algorithms for clustering data*. Prentice Hall, Upper Saddle River, USA.
- [7] Keogh, E. and Lin, J. (2005). Clustering of time series subsequence is meaningless: implications for previous and future search. *Knowledge and Information Systems*, Springer-Verlag, 8(2):154–177.
- [8] Last, M., Kandel, A., & Bunke, H. (2004). *Data Mining in Time Series Databases*, volume 57 of *Machine perception and artificial intelligence*. World Scientific, Massachusetts, USA.
- [9] Morais, M. B., Sdepanian, V. L., Tahan, S., Goshima, S., Soares, A. C. F., Motta, M. E. F. A. and Neto, U. F. (2005). A manometria ano-retal (método do balão) no diagnóstico diferencial da doença de Hirschsprung. *Revista da Associação Médica Brasileira*, 51(12):313–317.
- [10] Oliveira, L. (2006). Incontinência fecal. *Revista Brasileira de Gastroenterologia*. 6(1):35–37.
- [11] Pyle, D. (1999). *Data Preparation for Data Mining*. Morgan Kaufmann, California, USA.
- [12] R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- [13] Raskutti, B. and Leckie, C. (1999). An Evaluation of Criteria for Measuring the Quality of Clusters. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, pages 905–910, California, USA.
- [14] Rezende, S. O. (2003). *Sistemas Inteligentes: Fundamentos e Aplicações*. Editora Manole, São Paulo, Brasil.
- [15] Saad, L. H. C. (2002). *Quantificação da função esfíncteriana pela medida da capacidade de sustentação da pressão de contração voluntária do canal anal*. Tese de doutorado, Universidade Estadual de Campinas, São Paulo, Brasil.
- [16] Saad, L. H. C., Coy, C. S. R., Fagundes, J. J., Ariyazono, M. L., Shoji, N and Góes, J. R. N. (2002). Quantificação da função esfíncteriana pela medida da capacidade de sustentação da pressão de contração voluntária do canal anal. *Arquivos de Gastroenterologia*. 39(10):233–239.
- [17] Spolaôr, N., Lee, H. D., Ferrero, C. A., Coy, C. S. R., Fagundes, J. J., e Wu, F. C. (2008). Um estudo da aplicação de clustering de séries temporais em dados médicos. In *III Congresso da Academia Trinacional de Ciências*, pages 1–10, Foz do Iguaçu, Brasil.
- [18] Spolaôr, N. (2008). *Clustering de Séries Temporais: Aplicação em Dados Médicos*. Monografia de Graduação. Universidade Estadual do Oeste do Paraná. Paraná, Brasil.
- [19] Pinho, M. (2000). *Incontinência fecal*. Revinter, Rio de Janeiro, Brasil.
- [20] Witten, I. and Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufman Publishers Inc., San Francisco, Califórnia, EUA.

[21] Wu, W., Xiong, H. and Shekhar, S. (2003). *Clustering and Information Retrieval: Network Theory and Applications*. Kluwer Academic Publishers, Norwell, USA.

### **Dados de Contato**

*Carlos Andres Ferrero. Laboratório de Inteligência Computacional, Universidade de São Paulo. Av. Trabalhador São-carlense, 400, 13560-970, São Carlos, SP, Brasil. anfer86@gmail.com.*

*André Gustavo Maletzke. Laboratório de Inteligência Computacional, Universidade de São Paulo. Av. Trabalhador São-carlense, 400, 13560-970, São Carlos, SP, Brasil. andregustavom@gmail.com*

*Huei Diana Lee. Laboratório de Bioinformática, Universidade Estadual do Oeste do Paraná. Av. Presidente Tancredo Neves, 6731, 85866-900, Foz do Iguaçu, PR, Brasil. hueidianalee@gmail.com.*

*Newton Spolaôr. Centro de Matemática, Computação e Cognição, Universidade Federal do ABC. Rua Santa Adélia, 166, 09210-170, Santo André, SP, Brasil. newtonspolaor@gmail.com.*

*Cláudio Saddy Rodrigues Coy. Serviço de Coloproctologia da Faculdade de Ciências Médicas, Universidade Estadual de Campinas. Rua Carlos Chagas, 420, 13031-970, Campinas, SP, Brasil. ccoy@terra.com.br.*

*João José Fagundes. Serviço de Coloproctologia da Faculdade de Ciências Médicas, Universidade Estadual de Campinas. Rua Carlos Chagas, 420, 13031-970, Campinas, SP, Brasil. jjfagundes@mpcnet.com.br.*

*Feng Chung Wu, Laboratório de Bioinformática, Universidade Estadual do Oeste do Paraná. Fundação Parque Tecnológico Itaipu. Av. Presidente Tancredo Neves, 6731, 85866-900, Foz do Iguaçu, PR, Brasil. wufengchung@gmail.com.*