

# Avaliação de Algoritmos de *Clustering* Hierárquico de Séries Temporais Aplicados a Exames Médicos de Manometria Ano-retal

Newton Spolaôr<sup>1</sup>, Huei D. Lee<sup>1</sup>, André G. Maletzke<sup>2</sup>, Sidney B. Shiki<sup>1</sup>,  
Cláudio S. R. Coy<sup>3</sup>, João J. Fagundes<sup>3</sup>, Feng C. Wu<sup>1,3</sup>

<sup>1</sup>Laboratório de Bioinformática (LABI),

Universidade Estadual do Oeste do Paraná (UNIOESTE), Parque Tecnológico Itaipu (PTI)

<sup>2</sup>Laboratório de Inteligência Computacional (LABIC), Universidade de São Paulo (USP)

<sup>3</sup>Universidade Estadual de Campinas (UNICAMP)

## 1. Objetivo

Os fenômenos que ocorrem ao longo do tempo podem ser modelados como um tipo de dado denominado Séries Temporais – ST. Neste trabalho é apresentada uma avaliação de alguns algoritmos de clustering hierárquico de ST aplicados a exames de Manometria Ano-retal – MA.

## 2. Materiais e Métodos

A MA é um dos exames mais comuns na avaliação de anomalias como a Incontinência Fecal – IF –, a qual se caracteriza pelo descontrole da passagem de fezes e/ou gases. A IF pode ser classificada em três graus conforme a intensidade [1]: Grau I - incontinência para gases; Grau II - Grau I e incontinência para fezes líquidas; Grau III - Graus I e II e incontinência para fezes sólidas. Neste trabalho foram considerados 20 exames de MA elaborados pelo Serviço de Coloproctologia da UNICAMP. Esses exames descrevem a Pressão de Contração Voluntária – PCV – aplicada sobre um cateter de oito sensores, cada um dos quais modelado como uma ST. Em cada exame, a soma dos valores de PCV de cada sensor permitiu definir uma ST resultante, na qual identificou-se três seções de PCV, com duração individual de aproximadamente 40 segundos de acordo com o protocolo definido em [1]. Essas seções foram concatenadas para compor a ST correspondente ao exame. O processo de *clustering* de ST é geralmente composto por etapas: (1) Representação dos Dados, (2) Seleção da Medida de Similaridade e (3) Aplicação do Agrupamento [2]. A qualidade do agrupamento construído pode ser verificada por especialistas e por medidas objetivas de avaliação, as quais podem indicar os agrupamentos mais representativos em um domínio. A avaliação realizada neste trabalho envolveu a aplicação das técnicas de representação de ST a partir dos valores originais de PCV e da distância Euclidiana, respectivamente, nas Etapas 1 e 2 do processo de *clustering* de ST. Na

Etapa 3 foram utilizados os algoritmos *Average-link*, *Complete-link* e *Single-link*, os quais foram avaliados objetivamente a partir do Coeficiente de Correlação *Cophenetic* – CCC. A abordagem hierárquica possibilita a representação da similaridade envolvida na construção de cada *cluster* em uma árvore denominada dendograma.

## 3. Resultados e Discussão

A avaliação dos *clusters* a partir do CCC apresentou, para os algoritmos *Average-link*, *Complete-link* e *Single-link*, os valores de 0,8814, 0,7595 e 0,8828, respectivamente, o que indica que o dendograma correspondente ao *Single-link* é o mais representativo neste caso. A identificação dos conceitos descritos nesse dendograma está sendo realizada a partir da análise criteriosa, com o auxílio de especialistas do domínio, da construção de cada *cluster* em diferentes níveis de granularidade. Essa análise possibilita a aquisição de conhecimentos úteis, os quais poderão auxiliar especialistas em processos de tomada de decisão correspondentes ao diagnóstico de IF.

## 4. Conclusões

Neste trabalho constatou-se que o algoritmo *Single-link* permite a construção, segundo a medida CCC, de um modelo mais representativo. Como trabalho futuro, propõe-se a avaliação de outras medidas de similaridade, bem como a utilização de outras medidas objetivas para a avaliação dos agrupamentos.

## 5. Referências Bibliográficas

- [1] Saad LHC. Quantificação da função esfinteriana pela medida da capacidade de sustentação da pressão de contração voluntária do canal anal [Tese de Doutorado]. Universidade Estadual de Campinas; 2002.
- [2] Jain AK, Dubes RC. Algorithms for clustering data. New Jersey, USA: Prentice-Hall; 1988.