

Detecção do Congelamento de Marcha em Pacientes com Parkinson Usando Algoritmos de Similaridade

Daniil Sveshnikov¹, Gildo G. A. Oliveira², Sofia F. Martins²

¹Instituto de Ciências Matemáticas e de Computação - Universidade de São Paulo (USP)
São Carlos, SP - Brasil

²Instituto de Física de São Carlos - Universidade de São Paulo (USP)
São Carlos, SP - Brasil

daniil_s@usp.br, gildogutemberg2@usp.br, sofiafreitasmartins@usp.br

Abstract. *Freezing of gait (FOG) is characterized by the sudden and temporary loss of the ability to take steps, and is a debilitating symptom in patients with advanced stages of Parkinson's disease (PD). This study aims to apply time series similarity algorithms to detect FOG moments through matrix profile (MP) analysis. For this purpose, we used a public dataset containing information related to the gaits of 10 PD patients, collected by motion sensors.*

Resumo. *O congelamento de marcha (FOG) é caracterizado pela perda repentina e temporária da capacidade de produzir passos, sendo um sintoma debilitante em pacientes com a doença de Parkinson (PD) em estágios mais avançados. Este estudo visa aplicar algoritmos de similaridade em séries temporais para detectar momentos de FOG por meio da análise de matrix profile (MP). Para isso, utilizou-se um conjunto de dados público que possui informações relacionadas a marchas de 10 pacientes com PD, coletadas por sensores de movimentos.*

1. Introdução

A doença de Parkinson (PD) é um distúrbio neurológico caracterizado pela degeneração da substância negra, região localizada no sistema nervoso central responsável pela produção do neurotransmissor dopamina [Silva et al. 2021]. Apresentando-se de forma crônica e progressiva, essa doença compromete a atividade da alça frontoestriatal e acarreta na disfunção do lobo frontal pela perda de neurônios dopaminérgicos da substância negra [Balestrino and Schapira 2019]. Uma manifestação inicial da PD é o declínio do desenvolvimento motor, que afeta as atividades cotidianas dos pacientes de forma a limitá-las ou até mesmo impedi-las, prejudicando gravemente a qualidade de vida dessas pessoas.

Pacientes com PD em estágios mais avançados podem apresentar um sintoma conhecido como congelamento de marcha ou freezing (FOG), definido pela perda repentina e temporária da capacidade de produzir passos. Os episódios de perda podem ser mais breves ou excederem 30 segundos [Schaafsma et al. 2003]. Esses episódios são mais prováveis de ocorrer ao iniciar uma caminhada, virar ou passar por passagens estreitas. Curiosamente, dicas visuais e auditivas podem mudar o foco da atenção de um paciente e reduzir o FOG [Rahimpour et al. 2021]. A estimulação auditiva rítmica (RAS) foi uma

técnica utilizada por [Bächlin et al. 2010] na coleta de dados do *dataset* que será utilizado neste trabalho, feito com o objetivo de detectar momentos de FOG de pacientes por meio da utilização de algoritmos de similaridade aplicados em séries temporais.

2. Trabalhos relacionados

O conjunto de dados de [Bächlin et al. 2010] foi resultado de uma colaboração entre o Laboratory for Gait and Neurodynamics, Tel Aviv Sourasky Medical Center, Israel e o Wearable Computing Laboratory, ETH Zurich, Suíça. Os dados foram coletados para o projeto Daphnet (Dynamic Analysis of Physiological Networks), financiado pela União Europeia com o intuito de investigar redes fisiológicas humanas por meio da análise dinâmica de sinais como frequência cardíaca, controle motor e eletroencefalograma. Para o FOG, a equipe utilizou um sistema de sensores de aceleração para medir os movimentos dos pacientes, detectando automaticamente momentos de freezing e fornecendo um RAS que estimula o paciente a retomar a caminhada.

Os sinais coletados pelo grupo de pesquisadores podem ser analisados no âmbito de séries temporais, ao utilizar uma técnica que permite identificar padrões repetidos, anomalias, ou mesmo eventos de interesse, como o FOG. Tal técnica é conhecida como Matrix Profile (MP) e foi proposta por [Yeh et al. 2016] no contexto de oferecer um método unificado para análise de séries temporais.

A partir dos dados do Daphnet Freezing of Gait e utilizando a técnica de Matrix Profile, pretendemos realizar a análise dessas séries temporais para identificar e caracterizar eventos de FOG.

3. Metodologia

4. Conjunto de Dados

O conjunto de dados utilizado neste projeto consiste em sinais de aceleração 3D coletados de pacientes com doença de Parkinson, realizando tarefas de caminhada e atividades do dia a dia. Esses dados foram coletados para estudar o fenômeno de *freezing of gait* (FOG) e desenvolver métodos para sua detecção e análise.

4.1. Descrição dos Dados

Os dados foram coletados de 10 pacientes diagnosticados com a Doença de Parkinson (idade média de 68 anos, com 7 homens e 3 mulheres). Durante os experimentos, cada participante realizou tarefas:

1. Caminhar em linha reta ao longo do corredor, com voltas de 180 graus.
2. Caminhada aleatória no espaço de recepção, com paradas e voltas de 360 graus seguindo instruções do experimentador.
3. Caminhada simulando atividades diárias, como entrar e sair de salas, ir à cozinha e retornar com um copo de água.

Os sinais foram registrados por acelerômetros 3D posicionados no tornozelo, coxa e parte inferior do tronco, com uma taxa de amostragem de 64 Hz. Cada registro contém informações sobre os eixos horizontal, vertical e lateral.

A duração de cada sessão de caminhada foi de cerca de 10 a 15 minutos. Pacientes caminharam em seu próprio ritmo natural sem ajuda, mas um O terapeuta permaneceu por perto por razões de segurança

Apresentados dados sobre o tempo decorrido desde o início da medição - em milissegundos e um parâmetro que determina o estado-chave do paciente, determinado pelo especialista que observa o experimento:

- 0: não faz parte do experimento. Por exemplo, os sensores são instalados no usuário
- 1: experimento, sem Freezing
- 2: Freezing

Ao mesmo tempo, os pacientes 4 e 10 não apresentaram sinais de Freezing durante o experimento.

4.2. Processamento de Dados

Os dados foram processados para remover ruídos e preencher valores ausentes.

Este conjunto de dados é o resultado de uma colaboração entre o Laboratório de Marcha e Neurodinâmica, Centro Médico Sourasky de Tel Aviv, Israel e o Laboratório de Computação Vestível, ETH Zurique, Suíça. As gravações foram realizadas no Centro Médico Sourasky de Tel Aviv em 2008.

4.3. Considerações Éticas

Todos os participantes forneceram consentimento informado, e o estudo foi aprovado pelo Comitê de Revisão de Seres Humanos local e foi realizado de acordo com os padrões éticos da Declaração de Helsinque.

4.4. Disponibilidade dos Dados

O conjunto de dados está disponível no repositório público [UC Irvine Machine Learning Repository](#).

4.5. FLUSS

O FLUSS (Fast Low-cost Unipotent Semantic Segmentation) é um algoritmo desenvolvido por [Gharghabi et al. 2017] e projetado para identificar mudanças de regime em séries temporais, isto é, pontos onde há transições significativas entre estados diferentes. Ele opera criando uma curva auxiliar chamada Arc Curve (AC), que mede a quantidade de arcos (ligações entre subsequências que são vizinhas mais próximas) que cruzam cada ponto da série. Cada subsequência da série temporal é comparada com todas as outras subsequências para encontrar sua vizinha mais próxima (usando o conceito de Matrix Profile). Quando a série temporal passa por uma mudança de regime, as subsequências mais próximas ainda tendem a se conectar com outras subsequências do regime anterior, porque ainda são parecidas. Isso resulta em uma diminuição dos arcos entre subsequências de regimes diferentes. Esse ponto de transição mostra a mudança, onde o número de arcos começa a diminuir, indicando que o padrão está se alterando. Em suma, os valores da AC são baixos em fronteiras de regime e altos dentro de regimes estáveis.

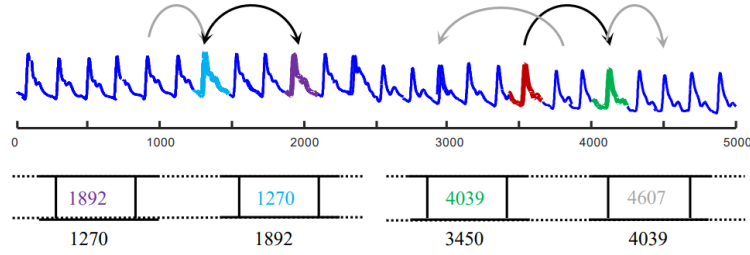


Figura 1. Ilustração do processo de segmentação de regimes usando o algoritmo FLUSS. Fonte: [Gharghabi et al. 2017].

A AC apresenta um viés nas bordas da série temporal. Isso acontece porque há menos subsequências candidatas para formar arcos nas extremidades, o que reduz artificialmente os valores da AC nessas regiões. A fim de corrigir isso, o FLUSS utiliza uma abordagem baseada em uma curva de referência chamada Curva de Arcos Idealizada (IAC). A IAC representa o número esperado de cruzamentos de arcos em cada posição caso as subsequências fossem distribuídas de forma completamente aleatória. Essa curva assume a forma de uma parábola invertida com altura máxima de $\frac{1}{2}n$ em que n é o comprimento da série temporal.

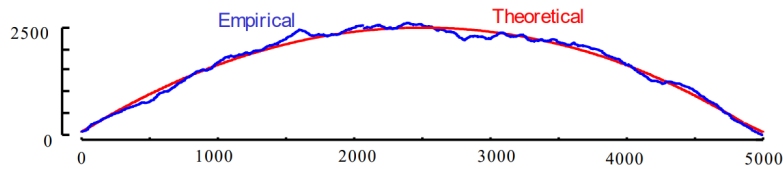


Figura 2. A Curva de Arcos Idealizada (IAC) para uma série temporal sem estrutura de similaridade localizada é uma parábola invertida com uma altura $\frac{1}{2}n$. Fonte: [Gharghabi et al. 2017].

A Curva de Arcos Corrigida (CAC) é obtida normalizando a AC pela IAC:

$$CAC_i = \min \left(\frac{AC_i}{IAC_i}, 1 \right) \quad (1)$$

Essa correção elimina o viés nas bordas, limitando os valores entre 0 e 1. Além disso, ela possibilita comparações entre fluxos monitorados em taxas de amostragem diferentes e possibilidade de aprender valores de limiar específicos para domínios.

Procedure CAC (<i>MPIndex</i>, <i>L</i>)	
Input- <i>MPIndex</i> : the matrix profile index for the time series of interest	
<i>L</i> : the subsequence length	
Output- <i>CAC</i> : a corrected Arc Curve	
1	<i>n</i> = length(<i>MPIndex</i>)
2	<i>AC</i> = <i>CAC</i> = <i>nnmark</i> = zero initialize array of size <i>n</i>
3	for <i>i</i> =1: <i>n</i>
4	<i>j</i> = <i>MPIndex</i> [<i>i</i>]
5	<i>nnmark</i> [min(<i>i,j</i>)] = <i>nnmark</i> [min(<i>i,j</i>)] + 1
6	<i>nnmark</i> [max(<i>i,j</i>)] = <i>nnmark</i> [max(<i>i,j</i>)] - 1
7	end
8	
9	<i>numArcs</i> = 0
10	for <i>i</i> =1: <i>n</i>
11	<i>numArcs</i> = <i>numArcs</i> + <i>nnmark</i> [<i>i</i>]
12	<i>AC</i> [<i>i</i>] = <i>numArcs</i>
13	end
14	
15	<i>IAC</i> = parabolic curve of length <i>n</i> and height $\frac{1}{2}n$
16	<i>CAC</i> = min (<i>AC</i> / <i>IAC</i> , 1)
17	Set the <i>L</i> length in the beginning and end of <i>CAC</i> to 1
18	return <i>CAC</i>

Figura 3. Algoritmo para a construção de uma Curva de Arcos Corrigida (CAC).
Fonte: [Gharghabi et al. 2017].

Uma vez calculada a CAC, o próximo passo é identificar os pontos de mudança de regime. Para isso, o algoritmo busca os k pontos de vale mais baixos na CAC. A identificação baseia-se na ideia de que os vales correspondem a locais onde há menos cruzamentos de arcos, indicando uma transição entre regimes. Para evitar mínimos triviais consecutivos (se x for o ponto mais baixo, é quase certo que $x + 1$ ou $x - 1$ seja o segundo ponto mais baixo), o FLUSS aplica uma zona de exclusão ao redor de cada vale identificado. Essa zona tem um tamanho de $5 \times L$ sendo L o comprimento da subsequência, impedindo a detecção de mudanças muito próximas. O algoritmo continua a identificar vales até encontrar todos os k pontos.

Procedure ExtractRegimes (<i>CAC</i>, <i>numRegimes</i>, <i>L</i>)	
Input- <i>CAC</i> : a corrected Arc Curve	
<i>numRegimes</i> : number of regime changes	
<i>L</i> : length of the subsequence	
Output- <i>locRegimes</i> : the locations of the regimes	
1	<i>locRegimes</i> = empty array of length <i>numRegimes</i>
2	for <i>i</i> =1: <i>numRegimes</i>
3	<i>locRegimes</i> (<i>i</i>) = indexOf(min(<i>CAC</i>))
4	Set exclusion zone of $5 \times L$
5	end
6	return <i>locRegimes</i>

Figura 4. Algoritmo para extração de regimes. Fonte: [Gharghabi et al. 2017].

4.6. Amostragem dos Dados

Cada acelerômetro foi posicionado para capturar a aceleração nos três eixos (X, Y e Z) em diferentes partes do corpo: calcanhar, coxa e tronco. Dessa forma, cada paciente

gera nove séries temporais que podem ser analisadas de maneira independente ou em conjunto.

Nosso objetivo é verificar a capacidade do algoritmo FLUSS de identificar os momentos de transição entre estados de "freezing" e "no-freezing". Conforme [Gharghabi et al. 2017], o método de avaliação mais comum no domínio da segmentação semântica é a análise visual. Assim, como um primeiro passo, realizamos uma inspeção visual da aplicação do FLUSS em uma das séries temporais de um paciente.

Para analisar as regiões de transição, selecionamos 500 pontos antes e 500 pontos após cada transição. Essas regiões foram submetidas à análise do FLUSS. É importante observar que, em algumas regiões, múltiplas transições podem estar presentes. Isso não é ideal, pois em uma mesma região de transição pode ocorrer de ter mais de um ponto de transição, isso dificulta a precisão do algoritmo. Para mitigar esse problema, optamos por limitar o intervalo de análise a um máximo de 1000 pontos. Além disso, configuramos os parâmetros L e M (L é o tamanho da subsequência para o algoritmo do FLUSS e M para o do MP) com valores moderados, ambos iguais a 50.

Dada a limitação do método visual na análise das quebras das séries temporais, desenvolvemos uma métrica para comparar a performance do algoritmo FLUSS nos nove tipos de séries temporais disponíveis.

Dado um determinado paciente que possui N regiões de transição, para cada uma dessas regiões, podemos calcular o erro associado, $Erro_n$, definido como:

$$Erro_n = \frac{\sum_{j=1}^i |\text{Previsto}_j - \text{Real}_j|}{i}, \quad (2)$$

onde j é o índice das transições que ocorrem dentro do n -ésimo intervalo, e i representa o número total de transições no n -ésimo intervalo.

Com base nos valores de $Erro_n$, podemos calcular o erro médio para uma dada série temporal como:

$$Erro_{\text{série}} = \frac{\sum_{n=1}^N Erro_n}{N}. \quad (3)$$

Essa métrica nos permite avaliar a precisão do FLUSS em detectar regiões de transição, ao calcular a média dos erros para todos os intervalos de uma série temporal.

5. Resultados e Discussões

5.1. Resultados Visuais

Para realizar análises visuais, selecionamos arbitrariamente o paciente 8 e, dentre as séries disponíveis, escolhemos aquela correspondente ao *Trunk_2*

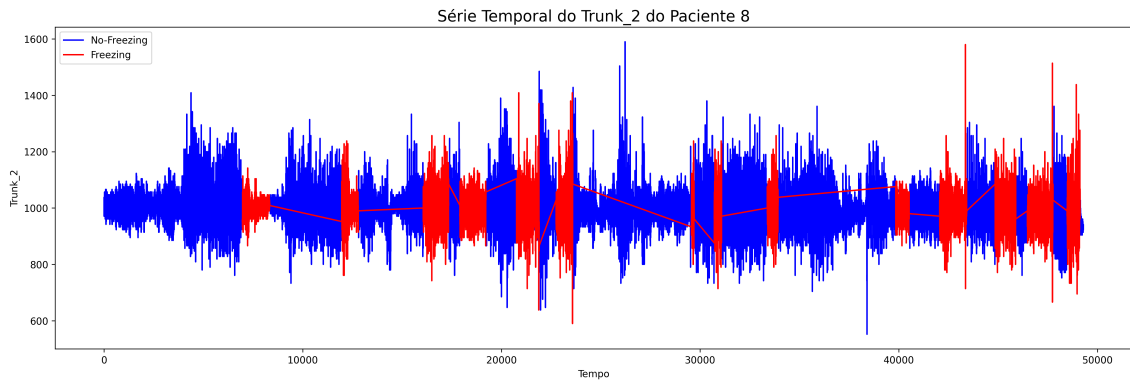


Figura 5. Série temporal do Trunk_2 do paciente 8. Fonte: Elaborado pelos autores.

Na Figura 5, apresentamos o gráfico desta série, com o tratamento básico para remoção dos momentos em que os experimentos não estavam sendo realizados. No gráfico, os trechos em azul correspondem aos períodos em que o *freezing of gait* (FOG) não ocorre, enquanto os trechos em vermelho indicam os períodos de ocorrência de FOG.

Para este paciente, identificamos um total de 28 regiões de transição. Selecionamos a primeira dessas regiões para uma análise inicial da técnica FLUSS, ilustrada na Figura 6.

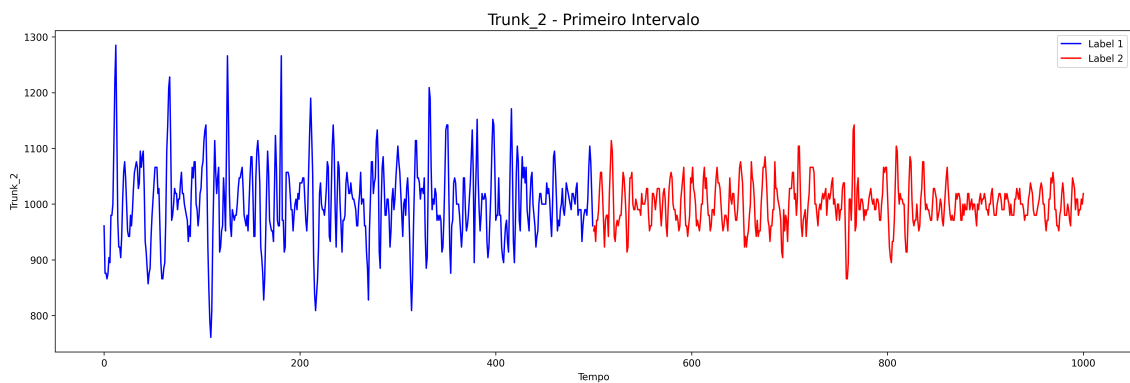


Figura 6. Primeiro intervalo de transição para o Trunk_2 do paciente 8. Fonte: Fonte: Elaborado pelos autores.

Ao aplicarmos o FLUSS, obtivemos um resultado interessante, como mostrado na Figura 7. A localização prevista para a transição ficou a uma distância de apenas 66 pontos em relação à posição real, o que consideramos razoável, dado o tamanho da série analisada, que é de 1000 pontos.

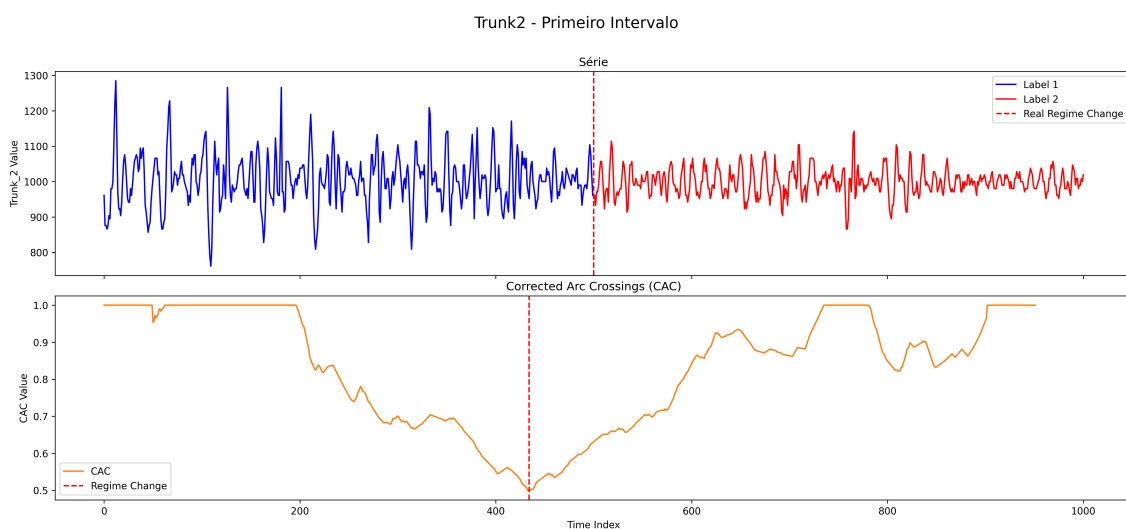


Figura 7. Aplicação do FLUSS para identificar o primeiro intervalo de transição na série Trunk_2 do paciente 8. Fonte: Elaborado pelos autores.

Apesar do bom desempenho na primeira região, a análise da segunda região revelou um resultado menos satisfatório. Visualmente, observamos que o ponto previsto estava a uma distância de 401 pontos do ponto real, como destacado na Figura 8.

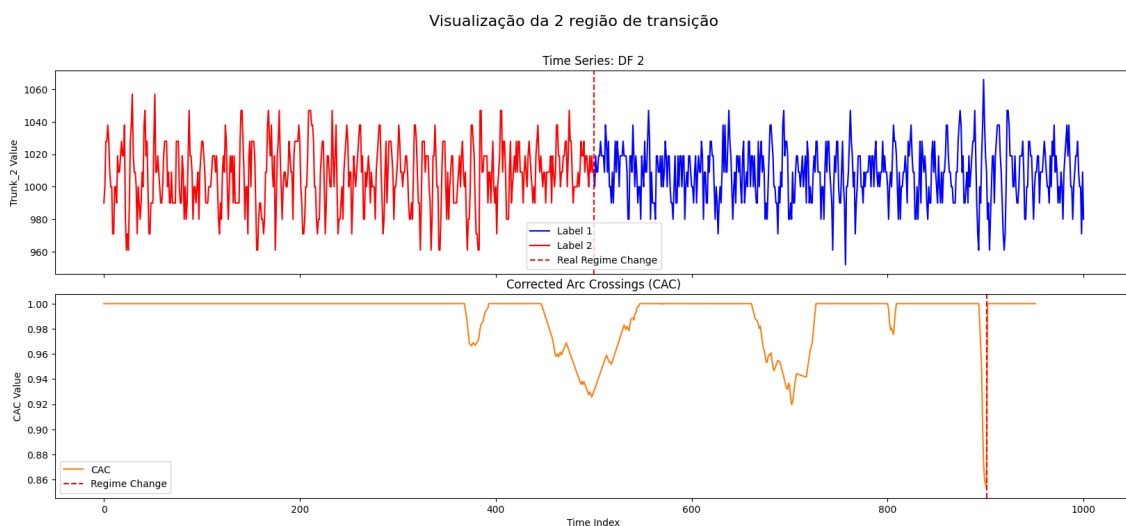


Figura 8. Aplicação do FLUSS para identificar o segundo intervalo de transição na série Trunk_2 do paciente 8. Fonte: Elaborado pelos autores.

Este comportamento sugere que algumas regiões de transição apresentam maior complexidade e que, nessas condições, o FLUSS pode não capturar adequadamente os pontos de transição.

Para o paciente 8, identificamos 28 regiões de transição. Ao considerarmos as 9 séries temporais disponíveis, isso resulta em um total de 252 análises potenciais. Esse

volume demonstra claramente que uma abordagem puramente visual não é viável para o nosso problema, reforçando a relevância da métrica de erro proposta.

5.2. Erro Médio

Ao aplicarmos a métrica de erro proposta anteriormente às séries temporais analisadas no Paciente 8, construímos a Tabela 1, que apresenta o erro médio e o desvio padrão associados ao uso do algoritmo FLUSS na detecção dos pontos de transição nas diferentes regiões anatômicas.

Os valores médios representam a precisão geral do FLUSS na identificação dos momentos de mudança, enquanto o desvio padrão indica a variabilidade dos erros dentro de cada categoria de séries temporais. Essa análise é fundamental para avaliar a robustez e a consistência do método em regiões com características distintas,

Tabela 1. Médias e desvios padrão da aplicação do FLUSS no Paciente 8

Coluna	Média	Desvio Padrão
Ankle_1	188.46	120.76
Ankle_2	198.68	143.75
Ankle_3	226.23	124.29
Thigh_1	193.34	139.32
Thigh_2	181.34	129.30
Thigh_3	176.41	135.14
Trunk_1	214.21	137.18
Trunk_2	203.66	137.99
Trunk_3	228.68	136.89

Dado os resultados, que apresentaram uma distância média elevada e um desvio padrão significativo, identificou-se a necessidade de investigar se esse comportamento também se manifesta em outros pacientes.

Tabela 2. Médias e desvios padrão da aplicação do FLUSS no Paciente 9

Coluna	Média	Desvio Padrão
Ankle_1	211.35	104.39
Ankle_2	205.37	114.11
Ankle_3	198.50	114.62
Thigh_1	211.70	123.28
Thigh_2	214.91	117.03
Thigh_3	185.05	111.28
Trunk_1	198.52	105.07
Trunk_2	204.48	128.74
Trunk_3	213.05	125.21

A Tabela 2 apresenta os resultados para o paciente 9, e, de fato, observamos que ambas as colunas indicam valores elevados.

Além disso, em ambas as tabelas, não há uma diferença significativa em relação à parte do corpo analisada, sugerindo que o comportamento é consistente independentemente da região.

Por fim, podemos concluir que, pelo menos com a segmentação dos dados utilizando uma janela de análise de 1000 pontos — em que, em alguns casos, há mais de uma região de transição — o FLUSS não se mostrou uma técnica eficaz para a detecção dessas regiões.

6. Conclusão

Dada a forma como segmentamos nossos dados, utilizando uma janela de 1000 pontos que pode conter mais de um ponto de transição, tanto nossa análise visual quanto a métrica proposta indicam que o FLUSS não é um método adequado para essa análise. O número de regiões de análise limita a aplicação visual, e, a partir das tabelas geradas para os pacientes 8 e 9, observamos um erro elevado associado ao método.

Uma possível direção para análises futuras é avaliar o desempenho do algoritmo FLOSS (Fast Lowcost Online Semantic Segmentation) em um cenário de detecção online, em vez de utilização do FLUSS que é um algoritmo para análise em batch. Outra abordagem interessante seria explorar diferentes formas de segmentar os dados, como a modificação do tamanho das janelas de análise, a fim de verificar se isso poderia melhorar os resultados obtidos.

7. Repositório

Todos os códigos utilizados na realização deste trabalho podem ser encontrados no seguinte repositório: https://github.com/gildex174/fog_detection.

Referências

- [Bächlin et al. 2010] Bächlin, M., Plotnik, M., Roggen, D., Maidan, I., Hausdorff, J. M., Giladi, N., and Tröster, G. (2010). Wearable assistant for parkinson's disease patients with the freezing of gait symptom. *IEEE Transactions on Information Technology in Biomedicine*, 14(2):436–446.
- [Balestrino and Schapira 2019] Balestrino, R. and Schapira, A. H. V. (2019). Parkinson disease. *European Journal of Neurology*, 27(1):27–42.
- [Gharghabi et al. 2017] Gharghabi, S., Ding, Y., Yeh, C.-C. M., Kamgar, K., Ulanova, L., and Keogh, E. (2017). Matrix profile viii: Domain agnostic online semantic segmentation at superhuman performance levels. In *2017 IEEE International Conference on Data Mining (ICDM)*, pages 117–126.
- [Rahimpour et al. 2021] Rahimpour, S., Gaztanaga, W., Yadav, A. P., Chang, S. J., Krucoff, M. O., Cajigas, I., Turner, D. A., and Wang, D. D. (2021). Freezing of gait in parkinson's disease: Invasive and noninvasive neuromodulation. *Neuromodulation*, 24(5):829–842. Epub 2020 Dec 26. PMID: 33368872; PMCID: PMC8233405.
- [Schaafsma et al. 2003] Schaafsma, J. D., Balash, Y., Gurevich, T., Bartels, A. L., Hausdorff, J. M., and Giladi, N. (2003). Characterization of freezing of gait subtypes and the response of each to levodopa in parkinson's disease. *European Journal of Neurology*, 10(4):391–398.

- [Silva et al. 2021] Silva, A. B. G., Pestana, B. C., Hirahata, F. A. A., Horta, F. B. d. S., and Oliveira, E. S. B. E. (2021). Doença de parkinson: revisão de literatura / parkinson's disease: literature review. *Brazilian Journal of Development*, 7(5):47677–47698.
- [Yeh et al. 2016] Yeh, C.-C. M., Zhu, Y., Ulanova, L., Begum, N., Ding, Y., Dau, H. A., Silva, D. F., Mueen, A., and Keogh, E. (2016). Matrix profile i: All pairs similarity joins for time series: A unifying view that includes motifs, discords and shapelets. In *2016 IEEE 16th International Conference on Data Mining (ICDM)*, pages 1317–1322.