

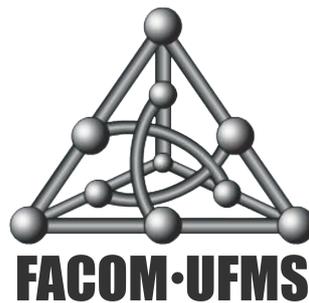
Dissertação de Mestrado

Uma Abordagem para Classificação de
Séries Temporais Baseada em Modelo
Autorregressivo e Análise de Recorrência

Hudson Fujikawa de Paula

Orientação: Prof. Dr. Renato Ishii

Área de Concentração: Sistemas de Computação



Faculdade de Computação
Universidade Federal de Mato Grosso do Sul
24 de Junho de 2016.

Pesquisa desenvolvida com auxílio financeiro da Coordenação de Aperfeiçoamento de
Pessoal de Nível Superior (CAPES).

Uma Abordagem para Classificação de Séries Temporais Baseada em Modelo Autorregressivo e Análise de Recorrência

Este exemplar corresponde à versão final da dissertação adequada às correções e sugestões apontadas pela Banca Examinadora.

Campo Grande, 24 de Junho de 2016.

Banca Examinadora:

- Prof. Dr. Renato Porfirio Ishii (FACOM/UFMS)- Orientador
- Prof. Dr. Eraldo Luis Rezende Fernandes (FACOM/UFMS)
- Prof. Dr. Ricardo Araújo Rios (DCC/UFBA)

Agradecimentos

Às minhas duas mães, dona Cristina, que sempre me incentivou enquanto estive por perto e, dona Beth, que sempre garantiu o suporte incondicional para o cumprimento de mais uma etapa desta jornada!

Ao professor Renato Ishii, pois sua orientação técnica e “parceira”, e sua amizade, foram determinantes para o cumprimento desta etapa, com tantos obstáculos, desta jornada!

A todos os meus amigos que, de alguma forma, prestaram suporte para o cumprimento desta etapa, em especial aos amigos do mestrado!

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro fundamental!

Conteúdo

Lista de Figuras	vi
Lista de Tabelas	viii
Lista de Abreviaturas	ix
1 Introdução	1
1.1 Contexto e Motivação	1
1.2 Objetivo	2
1.3 Estrutura do Texto	3
2 Análise de Séries Temporais	4
2.1 Elementos e Terminologias	5
2.2 Processos Geradores e Modelos	6
2.2.1 Séries temporais e ferramentas determinísticas	7
2.2.2 Séries temporais e ferramentas estocásticas	8
3 Classificação de Séries Temporais	10
3.1 Mineração de Dados - Síntese	10
3.2 Mineração de Dados de Séries Temporais - Síntese	14
3.2.1 Representação de séries temporais	15
3.2.2 A tarefa de classificar séries temporais	17
4 Uma Abordagem para Classificação de Séries Temporais Baseada em Modelo Autorregressivo e Gráfico de Recorrência	22
4.1 Abordagem Proposta	22

4.1.1	Representações autorregressiva e Gráfico de Recorrências para extração de características	24
4.1.2	Classificação por <i>Support Vector Machine</i> - SVM	28
5	Experimentos e Resultados	31
5.1	Metodologia	31
5.2	Experimentos	34
5.2.1	Classificação sobre características separadas	36
5.2.2	Classificação sobre características combinadas	36
5.2.3	Comparação com a abordagem referencial 1-NN	39
5.3	Análise dos Resultados	41
5.3.1	Influências das características na classificação	41
5.3.2	Sobre a comparação à abordagem 1-NN da literatura	43
6	Conclusões	46
6.1	Trabalhos Futuros	47
	Referências Bibliográficas	48

Lista de Figuras

2.1	Classificação dos processos geradores de séries temporais	6
2.2	Sistema de Rössler: a) como série temporal e b) reconstruído no espaço de fases. A noção de estados vizinhos no espaço de fases pode ser visualizada em b), onde os dois pontos pretos são vizinhos. O ponto cinza não é vizinho deles, considerando a distância imposta pelo círculo. Adaptado de [8, 19].	7
3.1	Taxonomia para representação de séries temporais. Adaptado de [31]	16
3.2	Diferença entre as séries temporais $X(t)$ e $Z(t)$ através das funções de distância (a) Euclidiana (<i>lockstep</i>), e (b) <i>Dynamic Time Warping</i> (elástica). Adaptado de [60] .	20
4.1	Representação e extração de características da abordagem proposta. Dentro do novo espaço de características a série temporal $X(t)$ é representada por $\{\Lambda, RQA\}$, onde Λ é o conjunto dos p coeficientes do modelo $AR(p)$ e, RQA , o conjunto de medidas extraídas pela Análise de Quantificação de Recorrência.	24
4.2	Sistema de Rössler: a) como série temporal, b) reconstruído no espaço de fases, com exemplo de um par de estados ε -vizinhos e c) no <i>Recurrence Plot</i> . Adaptado de [8, 19].	25
4.3	SVMs sobre dados distribuídos linearmente e não linearmente. a) SVM de margem suave sobre dados linearmente separáveis; b) dados não separáveis linearmente, e; c) SVM sobre dados não separáveis linearmente. Adaptado de [70].	29
5.1	Classificação de Séries Temporais em dois estágios.	31
5.2	Desempenho de classificação para cada conjunto de características.	36
5.3	Desempenho da classificação para conjuntos de características combinadas.	37
5.4	Desempenho da classificação sobre todas as características em estudo.	37
5.5	Relação entre desempenhos da melhor configuração de características da abordagem proposta e abordagens de referência.	39
5.6	Relação entre o desempenho das características propostas e o número de classes por conjunto de dados. As taxas de erros estão alinhadas com o eixo vertical esquerdo.	41

-
- 5.7 Variação de RQA-DET entre as séries temporais de cada conjunto de dados. Séries temporais determinísticas exibem um valor $DET \geq 95$, enquanto que séries temporais estocásticas exibem $DET < 95$ [8]. 42
- 5.8 Comparação entre a melhor configuração da abordagem proposta (eixos x) e as técnicas de referência (eixos y), onde os pontos representam os conjuntos de dados. As relações são $SVM(\Lambda_{max}, RQA)$ versus (a) 1-NN(Euclidiana), (b) 1-NN(DTW), e (c) 1-NN(DTW-W) 44

Lista de Tabelas

5.1	Conjuntos de dados do Repositório utilizados neste trabalho. UCR [10].	33
5.2	Desempenho da classificação sobre características individuais e combinadas. . . .	38
5.3	Comparação entre taxas de erros da abordagens proposta e de referência [10]. . .	40

Lista de Abreviações

AIC	<i>Akaike Information Criterion</i>
ANN	<i>Artificial Neural Network</i>
APCA	<i>Adaptative Piecewise Constant Approximations</i>
AR	<i>Auto Regressive process</i>
ARIMA	<i>Auto Regressive Integrated Moving Average</i>
ARMA	<i>Auto Regressive Moving Average</i>
BIC	<i>Bayesian Information Criterion</i>
CWT	<i>Continuous Wavelet Transform</i>
DBSCAN	<i>Density Based Spatial Clustering of Applications with Noise</i>
DET	<i>Determinism Rate</i>
DFT	<i>Discrete Fourier Transform</i>
DTW	<i>Dynamic Time Warping</i>
DWT	<i>Discrete Wavelet Transform</i>
ECG	<i>Electrocardiogram signal</i>
EEG	<i>Electroencephalogram</i>
EMD	<i>Empirical Mode Decomposistion</i>
FFT	<i>Fast Fourier Transform</i>
GLM	<i>Generalized Linear Model</i>
HHT	<i>Hilbert-Huang Transform</i>
HMM	<i>Hidden Markov Models</i>
kNN	<i>k-Nearest Neighbor</i>
LCSS	<i>Longest Common Subsequence</i>
LDA	<i>Linear Discriminant Analysis</i>
LOI	<i>Line of Identity</i>
MLP	<i>Multilayer Perceptron</i>
NLG	<i>Natural Language Generation</i>
P-SVM	<i>Potential Support Vector Machine</i>
PAA	<i>Piecewise Aggregate Approximation</i>
PLA	<i>Piecewise Linear Approximation</i>
RBF	<i>Radial Basis Funcions</i>

RP *Recurrence Plot*

RQA *Recurrence Quantification Analysis*

RR *Recurrence Rate*

SAX *Symbolic Aggregate Approximation*

SVD *Singular Value Decomposition*

SVM *Support Vector Machine*

WDTW *Weighted Dynamic Time Warping*

WNN *White Neural Network*

Resumo

Atualmente, os estudos sobre o problema da classificação de séries temporais tem se concentrado em elaborar abordagens em dois formatos: baseado em funções de distância entre itens do conjunto de dados, e baseado em um procedimento de dois estágios, onde as séries temporais são transformadas em vetores de características, permitindo o emprego de técnicas de classificação convencionais. Neste contexto, não tem sido notado na literatura estudos baseados na análise de propriedades intrínsecas do processo gerador da série temporal como, por exemplo, o determinismo. Neste trabalho de mestrado, é proposta uma abordagem para o problema da classificação de séries temporais, projetada em dois estágios e baseada na análise de propriedades intrínsecas de determinismo e de estocasticidade. Primeiramente, cada série temporal é processada pelo modelo autorregressivo (AR) e pelo Gráfico de Recorrências, para modelar as influências estocásticas e determinísticas, presentes nas séries temporais. Posteriormente, são extraídas características, a partir da nova representação, que compõem o novo espaço característico. Para a classificação em si, optou-se pelo SVM em seu formato convencional. Tomou-se como abordagem de referência da literatura, o classificador 1-NN com funções de distâncias Euclidiana, DTW e DTW otimizado por janela de busca. Os experimentos foram executados sobre os conjuntos de dados do repositório UCR. Os resultados finais mostram que o desempenho de classificação é competitivo, ou superior, à melhor configuração 1-NN em 19 de 41 conjuntos de dados. Não obstante, os resultados evidenciam, também, a necessidade de uma investigação mais aprofundada sobre as influências das propriedades intrínsecas, e outras técnicas da área de análise de séries temporais, quando aplicadas na tarefa de classificação.

Palavras-chave: classificação de séries temporais, análise de séries temporais, modelo autorregressivo, Gráfico de Recorrência, determinismo, Análise de Quantificação de Recorrência.

Abstract

Recently, the research on time series classification problem has focused on designing approaches based on distance functions (similarity measures) or two-step procedure. Within this context, we have found no study that considers the intrinsic properties of the time series generation process, namely, determinism and stochasticity, when discriminative modeling. This work presents an approach to the time series classification problem. This approach, which uses a two-step procedure, is based on influences analysis of the determinism and stochasticity intrinsic properties. Firstly, each time series of the dataset is fitted by autoregressive model (AR) and plotted in recurrence plot in order to analyze the stochasticity and determinism influences, respectively, embedded into the time series. Then, the feature extraction is performed by AR coefficients extraction and Recurrence Quantification Analysis. The feature vectors are used as input to a standard SVM classifier. In order to evaluate the effectiveness of the proposed approach, we compare the experimental results with results of the literature reference approaches, the 1-NN classifier with Euclidean distance and Dynamic Time Warping functions. The experiments were performed with datasets of the UCR repository. The results show that the approach performance is similar to (or better than) the best configuration 1-NN, in 19 of 41 datasets. Nevertheless, the achieved results reveal the need for further research on the influences analysis of generating process intrinsic properties, as well as other time series analysis techniques, when employing them in the time series classification task.

Keywords: time series classification, time series analysis, autoregressive model, Recurrence Plots, determinism, Recurrence Quantification Analysis.

Capítulo 1

Introdução

1.1 Contexto e Motivação

A facilidade em reconhecer rostos, compreender palavras pronunciadas e manuscritas, ou identificar se uma fruta está madura pelo seu cheiro, por exemplo, demonstra a habilidade do ser humano em reconhecer padrões e características em formas e objetos. De fato, essa é uma habilidade pela qual o ser humano se destaca, principalmente pela capacidade de explorá-la e desenvolvê-la facilmente [1]. Naturalmente, com o advento dos computadores, surgiu a possibilidade de se automatizar tais tarefas, uma vez que o processo de reconhecimento de padrões, eficiente e confiável, seria imensamente útil em diversos domínios de aplicações [2].

Por outro lado, a facilidade com a qual o ser humano lida com tarefas de reconhecimento de padrões camufla processos complexos inerentes à identificação, ao reconhecimento e, em geral, à análise dos dados em si. Lidar com tal complexidade computacionalmente é um grande desafio que atrai cientistas de diversas áreas do conhecimento. Na época em que grandes empresas passaram a investir efetivamente em pesquisas nessa área, notavelmente a partir de 1960, soluções eram fortemente vinculadas ao domínio da aplicação.

À medida que a evolução computacional contínua, em processamento e em produção e consumo de dados, ganhou proporções sem precedentes, pesquisadores, principalmente cientistas na área da computação e estatística, passaram a almejar algo diferente: soluções generalistas e inovadoras, e não somente resoluções especialistas de domínio [3–5]. De fato, a capacidade computacional disponível atualmente viabiliza procedimentos de alta complexidade, de análise e extração de informações tácitas, inviáveis manual e visualmente.

Além disso, a quantidade massiva de dados sendo produzida hoje em redes sociais, censos demográficos, monitoramento de eventos da natureza, comércio, mercado financeiro, sistemas de vigilância, dentre outros, reafirma o grande desafio de se descobrir e utilizar informações de forma eficiente. Em muitas das circunstâncias previamente apontadas, constitui-se um tipo diferente de dado, a *série temporal*.

As séries temporais são observações tomadas sequencialmente no tempo que podem, em geral, apresentar estruturas de correlação sobre observações adjacentes. Tal característica restringe a utilização de técnicas estatísticas convencionais. A área de Análise de Séries

Temporais fornece um ferramental matemático, capaz de identificar e modelar tal correlação [6, 7].

Em geral, o objetivo da modelagem é identificar propriedades de interesse para aproximar o processo gerador dos dados. Assim, é possível estimar valores ainda não conhecidos e tomar decisões antecipadamente. Nesse contexto, em [8], é apresentada uma abordagem para predição de eventos futuros através da modelagem de séries temporais. A abordagem propõe uma taxonomia que conduz o procedimento de escolha da técnica de modelagem mais adequada, considerando propriedades como o determinismo, a linearidade, e a estacionariedade, intrínsecas do processo gerador da série temporal. Com o mesmo propósito, o trabalho realizado em [9], propõe uma abordagem para modelagem híbrida, cuja propriedade de determinismo é analisada como uma componente da série temporal.

Por outro lado, a modelagem no contexto de Mineração de Dados de Séries Temporais está relacionada a tarefas como a representação e a classificação. Basicamente, neste contexto, os trabalhos buscam por representações com maior potencial informativo, através da redução de itens redundantes do conjunto de dados, através de transformações que reduzem o espaço de características, ou através da “expansão” dos dados para evidenciar informações implícitas relevantes.

Durante a análise da literatura neste trabalho de mestrado, buscou-se por estudos que relacionassem a análise das propriedades intrínsecas do processo gerador, principalmente dirigidos pela taxonomia proposta em [8] ou similares, com a tarefa de classificação de séries temporais. Entretanto, nenhum trabalho foi encontrado. Nesse sentido, nota-se na literatura, que há uma lacuna entre a tarefa de classificação de séries temporais e a análise das propriedades intrínsecas do processo gerador, como realizado em [8, 9].

1.2 Objetivo

O principal objetivo deste Trabalho de Mestrado, é propor uma abordagem para o problema da classificação de séries temporais, inspirando-se nas metodologias apresentadas em [8,9]. Especificamente, a abordagem proposta é baseada na classificação em dois estágios: as séries temporais são modeladas através de técnicas apropriadas para as propriedades intrínsecas de determinismo e estocasticidade, Gráfico de Recorrência e modelo autorregressivo (AR), respectivamente. Em seguida, características inerentes às novas representações são extraídas, compondo, assim, o espaço característico. No segundo estágio, o classificador SVM é empregado para realizar a classificação, propriamente dita.

Para analisar empiricamente a abordagem proposta, é necessária a definição de alguns objetivos específicos:

1. Analisar o desempenho sobre os conjuntos de dados do Repositório UCR [10];
2. Analisar estratégias para a definição da ordem do modelo AR;
3. Analisar o potencial discriminativo das características extraídas através da observação das influências determinística e estocástica, presentes nas séries temporais;
4. Analisar os efeitos da combinação das características propostas;

5. Comparar as abordagens proposta e de referência da literatura.

A partir dos objetivos específicos, notou-se que a abordagem proposta alcança seu melhor desempenho sobre uma combinação dos conjuntos de características proposta, e que é possível verificar que a abordagem proposta é superior, ou similar, em 19 de 41 conjuntos de dados. Dessa forma, qualificou-se a abordagem proposta, como uma abordagem promissora.

1.3 Estrutura do Texto

Esta Dissertação de mestrado se organiza da seguinte maneira: no Capítulo 2 são introduzidos os fundamentos sobre a análise de séries temporais, apresentando elementos básicos e uma breve explicação sobre processos geradores e suas propriedades intrínsecas.

No Capítulo 3, a classificação é tomada como uma tarefa de Mineração de Dados. Dessa forma, alguns dos principais conceitos de mineração são brevemente relacionados.

A abordagem proposta neste trabalho, bem como as técnicas envolvidas são apresentadas no Capítulo 4.

O Capítulo 5 é reservado para a análise experimental, descrevendo-se a metodologia, uma breve descrição dos conjunto e análise dos resultados.

No Capítulo 6 são apresentadas as conclusões gerais e direções para trabalhos futuros.

Capítulo 2

Análise de Séries Temporais

Uma série temporal é uma coleção de observações tomadas sequencialmente ao longo do tempo [11]. Tais observações, capturadas e analisadas em ordem cronológica, já eram realizadas desde os primórdios das ciências naturais, quando a astronomia babilônica predizia eventos astronômicos, observando as posições das estrelas e planetas [6]. Atualmente, observações desse tipo são realizadas em diversos domínios, por exemplo: na economia, observando-se exportações mensais; no estudos de fenômenos físicos, como na meteorologia e geofísica; em estudos demográficos; no controle de processos de produção industrial; e na medicina, com sinais biológicos como eletroencefalograma e eletrocardiograma [12, 13].

A característica que distingue uma série temporal de qualquer outra coleção de dados é a dependência que normalmente existe entre suas observações adjacentes no tempo. Tal dependência restringe o uso de métodos estatístico convencionais, fundamentados sobre dados independentes e identicamente distribuídos, de modo que a eficácia dos resultados obtidos sobre dados temporais não é garantida. Dessa forma, tem-se novos desafios conceituais e técnicos ao se trabalhar com dados temporais. Tais desafios são sistematicamente abordados pela Análise de Séries Temporais [6]. Os principais objetivos nessa área, segundo [11, 12, 14], são:

1. Sumarizar os dados, como uma análise inicial das propriedades explícitas das séries temporais. Por exemplo, tendência e sazonalidade. Construir gráficos, histogramas e diagramas de dispersão também auxilia na compreensão das variações nos dados;
2. Controlar o processo gerador. Considerando conceitos de sistemas dinâmicos, é possível determinar um modelo entrada-saída e, se necessário, adequar seus parâmetros e a série temporal de entrada $X(t)$, para se obter a série temporal de saída desejada $Y(t)$;
3. Investigar e modelar o processo gerador da série temporal;
4. Compreender as influências que uma série temporal pode exercer sobre outras. Por exemplo, relações de causalidade;
5. Classificar, como uma tarefa de aprendizado de máquina baseada em reconhecimento de padrões;
6. Predizer valores futuros da série temporal, através de valores atuais e passados.

Em todos os objetivos, modelos matemáticos probabilísticos e/ou determinísticos são empregados para representar o comportamento dos dados que evoluem com o tempo.

2.1 Elementos e Terminologias

Uma série temporal $X(t) = \{x_1, x_2, \dots, x_n\}$ representa a realização de um processo cujos valores observados formam uma trajetória no tempo [7]. Se T é um conjunto discreto de instantes de tempo equidistantes, tal que $t \in T$, então $X(t)$ é dita ser discreta. Diferentemente, uma série temporal é dita ser contínua, quando suas observações são tomadas continuamente no tempo, por exemplo, quando $T = [0, 1]$.

Naturalmente, o uso do computador para processar traços e sinais contínuos (por exemplo, eletrocardiograma) exige a digitalização dos valores de modo que uma série temporal discreta possa surgir da amostragem dos valores no tempo contínuo [13]. Uma série temporal discreta também surge com a agregação de valores de um certo período (por exemplo, total de vendas mensais). Tanto a amostragem quanto a agregação, no entanto, resultam em perda de informação, um vez que dados são descartados ou abstraídos, respectivamente [15].

Um outro aspecto importante é o relacionamento encontrado, frequentemente, entre duas (ou mais) séries temporais em um dado domínio. Nesse sentido, define-se $M(t) = \{X_1(t), X_2(t), \dots, X_q(t)\}$ como uma série temporal multivariada, sendo q o número de componentes univariadas. Por exemplo, é possível relacionar mudanças na temperatura dos mares com a quantidade de peixes encontrados neles, ao longo de alguns anos [6]. Nesse caso, $q = 2$ corresponde às componentes univariadas temperatura das águas do mar e quantidade de peixes observados, ambos ao longo do tempo.

A partir dos elementos de terminologia previamente expostos, e restringindo o escopo deste trabalho de mestrado a séries temporais univariadas, é possível explorar e buscar por informações que podem, ou não, estar presentes nos dados de séries temporais. Uma análise clássica sobre o comportamento dos dados propõe um modelo para séries temporais no qual é possível analisar três comportamentos inerentes a $X(t)$: a tendência, a sazonalidade e a aleatoriedade [7, 16]. Cada comportamento constitui uma componente do modelo, tal que

$$X(t) = T(t) + S(t) + Z(t).$$

A tendência $T(t)$, representa a trajetória da série temporal a longo prazo, ou seja, uma função suave do tempo influenciada por fatores observados em períodos longos de tempo e que altera a média. A componente sazonal $S(t)$ expõe o comportamento cíclico (periódico) observado em períodos curtos de tempo. Por fim, se removidas $T(t)$ e $S(t)$ de $X(t)$, resta a componente aleatória $Z(t)$, que é caracterizada pelo comportamento puramente aleatório, ou um ruído branco¹. Um material detalhado sobre a aplicação deste modelo e suas variações, pode ser encontrado em [7].

¹Uma série de valores aleatórios não correlacionados, com média zero ($\mu = 0$) e variância finita ($var(Z(t)) = \sigma^2$).

2.2 Processos Geradores e Modelos

As séries temporais possuem outras propriedades além daquelas exploradas pelo modelo clássico. Essas propriedades ficam evidentes quando se investiga os processos geradores dos dados. Ao reconhecer o processo gerador de uma série temporal, é possível escolher técnicas de análise mais adequadas aos comportamentos que seus dados podem exibir.

No estudo apresentado em [8] são apontadas três propriedades intrínsecas, o determinismo, a linearidade, e a estacionariedade. Nesse estudo, foi proposto um roteiro, apresentado na Figura 2.1, que direciona a análise das propriedades intrínsecas da série temporal. Através deste roteiro, é possível selecionar técnicas de modelagem que melhor se adequam às propriedades intrínsecas reveladas nos dados.

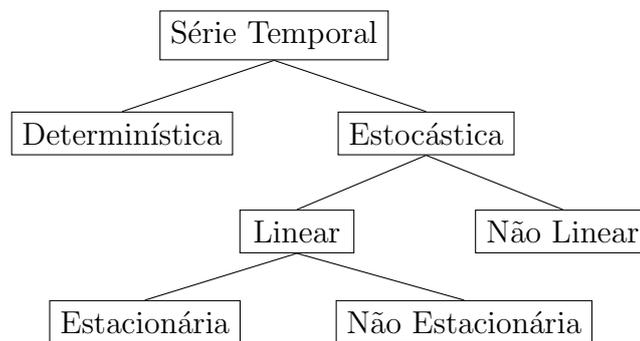


Figura 2.1: Classificação dos processos geradores de séries temporais. Adaptado de [8].

O contexto deste trabalho de mestrado abrange o estudo dos comportamentos determinístico e estocástico intrínsecos das séries temporais. Não obstante, há a necessidade em compreender as demais propriedades, a linearidade e a estacionariedade, uma vez que também são fundamentais no estudo de séries temporais quando se emprega técnicas de análise oriundas da área de sistemas dinâmicos.

Um sistema é um conjunto de “elementos” que interagem entre si. O conceito de “dinâmica” implica em movimento ou mudança. Nesse sentido, um sistema dinâmico pode ser definido como um conjunto de elementos que interagem entre si ao longo do tempo [17].

O “elemento” do sistema pode ser tomado como um conjunto de variáveis que compõem um “estado”. Este “estado” representa um objeto (seus atributos) em um determinado instante de tempo, em um dado problema. A “dinâmica”, por sua vez, implica em uma regra que estabelece alguma estrutura de correlação entre estados adjacentes no tempo. Assim, um sistema dinâmico é composto por um conjunto de estados e uma regra que determina um estado atual em função dos estados passados. Esta definição deve ser considerada quando o sistema dinâmico for tratado em espaço de fases.

As definições prévias de sistema dinâmico, apesar de breves, são suficientes para seguir no contexto deste trabalho de mestrado. Nas seções seguintes são apresentadas, sucintamente, as propriedades intrínsecas determinismo e estocasticidade, e as técnicas de análise e modelagem mais adequadas de interesse deste trabalho de mestrado. Tais propriedades e técnicas são expostas sob o ponto de vista depreendido da fundamentação teórica para o roteiro de análise apresentado em [8].

2.2.1 Séries temporais e ferramentas determinísticas

Uma série temporal determinística possui comportamentos recorrentes, cujos valores futuros são determinados exatamente por um modelo matemático, sem incertezas [13]. Tais sistemas, por exemplo a movimentação das galáxias, se comportam deterministicamente de modo a ser possível prever um estado futuro a qualquer momento, uma vez que as condições iniciais são exatamente conhecidas.

Na prática, no entanto, alguns desses sistemas são muito sensíveis a flutuações e pequenas perturbações nas condições iniciais. Nesses sistemas, referidos como caóticos, previsões de longo prazo podem ser impossíveis [18, 19]. Além disso, a análise através de modelos lineares, ou seja, da parcela temporal e de propriedades de segunda ordem (por exemplo, o correlograma) poderá mostrar uma série temporal de comportamento completamente aleatório [15]. Para compreender tais variações, técnicas da área de sistemas dinâmicos e da Teoria do Caos podem auxiliar na modelagem e análise dos comportamentos embutidos nas séries temporais [20]. Ao revelar essas informações tácitas, é possível identificar padrões (repetições) e, assim, realizar tarefas como classificação ou agrupamento.

Uma abordagem bem conhecida é a imersão no espaço de fases baseada em atrasos de tempo, proposta em [21]. Basicamente, o objetivo é transformar os dados de uma série temporal univariada, em um conjunto de estados (pontos) dentro de um espaço Euclidiano multidimensional: a série temporal $X(t) = \{x_1, x_2, \dots, x_n\}$ pode ser reconstruída num espaço com m dimensões, onde a observação $x_i \in X(t)$ passa a ser representada pelo ponto $x_i(m, \tau) = (x_i, x_{i+\tau}, \dots, x_{i+(m-1)\tau})$, tal que τ representa a fase (atrasos de tempo – *time delay*). Tal abordagem foi motivada pela necessidade em desdobrar os dados da série temporal, no espaço de fases, para a análise do atrator do sistema dinâmico em busca de comportamentos somente revelados em um espaço de maior dimensão.

Para a aplicação dessa abordagem, é necessário ter em mãos as dimensões de separação τ e de imersão m . Neste trabalho de mestrado, são utilizadas as técnicas *Auto-Mutual Information* (AMI) e *False Nearest Neighbors* (FNN), para estimar tais dimensões, respectivamente.

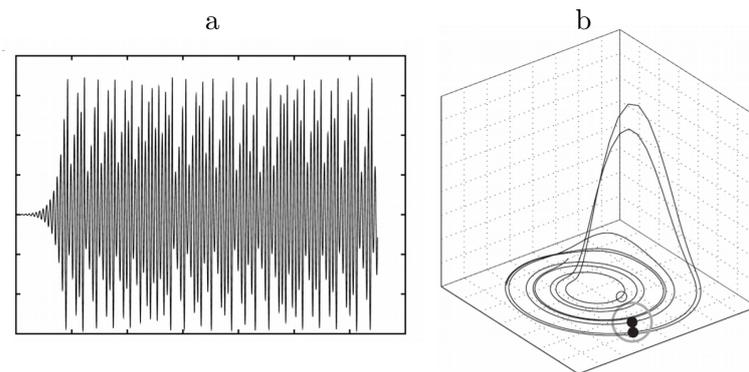


Figura 2.2: Sistema de Rössler: a) como série temporal e b) reconstruído no espaço de fases. A noção de estados vizinhos no espaço de fases pode ser visualizada em b), onde os dois pontos pretos são vizinhos. O ponto cinza não é vizinho deles, considerando a distância imposta pelo círculo. Adaptado de [8, 19].

Em [22] é proposta a abordagem que se utiliza da técnica AMI para estimar τ . Sumariamente, AMI é aplicada sobre a série temporal diversas vezes, com deslocamentos de

tempo (*lags*) incrementados sequencialmente. A dimensão de separação τ é determinada pelo *lag* responsável pelo primeiro valor mínimo de AMI calculado sobre a série temporal. Em [23] é proposta a abordagem para se estimar a dimensão de imersão m baseada na técnica FNN. Basicamente, calcula-se os vizinhos mais próximos para cada ponto da trajetória no espaço de fases, iterativamente, incrementando o número de dimensões do espaço. Quando o incremento de dimensão altera (aumenta) a distancia entre pontos, tem-se pontos falsos vizinhos, indicando que são necessários mais incrementos no número de dimensões. Se, porém, o incremento na dimensão não revela nenhum falso vizinho, então a dimensão atual é a dimensão de imersão adequada.

Um exemplo de uma série temporal desdobrada no espaço de fases pode ser visualizado nas Figuras 2.2 (a) e (b). A série temporal exibida na Figura 2.2 (a) é obtida do sistema de Rössler configurado com os parâmetros $a = 15$, $b = 0,2$, e $c = 10$.

2.2.2 Séries temporais e ferramentas estocásticas

Uma série temporal é dita estocástica se representa uma realização de um processo estocástico, $\{X(t), t \in \mathbb{N}\}$. Tal processo é considerado como um fenômeno estatístico que evolui ao longo do tempo, de modo que as observações podem ser descritas em termos de distribuição de probabilidade. Os dados futuros são parcialmente determinados por valores passados e/ou determinados de forma aleatória [11].

Um sistema estocástico é caracterizado como linear ou não linear. Se a série temporal exibe um comportamento linear, então existe uma função $v(j)$ que determina um relacionamento linear entre a entrada $X(t)$ e a saída $Y(t)$, na forma $Y(t) = \sum_{j=0}^{\infty} v(j)X(t-j)$ [11]. Os objetivos da análise sobre tais séries temporais se concentram em estimar $v(j)$ supondo que $X(t)$ e $Y(t)$ sejam conhecidas; prever valores de $Y(t)$; e controlar $Y(t)$ ajustando-se convenientemente $X(t)$ a fim de evitar os efeitos de perturbações no sistema [7, 15].

Em diversas situações, no entanto, principalmente em sistemas reais, as séries temporais podem ser geradas por processos não lineares. As observações de tais séries apresentam variações no comportamento durante a coleta, como mudanças repentinas, não invertibilidade e variância condicional. Nesses casos, as observações são obtidas através de combinações não lineares entre valores passados e ruídos, compondo, portanto, séries temporais mais difíceis de se estudar e modelar [8]. Tais séries temporais podem ser modeladas por técnicas não lineares, como o modelo autorregressivo com heteroscedasticidade condicional (ARCH) e por filtros que removem a componente não linear, tornando a série linear [7].

As séries temporais lineares podem, ainda, serem caracterizadas como estacionárias ou não estacionárias (Figura 2.1). As séries estacionárias exibem um estado particular de equilíbrio estatístico, onde a trajetória da série temporal evolui aleatoriamente ao redor de uma média constante. Se a distribuição de probabilidade associada aos n valores de $X(t)$, dado o conjunto de instantes $T = \{t_1, t_2, \dots, t_n\}$, for a mesma considerando um deslocamento de k sobre T , tal que $T = \{t_{1+k}, t_{2+k}, \dots, t_{n+k}\}$, de modo que ambas média e variância permaneçam constantes, então a série é dita estritamente estacionária. De modo menos restritivo, se uma série temporal possui média constante e autocovariância que dependa somente de k , então a série temporal é dita fracamente estacionária, ou estacionária de segunda ordem. Tais séries temporais são modeladas através de técnicas estatísticas, como a

família clássica de modelos de erro (ou de regressão), modelos autorregressivos (AR), modelos de médias móveis (MA), e modelos autorregressivos e de médias móveis (ARMA) [7, 11].

Diferentemente, uma série não estacionária geralmente apresenta comportamento evolutivo, principalmente devido à presença de tendência. Nesses sistemas, a média tende a aumentar ou diminuir quando novas observações subjacentes são consideradas. Por exemplo, as séries temporais econômicas apresentam tendências, sendo o caso mais simples aquele em que a série evolui ao redor de uma reta, de inclinação positiva ou negativa (tendência linear). Outro caso é o de sistemas com comportamentos explosivos, como uma colônia de bactérias que cresce muito rapidamente [7, 11]. As séries não estacionárias podem ser modeladas por métodos como autorregressivos integrados e de médias móveis (ARIMA) ou, simplesmente, através de filtros que removem a componente tendência, tornando a série estacionária [24]

Como visto previamente, o comportamento estocástico e suas variações podem ser analisados com melhor efetividade, empregando-se técnicas adequadas, considerando o roteiro de análise proposto em [8]. Uma dessas técnicas é de especial interesse para este trabalho de mestrado, o modelo autorregressivo (AR).

O modelo AR é baseado na ideia de que o valor atual x_i da série temporal $X(t)$ pode ser explicado através da relação entre p valores anteriores, $x_{i-1}, x_{i-2}, \dots, x_{i-p}$ e um termo aleatório ϵ_i . Formalmente, uma série temporal $X(t)$ pode ser modelada através de um processo autorregressivo de ordem p , referido por $AR(p)$ e definido pela soma ponderada

$$x_i = \phi_1 x_{i-1} + \phi_2 x_{i-2} + \dots + \phi_p x_{i-p} + \epsilon_i$$

onde ϵ_i denota um processo puramente aleatório de média zero e variância constante, e $\{\phi_p\}$ o conjunto dos coeficientes [6, 15]. O modelo assemelha-se ao modelo de regressão múltipla, porém $X(t)$ é regredida sobre suas p observações passadas, ao invés de variáveis independentes.

A aplicação do modelo sobre uma série temporal exige a definição da ordem p . Esse valor pode ser estimado através das abordagens que utilizam critério de informação, como o *Akaike Information Criteria* (AIC) [25] e o *Bayesian Information Criteria* (BIC) [26]. Após definir a ordem p , pode-se seguir à estimação dos coeficientes, para o ajuste do modelo. Para esse fim, geralmente são empregados os estimadores de Máxima Verossimilhança e de Yule-Walker [11]. Além destes, a estimação por meio do método dos Mínimos Quadrados também é empregada embora, em teoria, não seja adequado no contexto de autorregressão [15]. Em [11, 15] é possível encontrar descrições detalhadas sobre a aplicação e vantagens/desvantagens de cada um dos métodos previamente citados, tanto para a seleção do modelo quanto para a estimação dos coeficientes.

Neste capítulo foram apresentados, brevemente, alguns conceitos e terminologias abrangidos pela Análise de Séries Temporais e suas intersecções com outras grandes áreas de estudo, como Sistemas Dinâmicos e Estatística. Essa fundamentação teórica é suficiente para prosseguir no escopo deste trabalho. No Capítulo 3 é apresentada a segunda parte da teoria envolvida no escopo deste trabalho, a Classificação como uma tarefa da Mineração de Dados de Séries Temporais.

Capítulo 3

Classificação de Séries Temporais

Em seu aspecto amplo, a mineração de dados pode ser vista como um arcabouço que fornece diversas tarefas para um roteiro de análise. Cada tarefa, por si só, compõe grandes campos de estudos. Neste trabalho de mestrado, o procedimento de classificação é tomado como uma tarefa que compõe a mineração de dados. Especificamente, neste Capítulo apresenta-se alguns referenciais na área de classificação de séries temporais e a relação com outras tarefas, como a representação e o agrupamento de séries temporais. Adicionalmente, são apresentados trabalhos da literatura que exploram alguns dos conceitos expostos, e que se relacionam a este trabalho.

3.1 Mineração de Dados - Síntese

Desde o advento das tecnologias computacionais e de informação, grande esforço tem sido empregado para tornar os computadores tomadores de decisões. O primeiro sistema a simular, realmente, protocolos humanos de resolução de problemas foi, provavelmente, o *General Problem Solving*, em 1959 [27]. De fato, o grande objetivo, então, era fazer o computador “pensar humanamente” [28]. Decorrente disso, a pesquisa em reconhecimento de padrões se desenvolveu significativamente a partir dos anos de 1960, por uma perspectiva mais multidisciplinar. Pesquisadores ingressavam nesse campo com problemas reais, de domínios diversos. Tais esforços, somados às teorias estatísticas e da probabilidade, contribuíram para a formação de uma base sólida que se desenvolve continuamente [4, 5].

Mais recentemente, devido principalmente aos avanços de recursos computacionais, pesquisadores têm se voltado mais para o desenvolvimento de metodologias, ou seja, mais enfoque sobre os meios, buscando generalizações e metodologias inovadoras, do que resoluções especialistas de domínio [3, 5]. De fato, a capacidade computacional disponível atualmente viabiliza procedimentos de alta complexidade, de análise e extração de informações tácitas, inviáveis manual e visualmente.

Além disso, a quantidade massiva de dados sendo produzida hoje em redes sociais, censos demográficos, monitoramento de eventos da natureza, comércio, mercado financeiro, sistemas de vigilância, dentre outros, reafirma o grande desafio de se descobrir e utilizar informações de forma eficiente. A mineração de dados é a grande área de pesquisa e aplicação que lida

com este desafio. Esta compõe um arcabouço de tarefas e define, no contexto deste trabalho, o campo de Aprendizado de Máquina [29]. Para uma fundamentação conceitual e teórica completa sobre o campo de aprendizado de máquina, recomenda-se [28, 30].

Ao minerar dados, busca-se basicamente por padrões através da descoberta de informações ou estruturas implícitas neles [31]. Para isso, a Mineração de Dados fornece uma coleção de tarefas para análise, podendo ser empregadas conforme a adequabilidade de cada uma ao problema em questão. Em [32], as tarefas são enquadradas em algumas categorias:

Análise Exploratória: são técnicas que visam explorar os dados, visualmente e interativamente, como análise através de gráficos e observações tabuladas;

Modelagem Descritiva: as tarefas revelam informações sobre os dados ou sobre o respectivo processo gerador, como segmentação, análise de grupos e agrupamento;

Modelagem Preditiva: abrange tarefas que permitem construir modelos com o objetivo de estimar valores de variáveis, a partir do conhecimento dos valores de outras. São tarefas de regressão/classificação e predição de valores futuros (previsão);

Descoberta de Padrões e Regras: abrange técnicas cujo objetivo é detectar padrões e comportamentos nos dados.

Dentre as tarefas enquadradas nas categorias listadas previamente, algumas são relevantes para este trabalho de mestrado, como a regressão, o agrupamento e a classificação. O pré-processamento pode ser considerado como uma etapa onipresente na mineração de dados. Detalhes sobre outras tarefas, tais como descoberta de regras de associação, consultas por conteúdo, visualização, e a mineração no contexto de grandes bases de dados (*Data Warehouse*) podem ser encontrados em [33, 34].

Pré-processamento

O pré-processamento dos dados é um passo necessário para se obter resultados efetivos da análise. Os dados são provenientes de observações de fenômenos que estão sujeitos a ruídos, observações discrepantes ao fluxo natural (*outliers*), observações coletadas em escalas distintas, além de falha durante a coleta (*missing values*). Classificadores e técnicas de agrupamento podem ter seus resultados viesados, incoerentes ou, ainda, sem sentido, se a tarefa de pré-processamento não for empregada.

Os dados de séries temporais também estão sujeitos a ruídos, *outliers*, diferença de escala, e valores faltantes. Algumas técnicas de processamento de sinais, como filtros digitais, transformadas ou técnicas de decomposição que propiciam a manipulação de funções de base [35, 36], podem ser empregadas para lidar com séries temporais ruidosas. Além disso, técnicas de interpolação podem ser empregadas para lidar com valores faltantes [31].

A diferença em escalas de amplitude entre séries temporais, ou mesmo entre atributos de dados estáticos, é um problema que pode ser resolvido com técnicas de normalização. As mais empregadas são *min-max* e de média zero (*z-core*) [33].

Muitas técnicas de mineração se baseiam no conceito de distância, ou similaridade, entre os dados. A distância Euclidiana, dentre outras medidas, é uma técnica que exige normalização [33] e filtragem de ruídos [37,38]. Uma revisão detalhada em pré-processamento de dados pode ser encontrada em [39].

Regressão

A tarefa de regressão tem como objetivo modelar o relacionamento entre variáveis de um determinado problema. Com tal modelo, é possível analisar comportamentos e tendências sobre os dados e, principalmente, prever o valor de alguma variável de interesse.

Uma técnica bem conhecida é a regressão linear, onde as dependências são modeladas a partir da combinação linear entre observações e parâmetros auxiliares. O caso comum é quando o problema pode ser definido na forma $y = f(x)$, onde y pode ser estimado por $y = \varphi_1 x + \varphi_2 + \xi$. Nesse caso, o processo de regressão se resume em estimar os parâmetros (coeficientes) φ_{1i} e φ_{2i} para cada atributo preditivo x_i , representando o efeito que esse atributo tem sobre a variável de resposta y [40]. Com os parâmetros estimados, é possível prever valores desconhecidos de y .

Outra técnica muito conhecida, a regressão não linear, envolve combinações não lineares entre atributos preditivos e parâmetros. Redes Neurais Artificiais de propagação direta (*feed-forward Artificial Neural Networks* - ANN) e Máquinas de Vetores de Suporte (*Support Vector Machines* - SVM) são exemplos de técnicas que empregam regressão não linear para prever valores de interesse [31].

Agrupamento (*Clustering*)

Clustering é uma tarefa descritiva empregada para revelar automaticamente grupos (*clusters*) inerentes ao conjunto de dados. Esta tarefa é realizada sem conhecimento prévio dos dados, baseando-se em identificação de padrões ou funções de custo [2]. Sem rótulos para apoiar a discriminação dos itens do conjunto de dados, caracteriza o que se define como Aprendizado de Máquina Não Supervisionado.

Em geral, os grupos são revelados a partir de funções de custo, como densidade ou distância entre os itens do conjunto de dados. O agrupamento através da densidade encontra aglomerados densos em regiões com população relativamente esparsa. Uma técnica muito conhecida é o algoritmo *Density Based Spatial Clustering on Applications with Noise* (DBSCAN). Quando se emprega medidas de distância, os *clusters* revelados são distantes entre si, mas cada um é composto por itens próximos no conjunto de dados. O algoritmo mais utilizado é o *k*-Médias (*k-Means*). Adicionalmente, existem outras técnicas, como *clustering* hierárquico, de particionamento, e o baseado em modelos. Para mais detalhes, recomenda-se [33].

Classificação

A tarefa de classificação tem como objetivo construir um modelo discriminativo que seja capaz de inferir, ou atribuir, rótulos de classes a itens de dados. O treinamento é o processo pelo qual algoritmos de aprendizado de máquina são capazes de construir, ou induzir, tal modelo considerando dados previamente rotulados. Assim, basicamente, pode-se dizer que, em um determinado domínio de problema, a tarefa de classificação se caracteriza por induzir um modelo discriminativo sobre itens do conjunto de dados, de modo que este seja capaz de inferir a classe de um novo item de dado. Esta tarefa preditiva pode empregar técnicas de regressão para a construção do modelo, porém, a classificação é empregada para prever valores numéricos discretos ou nominais categóricos – ou seja, valores de rótulos de classes.

Formalmente, seja x um item de um conjunto de dados e $c \in \mathcal{C}$, a classe associada a x , sendo $\mathcal{C} = \{1, 2, \dots, l\}$ as classes definidas no problema. O conjunto de treinamento \mathcal{T} é o conjunto dos r pares definido por $\mathcal{T} = \{(x_1, c_1), (x_2, c_2), \dots, (x_r, c_r)\}$. Existe um conjunto \mathcal{H} de funções f que representam relacionamentos hipotéticos em cada par de \mathcal{T} . O problema é, portanto, definir \mathcal{H} e, principalmente, uma função $f^* \in \mathcal{H}$ que descreva (modele), com eficácia, o relacionamento $(x_i, c_i) \in \mathcal{T}$, para $\forall i \in \{1, 2, \dots, r\}$, minimizando, tanto quanto possível, o erro quando se atribui classes aos itens, durante o treinamento. Dessa forma, dado um novo item de dado $x_q \notin \mathcal{T}$, $f^*(x_q) = c_q$, onde $c_q \in \mathcal{C}$ é a classe mais adequada a x_q .

Em geral, as soluções têm como principal objetivo buscar por atributos, características, ou padrões, nos itens que melhor discriminem as classes entre si. Essa busca, direcionada por itens rotulados, compõe a definição de Aprendizado de Máquina Supervisionado. Em [41] é realizada uma revisão das técnicas mais empregadas sob a seguinte categorização:

Algoritmos baseados em lógica: abrange técnicas como Árvores de Decisão (*Decision Trees*) e baseadas em regras;

Técnicas baseadas em *perceptron*: abrange desde técnicas clássicas baseadas em *perceptron* até as Redes Neurais Artificiais atuais;

Algoritmos de aprendizado estatístico: compreende técnicas que modelam a probabilidade de um item pertencer a cada uma das classes, como a Análise de Discriminante Linear (*Linear Discriminant Analysis* - LDA), métodos de Entropia Máxima e Redes Bayesianas;

Aprendizado baseado em instâncias: abrange técnicas cuja indução, ou processo de generalização, é realizada durante a classificação, conhecido como aprendizado tardio (*lazy-learning*). O algoritmo dos k -Vizinhos Mais Próximos (*k-Nearest Neighbors* - k -NN) é um exemplo;

Máquina de vetores de suporte: esta técnica define um hiperplano através de vetores compostos por itens da fronteira entre classes. Tal hiperplano é capaz de separar linearmente os itens de classes distintas.

Os classificadores podem, ou não, modelar problemas de forma compreensível, ou significativa, para o entendimento humano. Por exemplo, as Redes Neurais podem aprender a atribuir classes aos itens, no entanto, os modelos construídos a partir do treinamento

utilizam recursos numéricos que não podem ser facilmente interpretados. Por outro lado, classificadores que constroem o modelo discriminativo na forma de Regras de Associação e Árvores de Decisão são mais compreensíveis, assim, legíveis ao humano [31].

3.2 Mineração de Dados de Séries Temporais - Síntese

No Capítulo 2, apresentou-se as características que diferenciam séries temporais de dados convencionais. Tais características restringem a utilização de técnicas convencionais de análise. Da mesma forma, a Mineração de Dados sobre dados temporais exige novas teorias e metodologias para lidar com a dependência temporal de maneira adequada.

Como uma ramificação da Mineração de Dados convencional [33], o arcabouço de Mineração de Dados de Séries Temporais é empregado para lidar com os grandes desafios atuais inerentes à análise de séries temporais. Segundo [42], os desafios são:

Representação dos dados: como representar as características fundamentais das séries temporais, considerando propriedades invariantes à transformação e dimensão da representação?

Medição da Similaridade: como estabelecer uma função de distância¹ que expresse a similaridade entre séries temporais? A eficácia é resultante de quais critérios, de percepção (heurísticos) ou puramente matemáticos?

Método de Indexação: em sistemas de armazenamento massivo, como indexar séries temporais para permitir consultas eficientes, considerando a dimensionalidade e problemas que incorrem da complexidade computacional?

Este trabalho de mestrado tem o enfoque sobre o desafio da representação dos dados. Mais do que extrair informações, o principal objetivo ao lidar com esse problema, é fornecer representações dos dados para que, assim, tarefas, propriamente ditas, possam ser aplicadas sobre a nova representação. Dentre essas tarefas, pode-se elencar as principais, relacionadas em [43, 44], que permeiam abordagens para lidar com os desafios previamente citados:

- **Segmentação:** dividir a série temporal em “segmentos significativos”;
- **Agrupamento (*Clustering*):** encontrar agrupamentos naturais de séries temporais (ou subsequências delas) no conjunto de dados;
- **Classificação:** atribuir classes predefinidas às séries temporais do conjunto de dados;
- **Indexação (*Query by Content*):** execução eficiente de consultas por séries temporais em base de dados;
- **Sumarização:** fornecer descrições breves às séries temporais que mantêm características essenciais para um determinado contexto;

¹No contexto deste trabalho, medidas/funções de distância/similaridade entre séries temporais são consideradas expressões equivalentes. Arbitrariamente, adotou-se “função de distância”.

- **Detecção de Anomalias (ou Novidades)**: encontrar padrões, ou comportamentos, não comuns ou inesperados em séries temporais;
- **Descoberta de Padrões Frequentes (ou *Motifs*)**: encontrar padrões, ou comportamentos, que ocorrem frequentemente em séries temporais;
- **Predição (*Forecasting*)**: prever valores de séries temporais, baseado em histórico ou conhecimento especialista;
- **Descoberta de Regras de Associação**: encontrar regras que relacionam padrões em séries temporais.

Tais tarefas podem ser complementares entre si. De fato, uma das abordagens mais comuns realiza a transformação das séries temporais para outro espaço de representação, de dimensão reduzida, seguida por mecanismos de indexação [45]. Adicionalmente, algumas soluções podem ser compartilhadas entre tarefas diferentes, como normalmente ocorre com *clustering* e classificação em relação à representação dos dados.

Uma vez que este trabalho concentra-se sobre a tarefa de classificação, por conveniência, não são detalhadas as demais tarefas acima relacionadas. Mais detalhes sobre elas podem ser encontrados em [42–45].

A representação adequada, medidas de similaridade, e a definição de padrões nos dados de séries temporais, influenciam a indução dos modelos, descritivos ou preditivos e, portanto, a acurácia dos resultados das aplicação que se utilizam destas. Nesse sentido, nas seções que se seguem, tais técnicas são discutidas, principalmente, a cerca da tarefa de classificação. Não obstante, a tarefa de agrupamento tem um relacionamento intrínseco com a classificação, permitindo, por vezes, traçar paralelos entre as tarefas, com relação às técnicas expostas.

3.2.1 Representação de séries temporais

Ao buscar uma nova representação para os dados de séries temporais, tenta-se evidenciar características implícitas nos dados que possam melhor descrevê-los. Somado a isso, é possível reduzir a dimensionalidade do conjunto de dados, acelerar o processamento, remover ruídos e, também, enfatizar características importantes. Para a classificação, isso implica em um modelo capaz de distinguir eficientemente as classes.

Em geral, as abordagens de classificação mais comuns induzem modelos considerando medidas de similaridade, ou distância, entre itens do conjunto de dados transformados em uma nova representação. Em [31] é definida uma taxonomia de representações, exposta na Figura 3.1. Não há intenção em exaurir a teoria acerca de todos os ramos da taxonomia. Para mais detalhes, recomenda-se [31, 46].

Representação baseada em modelos

Essa categoria de representação considera que uma série temporal é resultado da execução de um determinado modelo gerador. Em geral, tais modelos são ajustados às séries temporais e os parâmetros são utilizados nas tarefas de mineração.

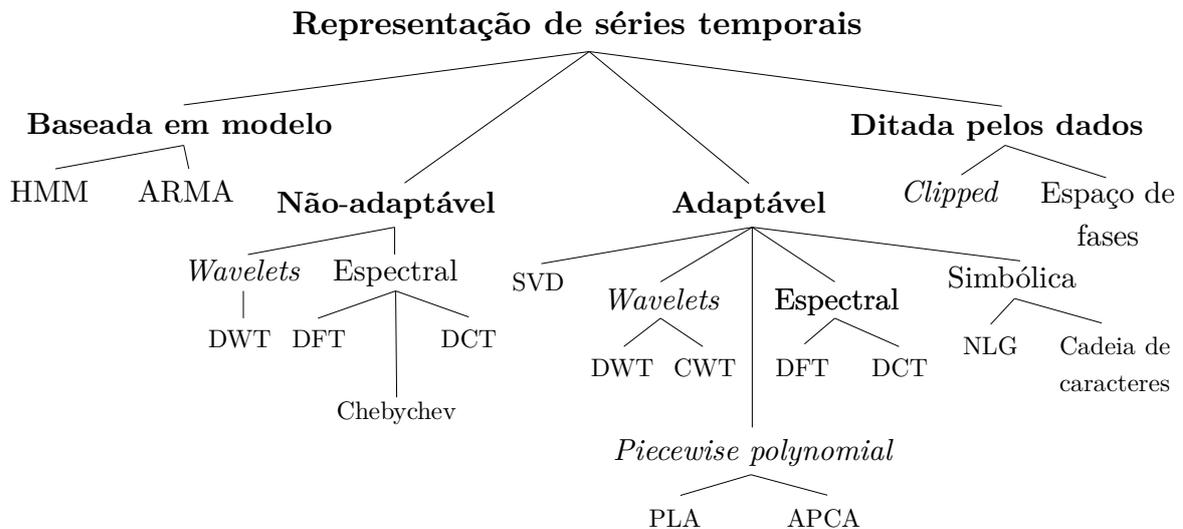


Figura 3.1: Taxonomia para representação de séries temporais. Adaptado de [31]

Um método de modelagem flexível, que pode ser empregado sobre dados que exibem dependência temporal, são os Modelos Ocultos de Markov (*Hidden Markov Models* - HMM). Outra técnica comum é o modelo ARMA.

Representação não adaptável aos dados

As técnicas nessa categoria transformam os dados das séries temporais de modo que a nova representação é independente daquela original. Por exemplo, em representações no domínio de frequência ou em ondaletas (*wavelets*), trabalha-se sobre coeficientes obtidos de um espaço funcional gerado por funções de base (obtidas pela decomposição da série temporal). Nesse contexto, é comum a obtenção de coeficientes a partir das técnicas de transformadas discretas de Fourier e de *wavelets* (respectivamente *Discrete Fourier Transform* - DFT, e *Discrete Wavelet Transform* - DWT).

Outro exemplo é a representação por médias segmentadas, conhecida como *Piecewise Aggregate Approximation* (PAA) [47]. Esta técnica expressa a série temporal através de valores médios de seus segmentos sequenciais de comprimento fixo.

Representação adaptável aos dados

O objetivo das técnicas nessa categoria é expressar os dados em uma representação alternativa, de modo que os novos itens carreguem as mesmas dependências, tácitas ou não, dos dados na representação original. Por exemplo, ao empregar as técnicas DWT ou DFT, ao invés de se utilizar todos os coeficientes extraídos, seleciona-se somente aqueles que carregam maior energia da série temporal [48]. A transformada de Hilbert-Huang (*Hilbert-Huang Transform* - HHT) também tem sido empregada para se aproveitar somente componentes que preservam mais energia na tarefa de classificação [49, 50].

Outro exemplo bem conhecido de transformação é a representação simbólica por discretização da técnica *Symbolic Aggregate Approximation* - SAX [46]. Basicamente, extrai-se

segmentos de uma série temporal através da técnica PAA que, posteriormente, são representados por símbolos equiprováveis.

Representação Ditada pelos Dados

Nas categorias anteriores, as técnicas são empregadas com nível de detalhamento arbitrário. Na representação ditada pelos dados, isso é automático. Por exemplo, a técnica conhecida como *clipping* realiza uma discretização binária dos dados, onde uma observação da série temporal é convertida para 0 ou 1, caso seja menor ou maior, respectivamente, que o valor médio da série temporal [51].

Outra técnica nesta categoria de representação é a imersão no espaço de fases. Como visto no Capítulo 2, é uma técnica de análise não linear que permite reconstruir as dinâmicas do sistema gerador da série temporal dentro de um espaço de estados. Embora forneça uma representação maior do que a original, permite que propriedades intrínsecas do processo gerador sejam reveladas para análise. Uma representação compacta desde espaço de fases é proposta em [52], onde as séries temporais de cada classe são desdobradas no espaço de fases. Posteriormente, são aprendidas as distribuições de probabilidade dos modelos das misturas Gaussianas para cada classe.

3.2.2 A tarefa de classificar séries temporais

A definição da tarefa de Classificação de Dados de Séries Temporais é similar à descrição da tarefa sobre dados convencionais, o objetivo é induzir classificadores sobre séries temporais rotuladas. Note que a definição a seguir não cobre o caso da classificação com janelas temporais.

Formalmente, seja $X_i(t) = \{x_{i1}, x_{i2}, \dots, x_{in}\}$ uma série temporal de n observações ordenadas, e $c_i \in \mathcal{C}$ a classe associada a $X_i(t)$, sendo $\mathcal{C} = \{1, 2, \dots, l\}$ as classes definidas no problema. O conjunto de treinamento \mathcal{T} é o conjunto dos r pares definido por $\mathcal{T} = \{(X_1(t), c_1), (X_2(t), c_2), \dots, (X_r(t), c_r)\}$. Existe um conjunto \mathcal{H} de funções f que representam relacionamentos hipotéticos em cada par de \mathcal{T} . O problema é, portanto, definir \mathcal{H} e, principalmente, uma função $f^* \in \mathcal{H}$ que descreva, de modo eficaz, o relacionamento $(X_i(t), c_i) \in \mathcal{T}$, para $\forall i \in \{1, 2, \dots, r\}$, minimizando, tanto quanto possível, o erro quando se atribui classes às séries temporais, durante o treinamento. Dessa forma, dada uma nova série temporal $X_q(t) \notin \mathcal{T}$, $f^*(X_q(t)) = c_q$, onde $c_q \in \mathcal{C}$ é a classe mais adequada a $X_q(t)$.

Na Seção 3.1, foram citadas as técnicas mais comumente empregadas nesta tarefa de aprendizado supervisionado. No contexto de séries temporais, tais técnicas também são empregadas, ou na forma convencional, ou com adaptações para lidar com a dependência temporal. O estudo realizado em [53] investiga as técnicas mais conhecidas de classificação para séries temporais, agrupando-as segundo a forma de operação. Neste trabalho de mestrado, é necessário e suficiente definir duas categorias: a classificação baseada em características e a baseada em distância. Em ambas categorias, é possível destacar trabalhos de cunho generalista, e trabalhos realizados em um domínio específico. Para cada categoria, são apresentados alguns trabalhos da literatura de classificação de séries temporais.

Classificação de séries temporais baseada em características

Nessa categoria, busca-se outra representação para a série temporal, transformando-a em um vetor de características. Isso implica que o conjunto de dados de séries temporais, então, seja transformado em um espaço de características multidimensional. Sobre esse novo espaço, tanto técnicas de classificação convencional, quanto as adaptadas para o contexto temporal podem ser empregadas. A seguir, são descritos alguns trabalhos da literatura de classificação que empregam essa abordagem e que exploram alguns conceitos relacionados neste trabalho de mestrado.

No estudo realizado em [29], propõe-se uma abordagem no contexto de controle de processos estatístico (*Control Chart*). Primeiramente, a série temporal é transformada em um vetor de características composto por estatísticas de primeira e segunda ordens, extraídas da própria série temporal. Posteriormente, o treinamento é realizado sobre os vetores de características com a rede neural *perceptron* de múltiplas camadas (*Multilayer Perceptron - MLP*). Comparada a um método de referência, que opera sobre séries temporais integrais, a abordagem proposta alcança maior acurácia e tempo de execução extremamente inferior.

No contexto de análise de sinais acústicos, o trabalho realizado em [49] busca a representação adaptativa, combinando as técnicas de Decomposição em Modos Empíricos (*Empirical Mode Decomposition - EMD*) e a Transformada Rápida de Fourier (*Fast Fourier Transform - FFT*). O objetivo é trabalhar sobre o domínio de tempo-frequência para lidar com não linearidade e não estacionariedade dos dados. A partir dessa representação, são extraídas características estatísticas e sonoras para o treinamento de diversos classificadores convencionais. O *k*-NN mostrou-se o mais eficaz para o problema.

Outras abordagens buscam representatividade através do ajuste de modelos geradores às séries temporais. Em [54] é empregada uma abordagem para classificação de tarefas mentais para Interface Cérebro Computador (*Brain-Computer Interface*). Sinais de eletroencefalograma (*electroencephalogram - ECG*) são modelados pelo processo AR, cujos coeficientes compõem o espaço de características para o treinamento convencional de uma Rede Neural e do classificador baseado em Análise de Discriminantes Lineares (*Linear Discriminant Analysis - LDA*). O mesmo procedimento para representação por modelo é empregado em [55], onde seis tipos de arritmias cardíacas de sinais ECG são eficientemente identificadas na classificação por modelo linear generalizado (*Generalized Linear Model - GLM*).

Nessa mesma linha de representação, em [56] foi proposto um método que representa cada classe de séries temporais através de um único modelo AR. O treinamento emprega conceitos de maximização de margens, também empregados no treinamento de SVMs. Em testes com dados simulados, o classificador alcança 100% de eficácia. Com dados reais de sinais ECG, a técnica proposta supera abordagens que utilizam os coeficientes do modelo AR como características extraídas. No entanto, conclui-se que algumas classes não podem ser representadas completamente por apenas um modelo AR. Nesse sentido, em outro trabalho publicado, os mesmos autores propõem a utilização de uma mistura de modelos ARs para representar cada classe do conjunto de dados [57]. Na análise comparativa entre o classificador adaptado proposto e o SVM convencional com características extraídas de modelos ARs, a técnica proposta supera a outra em três conjunto de dados reais. Tal noção, de se representar uma classe através da mistura de modelos ARs, também é empregada na tarefa de agrupamento, como apresentado em [58].

Em [36] é proposta uma abordagem que emprega modelos geradores sobre séries temporais transformadas, alcançando uma representação adaptativa híbrida. O objetivo é classificar danos causados por abalos terrestres em edificações, cuja série temporal é obtida por sensores de estruturas de amortecimento. Primeiramente, são extraídos os coeficientes do modelo AR ajustado sobre a série temporal, previamente processada pela DWT (para redução de ruídos). Em seguida, a partir desses coeficientes, são estimadas características de domínio² que alimentam o treinamento do classificador SVM. A abordagem proposta consegue eficientemente distinguir as classes em um espaço de características de baixa dimensionalidade.

Considerando os trabalhos acima relacionados, nota-se que o projeto de um classificador baseado em características pode ser definido, basicamente, em dois formatos: por adaptação do processo de treinamento, para lidar diretamente com dependência temporal; ou definido em dois estágios, onde representação e extração de características são procedimentos independentes do treinamento. Não é possível generalizar o desempenho de classificação considerando a forma pela qual as abordagens são projetadas, entretanto, nota-se que projetos no primeiro formato exibem maior complexidade inerente à dependência temporal, dificultando a compreensão humana sobre a modelagem discriminativa. Na classificação em dois estágios, a complexidade é isolada nos procedimentos de representação ou extração de características, simplificando o treinamento e o teste de classificação.

Em geral, a abordagem de classificação baseada em características exhibe tempos de treinamento e teste reduzidos e constantes, uma vez que a dimensionalidade do novo espaço de características é, quase sempre, muito menor que considerar a série temporal original. Ou seja, o comprimento dos itens do conjunto de dados não afetam os tempos de treino e teste.

Classificação de séries temporais baseada em distância

Métodos dessa categoria empregam uma função de distância que mede a similaridade entre duas séries temporais, determinando a qualidade da classificação. Após definir esta função, é possível empregar um classificador k -NN, por exemplo. As funções de distância são categorizadas pelo modo como realiza a comparação entre duas séries temporais [59].

Funções de distância *lockstep*: a comparação é realizada ponto a ponto, como as funções de distância Euclidiana e de Manhattan;

Funções de distância elástica: a comparação entre duas séries temporais pode ser realizada entre pontos nas formas um-para-muitos, como ocorre com a técnica *Dynamic Time Warping* - DTW, e um-para-muitos/nenhum, como é o caso da técnica de Maior Subsequência Comum (*Longest Common Subsequence* - LCSS).

A distância Euclidiana é a função mais simples no contexto de séries temporais, exibindo desempenho competitivo para a classificação, comparada a outras funções de distância, principalmente sobre conjuntos de dados massivos. Entretanto, por ser *lockstep*, o mapeamento entre as séries temporais é intransigente, tornando a comparação suscetível a ruídos, *outliers*, e desalinhamentos temporais [37, 59].

²Características que fazem sentido somente em um determinado problema, ou área.

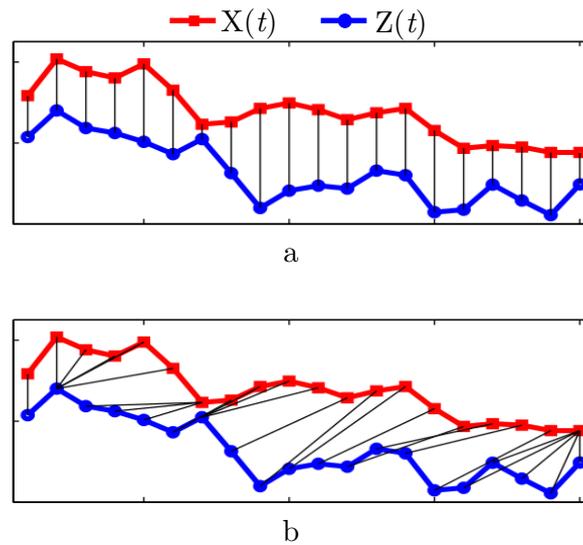


Figura 3.2: Diferença entre as séries temporais $X(t)$ e $Z(t)$ através das funções de distância (a) Euclidiana (*lockstep*), e (b) *Dynamic Time Warping* (elástica). Adaptado de [60]

Funções de distância elástica conseguem lidar com esses problemas. É possível visualizar a diferença entre os dois tipos de funções na Figura 3.2. A classificação realizada com 1-NN sobre DTW talvez seja a abordagem mais amplamente utilizada, devido suas baixas taxas de erro [42]. Isso tem motivado a proposição de extensões a essa função de distância. Em [61], é proposta uma variação da medida DTW que emprega uma função de penalidade. Essa nova função de distância, *Weighted Dynamic Time Warping* - WDTW, é testada sob classificação 1-NN e, também, na tarefa de agrupamento com o algoritmo k -Médias. Em ambas tarefas, a abordagem se mostrou competitiva a outras técnicas de 1-NN, sendo mais eficaz na tarefa de agrupamento.

Com o objetivo de tornar a classificação baseada em instâncias menos custosa computacionalmente, algumas abordagens buscam reduzir o número de itens do conjunto de dados. Em [62], é proposto um método baseado em DTW para encontrar séries temporais que consigam representar otimamente as classes. Mesmo com apenas uma série temporal por classe, consegue-se ser suficiente para manter taxas de erro competitivas quando submetido ao classificador 1-NN sobre DTW. Com o mesmo objetivo, em [63] é proposta uma abordagem para encontrar a série temporal padrão (*template*) para cada classe, de modo que a classificação por 1-NN sobre DTW também seja realizada com apenas uma serie temporal por classe.

Outras formas alternativas de classificação baseada em distância têm explorado a eficácia da função de distância DTW. Em [64], é apresentada uma abordagem cujas similaridades par a par entre séries temporais alimentam o treinamento do classificador SVM, exibindo melhores taxas de erro comparada ao 1-NN sobre DTW. Neste mesmo estudo, é proposta também uma abordagem que utiliza DTW como a função de distância para o *kernel* Gaussiano, utilizado pelo classificador *Potential Support Vector Machine* - P-SVM. Os resultados de classificação sobre os conjuntos de dados do repositório UCR [10] mostram que a abordagem é mais eficaz que diversas versões do classificador 1-NN.

Abordagens que empregam modelos geradores também conseguem mitigar o custo computacional da classificação baseada em instâncias. Alguns estudos são bem conhecidos na literatura de classificação, mesmo sendo voltados para a tarefa de agrupamento. Em [65] as séries temporais são representadas pelos seus coeficientes ARs. Segundo os autores, a função de distância deve ser sensível a coeficientes de maior ordem, para a tarefa de agrupamento. Nesse sentido, a distância Euclidiana é aplicada sobre coeficientes cepstrais extraídos dos coeficientes ARs de cada série temporal. Comparada a outras funções de distância, a técnica proposta é competitiva em alguns casos, e significativamente melhor em outros.

Considerando este breve apanhado de técnicas de classificação baseada em distância, é possível notar que, analogamente a classificação baseada em características, também busca-se uma representatividade que reduza os custos computacionais. A redução de itens do conjunto de dados lida com o *trade-off* entre eficiência na complexidade computacional e a eficácia da classificação. Por outro lado, reduzir a dimensionalidade dos dados através de transformações e extração de características, pode evidenciar padrões e comportamentos implícitos no dados, mas, também, negligenciar informações não capturadas por tais procedimentos.

Capítulo 4

Uma Abordagem para Classificação de Séries Temporais Baseada em Modelo Autorregressivo e Gráfico de Recorrência

Neste capítulo é apresentada a abordagem proposta neste trabalho de mestrado. Pretende-se abordar o problema da classificação de séries temporais, a partir da modelagem e da extração de características baseadas na propriedade intrínseca de determinismo do processo gerador das séries temporais. Primeiramente, são apresentados a motivação e o objetivo. Posteriormente, a abordagem é exposta, juntamente com as técnicas envolvidas.

4.1 Abordagem Proposta

A característica que distingue uma série temporal de qualquer outra coleção de dados é a dependência que normalmente existe entre suas observações adjacentes. Tal dependência compõe as leis que determinam, parcial ou totalmente, como os dados evoluem ao longo do tempo, caracterizando a correlação entre as observações adjacentes dentro da série temporal.

Essa dependência restringe a utilização de técnicas estatísticas convencionais, sendo necessárias novas abordagens para lidar efetivamente com os dados temporais. A área de Análise de Séries Temporais aborda este problema, fornecendo um ferramental matemático probabilístico e determinístico, capaz de modelar tal dependência [6, 7].

Por meio da modelagem de séries temporais é possível descobrir propriedades intrínsecas e, assim, compreender o comportamento embutido nos dados. No contexto da Análise de Séries Temporais, em geral, o objetivo da modelagem é identificar propriedades de interesse para aproximar o processo gerador dos dados. Assim, é possível estimar valores ainda não conhecidos e tomar decisões antecipadamente.

O estudo realizado em [8], apresenta uma abordagem para predição de eventos futuros através da modelagem de séries temporais. A abordagem propõe uma taxonomia que

conduz o procedimento de escolha da técnica de modelagem mais adequada, considerando propriedades de determinismo, de linearidade, e de estacionariedade, intrínsecas do processo gerador da série temporal. Ao selecionar a técnica mais adequada, é possível modelar o processo gerador e, então, estimar valores futuros com mais acurácia.

Tal ideia, de explorar propriedades do processo gerador, motivou outro trabalho no contexto da modelagem para predição de séries temporais. O trabalho realizado em [9], propõe uma abordagem para modelagem híbrida, cuja propriedade de determinismo do processo gerador é analisada como uma componente da série temporal. Utilizando técnicas de modelagem adequada para cada propriedade revelada da série temporal, e depois combinando-as aditivamente, é possível estimar valores futuros efetivamente.

A modelagem através da análise das propriedades intrínsecas do processo gerador da série temporal, em especial o determinismo, direcionam o estudo desenvolvido neste trabalho de mestrado. A modelagem no contexto de Mineração de Dados de Séries Temporais, por outro lado, está intimamente ligada à representatividade. O principal objetivo é transformar a representação original, tornando possível a descoberta de padrões que possam generalizar características em comum entre itens de um conjunto de dados. Dessa forma, grupos podem ser revelados, ou classes podem ser discriminadas com maior efetividade.

A literatura de Mineração de Dados de Séries Temporais é abundante em estudos sobre representação de dados, principalmente em tarefas de classificação e agrupamento [31, 46–50, 56, 57], como apresentado nas Seções 3.2.1 e 3.2.2. Basicamente, nota-se que estes trabalhos buscam representações com maior potencial informativo, através da redução de itens do conjunto de dados (descartando itens redundantes ou com baixo teor informativo), e através de transformações que reduzem o espaço de características ou “expandem” os dados para evidenciar informações implícitas relevantes.

Durante a análise da literatura neste trabalho de mestrado, buscou-se por estudos que relacionassem a análise das propriedades intrínsecas do processo gerador, principalmente dirigidos pela taxonomia proposta em [8] ou similares, com a tarefa de classificação de séries temporais. Entretanto, nenhum trabalho foi encontrado. Nesse sentido, nota-se na literatura, que há uma lacuna entre a tarefa de classificação de séries temporais e a análise das propriedades intrínsecas do processo gerador, como realizado em [8, 9].

Considerando este contexto, este trabalho de mestrado tem como objetivo abordar o problema da classificação de séries temporais, inspirando-se na análise de propriedades intrínsecas do processo gerador [8, 9]. Especificamente, deseja-se modelar as influências de comportamentos determinísticos, e não determinísticos (estocásticos), presentes em séries temporais. Para isso, são empregadas as técnicas de modelagem AR, e de imersão no espaço de fases baseado em atrasos de tempo. Tais técnicas são apropriadas para modelar comportamentos estocásticos, com baixo nível de determinismo, e comportamentos com altos níveis de determinismo [8, 9], respectivamente. Após a modelagem, características são extraídas para compor o novo espaço de características dos dados. A partir desse novo espaço, pode-se empregar técnicas de classificação convencionais para realização da tarefa. Neste trabalho, optou-se pelo classificador *Support Vector Machine*, uma que este exibe ótima capacidade de generalização do modelo discriminativo aprendido, e robustez sobre grandes conjuntos de dados.

A abordagem é avaliada experimentalmente, e posteriormente comparada a abordagens de referência da literatura. Na próxima seção, são apresentados os procedimentos de representação através da modelagem e de extração de características empregados pela abordagem proposta neste trabalho, bem como o classificador SVM.

4.1.1 Representações autorregressiva e Gráfico de Recorrências para extração de características

Como pode ser visto na Figura 4.1, a série temporal $X(t) = \{x_1, x_2, \dots, x_n\}$ é processada pelo modelo AR, tal que x_{n+1} passa a ser representada por $\sum_{i=1}^p \phi_i x_{t-i} + \epsilon_t$. Deste modelo, extrai-se o conjunto de seus coeficientes $\Lambda = \{\phi_1, \phi_2, \dots, \phi_p\}$, sendo p a ordem do modelo AR.

Como apontado em [8, 9], o modelo AR é indicado para se modelar séries temporais que exibem comportamento estocástico. Nesse sentido, considera-se o conjunto de coeficientes Λ como o conjunto de características que descreve a propriedade estocástica do processo gerador de $X(t)$.

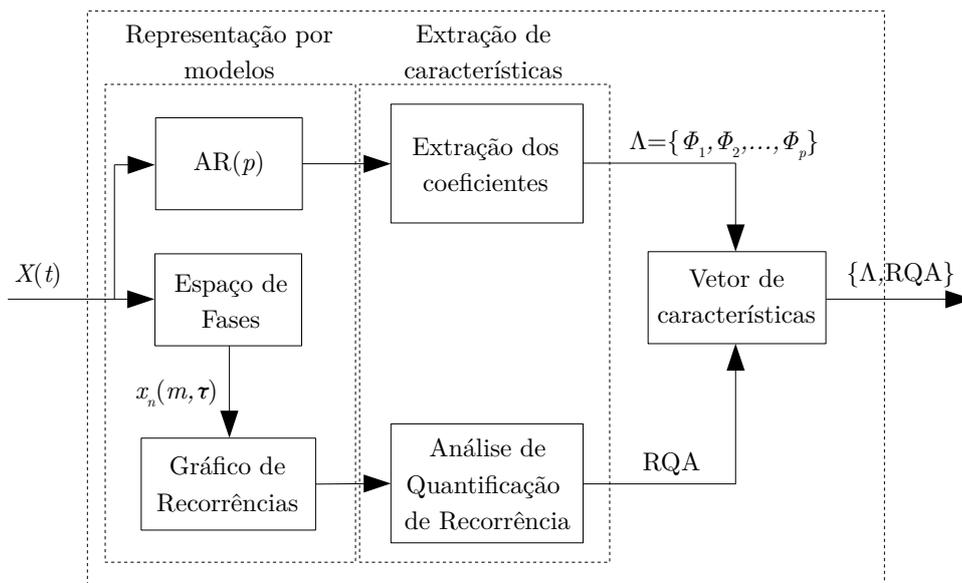


Figura 4.1: Representação e extração de características da abordagem proposta. Dentro do novo espaço de características a série temporal $X(t)$ é representada por $\{\Lambda, RQA\}$, onde Λ é o conjunto dos p coeficientes do modelo $AR(p)$ e, RQA , o conjunto de medidas extraídas pela Análise de Quantificação de Recorrência.

O modelo AR tem sido empregado para extração de características em outros estudos da literatura [36, 54, 55]. Neste trabalho de mestrado, no entanto, Λ compõe somente a primeira parte do espaço de características. A segunda parte é composta por características extraídas através de uma representação apropriada para se analisar o determinismo, o Gráfico de Recorrência (*Recurrence Plot* - RP) [8, 9]. Sobre o RP, é aplicada a Análise de Quantificação de Recorrência de onde se extrai as medidas que compõem o conjunto RQA , apontado na Figura 4.1.

O *Recurrence Plot* é uma ferramenta gráfica aplicada sobre séries temporais desdobradas no espaço de fases [66]. Primeiramente, a série temporal $X(t) = \{x_1, x_2, \dots, x_n\}$ deve ser

reconstruída no espaço de fases, onde a observação x_n é representada pelo estado $x_n(m, \tau) = (x_n, x_{n+\tau}, \dots, x_{n+(m-1)\tau})$, sendo m a dimensão de imersão e τ a dimensão de separação. Na Figura 4.2 (b) é apresentado um exemplo, o sistema de Rössler, desdobrado no espaço de fases. Nesse espaço multidimensional contendo N estados, define-se uma matriz de recorrência $R_{N \times N}$ e verifica-se a ε -vizinhança de cada um dos estados.

$$R_{i,j}(\varepsilon) = \Theta(\varepsilon - \|\vec{x}_i - \vec{x}_j\|) \quad i, j = 1, \dots, N \quad (4.1)$$

$$\Theta(\Delta_{i,j}) = \begin{cases} 0 & : \Delta_{i,j} < 0 \\ 1 & : \Delta_{i,j} \geq 1 \end{cases}$$

Por exemplo, após verificar que somente o estado $x_j(m, \tau)$ está próximo ao estado $x_i(m, \tau)$, $R_{i,j}$ recebe o valor 1 e as posições $R_{i,\{1,\dots,N\}-\{j\}}$ recebem valor 0. Essa relação é definida pela Equação 4.1, cujo ε é o raio de cobertura para vizinhança, $\Theta(\cdot)$ é uma função degrau (*heaviside*), e $\|\cdot\|$ a norma Euclidiana sobre os vetores de estados. Sobre a Figura 4.2 (b) é possível observar um estado que é coberto pela ε -vizinhança (círculo cinza) de outro estado, ambos os pontos são pretos. O ponto mais interno (sem preenchimento) está fora da ε -vizinhança.

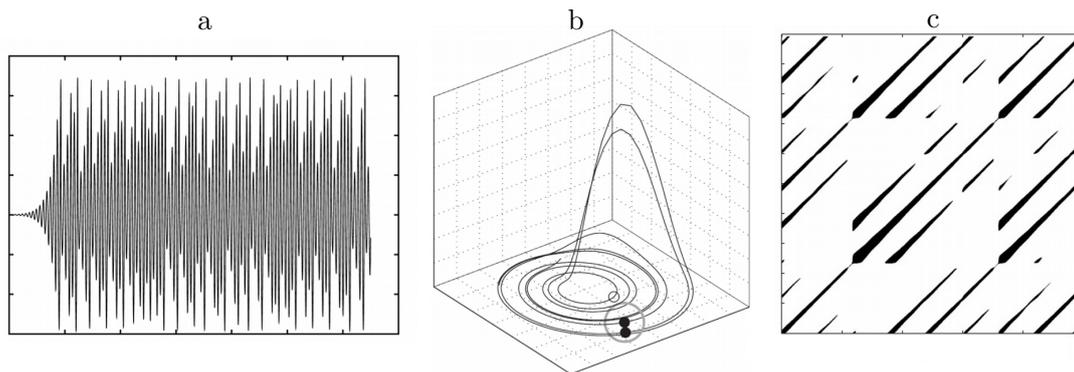


Figura 4.2: Sistema de Rössler: a) como série temporal, b) reconstruído no espaço de fases, com exemplo de um par de estados ε -vizinhos e c) no *Recurrence Plot*. Adaptado de [8, 19].

Após avaliar-se todos os N estados, a matriz de recorrência, completamente preenchida, é representada por um gráfico de pontos (o RP propriamente dito), onde valores de 1 em R são representados por pontos, e valores de 0 por espaços em branco. O RP completamente definido, então, apresenta algumas estruturas. Na Figura 4.2 (c) é apresentado o RP do sistema de Rössler, configurado com os parâmetros $a = 0,15$, $b = 0,2$ e $c = 10$. A série temporal deste sistema é exibida na Figura 4.2 (a) [19].

As estruturas que compõem a textura do RP descrevem algumas características importantes. Por exemplo, quanto maior o número de estruturas diagonais, maior o número de estados recorrentes na trajetória do espaço de fases. Tal recorrência de estados é um indicativo de comportamento determinístico na série temporal [19]. Outras estruturas podem surgir, como pontos isolados, estruturas horizontais e verticais, fornecendo ainda mais informação sobre a trajetória dos estados da série temporal no espaço de fases.

Analisar visualmente o RP, entretanto, é uma tarefa complexa e onerosa. Uma forma de lidar com esse problema é empregando-se a Análise de Quantificação de Recorrência

(*Recurrence Quantification Analysis* - RQA), uma metodologia para quantificar as estruturas do RP, através de medidas baseadas na densidade de pontos recorrentes e nas linhas verticais e diagonais. A seguir, é apresentada uma breve descrição de cada uma das medidas utilizadas neste trabalho de mestrado. Para maior riqueza de detalhes destas medidas, e de outras que compõem a metodologia RQA, recomenda-se [19].

A medida RQA mais simples é a taxa de recorrência (*recurrence rate*), que mensura a densidade dos pontos recorrentes no RP. Essa medida é definida pela Equação 4.2, onde $R_{i,j}$ é a matriz de recorrência, N é o número de estados da série temporal desdobrada e, ε , o raio de cobertura que determina a vizinhança.

$$RR(\varepsilon) = \frac{1}{N^2} \sum_{i,j=1}^N R_{i,j}(\varepsilon) \quad (4.2)$$

As medidas baseadas em estruturas de linhas diagonais DET e L_{mean} , utilizam em suas formulações o histograma $P(\varepsilon, l)$, definido pela Equação 4.3. Esse histograma expressa a distribuição entre comprimentos e quantidades de linhas diagonais no RP, cujo valor de l indica o comprimento da linha diagonal.

$$P(\varepsilon, l) = \sum_{i,j=1}^N (1 - R_{i-1,j-1}(\varepsilon))(1 - R_{i+l,j+l}(\varepsilon)) \prod_{k=0}^{l-1} R_{i+k,j+k}(\varepsilon) \quad (4.3)$$

A medida DET descreve a taxa de pontos recorrentes que formam linhas diagonais no RP. Considerando que séries temporais de comportamento determinístico exibem RPs com linhas diagonais compridas e poucos pontos isolados, pode-se considerar que DET representa o nível de determinismo (previsibilidade) de uma série temporal. Essa medida é definida pela Equação 4.4, onde l_{min} determina o comprimento mínimo para se considerar uma linha significativa. A análise sobre o histograma $P(\varepsilon, l)$ auxilia a seleção do valor l_{min} para que DET seja significativa.

$$DET = \frac{\sum_{l=l_{\text{min}}}^N lP(\varepsilon, l)}{\sum_{l=1}^N lP(\varepsilon, l)}, \quad (4.4)$$

No RP, uma linha diagonal de comprimento l modela a circunstância (no espaço de fases) em que um segmento da trajetória evolui dentro da ε -vizinhança de outro segmento, por um período de l instantes. A medida L_{mean} (Equação 4.5) calcula o comprimento médio dessas linhas diagonais presentes no RP. Além do comprimento médio, é possível obter o maior comprimento dentre as linhas diagonais, através da medida L_{max} (Equação 4.6). Esta última, compõe a taxa de divergência, calculada pela medida DIV (Equação 4.7). Tal medida expressa a relação entre o comprimento das linhas diagonais e a divergência entre os segmentos da trajetória. Por exemplo, uma divergência em ritmo exponencial modela um RP com linhas diagonais curtas, com um alto valor de DIV .

$$L_{\text{mean}} = \frac{\sum_{l=l_{\text{min}}}^N lP(\varepsilon, l)}{\sum_{l=l_{\text{min}}}^N P(\varepsilon, l)} \quad (4.5)$$

$$L_{\text{max}} = \max \left(\{l_i\}_{i=1}^{N_l} \right) \quad (4.6)$$

$$\text{DIV} = \frac{1}{L_{\max}} \quad (4.7)$$

A medida ENTR refere-se a entropia de Shannon para a probabilidade $p(\varepsilon, l) = P(\varepsilon, l)/N_l$, formulada pela Equação 4.8. Essa medida reflete a complexidade do RP em relação ao comprimento de suas linhas diagonais. Por exemplo, quanto menor a correlação entre as observações de uma série temporal, menor o valor ENTR, refletindo um RP de baixa complexidade.

$$\text{ENTR} = - \sum_{l=l_{\min}}^N p(\varepsilon, l) \ln p(\varepsilon, l) \quad (4.8)$$

A medida TREND (tendência) formulada pela Equação 4.9, calcula o coeficiente de regressão linear sobre taxas de recorrências de linhas diagonais τ -distantes da diagonal principal do RP (*line of identity* - LOI). Basicamente, tal medida expressa a distância temporal entre a LOI e cada linha diagonal.

$$\text{TREND} = \frac{\sum_{\tau=1}^{\tilde{N}} (\tau - \tilde{N}/2)(\text{RR}_{\tau} - \langle \text{RR}_{\tau} \rangle)}{\sum_{\tau=1}^{\tilde{N}} (\tau - \tilde{N}/2)^2} \quad (4.9)$$

As medidas baseadas em estruturas de linhas verticais LAM, TT, e V_{\max} , utilizam em suas formulações o histograma $P(v)$, definido pela Equação 4.10. Tal histograma expressa a distribuição entre comprimentos e quantidades de linhas verticais no RP, onde v representa o comprimento da linha vertical. Adicionalmente, o valor de v indica o intervalo de tempo em que um dado estado não se altera, ou se altera lentamente.

$$P(v) = \sum_{i,j=1}^N (1 - R_{i,j})(1 - R_{i,j+v}) \prod_{k=0}^{v-1} R_{i,j+k} \quad (4.10)$$

A medida LAM (Equação 4.11) fornece o nível de laminaridade, descrevendo a taxa de pontos recorrentes que formam linhas verticais no RP. Assim como l_{\min} serve como uma restrição para a definição do determinismo, v_{\min} também o é para laminaridade. Ao selecionar valores de v maiores que v_{\min} evita-se a influência do movimento tangencial¹ dos estados no espaço de fases. Os RPs que exibem mais pontos de recorrência isolados do que linhas verticais, possuem valores baixos de LAM.

$$\text{LAM} = \frac{\sum_{v=v_{\min}}^N vP(v)}{\sum_{v=1}^N vP(v)} \quad (4.11)$$

Similarmente a L_{mean} e L_{max} , as medidas TT e V_{\max} calculam o comprimento médio e o comprimento máximo das linhas verticais presentes no RP, respectivamente, definidas pelas Equações 4.12 e 4.13.

$$\text{TT} = \frac{\sum_{v=v_{\min}}^N vP(v)}{\sum_{v=v_{\min}}^N P(v)}, \quad (4.12)$$

¹O movimento tangencial é caracterizado por diversos pontos em uma ε -vizinhança que evoluem em conjunto no espaço de fases. A ε -vizinhança de um ponto no instante t pode abranger pontos que estão em instantes de tempo adjacentes a t . Tais vizinhos devem ser evitados, uma vez que expressam falsa recorrência [19].

$$V_{\max} = \max \left(\{v_l\}_{l=1}^{N_V} \right) \quad (4.13)$$

Basicamente, as medidas TT e V_{\max} estimam, respectivamente, o intervalo de tempo médio e o intervalo mais longo em que um estado permanece inalterado.

As medidas apresentadas acima compõem a metodologia RQA, que visa quantificar a recorrência dos estados no espaço de fases, modelada pelo RP. É sabido que tais medidas podem fornecer outras informações sobre comportamentos inerentes aos dados, como o caos, e que são capazes de lidar com outras propriedades intrínsecas, como a não linearidade e não estacionariedade. Neste trabalho de mestrado, tais medidas são tomadas como características que expressam propriedade intrínseca de determinismo do processo gerador da série temporal.

Nesse sentido, o conjunto de características RQA (Figura 4.1) é dado por $RQA = \{RR, DET, L_{\text{mean}}, L_{\text{max}}, DIV, ENTR, TREND, LAM, TT, V_{\text{max}}\}$, e compõe o conjunto de características global $\{\Lambda, RQA\}$. Tal conjunto, estimado para cada série temporal, compõe o novo espaço de características.

Para fechar o escopo da abordagem proposta neste trabalho, na seção seguinte, é apresentada a técnica escolhida para a tarefa de classificação.

4.1.2 Classificação por *Support Vector Machine* - SVM

O foco principal da abordagem proposta neste trabalho de mestrado, está na busca por representações para extração de características dos dados de séries temporais. Portanto, a tarefa de classificação pode ser realizada por uma técnica de classificação convencional. Nesse sentido, optou-se pela classificação através das Máquinas de Vetores de Suporte (*Support Vector Machines* - SVM). A teoria sobre o classificador SVM é ampla, conceitual e matematicamente bem estabelecida. Não é objetivo deste trabalho explorá-la em profundidade, sendo apresentada uma descrição breve sobre seus principais tópicos. Para maiores detalhes, recomenda-se [67–69].

A escolha por esse classificador é motivada pelas vantagens que ele fornece. Baseado na Teoria do Aprendizado Estatístico, o classificador SVM exhibe uma ótima capacidade de generalização, sendo robusto sobre conjunto de dados massivos onde, geralmente, outros classificadores acabam superajustados (*overfitting*). Embora não forneça modelos discriminativos de fácil compreensão, a convexidade da função objetivo a ser minimizada envolve um mínimo global, o que o difere das Redes Neurais, por exemplo, que implicam em mínimos locais na função objetivo [69].

O SVM utiliza um hiperplano ótimo, dado pela resolução de um problema de otimização quadrático, para separar os dados de classes distintas dentro de um espaço de características multidimensional. Esse hiperplano é determinado entre margens que visam maximizar a separação dos itens. Elas são empregadas na forma rígida, quando não se permite itens dentro das margens, ou na forma suave, quando tal restrição é relaxada (Figura 4.3 (a)).

Originalmente, a formulação do hiperplano garantia a definição das fronteiras somente sobre dados linearmente separáveis. Essa situação pode ser visualizada na Figura 4.3 (a): pode-se notar o hiperplano ótimo separando as duas classes com restrições relaxadas. Basi-

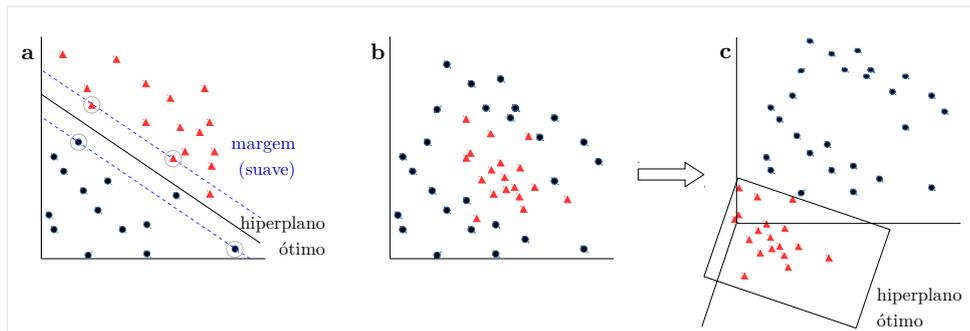


Figura 4.3: SVMs sobre dados distribuídos linearmente e não linearmente. a) SVM de margem suave sobre dados linearmente separáveis; b) dados não separáveis linearmente, e; c) SVM sobre dados não separáveis linearmente. Adaptado de [70].

camente, esse relaxamento permite a tolerância a ruídos e *outliers*. Os itens circulados em cinza, sobre as margens, são os vetores de suporte.

Dados reais, entretanto, possuem ruídos, *outliers*, ou pela própria natureza, não podem ser linearmente separados. O simples relaxamento da restrição das margens não garante a separação efetiva. Esse é o caso representado na Figura 4.3 (b). Para lidar com esse problema, o SVM mapeia os dados para um espaço de características de dimensão suficientemente maior, onde seja alta a probabilidade de separação linear dos dados.

A formulação desse mapeamento, porém, é altamente custosa e complexa. Para contornar isso, é necessário empregar uma função *kernel*, um tipo de função capaz de realizar tal mapeamento de modo implícito, eficientemente. Dentre os *kernels* mais utilizados para o classificador SVM, pode-se apontar os Polinomiais e de Funções de Base Radial (*Radial Basis Functions* - RBF).

Na Figura 4.3, a transição entre (b) e (c) representa a transformação do espaço de características, elevando sua dimensão. Com um espaço de dimensão suficientemente alto, os dados podem ser linearmente separados. Portanto, o SVM realiza a transformação do espaço de características através da função *kernel* para, em seguida, encontrar o hiperplano ótimo que separe linearmente os dados de treinamento.

A formulação original do SVM tem como objetivo resolver o problema da classificação binária (Figura 4.3). Mais recentemente, porém, surgiram propostas de extensão da metodologia para o problema multiclasse. Duas formas têm sido empregadas, ou as classes são todas incluídas na formulação da otimização no treinamento, ou, sem perda de generalidade, é realizada uma combinação de vários classificadores binários.

Claramente, a primeira forma incorre em maior complexidade para o modelo discriminativo. A segunda, de menor complexidade, é mais conveniente [68]. A combinação de classificadores binários é realizada empregando-se as políticas um-contra-todos ou um-contra-um.

Basicamente, a primeira política conduz a construção de um classificador que considera uma classe como a positiva e todas as outras como negativa. Dessa forma, são treinados tantos classificadores quanto o número de classes. Dado um novo item, o teste é realizado em todos os classificadores, sendo atribuída a classe referente ao classificador onde ele foi testado

como positivo. Na política um-contra-um, por outro lado, constrói-se tantos classificadores quanto a combinação simples das classes, tomadas duas a duas. Dado um novo item, o teste é realizado em todos os classificadores, de modo que a classe de decisão é aquela que mais vezes foi atribuída ao item (estratégia *Max Wins*).

Após definir a técnica de classificação, encerra-se o escopo da abordagem proposta neste trabalho. Sumariamente, a abordagem busca representar e extrair características, através de modelos que evidenciem as propriedades intrínsecas de determinismo e estocasticidade do processo gerador, para a tarefa de classificação de séries temporais. O próximo capítulo apresenta avaliação experimental, bem como a discussão dos resultados.

Capítulo 5

Experimentos e Resultados

5.1 Metodologia

Como foi apresentada na Seção 3.2.2, a tarefa de classificação de séries temporais pode ser realizada, em geral, diretamente sobre as séries temporais, através da classificação baseada em funções de distância, ou através da transformação dos dados para se utilizar técnicas de classificação convencionais (dois estágios). A abordagem proposta neste trabalho se baseia na metodologia em dois estágios, representada pela Figura 5.1.

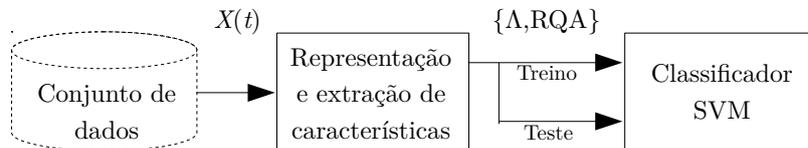


Figura 5.1: Classificação de Séries Temporais em dois estágios.

No primeiro estágio, a série temporal é processada pelo modelo AR, para a extração dos seus coeficientes. Paralelamente, a série temporal original é desdobrada no espaço de fases e modelada pelo *Recurrence Plot* (RP), para se extrair as medidas RQA. Em conjunto, tais valores definem a dimensão do espaço de características, aonde a série temporal é representada para a técnica de classificação. No entanto, durante o ajustamento do modelo AR sobre as séries temporais, estima-se uma ordem p por item do conjunto de dados, ou seja, dadas duas séries temporais $X_i(t)$ e $X_j(t)$, p_i pode não ser igual a p_j ainda que pertençam à mesma classe. Isso representa um problema quando considera-se a tarefa de classificação convencional por SVM, pois todos os itens do conjunto de dados devem ser representados por vetores de características de um único comprimento. Neste trabalho, foi proposto empregar duas formas para generalizar o valor de p : a ordem *moda* e a ordem *máxima* do conjunto de dados. Dessa forma, todas as séries temporais de um conjunto de dados são representadas por vetores de características de um mesmo comprimento, por exemplo, $X(t)$ é representada pelos conjuntos $\{\Lambda_{\text{moda}}, \text{RQA}\}$ e $\{\Lambda_{\text{max}}, \text{RQA}\}$.

No segundo estágio, é realizada a classificação através do SVM com *kernel* RBF. A estimação de erro foi realizada através da validação cruzada estratificada com 10 partições, amostrada em 10 execuções independentes. A métrica de desempenho utilizada foi a taxa

de erro (*misclassification*), definida pela Equação 5.1 (adaptada de [71])

$$\text{taxa de erro} = \frac{1}{n} \sum_{i=1}^n \Psi(X_i(t), c_i) \quad (5.1)$$

$$\Psi(X_i(t), c_i) = \begin{cases} 0 & : f(X_i(t)) = c_i \\ 1 & : f(X_i(t)) \neq c_i \end{cases}$$

onde $\Psi(\cdot)$ é uma função *heaviside*, $f(\cdot)$ é a função preditora de classe e, c_i , a classe real de $X_i(t)$. A taxa de erro é uma das medidas de desempenho mais empregadas para o problema da classificação de séries temporais [64, 72]. Adicionalmente, ela foi escolhida para facilitar a comparação de desempenho da abordagem proposta com outras abordagens de referência da literatura, os classificadores 1-NN sobre as funções de distância Euclidiana, DTW simples, e DTW com espaço de busca otimizado por janela [10, 73].

A classificação é realizada sobre as representações separadas Λ_{moda} , Λ_{max} , RQA, e conjuntas $\{\Lambda_{\text{moda}}, \text{RQA}\}$ e $\{\Lambda_{\text{max}}, \text{RQA}\}$. A configuração de melhor desempenho é comparada aos desempenhos dos classificadores 1-NN, com teste estatístico de Wilcoxon sobre as respectivas taxas de erros, para se verificar diferença significativa. Tal metodologia visa explorar os seguintes pontos:

- Análise das estratégias para definição da ordem do modelo AR;
- Análise do potencial discriminativo das características extraídas através da observação das influências determinísticas e estocásticas, presentes nas séries temporais;
- Análise dos efeitos da combinação das características propostas;
- Análise comparativa entre as abordagens proposta e de referência da literatura.

Em toda a implementação dos experimentos, utilizou-se a linguagem de *script* R [74], sobre o sistema operacional Debian GNU/Linux. A quantidade e o comprimento das séries temporais dos conjuntos de dados, tornaram inviável o processamento dos experimentos em apenas um computador. Dessa forma, foram utilizados três computadores (configurações distintas), para agilizar a avaliação experimental.

Conjuntos de Dados

Nos experimentos conduzidos neste trabalho, foram utilizados os conjuntos de dados de séries temporais UCR [10], gentilmente cedidos pelo professor Eamon Keogh¹. Este repositório disponibiliza 42 conjuntos de dados² de aplicações de domínios reais diversos, gerados artificialmente ou obtidos experimentalmente. Cada conjunto de dados é composto por séries temporais univariadas de único comprimento, rotuladas para a finalidade de classificação e agrupamento, sem valores faltantes (*missing values*).

¹<http://www.cs.ucr.edu/~eamonn/>. Último acesso em outubro de 2014

²Foram utilizados todos os conjuntos de dados, com exceção do *Gesture Library Z*. Este conjunto de dados apresentou problemas sob os algoritmos implementados.

A Tabela 5.1 exibe as características de cada conjunto: o número de classes, o número de itens, e o comprimento das séries temporais, respectivamente. Como relatado no estudo [75], que também utiliza tal repositório, é possível sumarizar os conjuntos agrupando-os conforme o método de aquisição:

Tabela 5.1: Conjuntos de dados do Repositório utilizados neste trabalho. UCR [10].

	Conjunto de Dados	Classes	Exemplos	Comprimento
1	OSU Leaf	6	442	427
2	In-line Skate	7	650	1882
3	AIBO Robot Surface	2	621	70
4	AIBO Robo tSurfaceII	2	980	65
5	Star Light Curves	3	9236	1024
6	ECG (5-days)	2	884	136
7	Two Lead ECG	2	1161	82
8	Gun Point	2	200	150
9	Haptics	5	463	1092
10	Swedish Leaf	15	1125	128
11	Wafer	2	7174	152
12	Symbols	6	1020	398
13	Yoga	2	3300	426
14	Trace	4	200	275
15	Diatom Size Reduction	4	322	345
16	CinC ECG Torso	4	1420	1639
17	Mallat	8	2400	1024
18	ECG (200)	2	200	96
19	Chlorine Concentration	3	4307	166
20	Lightning-2	2	121	637
21	Adiac	37	781	176
22	Mote Strain	2	1272	84
23	Face (all)	14	2250	131
24	Medical Images	10	1141	99
25	Olive Oil	4	60	570
26	Lightning-7	7	143	319
27	CBF	3	930	128
28	Coffee	2	56	286
29	Face (four)	4	112	350
30	Faces (UCR)	14	2250	131
31	Beef	5	60	470
32	Fish	7	350	463
33	Synthetic Control	6	300	60
34	Gesture Library X	8	4478	315
35	Gesture Library Y	8	4478	315
36	Words Synonyms	25	905	270
37	50 Words	50	905	270
38	Cricket X	12	780	300
39	Cricket Y	12	780	300
40	Cricket Z	12	780	300
41	Two Patterns	4	5000	128

compatibilidade de contornos: nesse grupo, imagens de objetos são transformados em séries temporais medindo-se a distância entre a extremidade (o contorno) e o centro do objeto. Os conjuntos desse grupo são *Face (four)*, *Face (all)*, *Faces (UCR)* [10], *Diatom Size Reduction*, *Adiac* [76], *Fish* [77], *OSU Leaf* [78], *Swedish Leaf* [79], e *Yoga* [80];

eletrocardiograma - ECG: os dados são obtidos de sinais ECG. O grupo é composto pelos conjuntos *ECG (200)*, *Two Lead ECG*, *CinC ECG Torso* e *ECG (5-days)* [10];

monitoramento de processo: os conjuntos *Trace* [81] e *Wafer* [82] são compostos por observações extraídas do monitoramento de processos em usina de energia nuclear e de fabricação de semicondutores, respectivamente;

projeções de manuscritos: ambas as coleções *50 Words* e *Words Synonyms* [83] são compostas por dados extraídos de projeções de palavras digitalizadas de documentos históricos;

classificação de sensor: os conjuntos de dados *Mote Strain* [84] e *Chlorine Concentration* [84, 85] são sinais de sensores para análise de concentração de cloro em água e de umidade no ar;

dados simulados: os conjuntos *Synthetic Control* [86], *CBF* [87], *Mallat* [88] e *Two Patterns* [89] são compostos por séries temporais artificiais, geradas para avaliação de técnicas;

rastreamento: os conjuntos de dados deste grupo são compostos por sinais obtidos de medições das coordenadas de um objeto alvo, através de acelerômetros ou sistemas de vigilância. São as coleções *Sony AIBO Robot* [90], *Symbols* [91], *Gun Point* [92], *Cricket* [93], *Haptics* [94], *Inline Skate* [10], e *Gesture Librarys* [95];

cromatografia: as coleções *Olive Oil* [96], *Coffee* [97] e *Beef* [98] são compostas por séries temporais extraídas de técnicas de cromatografia;

diversos: neste grupo estão os conjuntos *Lightning* [99], *Medical Images* [100], *Star Light Curves* [101], com séries temporais extraídas de análises sobre transientes eletromagnéticos, intensidades de pixels em imagens, e medições da magnitude de luz em imagens telescópicas espaciais, respectivamente.

A utilização de todos os conjuntos é uma recomendação do próprio repositório e, assim, seguiu-se neste trabalho. De fato, a quantidade e a diversidade dos dados tornam a avaliação da abordagem proposta mais abrangente, uma vez que poucos conjuntos de dados restringem a capacidade de se fazer afirmações em uma análise comparativa de técnicas [43].

5.2 Experimentos

A abordagem em dois estágios exigiu a configuração de diversos parâmetros. No primeiro estágio, a série temporal é transformada para a nova representação e, em seguida, realiza-se a extração de características.

O primeiro passo para se ajustar o modelo AR sobre a série temporal, é encontrar um valor adequado para a ordem p . Algumas estimativas, pouco subjetivas, podem ser empregadas como o BIC e o AIC, como visto no Capítulo 2. Nos experimentos, foram testadas as duas estimativas sobre os conjuntos de dados e, empiricamente, decidiu-se utilizar o BIC. A estimativa por AIC forneceu valores muito elevados para p , comparando-se ao BIC. Embora

pareça ser mais vantajoso ter muitos coeficientes em relação ao potencial discriminativo para o treinamento, um valor alto de p pode capturar ruídos e, assim, comprometer o modelo extraído da série temporal [7, 56].

Após obter a ordem p , os coeficientes do modelo AR devem ser ajustados. Essa tarefa pode ser realizada por estimadores de Yule-Walker, de Máxima Verossimilhança, ou o método dos Mínimos Quadrados. Nos experimentos, foram testados os estimadores por Mínimos Quadrados e Máxima Verossimilhança. Não foi observada, empiricamente, diferença significativa entre os valores dos coeficientes ajustados sobre os dados, entre os dois métodos. Dessa forma, por ser o método padrão da ferramenta utilizada, optou-se pelo estimador de Máxima Verossimilhança para toda a avaliação experimental.

As medidas RQA, por outro lado, são calculadas sobre o RP, que modela a trajetória dos estados no espaço de fases. Primeiramente, a série temporal é desdobrada no espaço de fases de m dimensões de imersão e a dimensão de separação τ , como visto no Capítulo 2. Em seguida, o RP é modelado exigindo-se a definição do parâmetro ε , que define o raio de cobertura da vizinhança (ver Figura 4.2 (b)). A definição deste parâmetro representa um *trade-off* entre definir um valor tão baixo quanto possível, e que garanta, porém, recorrências suficientes para a análise sobre o RP.

Existem diversas maneiras para se estimar o valor de ε , algumas delas são discutidas em [102]. Em [103], é apontado $\varepsilon = 0,1\mu$ como uma estimativa geral aceitável. Entretanto, neste Trabalho de Mestrado, tal estimativa não garantiu pontos de recorrência suficientes sobre alguns conjuntos de dados, para que as medidas RQA pudessem ser extraídas. Nesse sentido, definiu-se o valor deste parâmetro, empiricamente, como $\varepsilon = \eta\mu$, onde μ representa o valor médio da série temporal e $\eta \in (0,05; 1,1]$. Neste intervalo, os conjuntos de dados foram modelados pelo RP sem danos para a extração das medidas RQA. Por sua vez, a extração das medidas RQA exige a definição dos parâmetros l_{\min} e v_{\min} (ver Equações 4.4 e 4.11) que, neste trabalho, foram definidas pelos valores padrões da ferramenta utilizada.

Após a extração das características, através do modelo AR e do RQA, tem-se o espaço de características para o classificador. No entanto, no ajustamento do modelo AR sobre as séries temporais, estima-se uma ordem p por item do conjunto de dados. Neste trabalho, empregou-se duas formas para generalizar um valor p : a ordem moda e a ordem máxima do conjunto de dados. Dessa forma, todas as séries temporais de um conjunto de dados são representadas por vetores de característica de um mesmo comprimento.

No segundo estágio, é realizada a classificação propriamente dita. O classificador SVM com *kernel* RBF exige a definição dos parâmetros livres γ e C . O primeiro, está relacionado à variância da função Gaussiana (RBF), por exemplo, quanto maior seu valor, menor a variância do modelo treinado e maior o viés e, por outro lado, quanto menor o valor de γ , maior a variância do modelo treinado e menor o viés. O segundo parâmetro está relacionado ao custo do relaxamento das restrições das margens. Quanto maior o valor C , maior a penalidade sobre itens preditos erroneamente. Ou seja, valores altos de C podem resultar *overfitting*.

A maneira mais comum de se estimar tais parâmetros, é através da técnica *grid-search*. Através desta técnica, define-se um intervalo de valores para cada parâmetro, e testa-se cada par de valores (γ_i, C_j) durante a validação cruzada. Neste experimento, definiu-se a melhor combinação (γ_i, C_j) buscando-se em $C = \{2^{-8}, 2^{-7}, \dots, 2^8\}$ e $\gamma = \{2^{-8}, 2^{-7}, \dots, 2^8\}$.

Adicionalmente, o problema multiclasse para o SVM é tratado com a política um-contra-um (Seção 4.1.2). Essa é a política empregada pelo pacote do R utilizado, o e1071, que é uma interface para a ferramenta LIBSVM [104].

Definidas as parametrizações dos métodos envolvidos na avaliação experimental, pode-se seguir para os resultados.

5.2.1 Classificação sobre características separadas

Na Figura 5.2, são apresentados os desempenhos do classificador SVM, treinado e testado sobre cada conjunto de características separadamente. É possível verificar que, para o classificador modelado sobre as medidas RQA, alcançou-se as melhores taxas de erro em 27 dos 41 conjuntos de dados. Em apenas 6 deles, o classificador apresentou o pior desempenho, porém, sendo apenas 1 com diferença significativa (*AIBO Robot Surface II*).

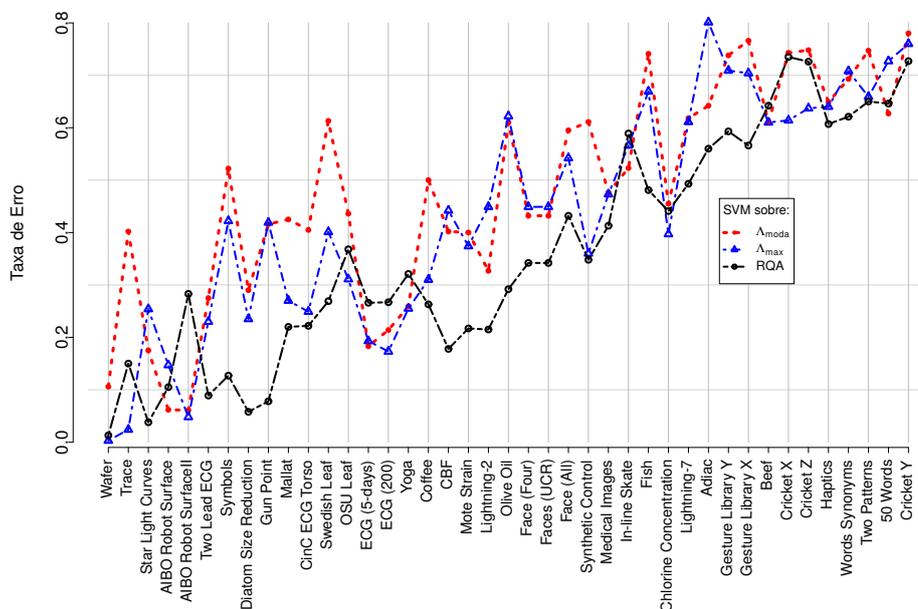


Figura 5.2: Desempenho de classificação para cada conjunto de características.

Ao comparar os classificadores modelados sobre Λ_{moda} e Λ_{max} , verifica-se que o segundo apresentou melhor desempenho em 26 conjuntos de dados e, em 10 deles, a vantagem foi acima de 0,1, na taxa de erro. Além disso, em apenas 4 conjunto de dados, o SVM sobre Λ_{moda} foi o melhor, com diferença de 0,1, na taxa de erro. Nesse sentido, os classificadores sobre Λ_{max} e RQA passam a ser mais interessantes, em relação a comparação com os classificadores de referência 1-NN.

5.2.2 Classificação sobre características combinadas

Na Figura 5.3, são apresentados os desempenhos dos classificadores modelados sobre as características combinadas, $\{\Lambda_{moda}, RQA\}$ e $\{\Lambda_{max}, RQA\}$. Como pode ser analisado

visualmente, os desempenhos são similares e, apenas sobre alguns conjuntos de dados, o SVM sobre $\{\Lambda_{\max}, RQA\}$ é melhor, com diferença visualmente significativa.

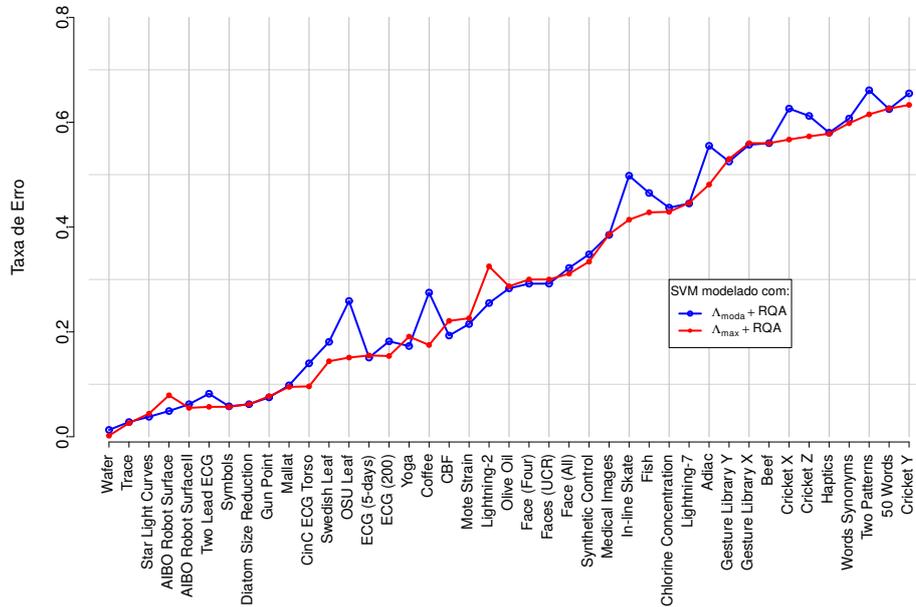


Figura 5.3: Desempenho da classificação para conjuntos de características combinadas.

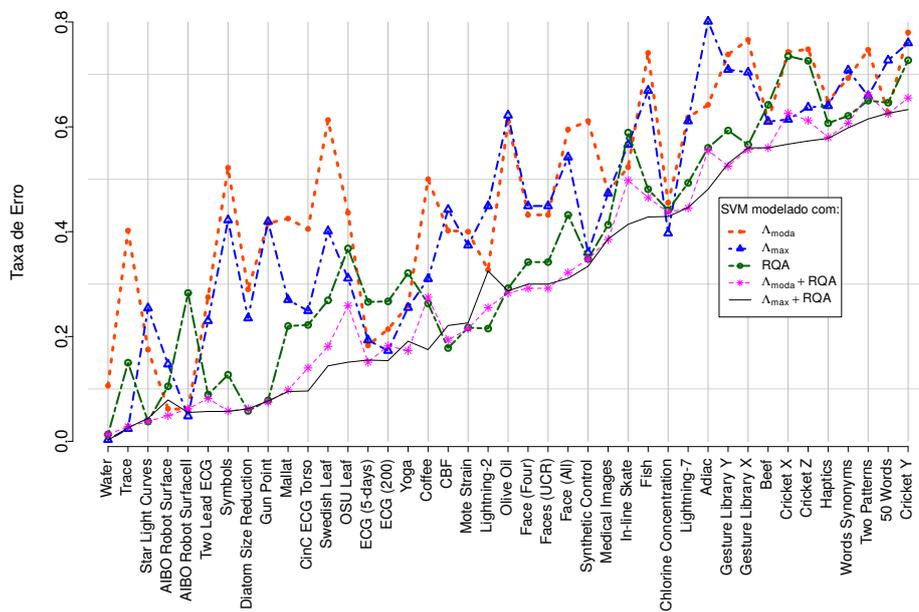


Figura 5.4: Desempenho da classificação sobre todas as características em estudo.

Os desempenhos dos classificadores sobre todas as configurações de características são exibidos na Figura 5.4. É possível visualizar que, em geral, o classificador SVM sobre as características combinadas, $\{\Lambda_{\max}, RQA\}$ e $\{\Lambda_{\text{moda}}, RQA\}$, possui desempenho superior às outras configurações. Na Tabela 5.2 são apresentadas as taxas de erro médias (com desvio

padrão³) referentes à classificação sobre as características propostas⁴. Sobre alguns conjuntos de dados, são notáveis os desempenhos obtidos.

Tabela 5.2: Desempenho da classificação sobre características individuais e combinadas.

Conjuntos de dados	Λ_{\max}	RQA	$\Lambda_{\text{moda}}+\text{RQA}$	$\Lambda_{\max}+\text{RQA}$
Wafer	0.003	0.013	0.013	0.002
Trace	0.024	0.15 \pm 0.003	0.028	0.026
Star Light Curves	0.254 \pm 0.003	0.038 \pm 0.003	0.038 \pm 0.003	0.044 \pm 0.003
AIBO Robot Surface	0.147 \pm 0.001	0.105 \pm 0.001	0.049	0.079
AIBO Robot SurfaceII	0.048	0.283	0.062	0.055
Symbols	0.422 \pm 0.001	0.127	0.058	0.057
Two Lead ECG	0.23	0.089	0.082	0.057
Diatom Size Reduction	0.235 \pm 0.003	0.058 \pm 0.002	0.062 \pm 0.001	0.062 \pm 0.001
Gun Point	0.419	0.078 \pm 0.001	0.075 \pm 0.002	0.077 \pm 0.001
Mallat	0.27	0.22	0.098	0.095
CinC ECG Torso	0.249 \pm 0.001	0.222	0.14	0.096
Swedish Leaf	0.401 \pm 0.002	0.269	0.181	0.144
OSU Leaf	0.311 \pm 0.003	0.368 \pm 0.002	0.259 \pm 0.001	0.151 \pm 0.002
ECG (5-Days)	0.193	0.266 \pm 0.001	0.151 \pm 0.001	0.155 \pm 0.001
ECG (200)	0.173 \pm 0.003	0.267 \pm 0.002	0.182 \pm 0.003	0.154 \pm 0.003
Yoga	0.255	0.321	0.173	0.191
Coffee	0.31 \pm 0.01	0.263 \pm 0.009	0.275 \pm 0.007	0.175 \pm 0.007
CBF	0.442 \pm 0.002	0.178	0.193 \pm 0.001	0.221 \pm 0.001
Mote Strain	0.374 \pm 0.001	0.217 \pm 0.001	0.215 \pm 0.001	0.226 \pm 0.001
Lighting-2	0.449 \pm 0.005	0.215 \pm 0.003	0.255 \pm 0.004	0.325 \pm 0.005
Olive Oil	0.622 \pm 0.004	0.292 \pm 0.008	0.283 \pm 0.01	0.287 \pm 0.01
Face (Four)	0.449 \pm 0.006	0.342 \pm 0.006	0.292 \pm 0.006	0.3 \pm 0.005
Faces (UCR)	0.449	0.342	0.292	0.3
Face (All)	0.542	0.432	0.322 \pm 0.001	0.311
Synthetic Control	0.36 \pm 0.003	0.348 \pm 0.003	0.348 \pm 0.003	0.334 \pm 0.003
Medical Images	0.473	0.413	0.385	0.386
In-line Skate	0.566 \pm 0.001	0.589 \pm 0.003	0.498 \pm 0.004	0.414 \pm 0.002
Fish	0.669 \pm 0.004	0.481 \pm 0.003	0.465 \pm 0.004	0.428 \pm 0.003
Chlorine Concentration	0.397	0.441	0.437	0.429
Lighting-7	0.611 \pm 0.006	0.493 \pm 0.003	0.445 \pm 0.004	0.446 \pm 0.001
Adiac	0.801 \pm 0.002	0.56 \pm 0.002	0.555 \pm 0.002	0.481 \pm 0.003
Gesture Library Y	0.709 \pm 0.001	0.593	0.525	0.53 \pm 0.001
Gesture Library X	0.704 \pm 0.001	0.566	0.557	0.56
Beef	0.61 \pm 0.01	0.642 \pm 0.01	0.56 \pm 0.010	0.56 \pm 0.01
Cricket X	0.614 \pm 0.002	0.735 \pm 0.001	0.626 \pm 0.001	0.567 \pm 0.002
Cricket Z	0.637 \pm 0.002	0.726 \pm 0.003	0.612 \pm 0.002	0.573 \pm 0.002
Haptics	0.64 \pm 0.003	0.607 \pm 0.003	0.58 \pm 0.003	0.578 \pm 0.003
Words Synonyms	0.708	0.621	0.607 \pm 0.002	0.598
Two Patterns	0.659	0.65 \pm 0.001	0.661 \pm 0.001	0.615 \pm 0.001
50 Words	0.727	0.646 \pm 0.001	0.625 \pm 0.001	0.626 \pm 0.002
Cricket Y	0.76 \pm 0.002	0.727 \pm 0.003	0.655 \pm 0.002	0.633 \pm 0.003

³Sobre a amostragem dos experimentos, foi calculado o desvio padrão da média. Por conveniência, os valores abaixo de 0,001 não foram expostos.

⁴As taxas de erro do SVM sobre Λ_{moda} foram omitidas por legibilidade, uma vez que o baixo poder discriminativo não foi aproveitado para comparações adicionais.

Por exemplo, sobre o conjunto de dados *Wafer*, todas as características exibiram ótimo poder discriminativo. Por outro lado, sobre o conjunto de dados *Crickets Y*, a classificação sobre todas as características, combinadas ou não, foram significativamente ruins. Outro exemplo notável, é o desempenho sobre o conjunto de dados *CinC ECG Torso*. A classificação sobre as características $\{\Lambda_{\max}, \text{RQA}\}$ é significativamente melhor que todas as demais.

5.2.3 Comparação com a abordagem referencial 1-NN

Os resultados apresentados na Figura 5.4, evidenciam o poder discriminativo das características $\{\Lambda_{\text{moda}}, \text{RQA}\}$ e $\{\Lambda_{\max}, \text{RQA}\}$ sobre as demais. Nesse sentido, tais configurações são tomadas para comparação com os classificadores 1-NN com as funções de distâncias Euclidiana, DTW com espaço de busca otimizado por janela, e com DTW convencional [10, 73].

Na Tabela 5.3, são apresentadas as taxas de erro das abordagens propostas e das abordagens da literatura. Os valores referentes às taxas de erro para os classificadores 1-NN foram todos coletados do repositório *online* UCR⁵. Nesta Tabela, os valores estão organizados em três partes. A parte superior exhibe os valores para os quais as abordagens propostas foram mais eficazes. A parte intermediária exhibe os valores para os quais as abordagens propostas se diferenciaram, no máximo, em 0,1 na taxa de erro, na comparação geral. Na parte inferior da tabela, são apresentados os valores cujas abordagens propostas foram significativamente inferiores na comparação geral.

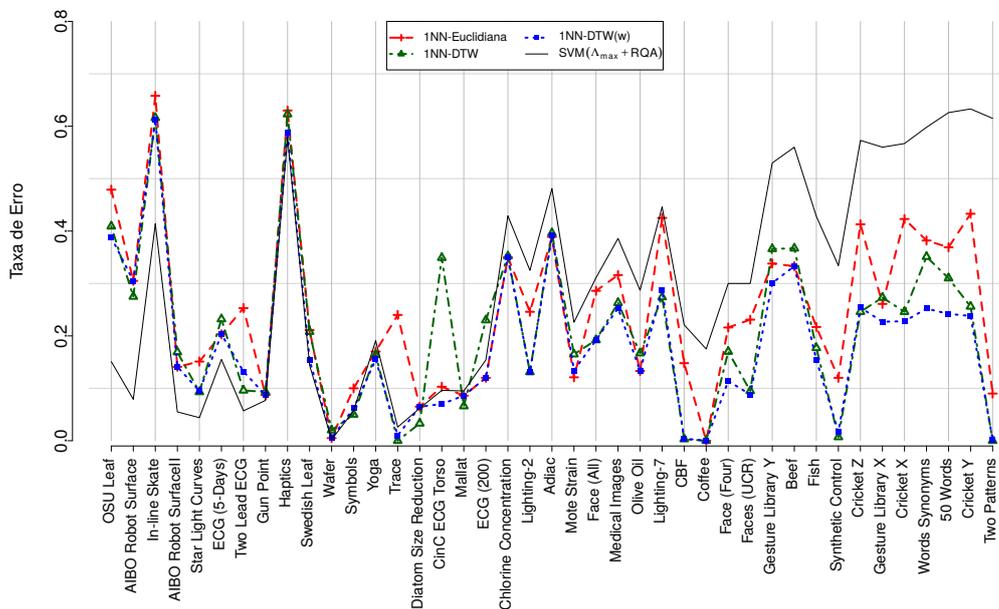


Figura 5.5: Relação entre desempenhos da melhor configuração de características da abordagem proposta e abordagens de referência.

⁵Para uma simples verificação das taxas de erro expostas no repositório UCR [10], implementou-se os classificadores 1-NN com as funções de distâncias DTW e Euclidiana. Sobre alguns conjuntos de dados aleatoriamente selecionados, os desempenhos obtidos foram similares, senão iguais, aos expostos no repositório.

A partir da Tabela 5.3, é possível notar que a configuração $\{\Lambda_{\max}, \text{RQA}\}$ é a melhor dentre as duas combinações possíveis. Portanto, esta é tomada para a comparação exibida na Figura 5.5. Visualmente, é possível notar que em quase metade dos conjuntos de dados, a abordagem proposta é similar ou melhor que as abordagens da literatura. Entretanto, sobre os conjuntos de dados cujo desempenho da abordagem é inferior na comparação, as taxas de erro são substancialmente superiores.

Tabela 5.3: Comparação entre taxas de erros da abordagens proposta e de referência [10].

Conjunto de Dados	SVM		1-NN		
	$\{\Lambda_{\max}, \text{RQA}\}$	$\{\Lambda_{\text{moda}}, \text{RQA}\}$	Euclid.	DTW(w)	DTW
OSU Leaf	0.151	0.259	0.479	0.388	0.409
AIBO Robot Surface	0.079	0.049	0.305	0.305	0.275
In-line Skate	0.414	0.498	0.658	0.613	0.616
AIBO Robot SurfaceII	0.055	0.062	0.141	0.141	0.169
Star Light Curves	0.044	0.038	0.151	0.095	0.093
ECG (5-Days)	0.155	0.151	0.203	0.203	0.232
Two Lead ECG	0.057	0.082	0.253	0.132	0.096
Gun Point	0.077	0.075	0.087	0.087	0.093
Haptics	0.578	0.58	0.630	0.588	0.623
Swedish Leaf	0.144	0.181	0.211	0.154	0.208
Wafer	0.002	0.013	0.005	0.005	0.020
Symbols	0.057	0.058	0.100	0.062	0.050
Yoga	0.191	0.173	0.170	0.155	0.164
Trace	0.026	0.028	0.240	0.010	0.000
Diatom Size Reduction	0.062	0.062	0.065	0.065	0.033
CinC ECG Torso	0.096	0.14	0.103	0.070	0.349
Mallat	0.095	0.098	0.086	0.086	0.066
ECG (200)	0.154	0.182	0.120	0.120	0.230
Chlorine Concentration	0.429	0.437	0.350	0.350	0.352
Lighting-2	0.325	0.255	0.246	0.131	0.131
Adiac	0.481	0.555	0.389	0.391	0.396
Mote Strain	0.226	0.215	0.121	0.134	0.165
Face (All)	0.311	0.322	0.286	0.192	0.192
Medical Images	0.386	0.385	0.316	0.253	0.263
Olive Oil	0.287	0.283	0.133	0.133	0.167
Lighting-7	0.446	0.445	0.425	0.288	0.274
CBF	0.221	0.193	0.148	0.004	0.003
Coffee	0.175	0.275	0.000	0.000	0.000
Face (Four)	0.300	0.292	0.216	0.114	0.170
Faces (UCR)	0.300	0.292	0.231	0.088	0.095
Gesture Library Y	0.530	0.525	0.338	0.301	0.366
Beef	0.560	0.56	0.333	0.333	0.367
Fish	0.428	0.465	0.217	0.154	0.177
Synthetic Control	0.334	0.348	0.120	0.017	0.007
Cricket Z	0.573	0.612	0.413	0.254	0.246
Gesture Library X	0.560	0.557	0.261	0.227	0.273
Cricket X	0.567	0.626	0.423	0.228	0.246
Words Synonyms	0.598	0.607	0.382	0.252	0.351
50 Words	0.626	0.625	0.369	0.242	0.310
Cricket Y	0.633	0.655	0.433	0.238	0.256
Two Patterns	0.615	0.661	0.090	0.002	0.000

5.3 Análise dos Resultados

5.3.1 Influências das características na classificação

Dos resultados de classificação obtidos na Seção 5.2, é possível depreender algumas informações.

- Comparando-se o desempenho de classificação sobre as características Λ_{moda} e Λ_{max} , a segunda configuração foi superior em 28 dos 41 conjuntos de dados;
- O desempenho de classificação sobre o conjunto de características RQA é superior aos desempenhos sobre Λ_{moda} e Λ_{max} , em 27 dos 41 conjuntos de dados;
- Em uma comparação geral entre as configurações de características separadas e combinadas, nota-se que as configurações combinadas resultam em desempenhos superiores às configurações separadas, em 34 dos 41 conjuntos de dados;
- Em 33 conjuntos de dados, o desempenho de classificação sobre $\{\Lambda_{max}, RQA\}$ é superior as respectivas configurações separadas Λ_{max} e RQA;

Na Figura 5.6, é exibida a relação entre os desempenhos de classificação das características propostas⁶ e o número de classes por conjunto de dados. Embora os melhores desempenhos tenham sido obtidos sobre conjuntos de dados com poucas classes, e os piores desempenhos tenham sido obtidos sobre conjuntos de dados com muitas classes, não é possível estabelecer relações entre desempenho e quantidade de classes, para a abordagem proposta.

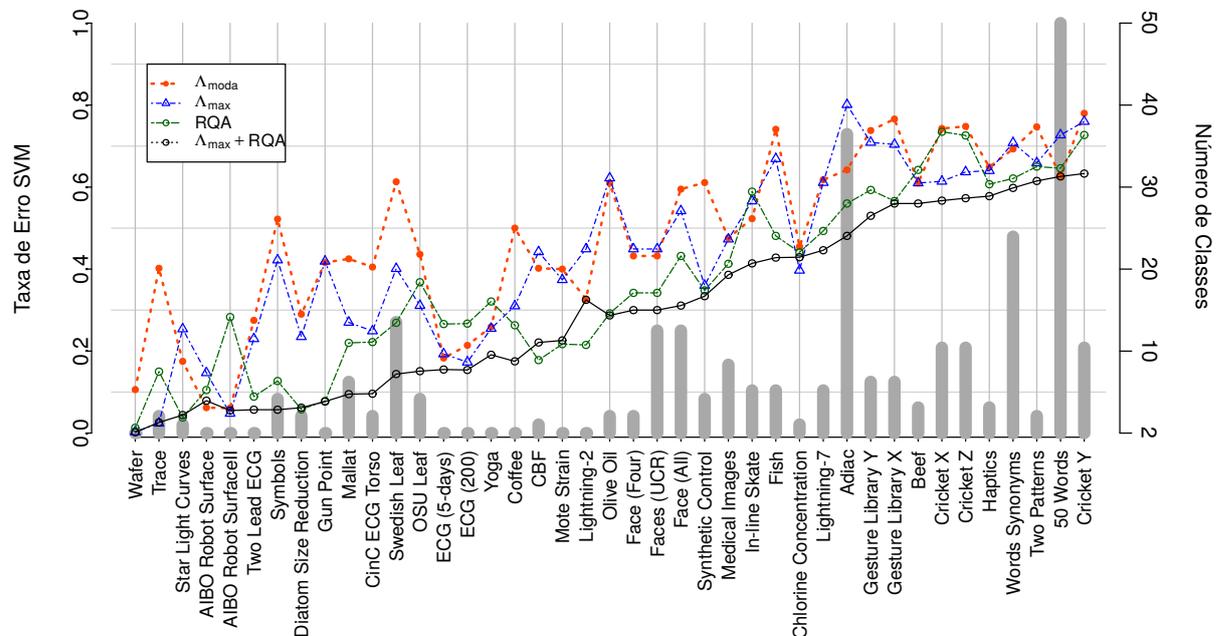


Figura 5.6: Relação entre o desempenho das características propostas e o número de classes por conjunto de dados. As taxas de erros estão alinhadas com o eixo vertical esquerdo.

⁶Convenientemente, omitiu-se a combinação $\{\Lambda_{moda}, RQA\}$ por ser similar à segunda combinação e para tornar o gráfico mais legível.

As técnicas utilizadas para representar a série temporal, o RP e o modelo AR, são apropriadas para lidar com níveis de determinismo altos e baixos (estocasticidade), respectivamente. O nível de determinismo dos conjuntos de dados é apresentado na Figura 5.7. Através do diagrama de caixas, consegue-se verificar a distribuição (empírica) do nível de determinismo (RQA-DET) entre as séries temporais de cada conjunto de dados.

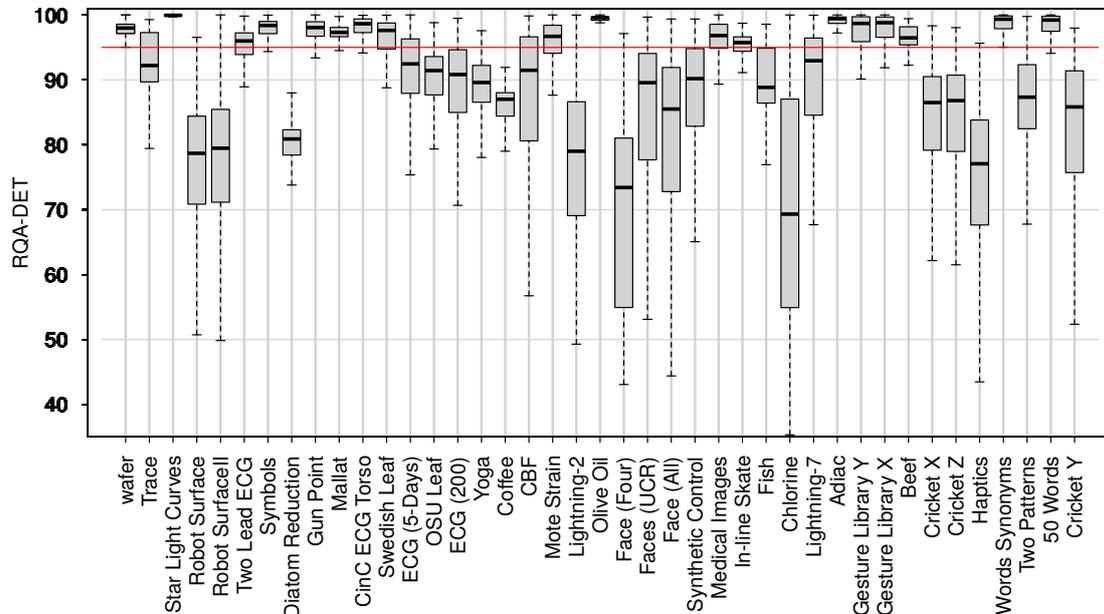


Figura 5.7: Variação de RQA-DET entre as séries temporais de cada conjunto de dados. Séries temporais determinísticas exibem um valor DET ≥ 95 , enquanto que séries temporais estocásticas exibem DET < 95 [8].

É possível analisar os desempenhos de classificação sobre os conjuntos de características, considerando a Figura 5.7 e os resultados apresentados na Figura 5.6. Não considerou-se os conjuntos de dados sobre os quais o classificador treinado com as características exibiu desempenho inferior a um classificador aleatório.

Dentre os 13 conjuntos de dados considerados determinísticos:

- O desempenho de classificação sobre as características RQA foi superior, ou igual, sobre as outras características;
- Em 4 deles, o desempenho do classificador treinado com as características Λ_{\max} foi superior, ou igual, às outras características;
- Em apenas 1 deles, o desempenho obtido sobre as características Λ_{moda} foi superior, ou igual, à classificação sobre as demais características.

Dentre os 18 conjunto de dados considerados estocásticos:

- Em 11 deles, o desempenho de classificação sobre as características RQA foi superior, ou igual, às demais características;

- Em 8 deles, o classificador treinado sobre as características Λ_{\max} foi superior, ou igual, às demais;
- Em 4 deles, o desempenho de classificação treinado sobre as características Λ_{moda} foi superior, ou igual, às outras características.

O que mais se destaca entre os números previamente apresentados, é o baixo desempenho para a classificação das características Λ_{moda} e Λ_{\max} , principalmente, sobre os conjuntos de dados estocásticos. Para Λ_{moda} , uma justificativa possível, é que $\text{AR}(p = \text{ordem } \textit{moda})$ tenha gerado menos coeficientes que o necessário, e tal generalização de ordem não seja uma boa estratégia.

Entretanto, para Λ_{\max} , onde $\text{AR}(p = \text{ordem } \textit{máxima})$, o desempenho de classificação também não foi satisfatório, dado que superou as demais características somente em 8 de 18 conjuntos de dados estocásticos. Uma justificativa possível, seria que a estratégia de generalização por ordem máxima tenha atribuído ordens elevadas para dados que poderiam ser ajustados com ordens inferiores, superajustando ou capturando ruídos e *outliers* [57]. Dessa forma, se fazem necessárias análises mais profundas para se investigar estratégias de generalização para a ordem do modelo AR, no contexto da classificação de séries temporais.

O desempenho de classificação sobre as características RQA, por outro lado, superou o desempenho sobre as demais características, para os conjuntos de dados determinísticos. Além disso, superou as demais características, também, em 11 dos 18 conjuntos de dados estocásticos. As características RQA vão além da identificação de propriedades determinísticas puras. Dessa forma, o bom desempenho sobre conjuntos de dados estocásticos pode ser ocasionado por outras propriedades presentes nos dados, que podem ser analisadas pela Análise de Quantificação de Recorrência, como o caos e não linearidades estocásticas, que não foram abordadas neste trabalho de mestrado.

Por fim, nota-se que em 83% dos conjuntos de dados, o desempenho de classificação sobre as características combinadas é superior ao desempenho sobre todas as características separadas, tanto em conjuntos de dados determinísticos quanto estocásticos. Isso pode ser observado na Figura 5.6, onde o desempenho sobre as características $\{\Lambda_{\max}, \text{RQA}\}$ é superior às respectivas características separadas. Assim, para 83% dos conjuntos de dados, o potencial discriminativo das características extraídas, considerando influências determinísticas e estocásticas, de fato, são complementares entre si, para melhoria de desempenho de classificação de séries temporais.

5.3.2 Sobre a comparação à abordagem 1-NN da literatura

Comparando-se a melhor configuração proposta, $\{\Lambda_{\max}, \text{RQA}\}$, com a abordagem de referência de melhor desempenho, o 1-NN com DTW otimizado, nota-se que:

- Em 11 deles a abordagem proposta alcançou desempenho superior;
- Em outros 11, o desempenho da abordagem proposta foi inferior em, no máximo, 0,1 na taxa de erro;

- Sobre os 19 conjuntos de dados restantes, o desempenho da abordagem proposta foi substancialmente inferior.

Para fins de comparação, realizou-se testes estatísticos para diferença significativa, através do teste pareado de Wilcoxon. Para o nível de 95% de confiança, a hipótese nula considera que não há diferença significativa entre as amostras avaliadas. Para os valores de $w\text{-}\rho < 0,05$, tal hipótese é descartada.

Ao tomar-se as características $\{\Lambda_{\max}, \text{RQA}\}$ como referência da abordagem proposta, e o 1-NN com DTW otimizado como a referência padrão, é possível verificar que há diferença significativa entre todos os desempenhos ($w\text{-}\rho = 0,0006$), mesmo notando-se visualmente na Figura 5.5, similaridades entre as taxas de erros sobre metade dos conjuntos de dados.

No teste realizado somente sobre os desempenhos cuja abordagem proposta foi superior (o quadrante superior da Tabela 5.3), apontou-se $w\text{-}\rho = 0.00384$, indicando que há diferença significativa entre os desempenhos.

Ao considerar-se os desempenhos da abordagem proposta que diferem em até 0,1 nas taxas de erro, em relação à abordagem de referência (o quadrante intermediário da Tabela 5.3), é possível verificar estatisticamente que são similares em 8 dos 11 conjuntos de dados.

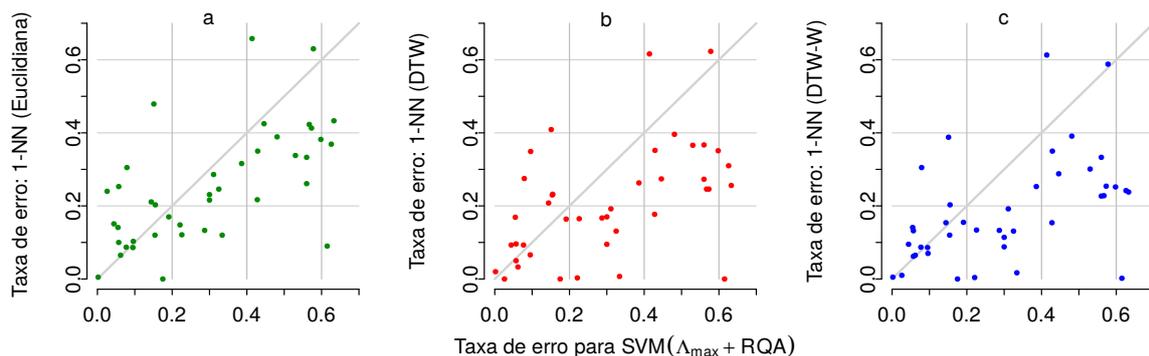


Figura 5.8: Comparação entre a melhor configuração da abordagem proposta (eixos x) e as técnicas de referência (eixos y), onde os pontos representam os conjuntos de dados. As relações são SVM(Λ_{\max} , RQA) *versus* (a) 1-NN(Euclidiana), (b) 1-NN(DTW), e (c) 1-NN(DTW-W)

Sobre a Figura 5.8, pode-se visualizar tal similaridade de desempenho entre as abordagens proposta e de referência. O eixo horizontal indica o desempenho da abordagem proposta, o eixo vertical, os desempenhos das abordagens de referência (uma em cada gráfico), e os pontos representam os conjuntos de dados. Basicamente, quanto mais pontos sobre a parte superior à diagonal, melhor o desempenho da abordagem proposta em relação a abordagem de referência. É possível notar que em todas as comparações (os três gráficos), o classificador proposto exibe desempenho inferior às abordagens de referência. Entretanto, quanto mais próximos os pontos estão da diagonal, menor a diferença entre os desempenhos de ambas abordagens, na comparação.

Dessa forma, considerando que, em 19 dos 41 conjuntos de dados, a abordagem proposta obteve desempenhos iguais ou superiores às abordagens de referência, é possível afirmar que a abordagem proposta é, de fato, promissora. Além disso, indica que a análise das influências

das propriedades intrínsecas são relevantes nos processos de representação e extração de características, e mutualmente complementares, para o problema da classificação.

Não obstante, os resultados obtidos mostram que técnicas bem estabelecidas da área de análise de séries temporais, exigem mais do que simples adaptações, para serem empregadas na tarefa de classificação.

Capítulo 6

Conclusões

O estudo realizado neste Trabalho de Mestrado, envolve duas grandes áreas de pesquisa: a Análise de Séries Temporais e a Mineração de Dados.

Este trabalho propõe uma abordagem para o problema da classificação de séries temporais, inspirando-se em uma metodologia apresentada em [8], para a modelagem baseada em propriedades intrínsecas do processo gerador dos dados.

A abordagem proposta é executada em dois estágios: primeiramente, as séries temporais são processadas através de técnicas apropriadas para modelar os comportamentos determinístico e estocástico. Nesse sentido, são empregados, respectivamente, o Gráfico de Recorrência (RP) sobre as séries temporais desdobradas no espaço de fases, e o modelo autorregressivo (AR). Em seguida, características inerentes às novas representações obtidas, os coeficientes do modelo AR e medidas de Análise de Quantificação de Recorrência (RQA), são extraídas para compor o novo espaço de características. No segundo estágio, a classificação, propriamente dita, é realizada sobre este novo espaço característico, empregando-se o classificador SVM.

Com base nos resultados obtidos da avaliação experimental, verificou-se que dentre os conjuntos de características propostos, o conjunto RQA demonstrou maior potencial discriminativo para a classificação, tanto em conjuntos de dados determinísticos, quanto estocásticos. Por outro lado, as características fornecidas pelos modelos AR não conseguiram modelar as influências estocásticas presente nos dados, de forma efetiva. Isso pode estar relacionado à estratégia de estimação do valor da ordem, para todo o conjunto de dados.

As medidas extraídas do RQA vão além de propriedades determinísticas puras. O bom desempenho sobre os conjuntos de dados estocásticos pode ser ocasionado por outras propriedades presentes nas séries temporais, que podem ser investigadas pela Análise de Quantificação de Recorrência, como o caos e não linearidades estocásticas, que não foram abordadas neste Trabalho de Mestrado.

O ponto importante a se destacar, é que as características propostas são mutualmente complementares, ou seja, quando combinadas, o classificador exibe o melhor desempenho dentre as outras configurações de características.

Nos resultados dos experimentos comparativos, verificou-se que em 19 dos 41 conjunto de dados, a melhor configuração da abordagem proposta foi igual ou superior, às abordagens

de referência da literaturam, o classificador 1-NN com as funções de distância Euclidiana, DTW, e DTW otimizado com janela de busca.

Considerando o exposto, portanto, é possível afirmar que a abordagem proposta é promissora. Além disso, indica que a análise das influências das propriedades intrínsecas são relevantes nos processos de representação e extração de características, e mutualmente complementares, para o problema da classificação.

Não obstante, os resultados obtidos mostram que técnicas bem estabelecidas da área de análise de séries temporais, exigem mais do que simples adaptações, para serem empregadas na tarefa de classificação.

6.1 Trabalhos Futuros

Os resultados da avaliação experimental da abordagem proposta neste Trabalho de Mestrado, revelam fraquezas que podem direcionar pesquisas futuras, como:

- Investigar técnicas de decomposição para auxiliar a descoberta de conhecimento em séries temporais, no contexto da classificação;
- Elaborar estratégias para estimação do valor ótimo (ou valores) da ordem do modelo AR, para o problema da classificação de séries temporais;
- Estender a abordagem para análise não linear e não estacionária nos procedimentos de representação e extração de características, no problema de classificação e/ou agrupamento.

Bibliografia

- [1] A. Jain, R. P. W. Duin, and J. Mao, “Statistical pattern recognition: a review,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 1, pp. 4–37, 2000.
- [2] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. John Wiley & Sons, 2012.
- [3] C. M. Antunes and A. L. Oliveira, “Temporal data mining: An overview,” in *KDD workshop on temporal data mining*, vol. 1, p. 13, 2001.
- [4] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. Academic Press, 2009.
- [5] A. R. Webb and K. D. Copsey, *Statistical Pattern Recognition*. John Wiley & Sons, 2011.
- [6] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and its Applications: with R Examples*. Springer, 2006.
- [7] P. A. Morettin and C. Toloi, *Análise de Séries Temporais*. Blucher, 2006.
- [8] R. P. Ishii, R. A. , and R. F. Mello, “Classification of time series generation processes using experimental tools: a survey and proposal of an automatic and systematic approach,” *International Journal of Computational Science and Engineering*, vol. 6, no. 4, pp. 217–237, 2011.
- [9] R. Rios, *Improving time series modeling by decomposing and analyzing stochastic and deterministic influences*. PhD thesis, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos - SP, Brasil, 2013.
- [10] E. Keogh, Q. Zhu, B. Hu, H. Y, X. Xi, L. Wei, and C. A. Ratanamahatana, “The UCR time series classification/clustering,” 2011. Disponível em http://www.cs.ucr.edu/~eamonn/time_series_data. Último acesso em Outubro de 2014.
- [11] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, 2008.
- [12] C. Chatfield, *The Analysis of Time Series: An Introduction*. CRC press, 2013.
- [13] S. S. Haykin and B. Van Veen, *Sinais e Sistemas*. Bookman, 2001.
- [14] E. Parzen, “Some recent advances in time series modeling,” *Automatic Control, IEEE Transactions on*, vol. 19, no. 6, pp. 723–730, 1974.

-
- [15] C. Chatfield, *Time-series Forecasting*. CRC Press, 2000.
- [16] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*. Springer, 2006.
- [17] G. P. Williams, *Chaos theory tamed*. Joseph Henry Press, 1997.
- [18] E. Ott, *Chaos in Dynamical Systems*. Cambridge University Press, 2002.
- [19] N. Marwan, M. Carmen Romano, M. Thiel, and J. Kurths, “Recurrence plots for the analysis of complex systems,” *Physics Reports*, vol. 438, no. 5, pp. 237–329, 2007.
- [20] K. T. Alligood, T. D. Sauer, and J. A. Yorke, *Chaos: An Introduction to Dynamical Systems*. Springer Berlin Heidelberg, 1997.
- [21] F. Takens, “Detecting strange attractors in turbulence,” *Dynamical Systems and Turbulence, Warwick 1980: Proceedings of a Symposium Held at the University of Warwick 1979/80*, pp. 366–381, 1981.
- [22] A. M. Fraser and H. L. Swinney, “Independent coordinates for strange attractors from mutual information,” *Physical review A*, vol. 33, no. 2, p. 1134, 1986.
- [23] M. B. Kennel, R. Brown, and H. D. Abarbanel, “Determining embedding dimension for phase-space reconstruction using a geometrical construction,” *Physical review A*, vol. 45, no. 6, p. 3403, 1992.
- [24] G. Kirchgässner, J. Wolters, and U. Hassler, *Introduction to Modern Time Series Analysis*. Springer, 2013.
- [25] H. Akaike, “Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes,” *Annals of the Institute of Statistical Mathematics*, vol. 26, no. 1, pp. 363–387, 1974.
- [26] G. Schwarz *et al.*, “Estimating the dimension of a model,” *The annals of statistics*, vol. 6, no. 2, pp. 461–464, 1978.
- [27] A. Newell, J. C. Shaw, and H. A. Simon, “Report on a general problem-solving program,” in *IFIP Congress*, vol. 256, p. 64, 1959.
- [28] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. Prentice Hall series in artificial intelligence, Prentice Hall, 2010.
- [29] A. Nanopoulos, R. Alcock, and Y. Manolopoulos, “Information processing and technology,” ch. Feature-based Classification of Time-series Data, pp. 49–61, Commack, NY, USA: Nova Science Publishers, Inc., 2001.
- [30] T. M. Mitchell, *Machine Learning*. McGraw-Hill Science/Engineering/Math, 1999.
- [31] F. Mörchen, *Time series knowledge mining*. PhD thesis, Phillips University, Marburg, Alemanha, 2002.
- [32] D. J. Hand, H. Mannila, and P. Smyth, *Principles of data mining*. MIT press, 2001.
- [33] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. Elsevier, 2011.

- [34] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, 2005.
- [35] N. Krupa, A. M. Mohd, E. Zahedi, S. Ahmed, and F. M. Hassan, “Antepartum fetal heart rate feature extraction and classification using empirical mode decomposition and support vector machine,” *Biomedical engineering online*, vol. 10, no. 1, p. 6, 2011.
- [36] Y. Kim, J. W. Chong, K. H. Chon, and J. Kim, “Wavelet-based AR-SVM for health monitoring of smart structures,” *Smart Materials and Structures*, vol. 22, no. 1, p. 015003, 2013.
- [37] R. Agrawal, K.-I. Lin, H. S. Sawhney, and K. Shim, “Fast similarity search in the presence of noise, scaling, and translation in time-series databases,” in *Proceedings of the 21th International Conference on Very Large Data Bases*, pp. 490–501, Morgan Kaufmann Publishers Inc., 1995.
- [38] S. Chu, E. Keogh, D. Hart, and M. Pazzani, “Iterative deepening dynamic time warping for time series,” in *Society for Industrial and Applied Mathematics. Proceedings of the SIAM International Conference on Data Mining*, p. 195, Society for Industrial and Applied Mathematics, 2002.
- [39] S. Kotsiantis, D. Kanellopoulos, and P. Pintelas, “Data preprocessing for supervised learning,” *International Journal of Computer Science*, vol. 1, no. 2, pp. 111–117, 2006.
- [40] R. Nisbet, G. Miner, and J. Elder, *Handbook of Statistical Analysis and Data Mining Applications*. Elsevier Science, 2009.
- [41] S. Kotsiantis, “Supervised machine learning: A review of classification techniques,” *Informatika (Ljubljana)*, vol. 31, no. 3, pp. 249–268, 2007.
- [42] P. Esling and C. Agon, “Time-series data mining,” *ACM Comput. Surv.*, vol. 45, no. 1, pp. 12:1–12:34, 2012.
- [43] E. Keogh and S. Kasetty, “On the need for time series data mining benchmarks: A survey and empirical demonstration,” *Data Mining and Knowledge Discovery*, vol. 7, no. 4, pp. 349–371, 2003.
- [44] I. López-Yáñez, L. Sheremetov, and C. Yáñez-Márquez, “A novel associative model for time series data mining,” *Pattern recognition letters*, vol. 41, pp. 23–33, 2014.
- [45] T.-c. Fu, “A review on time series data mining,” *Engineering Applications of Artificial Intelligence*, vol. 24, no. 1, pp. 164–181, 2011.
- [46] J. Lin, E. Keogh, S. Lonardi, and 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*, pp. 2–11, ACM, 2003.
- [47] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra, “Dimensionality reduction for fast similarity search in large time series databases,” *Knowledge and information Systems*, vol. 3, no. 3, pp. 263–286, 2001.

- [48] F. Mörchen and A. Ultsch, “Efficient mining of understandable patterns from multivariate interval time series,” *Data Mining and Knowledge Discovery*, vol. 15, no. 2, pp. 181–215, 2007.
- [49] D. Pandya, S. Upadhyay, and S. Harsha, “Fault diagnosis of rolling element bearing with intrinsic mode function of acoustic emission data using APF-KNN,” *Expert Systems with Applications*, vol. 40, no. 10, pp. 4137–4145, 2013.
- [50] F. Zhang and Q. Zhou, “HHT-SVM: An online method for detecting profile injection attacks in collaborative recommender systems,” *Knowledge-Based Systems*, vol. 65, pp. 96–105, 2014.
- [51] A. Bagnall and G. Janacek, “Clustering time series with clipped data,” *Machine Learning*, vol. 58, no. 2-3, pp. 151–178, 2005.
- [52] R. J. Povinelli, M. T. Johnson, A. C. Lindgren, and J. Ye, “Time series classification using gaussian mixture models of reconstructed phase spaces,” *Knowledge and Data Engineering, IEEE Transactions on*, vol. 16, no. 6, pp. 779–783, 2004.
- [53] Z. Xing, J. Pei, and E. Keogh, “A brief survey on sequence classification,” *SIGKDD Explor. Newsl.*, vol. 12, no. 1, pp. 40–48, 2010.
- [54] N.-J. Huan and R. Palaniappan, “Neural network classification of autoregressive features from electroencephalogram signals for brain-computer interface design,” *Journal of neural engineering*, vol. 1, no. 3, p. 142, 2004.
- [55] D. Ge, N. Srinivasan, and S. M. Krishnan, “Cardiac arrhythmia classification using autoregressive modeling,” *Biomedical engineering online*, vol. 1, no. 1, p. 5, 2002.
- [56] K. B. Venkataramana and C. C. Sekhar, “Large margin AR model for time series classification,” in *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pp. 1–4, 2008.
- [57] B. V. Kini and C. C. Sekhar, “Large margin mixture of AR models for time series classification,” *Applied Soft Computing*, vol. 13, no. 1, pp. 361–371, 2013.
- [58] Y. Xiong and D.-Y. Yeung, “Time series clustering with ARMA mixtures,” *Pattern Recognition*, vol. 37, no. 8, pp. 1675–1689, 2004.
- [59] X. Wang, A. Mueen, H. Ding, G. Trajcevski, P. Scheuermann, and E. Keogh, “Experimental comparison of representation methods and distance measures for time series data,” *Data Mining and Knowledge Discovery*, vol. 26, no. 2, pp. 275–309, 2013.
- [60] J. Serrà and J. L. Arcos, “An empirical evaluation of similarity measures for time series classification,” *Knowledge-Based Systems*, vol. 67, pp. 305–314, 2014.
- [61] Y.-S. Jeong, M. K. Jeong, and O. A. Omitaomu, “Weighted dynamic time warping for time series classification,” *Pattern Recognition*, vol. 44, no. 9, pp. 2231–2240, 2011.
- [62] X. Xi, E. Keogh, C. Shelton, L. Wei, and C. A. Ratanamahatana, “Fast time series classification using numerosity reduction,” in *Proceedings of the 23rd International Conference on Machine Learning*, pp. 1033–1040, ACM, 2006.

- [63] D. Srisai and C. A. Ratanamahatana, "Efficient time series classification under template matching using time warping alignment," in *Computer Sciences and Convergence Information Technology, 2009. ICCIT '09. Fourth International Conference on*, pp. 685–690, IEEE, 2009.
- [64] A. Jalalian and S. K. Chalup, "GDTW-P-SVMs: Variable-length time series analysis using support vector machines," *Neurocomputing*, vol. 99, pp. 270–282, 2013.
- [65] K. Kalpakis, D. Gada, and V. Puttagunta, "Distance measures for effective clustering of ARIMA time-series," in *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on*, pp. 273–280, IEEE, 2001.
- [66] J.-P. Eckmann, S. O. Kamphorst, and D. Ruelle, "Recurrence plots of dynamical systems," *EPL (Europhysics Letters)*, vol. 4, no. 9, p. 973, 1987.
- [67] C. J. Burges, "A tutorial on support vector machines for pattern recognition," *Data mining and knowledge discovery*, vol. 2, no. 2, pp. 121–167, 1998.
- [68] C. A. de Moraes Lima, *Comitê de Máquinas: uma abordagem unificada empregando máquinas de vetores-suporte*. PhD thesis, Universidade Estadual de Campinas, Campinas - SP, Brasil, 2004.
- [69] A. C. Lorena, *Investigação de estratégias para a geração de máquinas de vetores de suporte multiclases*. PhD thesis, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos - SP, Brasil, 2006.
- [70] B. Schölkopf and A. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Adaptive computation and machine learning, MIT Press, 2002.
- [71] J. A. Baranauskas and M. C. Monard, "Reviewing some machine learning concepts and methods," Technical Report 102, ICMC-USP, São Carlos - SP, 2000.
- [72] C. Orsenigo and C. Vercellis, "Combining discrete SVM and fixed cardinality warping distances for multivariate time series classification," *Pattern Recognition*, vol. 43, no. 11, pp. 3787–3794, 2010.
- [73] C. A. Ratanamahatana and E. Keogh, "Making time-series classification more accurate using learned constraints," SIAM, 2004.
- [74] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014.
- [75] H. Kaya and S. Gündüz-Ögüdücü, "A distance based time series classification framework," *Information Systems*, vol. 51, pp. 27–42, 2015.
- [76] Adiac, "Automatic diatom identification and classification," 2011. Disponível em <http://rbg-web2.rbge.org.uk/ADIAC/>. Último acesso em Outubro de 2014.
- [77] D.-J. Lee, R. B. Schoenberger, D. Shiozawa, X. Xu, and P. Zhan, "Contour matching for a fish recognition and migration-monitoring system," *Proc. SPIE*, vol. 5606, pp. 37–48, 2004.

- [78] A. Gandhi, “Content-based image retrieval: Plant species identification,” master’s thesis, Oregon State University, 2002.
- [79] O. Soderkvist, “Computer vision classification of leaves from swedish trees,” master’s thesis, Linkoping University, 2001.
- [80] L. Wei and E. Keogh, “Semi-supervised time series classification,” in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 748–753, ACM, 2006.
- [81] D. Roverso, “Multivariate temporal classification by windowed wavelet decomposition and recurrent neural networks,” in *3rd ANS international topical meeting on nuclear plant instrumentation, control and human-machine interface*, vol. 20, Citeseer, 2000.
- [82] R. T. Olszewski, “Generalized feature extraction for structural pattern recognition in time-series data,” tech. rep., DTIC Document, 2001.
- [83] T. M. Rath and R. Manmatha, “Word spotting for historical documents,” *International Journal of Document Analysis and Recognition (IJ DAR)*, vol. 9, no. 2, pp. 139–152, 2006.
- [84] J. Sun, S. Papadimitriou, and C. Faloutsos, “Online latent variable detection in sensor networks,” in *Data Engineering, 2005. ICDE 2005. Proceedings. 21st International Conference on*, pp. 1126–1127, April 2005.
- [85] L. Li, “Fast algorithms for mining co-evolving time series,” tech. rep., DTIC Document, 2011.
- [86] D. Pham and A. Chan, “Control chart pattern recognition using a new type of self-organizing neural network,” *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, vol. 212, no. 2, pp. 115–127, 1998.
- [87] N. Saito and R. R. Coifman, “Local discriminant bases and their applications,” *Journal of Mathematical Imaging and Vision*, vol. 5, no. 4, pp. 337–358, 1995.
- [88] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic press, 1999.
- [89] P. Geurts, *Contributions to decision tree induction: bias/variance tradeoff and time series classification*. PhD thesis, University of Liège Belgium, 2002.
- [90] D. Vail and M. Veloso, “Learning from accelerometer data on a legged robot,” in *Proceedings of the 5th IFAC/EURON symposium on intelligent autonomous vehicles*, 2004.
- [91] J. A. Brady, *Considering Complexity: Image Matching With Time Series*. PhD thesis, University of California Riverside, 2007.
- [92] C. A. Ratanamahatana and E. Keogh, *Three Myths about Dynamic Time Warping Data Mining*, ch. 50, pp. 506–510. 2005.

- [93] A. Mueen, E. Keogh, and N. Young, “Logical-shapelets: an expressive primitive for time series classification,” in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1154–1162, ACM, 2011.
- [94] B. Malek, M. Orozco, and A. El Saddik, “Novel shoulder-surfing resistant haptic-based graphical password,” in *Proc. EuroHaptics*, vol. 6, 2006.
- [95] J. Liu, L. Zhong, J. Wickramasuriya, and V. Vasudevan, “uwave: Accelerometer-based personalized gesture recognition and its applications,” *Pervasive and Mobile Computing*, vol. 5, no. 6, pp. 657–675, 2009.
- [96] H. S. Tapp, M. Defernez, and E. K. Kemsley, “Ftir spectroscopy and multivariate analysis can distinguish the geographic origin of extra virgin olive oils,” *Journal of Agricultural and Food Chemistry*, vol. 51, no. 21, pp. 6110–6115, 2003.
- [97] A. Bagnall, L. M. Davis, J. Hills, and J. Lines, “Transformation based ensembles for time series classification.,” *Proc. 12th SIAM Int. Conf. on Data Mining (SDM)*, vol. 12, pp. 307–318, 2012.
- [98] O. Al-Jowder, E. K. Kemsley, and R. H. Wilson, “Detection of adulteration in cooked meat products by mid-infrared spectroscopy,” *Journal of Agricultural and Food Chemistry*, vol. 50, no. 6, pp. 1325–1329, 2002.
- [99] D. R. Eads, D. Hill, S. Davis, S. J. Perkins, J. Ma, R. B. Porter, and J. P. Theiler, “Genetic algorithms and support vector machines for time series classification,” 2002.
- [100] J. C. Felipe, A. J. Traina, and C. Traina Jr, “A new similarity measure for histograms applied to content-based retrieval of medical images,” in *Proceedings of the 2006 ACM symposium on Applied computing*, pp. 258–259, ACM, 2006.
- [101] P. Protopapas, J. M. Giammarco, L. Faccioli, M. F. Struble, R. Dave, and C. Alcock, “Finding outlier lightcurves in catalogs of periodic variable stars,” *Monthly Notices of the Royal Astronomical Society*, pp. 677–696, 2006.
- [102] N. Marwan, “How to avoid potential pitfalls in recurrence plot based data analysis,” *International Journal of Bifurcation and Chaos*, vol. 21, no. 04, pp. 1003–1017, 2011.
- [103] J. P. Zbilut and C. L. Webber, “Embeddings and delays as derived from quantification of recurrence plots,” *Physics letters A*, vol. 171, no. 3, pp. 199–203, 1992.
- [104] C.-C. Chang and C.-J. Lin, “LIBSVM: A library for support vector machines,” *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011. Software disponível em <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.