



**Thiago Medeiros Carvalho**

**e-AutoMFIS: Modelo interpretável para  
previsão de séries multivariadas usando comitês  
de Sistemas de Inferência Fuzzy**

**Dissertação de Mestrado**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação em Engenharia Elétrica, do Departamento de Engenharia Elétrica da PUC-Rio.

Orientador : Prof. Marley Maria Bernardes Rebuszi Vellasco  
Coorientador: Prof. José Franco Machado do Amaral

Rio de Janeiro  
Abril de 2021



**Thiago Medeiros Carvalho**

**e-AutoMFIS: Modelo interpretável para  
previsão de séries multivariadas usando comitês  
de Sistemas de Inferência Fuzzy**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação em Engenharia Elétrica da PUC-Rio. Aprovada pela Comissão Examinadora abaixo:

**Prof. Marley Maria Bernardes Rebuzzi Vellasco**

Orientador

Departamento de Engenharia Elétrica – PUC-Rio

**Prof. José Franco Machado do Amaral**

Coorientador

Faculdade de Engenharia – UERJ

**Prof. Ricardo Tanscheit**

Departamento de Engenharia Elétrica – PUC-Rio

**Prof. Karla Tereza Figueiredo Leite**

Universidade do Estado do Rio de Janeiro – UERJ

**Prof. Daniel Furtado Leite**

Universidade Federal de Lavras – UFLA

Rio de Janeiro, 7 de Abril de 2021

Todos os direitos reservados. A reprodução, total ou parcial do trabalho, é proibida sem a autorização da universidade, do autor e do orientador.

### Thiago Medeiros Carvalho

Graduado em Engenharia Elétrica com ênfase em Sistemas Eletrônicos pela UERJ, teve como interesse durante a graduação o estudo sobre técnicas de inteligência computacional, especialmente na área de síntese automática de circuitos eletrônicos e otimização. Suas principais áreas de interesse são *Explainable AI* e robótica inteligente.

#### Ficha Catalográfica

Medeiros Carvalho, Thiago

e-AutoMFIS: Modelo interpretável para previsão de séries multivariadas usando comitês de Sistemas de Inferência Fuzzy / Thiago Medeiros Carvalho; orientador: Marley Maria Bernardes Rebuszi Vellasco; coorientador: José Franco Machado do Amaral. – 2021.

132 f: il. color. ; 30 cm

Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica, 2021.

Inclui bibliografia

1. Engenharia Elétrica – Teses.  
2. Previsão de séries multivariadas. 3. Sistema de Inferência Fuzzy. 4. Interpretabilidade. 5. Comitê de previsores. 6. Base de dados. I. Vellasco, Marley Maria Bernardes Rebuszi. II. Amaral, José Franco Machado do . III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. IV. Título.

CDD: 621.3

## Agradecimentos

Durante a jornada da vida no curso de mestrado, algumas pessoas foram responsáveis para que este momento final fosse possível, e aqui presto toda a minha gratidão àqueles que foram fundamentais nesta etapa.

Primeiramente, agradeço à minha família por todo o encorajamento fornecido para continuar na área de pesquisa, suporte fornecido no dia-a-dia e amor incondicional.

À minha orientadora, prof<sup>a</sup> Marley Vellasco, agradeço pela amizade e todo o apoio acadêmico cedido durante o mestrado. Sou extremamente grato pela orientação exemplar, que certamente ajudou para que o andamento da dissertação fosse o mais tranquilo possível.

Também agradeço ao meu coorientador, prof. José Franco, por todo o apoio realizado, desde a época de graduação na UERJ até os dias de hoje. Tenho bastante apreço por essa amizade, que também vai além do ambiente acadêmico.

Agradeço ao professores da PUC-Rio que tive contato durante o mestrado, por todo o conhecimento compartilhado e estímulo à pesquisa. Neste mesmo sentido, também agradeço à UERJ, instituição que tenho muito carinho e que foi fundamental na minha vida.

À Adriane França, companheira de longa data, meu profundo agradecimento pelo amor, carinho e suporte em todos os momentos desta jornada. Certamente o apoio cedido foi muito importante para que eu buscasse meus objetivos, e por isso sou extremamente grato.

Aos meus amigos que estiveram e estão presentes na minha vida, deixo registrado o meu agradecimento.

Por fim, gostaria de agradecer ao CNPq e à FAPERJ pelo apoio financeiro durante o curso de mestrado, que possibilitou a minha dedicação exclusiva ao Programa de Engenharia Elétrica da PUC-Rio.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

## Resumo

Medeiros Carvalho, Thiago; Vellasco, Marley Maria Bernardes Rebuzzi; Amaral, José Franco Machado do . **e-AutoMFIS: Modelo interpretável para previsão de séries multivariadas usando comitês de Sistemas de Inferência Fuzzy**. Rio de Janeiro, 2021. 132p. Dissertação de Mestrado – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

Por definição, a série temporal representa o comportamento de uma variável em função do tempo. Para o processo de previsão de séries, o modelo deve ser capaz de aprender a dinâmica temporal das variáveis para obter valores futuros. Contudo, prever séries temporais com exatidão é uma tarefa que vai além de escolher o modelo mais complexo, e portanto a etapa de análise é um processo fundamental para orientar o ajuste do modelo. Especificamente em problemas multivariados, o AutoMFIS é um modelo baseado na lógica *fuzzy*, desenvolvido para introduzir uma explicabilidade dos resultados através de regras semanticamente compreensíveis. Mesmo com características promissoras e positivas, este sistema possui limitações que tornam sua utilização impraticável em problemas com bases de dados com alta dimensionalidade. E com a presença cada vez maior de bases de dados mais volumosas, é necessário que a síntese automática de sistemas *fuzzy* seja adaptada para abranger essa nova classe de problemas de previsão. Por conta desta necessidade, a presente dissertação propõe a extensão do modelo AutoMFIS para a previsão de séries temporais com alta dimensionalidade, chamado de e-AutoMFIS. Apresenta-se uma nova metodologia, baseada em comitê de previsores, para o aprendizado distribuído de geração de regras *fuzzy*. Neste trabalho, são descritas as características importantes do modelo proposto, salientando as modificações realizadas para aprimorar tanto a previsão quanto a interpretabilidade do sistema. Além disso, também é avaliado o seu desempenho em problemas reais, comparando-se a acurácia dos resultados com as de outras técnicas descritas na literatura. Por fim, em cada problema selecionado também é considerado o aspecto da interpretabilidade, discutindo-se os critérios utilizados para a análise de explicabilidade.

## Palavras-chave

Previsão de séries multivariadas; Sistema de Inferência Fuzzy; Interpretabilidade; Comitê de previsores; Base de dados.

## Abstract

Medeiros Carvalho, Thiago; Vellasco, Marley Maria Bernardes Rebuzzi (Advisor); Amaral, José Franco Machado do (Co-Advisor). **e-AutoMFIS: Interpretable model for time series forecasting using ensemble learning of Fuzzy Inference System**. Rio de Janeiro, 2021. 132p. Dissertação de Mestrado – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

By definition, the time series represents the behavior of a variable as a time function. For the series forecasting process, the model must be able to learn the temporal dynamics of the variables in order to obtain consistent future values. However, an accurate time series prediction is a task that goes beyond choosing the most complex (or promising) model that is applicable to the type of problem, and therefore the analysis step is a fundamental procedure to guide the adaptation of a model. Specifically, in multivariate problems, AutoMFIS is a model based on fuzzy logic, developed not only to give accurate forecasts but also to introduce the explainability of results through semantically understandable rules. Even with such promising characteristics, this system has shown practical limitations in problems that involve datasets of high dimensionality. With the increasing demand for methods to deal with large datasets, it should be great that approaches for the automatic synthesis of fuzzy systems could be adapted to cover a new class of forecasting problems. This dissertation proposes an extension of the base model AutoMFIS modeling method for time series forecasting with high dimensionality data, named as e-AutoMFIS. Based on the Ensemble learning theory, this new methodology applies distributed learning to generate fuzzy rules. The main characteristics of the proposed model are described, highlighting the changes in order to improve both the accuracy and the interpretability of the system. The proposed model is also evaluated in different case studies, in which the results are compared in terms of accuracy against the results produced by other methods in the literature. In addition, in each selected problem, the aspect of interpretability is also assessed, which is essential for explainability evaluation.

## Keywords

Multivariate time-series forecasting; Fuzzy Inference System; Interpretability; Ensemble method; Big data.

# Sumário

<b>1</b>	<b>Introdução</b>	<b>13</b>
1.1	A vantagem de interpretabilidade em séries temporais	14
1.2	Objetivos	15
1.3	Contribuições	16
1.4	Organização	17
<b>2</b>	<b>Previsão de séries temporais</b>	<b>19</b>
2.1	Conceitos Fundamentais	20
2.1.1	Componentes de uma série temporal	20
2.1.2	Classificação de problemas de séries temporais	23
2.2	Técnicas de análise e processamento de dados em séries temporais	27
2.2.1	Método de Média Móvel	28
2.2.2	Remoção de tendências	28
2.2.3	Diferenciação de séries	29
2.2.4	Autocorrelação	31
2.2.5	Correlação linear	32
2.3	Revisão bibliográfica de modelos para previsão de séries temporais	32
2.3.1	Primeira fase - Modelos estatísticos	33
2.3.2	Segunda fase - Modelos de <i>Machine Learning</i>	34
2.3.3	Terceira fase - Deep Learning	36
2.3.4	Considerações finais sobre modelos de previsão	37
<b>3</b>	<b>Sistemas de Inferência Fuzzy em Previsão de Séries Temporais</b>	<b>39</b>
3.1	Componentes de um SIF	40
3.1.1	Fuzzificação	40
3.1.2	Base de regras	42
3.1.3	Inferência	42
3.1.4	Defuzzificação	43
3.2	Interpretabilidade	43
3.3	Previsão de séries temporais usando SIF	47
3.4	Modelos	48
3.4.1	Revisão de modelos preditores <i>fuzzy</i> usados na literatura	49
3.4.2	Modelo AutoMFIS	51
<b>4</b>	<b>Modelo proposto: e-AutoMFIS</b>	<b>55</b>
4.1	Motivação para a extensão do modelo AutoMFIS	56
4.1.1	Necessidade e vantagem de um modelo <i>ensemble</i>	56
4.1.2	Usabilidade em grandes bases de dados	57
4.2	Arquitetura do modelo e-AutoMFIS	58
4.2.1	Definição e organização do problema	59
4.2.2	Formação da base de regras	66
4.2.2.1	Formulação	66
4.2.2.2	Associação	68

4.2.2.3	Considerações sobre o módulo de formação de regras	70
4.2.3	Filtragem e agregação de regras	72
4.2.3.1	Filtragem Interna	72
4.2.3.2	Filtragem Externa	75
4.2.4	Defuzzificação	77
4.2.5	Previsão <i>multi-step</i>	78
4.3	Resumo do modelo e-AutoMFIS	79
<b>5</b>	<b>Estudos de casos</b>	<b>82</b>
5.1	Competição M3	83
5.2	Air Quality	89
5.3	Traffic-Rate	97
5.4	Taxa de câmbio	103
5.5	CIS-Challenge	108
5.6	Considerações sobre os estudos de casos	115
<b>6</b>	<b>Conclusão e trabalhos futuros</b>	<b>117</b>
	<b>Referências bibliográficas</b>	<b>120</b>



## Lista de figuras

Figura 2.1	Exemplo de decomposição de séries temporais. (a) Série temporal sintética. (b) Componente de tendência linear da série. (c) Componente de sazonalidade. (d) Componente cíclica. (e) Flutuações irregulares.	21
Figura 2.2	Inspeção visual para séries não-estacionárias (sup.) e possivelmente estacionária (inf.)	27
Figura 2.3	Exemplo de remoção de tendências em séries temporais	29
Figura 2.4	Preço da ação Google (sup.), e preços usando diferenciação de valores consecutivos da série (inf.)	30
Figura 2.5	Exemplo gráfico de autocorrelação	31
Figura 3.1	Diagrama de blocos de um Sistema de Inferência <i>fuzzy</i>	40
Figura 3.2	Exemplo de Fuzzificação	41
Figura 3.3	Exemplo de conjuntos <i>fuzzy</i> interpretáveis e não interpretáveis	45
Figura 3.4	Exemplo de configuração de um SIF para previsão de séries	48
Figura 3.5	Exemplo de previsão multivariada usando SIF	49
Figura 3.6	Arquitetura modular do AutoMFIS	52
Figura 3.7	Exemplo de geração semi-exaustiva	53
Figura 4.1	Arquitetura simplificada do modelo e-AutoMFIS	58
Figura 4.2	Exemplo de divisão de dados entre treinamento, validação e teste em uma aplicação de previsão do volume de tráfego	60
Figura 4.3	Conjuntos <i>fuzzy</i> com baixa interpretabilidade no domínio de X	61
Figura 4.4	Representação da partição <i>fuzzy</i> forte	62
Figura 4.5	Conjuntos <i>fuzzy</i> gerados via particionamento uniforme	63
Figura 4.6	Conjuntos <i>fuzzy</i> gerados pelo método baseado em percentis	64
Figura 4.7	Comparação de geração dos conjuntos <i>fuzzy</i> usando as metodologias disponíveis para o e-AutoMFIS	65
Figura 4.8	Exemplo de subamostragem proposta	70
Figura 4.9	Diagrama de blocos da etapa de filtragem	73
Figura 4.10	Efeito da ponderação na defuzzificação do e-AutoMFIS	77
Figura 4.11	Comparação de métodos de defuzzificação	78
Figura 4.12	Arquitetura completa do modelo e-AutoMFIS	80
Figura 5.1	Registro histórico de 5 séries financeiras da base de dados Competição M3	84
Figura 5.2	Matriz de correlação para o estudo de caso Competição M3	85
Figura 5.3	Distribuição da quantidade de antecedentes para a base de regras do estudo de caso Competição M3	88
Figura 5.4	Exemplo de fatores de influência na qualidade do ar	89

Figura 5.5	Taxa de valores faltantes na base de dados Qualidade do Ar	92
Figura 5.6	Matriz de correlação de Pearson para a base de dados Qualidade do Ar	92
Figura 5.7	Amostra das séries do estudo de caso 'Qualidade do ar'	94
Figura 5.8	Função de autocorrelação para as variáveis selecionadas na base de dados Qualidade do Ar	95
Figura 5.9	Exemplo de padrões temporais no estudo de caso	99
Figura 5.10	Exemplo de previsão para o estudo de caso Taxa de ocupação de ruas	102
Figura 5.11	Comportamento das séries do estudo de caso - Taxa de Câmbio	104
Figura 5.12	Matriz de correlação de Pearson - Taxa de Câmbio	105
Figura 5.13	Diferenciação das séries do estudo de caso Taxa de Câmbio	106
Figura 5.14	Valores faltantes na base de dados CIS-Challenge	110
Figura 5.15	Grupos gerados na base de dados CIS-Challenge	112
Figura 5.16	Resultado de uma série no estudo de caso CIS-Challenge	114

## Lista de tabelas

Tabela 3.1	Métricas de interpretabilidade	44
Tabela 4.1	Lista de parâmetros do modelo e-AutoMFIS	79
Tabela 4.2	Lista de métodos do modelo e-AutoMFIS	81
Tabela 5.1	Informações sobre as bases de dados dos estudos de casos	82
Tabela 5.2	Parâmetros testados no estudo de caso Competição M3	86
Tabela 5.3	Resultado do Estudo de Caso 1	86
Tabela 5.4	Critérios de interpretabilidade do estudo de caso Competição M3	87
Tabela 5.5	Variáveis da base de dados Qualidade do ar	90
Tabela 5.6	Importância das variáveis usando o método RFE	93
Tabela 5.7	Parâmetros testados no estudo de caso Qualidade do Ar	95
Tabela 5.8	Resultado do estudo de caso Qualidade de Ar	95
Tabela 5.9	Critérios de interpretabilidade do estudo de caso Qualidade do Ar	96
Tabela 5.10	Parâmetros testados no estudo de caso Taxa de ocupação de ruas	100
Tabela 5.11	Abordagens testadas no e-AutoMFIS para a base de dados Taxa de ocupação de ruas	100
Tabela 5.12	Métodos usados para a base de dados Taxa de ocupação de ruas	101
Tabela 5.13	Resultado obtido para a base de dados Taxa de ocupação de ruas	101
Tabela 5.14	Critérios de interpretabilidade do estudo de caso Taxa de ocupação de ruas	103
Tabela 5.15	Parâmetros testados no estudo de caso Taxa de Câmbio	107
Tabela 5.16	Métodos usados na base de dados Taxa de Câmbio	107
Tabela 5.17	Critérios de interpretabilidade do estudo de caso Taxa de Câmbio	108
Tabela 5.18	Parâmetros testados no estudo de caso Competição CIS	114
Tabela 5.19	Modelos usados para a base de dados Competição CIS	115

## Lista de Abreviaturas

SIF – Sistema de Inferência Fuzzy

ML – *Machine Learning*

DL – *Deep Learning*

AR – Autoregressivo

LSTM – *Long Short-Term Memory*

MQR – Mínimos Quadrados Restrito

AutoMFIS – *Automatic Multivariate Fuzzy Inference System*

e-AutoMFIS – *Ensemble of Automatic Multivariate Fuzzy Inference System*

# 1

## Introdução

Durante o curso da humanidade, diferentes materiais foram considerados bens com grande valor agregado, como especiarias, ouro, diamante e outros minérios. Contudo, no período atual em que vivemos, um dos recursos mais desejados não é um material físico: trata-se da informação. O aprimoramento de tecnologias de medição de valores e a permissividade de usuários na Internet em fornecer informações, mesmo que anonimamente, são exemplos que contribuem na formação de bases de dados mais volumosas.

A facilidade de obtenção e armazenamento de dados tem um grande impacto em todos os setores da sociedade. Na área comercial, por exemplo, as informações coletadas auxiliam a classificar o perfil de consumidor e recomendar itens mais prováveis de serem comprados. Além disso, outra grande vantagem decorrente do aumento do volume de dados é a análise temporal de variáveis de interesse, como a venda diária de produtos ou a quantidade de acessos semanais a um site de notícias. Todas estas informações temporais ajudam a fundamentar estratégias com base em projeções futuras, permitindo um melhor ajuste em função da dinâmica observada.

A realização de projeções futuras não é uma novidade contemporânea. Elas existiam mesmo sem todo o aparato existente nos dias atuais. Talvez uma das análises mais conhecidas que pode ilustrar esta atividade foi concebida por Malthus, no século XVIII. Segundo a teoria malthusiana, o crescimento populacional poderia ser descrito por uma função exponencial, enquanto a produção de alimentos acompanharia um crescimento linear em função do tempo. Consequentemente, essas projeções indicavam que em algum momento no futuro a população poderia sofrer por escassez de alimentos e, portanto, novas políticas deveriam ser criadas para antecipar este problema, como o desenvolvimento de tecnologias para tornar a agricultura mais eficiente. Mesmo com fortes críticas sobre seu estudo [1] e suas implicações, este exemplo serve para mostrar a importância da projeção de séries temporais, que vai muito além de prever valores futuros. Deste exemplo também é possível perceber a dificuldade de conceber um modelo matemático para problemas reais, visto que existem comportamentos imprevisíveis que são capazes de modificar a dinâmica das séries.

Por conta da sua grande relevância e grande aplicabilidade em situações reais, o estudo de séries temporais é amplamente divulgado na área científica, tendo conceitos e metodologias já consolidados na literatura. Um dos tópicos de maior interesse desta área é a previsão, cujo objetivo é generalizar de forma acurada o comportamento futuro em função de dados históricos. Com a ausência de um modelo único que represente com fidelidade todas as dinâmicas temporais, a área de previsão de séries é bastante diversa no sentido de modelagem da solução, onde tanto abordagens estatísticas quanto de Inteligência Computacional podem produzir bons resultados, dependendo do problema. Nesta dissertação, o desenvolvimento do modelo proposto tem como base a lógica *fuzzy*. Conhecido na literatura por introduzir o aspecto da interpretabilidade através de termos linguísticos e regras semânticas, um Sistema de Inferência *Fuzzy* (SIF) tem o objetivo de tornar a modelagem matemática mais compreensível para seres humanos.

## 1.1

### A vantagem de interpretabilidade em séries temporais

Não é simples desenvolver uma metodologia adequada para um problema específico de previsão de séries temporais. Normalmente, a acurácia é a característica mais importante para determinar o modelo mais adequado, mas não é o único critério possível. Entretanto, tão importante quanto prever com exatidão os valores futuros é entender o porquê do resultado gerado, compreendendo os fatores que contribuíram para a previsão. Neste sentido, a interpretabilidade tem o objetivo de quantificar aspectos que tornam a leitura e análise dos resultados mais simples para o entendimento humano.

A interpretabilidade é uma característica que não é inerente a qualquer modelo matemático da literatura, e portanto este critério não é sempre utilizado para avaliação dos resultados. As abordagens inspiradas em Redes Neurais, por exemplo, não possuem uma nítida explicação dos resultados e a depuração dos resultados torna-se mais complexa do que nos modelos baseados na lógica *fuzzy*. Portanto, mais do que um critério de avaliação, a interpretabilidade também é uma característica interessante a ser considerada para a escolha do modelo.

A aplicação de modelos que forneçam explicações para o resultado obtido resulta em consequências práticas em diferentes âmbitos. Primeiro, a interpretabilidade contribui para o propósito principal dos algoritmos desenvolvidos, que é servir como um método de apoio a decisão. Além disso, o fator de interpretabilidade serve de auxílio para tornar o resultado mais auditável, facilitando o ajuste do modelo em conjunto com o conhecimento técnico sobre o

problema. Por fim, esta característica também facilita a extração de conhecimento do problema, contribuindo não apenas para a previsão, mas também para a análise de séries temporais.

Para a síntese automática de modelos que utilizem a lógica *fuzzy* para previsão de séries temporais, tanto o sistema *fuzzy* evolucionário quanto o SIF evolutivo são opções consolidadas na literatura. Ambas definem as estratégias que serão utilizadas para o ajuste de parâmetros e criação da base de regras. Enquanto a primeira alternativa utiliza uma técnica de otimização baseada em algoritmos evolucionários, aplicável em problemas cujo espaço de busca é relativamente alto (como é para a configuração de um sistema *fuzzy*), a abordagem evolutiva ajusta o modelo de acordo com o fluxo de dados, tornando-se aplicável inclusive em previsão de séries não-estacionárias. Visando a agregar as vantagens destas duas técnicas, em [2] é proposta uma nova arquitetura para o desenvolvimento automático de sistemas *fuzzy* em previsão de séries temporais, chamado AutoMFIS.

O modelo AutoMFIS faz uso de técnicas de estatística e heurísticas para o ajuste dos parâmetros do sistema *fuzzy*, além do desenvolvimento de regras usando a estratégia de geração semi-exaustiva. Para uma base de dados com dimensionalidade razoável, este modelo possui um bom *trade-off* entre a exatidão da previsão e a interpretabilidade, como nos problemas mostrados em [2]. Contudo, os estudos conduzidos sobre o AutoMFIS mostraram uma grande limitação em base de dados com uma quantidade alta de variáveis ou registros, relacionado ao fenômeno da maldição da dimensionalidade. Este fato, inclusive, torna o modelo computacionalmente inviável para este tipo de problema.

Neste trabalho, é proposta uma nova arquitetura para previsão multivariada, denominada e-AutoMFIS. Ela estende o modelo básico para tarefas que envolvem grandes bases de dados. Para isso, são aplicados os conceitos de subamostragem e *Ensemble Learning* na geração do sistema *fuzzy*, de modo a tornar o algoritmo mais adequado para problemas com uma extensa base de dados e capaz de extrair com maior facilidade o conhecimento. Conforme será demonstrado no decorrer deste trabalho, estas novas modificações proporcionam ao modelo uma maior facilidade no aprendizado e melhor capacidade de exploração de parâmetros, além de fornecer a capacidade de extração de regras *fuzzy* em diferentes intervalos de tempo.

## 1.2 Objetivos

Este trabalho tem como objetivo principal desenvolver uma abordagem interpretável para previsão de séries multivariadas que seja aplicável a bases

de dados com alta dimensionalidade. Para isso, é proposto um modelo *fuzzy* baseado no sistema AutoMFIS utilizando o conceito de comitê de previsores para geração da base de regras. Também é proposto neste trabalho avaliar o impacto desta nova metodologia em termos de acurácia e interpretabilidade em diferentes estudos de caso.

Objetivos secundários também são apresentados neste trabalho para auxiliar o estudo teórico sobre as características do modelo proposto. Deste modo, esta dissertação também propõe estudar as limitações desta abordagem e os parâmetros que influenciam o processo de aprendizagem, a acurácia e a interpretabilidade do modelo resultante.

### 1.3

#### Contribuições

Este trabalho possui três vertentes principais como contribuições, a saber:

- **Teórica:** por se tratar de uma extensão de um modelo já existente, este trabalho apresenta uma contribuição ao investigar as limitações do modelo AutoMFIS e propor uma solução adequada para tornar esta técnica viável para lidar com grandes bases de dados. Além disso, este trabalho discorre sobre a utilidade de uma abordagem *Ensemble* na extração de conhecimento em forma de regras *fuzzy*.
- **Prática:** o modelo e-AutoMFIS possui uma forte contribuição prática, principalmente sob a perspectiva computacional. A abordagem proposta oferece um menor custo computacional necessário para a extração de conhecimento em forma de regras, tornando o seu uso viável. Além disso, este modelo potencialmente oferece um menor tempo de treinamento, por conta do paralelismo possível durante o ajuste dos comitês de previsores.
- **Literária:** com a aplicação do modelo proposto em estudos de casos conhecidos, este trabalho tem como objetivo registrar os resultados de um sistema *fuzzy* nos problemas selecionados, tanto em termos de acurácia quanto interpretabilidade. Especificamente, quanto a esta, a falta de resultados publicados na literatura dificulta a análise de um aspecto importante inerente ao modelo proposto. Portanto, esta dissertação contribui fornecendo resultados das principais métricas de interpretabilidade de um sistema *fuzzy*, de modo que futuros trabalhos nesta área tenham mais um recurso para uma análise comparativa.



## 1.4

### Organização

Esta dissertação está organizada em duas partes principais. A primeira parte descreve a teoria fundamental e básica para o trabalho proposto sobre sistemas *fuzzy* e séries temporais. Já na segunda parte, o foco principal é o detalhamento do modelo proposto e suas aplicações. Todos estes assuntos estão organizados da seguinte forma:

- **Capítulo 2** - Conceitos fundamentais de séries temporais. Neste capítulo são descritos os fundamentos de séries temporais, úteis para o desenvolvimento teórico e prático deste trabalho. São apresentadas técnicas para análise de séries temporais e processamento de dados, com o objetivo de embasar as estratégias utilizadas no estudo de caso. Por fim, são citados os modelos relevantes na literatura para previsão de séries temporais, citando motivos de sua popularidade e ampla aplicação prática.
- **Capítulo 3** - Apresentação de sistemas *fuzzy* para previsão de séries temporais. Com o objetivo de embasar a descrição do modelo proposto, este capítulo introduz conceitos básicos importantes para sistemas de inferência *fuzzy* e sua aplicação em modelagem e previsão de séries temporais. Em seguida, é discutido o aspecto da interpretabilidade, descrevendo alguns critérios que serão úteis para a análise de resultados. Por fim, neste capítulo também são apresentados alguns modelos baseados em lógica *fuzzy* para previsão de séries temporais, destacando o funcionamento do AutoMFIS devido à sua correlação direta com o modelo proposto nesta dissertação.
- **Capítulo 4** - Apresentação do e-AutoMFIS, modelo proposto nesta dissertação. Neste capítulo é apresentada a sua arquitetura, mostrando-se detalhadamente o seu funcionamento através de uma explicação modular. Os critérios de interpretabilidade também são analisados ao decorrer da descrição do modelo.
- **Capítulo 5** - Estudos de caso propostos para avaliação do e-AutoMFIS. Neste capítulo são apresentados problemas selecionados da literatura para previsão de séries temporais multivariadas, comparando o desempenho do modelo proposto. Em cada estudo de caso, também é avaliado o aspecto da interpretabilidade do sistema, analisando o seu potencial para diferentes aplicações.
- **Capítulo 6** - Conclusão da dissertação, ponderando as vantagens e desvantagens da nova abordagem *fuzzy*. Neste capítulo também são

discutidos alguns tópicos abertos que podem servir de auxílio para investigações futuras. Por fim, são sugeridas perspectivas que podem ser úteis para tornar o modelo proposto mais robusto em variações futuras.

Segundo o dicionário, o termo previsão é definido como a antecipação de algo que ainda não aconteceu. Embora este termo seja popularmente vinculado a achismos, a previsão é uma prática de extrema importância para a ciência, com metodologias desenvolvidas para que a inferência seja sempre mais acurada. Em problemas do cotidiano, a tarefa de previsão é comumente relacionada a uma componente temporal, como antever a temperatura de um ambiente daqui a uma hora. No caso de séries temporais, a previsão é a determinação de valores futuros de variáveis em função do tempo, geralmente baseada em dados históricos.

A previsão de séries temporais é uma das áreas mais relevantes na área de ciência de dados, com aplicações importantes na economia, indústria e logística. Para isso, boas previsões são necessárias para que as pessoas tenham confiança nos resultados e auxiliem no planejamento de futuras ações. Um exemplo bastante útil pode ser encontrado na geração de energia eólica, com o objetivo de otimização do consumo energético de parques eólicos. O desenvolvimento de um modelo confiável que realize a previsão de velocidade do vento pode contribuir para o gerenciamento de acionamento dos geradores de turbinas eólicas e, conseqüentemente, prover maior eficiência energética [3, 4].

A modelagem matemática para o desenvolvimento de um sistema de previsão é bastante diversificada, contemplando desde métodos mais tradicionais, como métodos estatísticos, até técnicas mais avançadas de Inteligência Computacional. Todas estas abordagens são relevantes nos dias atuais e presentes em recentes trabalhos acadêmicos, visto que não existe uma unanimidade quanto à técnica mais precisa para todo tipo de problema [5]. Independentemente do método utilizado, a análise de séries temporais é fundamental para entender o comportamento da variável em questão. Portanto, para que o processo de previsão de séries temporais seja realizado de uma forma mais adequada, é necessário analisar e extrair informações importantes de dados históricos, tratá-los e ajustar o modelo de previsão.

Este capítulo apresenta uma perspectiva teórica de séries temporais, com o objetivo de descrever conceitos importantes para o desenvolvimento do trabalho e técnicas consideradas importantes na parte experimental. Ao

final deste capítulo, é apresentada uma revisão da literatura de modelos de previsão de séries temporais.

## 2.1

### Conceitos Fundamentais

Uma série temporal é definida como um conjunto de valores registrados de uma variável com o passar do tempo, podendo (ou não) ter um significado semântico. Estes dados podem ser representados continuamente em função do tempo, como uma função  $f(t)$ , ou coletados em intervalos discretos de tempo. Exemplos concretos são o Produto Interno Bruto de um país, levantamento diário de infectados por uma doença ou a umidade em um vaso de planta conforme mensurada por sensores.

Compreender o comportamento de uma série temporal e identificar as características do problema de previsão são fundamentais para o desenvolvimento de uma abordagem que produza bons resultados. Com o objetivo de descrever os principais aspectos teóricos sobre esta área, esta seção apresenta o funcionamento básico da dinâmica de séries temporais e a categorização do problema de previsão.

#### 2.1.1

##### Componentes de uma série temporal

O estudo de modelagem de séries temporais é importante para entender sua dinâmica e planejar a metodologia desejada para realizar a previsão. Um dos conceitos mais importantes é o de decomposição de séries. Tradicionalmente, uma variável temporal é afetada por quatro componentes principais: tendência, flutuações cíclicas, variações sazonais e variações irregulares [6]. Neste contexto, a Figura 2.1 exemplifica a decomposição de séries temporais e ilustra os elementos.

A tendência (Figura 2.1-b) é uma observação, em longo período, de mudança de uma série, demonstrando o crescimento, decrescimento ou estagnação ao decorrer do tempo. Em termos gerais, esta componente capta o movimento gradual de uma série, podendo ser útil para a investigação do comportamento de uma série a longo prazo e até extrapolação de tendências para projeções futuras. Contudo, a percepção de tendência é quase sempre dependente do intervalo de tempo das observações [7]. Exemplos de tendências podem ser encontrados com facilidade, como a população mundial e a taxa de mortalidade infantil, que possuem inclinações positivas e negativas, respectivamente. A tendência pode ser simples, como uma tendência linear, ou complexa, como uma tendência não-linear. Por conta disso, a modelagem matemática da compo-

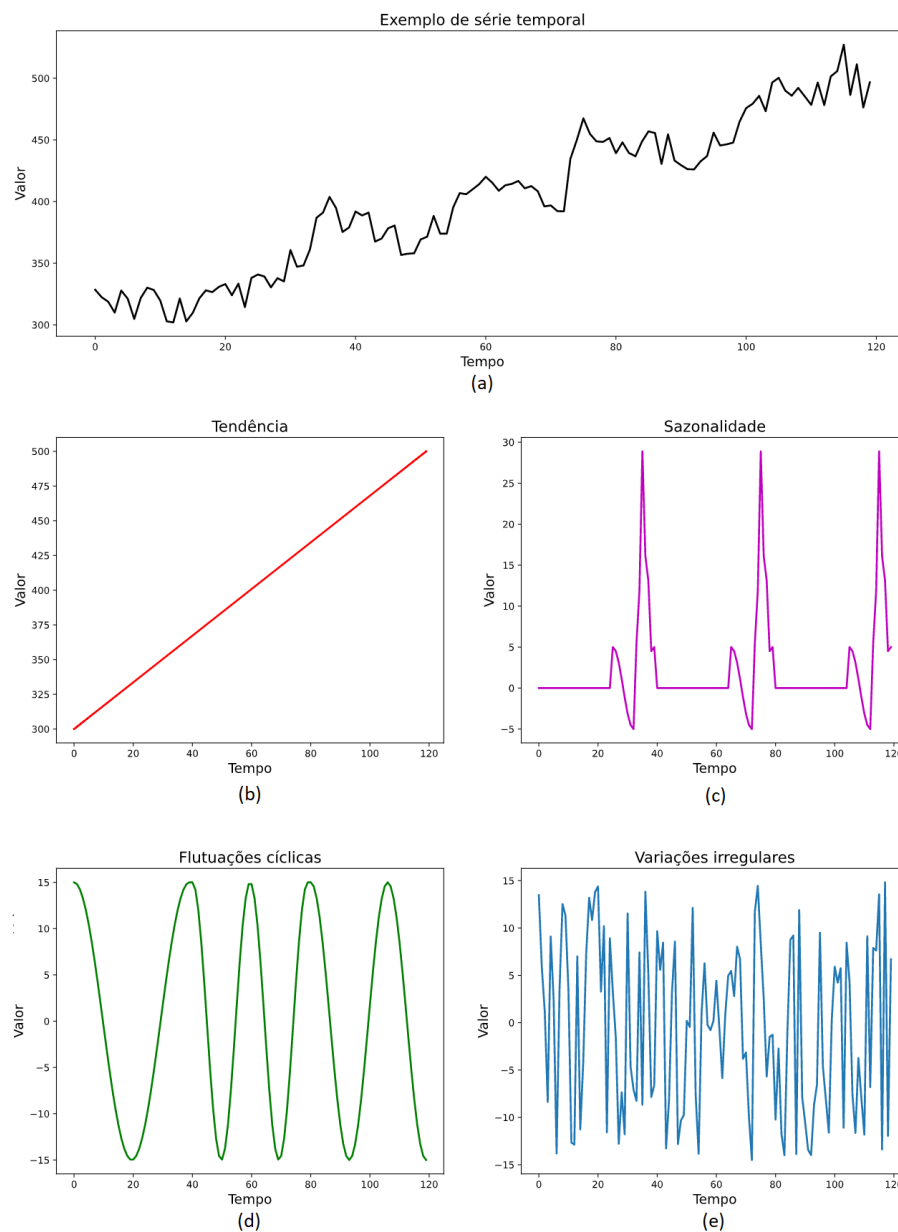


Figura 2.1: Exemplo de decomposição de séries temporais. (a) Série temporal sintética. (b) Componente de tendência linear da série. (c) Componente de sazonalidade. (d) Componente cíclica. (e) Flutuações irregulares.

nente de tendência não é única e depende da análise utilizada para o ajuste de curva. Conforme será discutido futuramente, técnicas de identificação e remoção de tendências são importantes para tornar possível a utilização de diversos modelos de previsão [8, 9].

Além do comportamento de tendência, as variações periódicas também são componentes importantes que estão presentes em séries temporais, decorrentes de eventos regulares presentes na natureza e no comportamento humano, por exemplo. Estes efeitos periódicos são modelados por efeitos sazonais e flutuações cíclicas, cuja principal diferença é a regularidade dos eventos no período observado.

A sazonalidade (Figura 2.1-c) é frequentemente conhecida pela presença de eventos existentes em específicas estações do ano, como o período de semeadura e colheita de uma plantação. Contudo, por definição, a sazonalidade é uma característica que reflete a presença de repetições de certos padrões em períodos regulares, possivelmente em um intervalo anual. Um exemplo bem comum de variações sazonais é o de sensação térmica, onde períodos mais quentes são geralmente registrados no verão e temperaturas mais baixas, no inverno.

O estudo da sazonalidade é de grande interesse para a área de séries temporais, pois contribui para a análise de padrões temporais e possibilita a descoberta de conhecimento de causalidade de eventos periódicos, como aqueles oriundos de festividades (e.g. Natal, Páscoa e carnaval). Especificamente na economia, a sazonalidade tem um grande impacto nas estratégias de negócios, onde datas comemorativas são conhecidas por alavancar as vendas no comércio.

As variações cíclicas (Figura 2.1-d) correspondem a flutuações periódicas em uma série temporal que não sejam consideradas sazonais. Estas variações podem ter um longo período, como em problemas envolvendo negócios [10], ou em um curto período, como as variações diárias existentes por conta do comportamento social e biológico da população.

Por fim, as flutuações irregulares (Figura 2.1-e) indicam qualquer variação que não pode ser modelada por componentes periódicas ou de tendência. Estas inconstâncias podem acontecer por influências imprevisíveis ou aleatórias, que podem tornar mais difícil obter um desempenho satisfatório. Contudo, em alguns casos é possível extrair correlações em curto-prazo para melhorar a capacidade preditiva de séries temporais [7]. Portanto, diferentemente das outras componentes que possuem técnicas estatísticas bem definidas para análise, essas variações aleatórias são importantes de serem captadas pelo modelo de previsão para que o resultado seja o mais exato possível.

Os elementos que compõem uma série temporal são relevantes para a

compreensão do comportamento e podem servir como indicadores de padrões existentes. A relação entre as componentes define a modelagem matemática mais adequada para cada ocasião. O modelo aditivo, por exemplo, pressupõe que todas as componentes são independentes, como mostra a Eq. 2-1, onde a tendência  $T(t)$ , sazonalidade  $S(t)$ , a componente cíclica  $C(t)$  e as flutuações irregulares  $I(t)$  podem ser desacopladas facilmente [11]. Em contrapartida, o modelo multiplicativo representa uma maior interdependência entre as componentes. Por exemplo, a magnitude do padrão de sazonalidade pode aumentar com a tendência positiva, ou vice-versa.

$$Y(t) = T(t) + S(t) + C(t) + I(t) \quad (2-1)$$

Conforme será discutido neste capítulo e oportunamente no Capítulo 5, estas definições são importantes para compreender o comportamento do problema e ajustar a série para o uso correto do modelo de previsão, utilizando-se para isso técnicas de remoção de componentes indesejados, como a tendência.

### 2.1.2

#### Classificação de problemas de séries temporais

O processo de previsão de séries temporais envolve o desenvolvimento de modelos que usem valores do passado e do presente para inferir observações futuras. Os problemas de previsão podem ser classificados de acordo com diferentes aspectos:

- **Quanto ao número de variáveis:** univariado, multivariado e julgamento.
- **Quanto ao período de previsão:** longo-prazo, médio-prazo, curto-prazo e curtíssimo-prazo.
- **Quanto às propriedades estatísticas:** estacionárias e não-estacionárias.

Além de auxiliar na classificação do problema, o conhecimento destas características serve de auxílio para a escolha apropriada do modelo a ser utilizado. Em uma tarefa de previsão de séries multivariadas, por exemplo, o uso de um modelo univariado não é o mais adequado.

O termo mais importante que define o modelo proposto é o relativo ao número de variáveis utilizadas, podendo este ser classificado como univariado, multivariado ou julgamento [7]. Esta classificação define como o problema pode ser definido e as relações que são extraídas do modelo matemático. O método

de previsão univariado define a previsão de uma série unicamente em função de valores do presente e do passado desta mesma variável. A Eq. 2-2 mostra, em termos genéricos, a função de previsão de um caso univariado.

$$\hat{y}_{t+h} = f(y_t, y_{t-1}, y_{t-2}, \dots) \quad (2-2)$$

O método de previsão univariado é muito útil quando é esperado que apenas a própria variável seja suficiente para prever seus valores futuros. Ainda assim, em casos onde não são conhecidas ou disponíveis variáveis exógenas que auxiliem na previsão, o processo univariado torna-se uma boa alternativa a ser utilizada.

Em problemas do cotidiano, é comum que a previsão possa depender de informações adicionais que vão além dos valores da própria variável. No método multivariado, a previsão de uma série depende de variáveis adicionais. Estes termos adicionais, também conhecidos por variáveis explicativas ou exógenas, têm como objetivo contribuir para a obtenção de um melhor modelo de previsão. O consumo de energia elétrica residencial, por exemplo, geralmente é afetado por características climáticas, principalmente a temperatura [12]. Matematicamente, a Eq. 2-3 apresenta, em termos genéricos, a função de previsão de um caso multivariado.

$$\hat{y}_{i,t+h} = f(y_{1,t}, y_{2,t}, \dots, y_{S,t}, y_{1,t-1}, \dots) \quad (2-3)$$

onde  $y_{s,t}$  representa a  $s$ -ésima série no tempo  $t$  e  $\hat{y}_{i,t+h}$  representa o valor estimado para a  $i$ -ésima série  $h$  passos a frente. O processo multivariado tenta combinar variáveis endógenas e exógenas para extrair as relações de interdependência entre elas, com o objetivo de que variáveis explicativas sejam úteis para uma previsão mais acurada. A adição de novas variáveis deve ser feita com cautela, de modo que estas contribuam positivamente para a previsão. Caso contrário, a inserção de variáveis exógenas não só aumenta desnecessariamente a dimensionalidade do modelo, como também pode contribuir para a inserção de erro ao sistema por diversos motivos. Em um problema de previsão *multi-step*, por exemplo, as variáveis exógenas geralmente precisam ser estimadas e suas incertezas podem prejudicar a previsão da série. Em suma, a utilização do método multivariado não garante um melhor desempenho do que os de abordagens univariadas [13, 14].

Uma outra abordagem de previsão é conhecida por julgamento. Este método é baseado em informações importantes e geralmente subjetivas, como



a noção pessoal de mercado financeiro, intuição ou conhecimento relevante [7]. Embora esta abordagem fosse estigmatizada para alguns autores [15, 16], hoje em dia é utilizada principalmente em conjunto com outros métodos (univariados ou multivariados) para ajustar o modelo e aumentar a sua acurácia [17, 18]. Ainda assim, este método não será explorado futuramente, pois o trabalho se concentra na previsão multivariada de séries temporais.

Além da classificação quanto à quantidade de variáveis usadas, o problema também pode ser definido quanto ao intervalo de tempo de previsão. Não existe um critério quantitativo para este tipo de classificação e, portanto, em alguns problemas pode ser difícil a sua caracterização. Contudo, essa definição não só favorece a compreensão do problema, como também serve de auxílio para a escolha adequada do modelo.

A previsão a longo prazo é definida como uma projeção a um longo período em relação ao intervalo de tempo considerado. Um exemplo clássico é a previsão de indicadores de um país, onde geralmente são feitos estudos de previsão para compreender melhor o cenário econômico. Por mais que este tipo de previsão acabe não sendo extremamente acurado, devido a diversas incertezas difíceis de serem modeladas, este tipo de problema é de grande utilidade em captar o comportamento geral da série [19]. Neste tipo de previsão, geralmente são utilizados modelos físicos (como para previsão do clima [20]) ou modelos estatísticos.

Do lado oposto ao citado anteriormente, a previsão a curto-prazo é caracterizada por um horizonte de previsão curto, geralmente no intervalo de dias ou meses. Neste tipo de problema, é frequente o uso de diferentes metodologias, como modelos de *Machine Learning* (ML), estatísticos ou modelagem analítica. Além disso, existem problemas de inferência de prazo ainda mais curto, chamados de curtíssimo prazo. A maior característica deste tipo de previsão é o horizonte bastante reduzido, geralmente com previsão na escala de horas. Problemas de previsão de curtíssimo prazo são corriqueiros, como previsão de velocidade do vento [21] e geração de energia [22]. Os modelos mais utilizados são baseados em ML ou metodologia estatística, visto que são abordagens que captam com maior facilidade as rápidas variações de séries temporais [23].

O problema também pode ser classificado em função das propriedades estatísticas das séries. Antes de uma definição formal de estacionariedade, vale ressaltar que este trabalho não tem a pretensão de se aprofundar no desenvolvimento matemático das propriedades estatísticas necessárias para esta classificação, pois acredita-se que o aprofundamento deste tópico está fora do seu escopo. Entretanto, é interessante que estas propriedades sejam definidas para que a explicação torne-se mais coerente. Assim, para uma de

série temporal  $x$  com  $T$  observações, é possível obter informações estatísticas, como o valor esperado  $E[X]$  (Eq. 2-4), variância (Eq. 2-5) e autocorrelação. Como em séries temporais a probabilidade  $p(x)$  é uniforme, o valor esperado da série  $x$  é equivalente a média  $\mu_x$ . Além disso, para duas variáveis  $x$  e  $y$  é possível calcular a covariância, dada pela Eq. 2-6.

$$E[x] = \sum_{t=1}^T x(t)p(x) \quad (2-4)$$

$$Var(x) = E[x - \mu_x] \quad (2-5)$$

$$Cov(x, y) = E[(x - \mu_x)(y - \mu_y)] \quad (2-6)$$

Com essas informações, é possível estabelecer o que é uma série estacionária. Uma série temporal é dita estacionária se as suas propriedades estatísticas permanecem constantes com o decorrer do tempo. Como geralmente esta definição é muito restritiva para séries temporais, este conceito de estacionariedade é atribuída a séries fortemente estacionárias, expandindo a classificação quanto às características estatísticas [24]. Por exemplo, uma série é definida como fracamente estacionária se a média e a autocovariância não mudam em função do tempo e a sua variância é finita [25]. Na literatura, é possível também encontrar definições para séries estacionárias de primeira ordem e de  $n$ -ésima ordem [26].

Intuitivamente, é possível interpretar uma série estacionária como uma variação próximo de um valor médio ao longo do tempo. Isto significa que componentes como tendência e sazonalidade são características que tornam uma série não-estacionária, visto que a média e covariância mudam em função do tempo. Portanto, a inspeção visual do gráfico de uma série temporal pode servir de auxílio para uma primeira avaliação qualitativa de estacionariedade, como exemplifica a Figura 2.2. Ainda assim, para verificar esta propriedade, é recomendado o uso de testes estatísticos.

Esta classificação é fundamental para análise de séries temporais e a escolha apropriada do modelo de previsão, visto que algumas abordagens pressupõem que as variáveis sejam estacionárias. Contudo, em problemas reais é natural a ocorrência de comportamentos como tendências e sazonalidades, que tornam o problema não-estacionário. Para contornar esta questão, podem

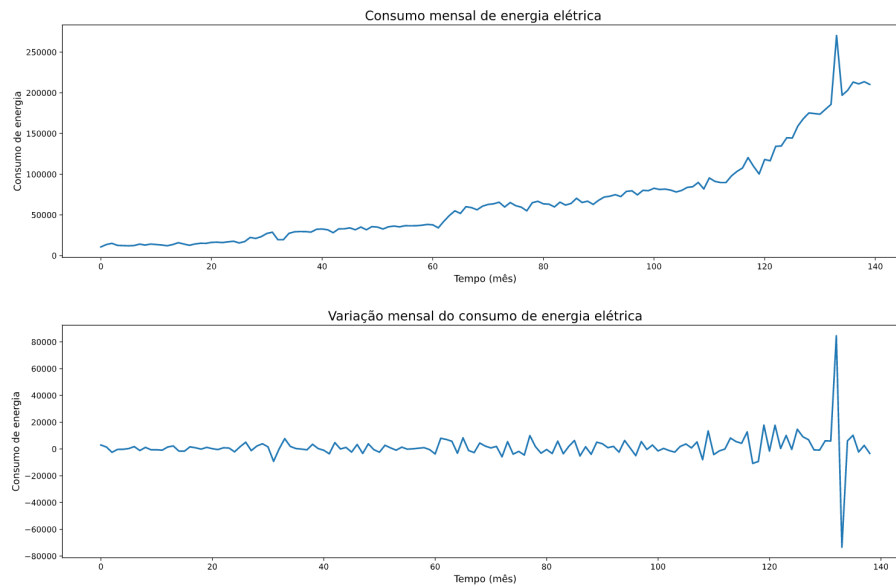


Figura 2.2: Inspeção visual para séries não-estacionárias (sup.) e possivelmente estacionária (inf.)

ser aplicadas transformações às séries temporais para que passem a apresentar um comportamento estacionário, como remoção de tendências e diferenciação de séries, que serão apresentadas neste capítulo.

## 2.2

### Técnicas de análise e processamento de dados em séries temporais

Em problemas relacionados a séries temporais, é comum que os dados não estejam originalmente preparados para serem utilizados diretamente em modelos de previsão, sendo necessária uma etapa intermediária de processamento. Diferentes obstáculos podem ser encontrados em base de dados, como a presença de pontos fora da curva e valores faltantes, ocasionados por problemas de naturezas distintas, seja por falha de sensores, erro humano ou eventos atípicos. Portanto, a análise e processamento de dados é uma etapa essencial para qualquer metodologia de previsão.

As abordagens para avaliação e processamento de dados servem não apenas para compreender melhor a dinâmica temporal das variáveis, como também para modificar as séries para usar adequadamente o modelo proposto. Estas técnicas servem, por exemplo, para auxiliar na escolha de parâmetros ou tornar as séries estacionárias.

Nesta seção serão descritas algumas técnicas importantes para análise de séries temporais. Estas abordagens foram escolhidas por se tratarem de métodos tradicionais e por sua aplicação nos estudos de caso apresentados no Capítulo 5.

### 2.2.1

#### Método de Média Móvel

Nem todas as séries possuem um comportamento que favorecem o desenvolvimento do modelo. Séries que possuem ruídos na aquisição dos dados, seja por falha humana ou leitura dos sensores, geralmente são prejudiciais para o processo de previsão. Na literatura, existem diferentes técnicas propostas para a eliminação destes ruídos, como a implementação de filtros [27] e média móvel.

A média móvel é uma técnica utilizada para suavização de curvas, removendo as variações em curto período. Ao definir o tamanho de uma janela  $M$ , o método da média móvel realiza a média das  $M$  observações seguidas para o cálculo do novo valor, conforme mostra a Eq 2-7. Portanto, calcular a média móvel consiste na criação de uma nova série, onde os valores correspondem à média das observações originais dentro da janela [28].

$$y_i = \frac{1}{M} \sum_{k=i}^{i+M} y_k \quad (2-7)$$

Esta estratégia tem grande importância na redução de ruídos na série, que podem ser interpretadas como variações aleatórias e difíceis de serem previstas. É esperado que a visualização e análise desta variável sejam mais compreensíveis de se observar (como analisar a tendência de uma variável em função do tempo), além de tornar o ajuste do modelo de previsão mais simples.

### 2.2.2

#### Remoção de tendências

Conforme citado, a tendência é um dos quatro principais elementos que compõem uma série temporal. Contudo, a presença deste termo pode se tornar indesejada, visto que uma gama de modelos podem não ser adequadas para previsão de séries não-estacionárias (ou pelo menos com tendência, como em algumas modelagens de Redes Neurais (RN) e sistemas baseados em lógica *fuzzy*). Portanto, a identificação e remoção de tendências é uma etapa importante de pré-processamento de dados.

Existem diferentes técnicas que buscam identificar e remover as tendências, que vão desde métodos mais simples, como remoção de tendência linear ou não-linear por ajuste polinomial, até a utilização de abordagens iterativas [29, 30, 31]. De forma geral, tendências lineares são muito comuns em problemas reais e, com isso, soluções simples podem ser utilizadas para o tratamento da informação. Assim, neste trabalho será apresentada a remoção de tendências

usando regressão linear. O objetivo desta técnica é determinar os parâmetros de uma reta  $\alpha$  e  $\beta$  da Eq. 2-8 que melhor se adapte à série temporal. Além do método tradicional de regressão linear, podem ser utilizados métodos mais robustos a pontos fora da curva, como *Random Sample Consensus* (RANSAC) [32].

$$y(t)_{trend} = \alpha x(t) + \beta \quad (2-8)$$

Após o ajuste desta reta, a remoção da tendência é feita subtraindo-se a tendência do valor original da série, conforme exemplificado na Figura 2.3. A aplicação desta técnica pode não ser suficiente para tornar a série estacionária, mas pode ser útil para manter os valores da variável dentro de um intervalo fixo de valores. No modelo proposto neste trabalho, por exemplo, é importante que as séries temporais não tenham deslocamentos temporais, já que a modelagem do sistema exige que os valores estejam em torno de um valor médio.

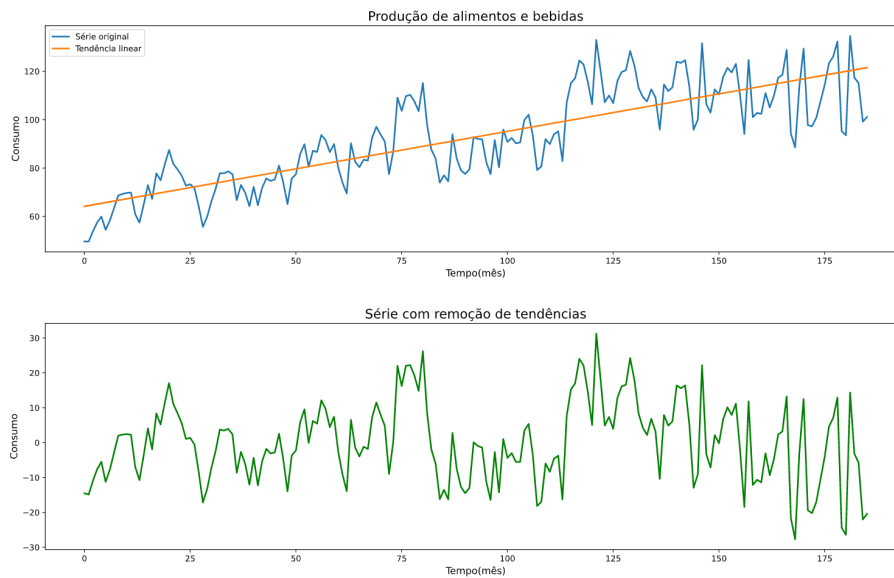


Figura 2.3: Exemplo de remoção de tendências em séries temporais

### 2.2.3

#### Diferenciação de séries

A propriedade de estacionariedade em séries temporais é, em geral, requisito necessário para a tarefa de previsão, e, por conta disso, técnicas de transformação de séries são de grande valia. A remoção de tendências, citada anteriormente, tem o objetivo de remover apenas uma das características, mas ainda não é garantido que a variável seja estacionária. Neste contexto, a diferenciação é um processo típico de pré-processamento de dados.

A diferenciação, como o próprio nome sugere, é a técnica de transformação de séries através da diferença de valores consecutivos, cujo objetivo é tentar eliminar os efeitos de tendência e sazonalidade do problema. Por simplicidade, neste trabalho será utilizada uma abordagem mais simples, conhecido por diferenciação de primeira ordem. Ela consiste em subtrair do valor no instante  $t$  o seu valor anterior, como mostrado na Eq. 2-9.

$$y'(t) = y(t) - y(t - 1) \quad (2-9)$$

A Figura 2.4 mostra um exemplo de diferenciação de séries de um caso real do valor diário das ações da empresa Google, sendo perceptível a diferença de comportamento temporal. É possível recuperar facilmente os valores originais invertendo a operação de diferenciação e retornando, assim, a informação para a escala original.

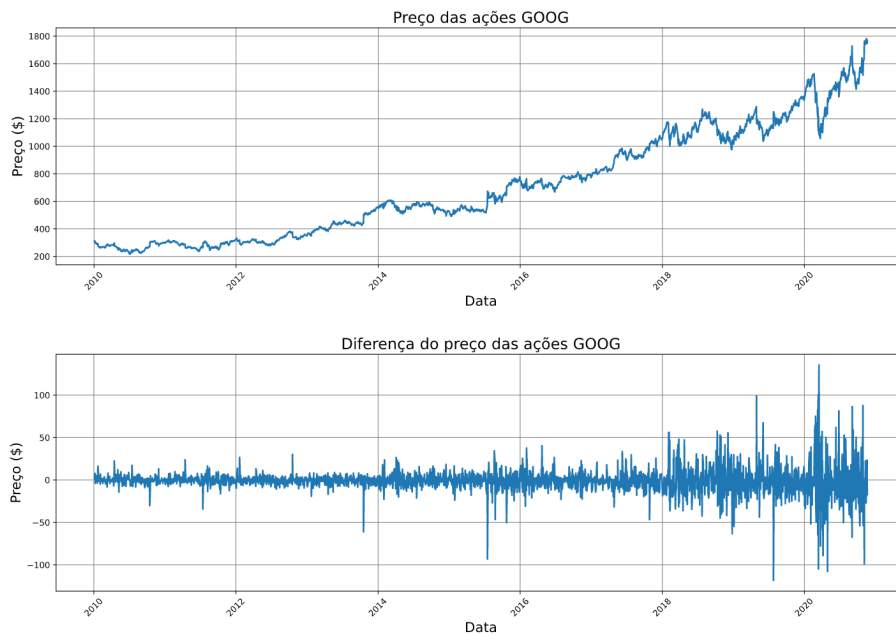


Figura 2.4: Preço da ação Google (sup.), e preços usando diferenciação de valores consecutivos da série (inf.)

A técnica de diferenciação de primeira ordem possui diversas aplicações, principalmente em problemas relacionados a séries financeiras, mas existem diferentes abordagens. A diferenciação de segunda ordem, por exemplo, realiza a diferença da série já diferenciada, e é utilizada quando a técnica de primeira ordem não é suficiente para tornar o problema estacionário. Além disso, aplicações mais específicas podem ser encontradas na literatura, como diferenciação de sazonalidade [33] e diferenciação de ordem  $N$  [34].

### 2.2.4

#### Autocorrelação

Em análise de séries temporais, é importante conhecer as defasagens que mais podem influenciar os valores atuais.

A autocorrelação  $\rho(k)$  é uma técnica que calcula a correlação entre a série e a  $k$ -ésima defasagem. Matematicamente, esta função é calculada pelo quociente entre a covariância entre  $y_t$  e  $y_{t-k}$  e a variância de  $y_t$ , conforme mostra a Eq. 2-10.

$$\rho(k) = \frac{Cov(y_t, y_{t-k})}{Var(y_t)} \quad (2-10)$$

Esta função varia entre -1 e 1, no qual o valor máximo representa uma máxima correlação e o valor mínimo a máxima anti-correlação. A autocorrelação igual a zero demonstra a ausência de correlação. Graficamente, a função de autocorrelação é capaz de identificar as melhores defasagens, além de identificar padrões sazonais. A Figura 2.5 mostra o gráfico de autocorrelação para diferentes defasagens para a série de energia (Figura 2.2).

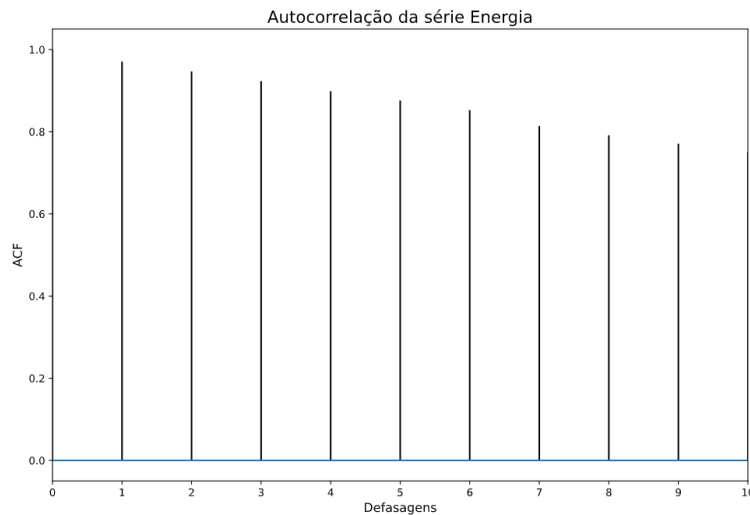


Figura 2.5: Exemplo gráfico de autocorrelação

A técnica de autocorrelação é amplamente utilizada na etapa de pré-processamento de séries. Provavelmente a aplicação mais conhecida seja a descoberta de defasagens mais relevantes para o processo de previsão, servindo como um primeiro indicador para a seleção de variáveis.

### 2.2.5

#### Correlação linear

Especificamente em problemas de previsões multivariadas, a correlação cruzada também é uma etapa relevante para a seleção de variáveis. Esta técnica calcula a dependência entre duas variáveis de um problema. Um dos métodos mais populares é a correlação de Pearson, definida pela Eq. 2-11.

$$\rho(x_{t-i}, y_{t-j}) = \frac{Cov(x_{t-i}, y_{t-j})}{\sqrt{Var(x_{t-i})Var(y_{t-j})}} \quad (2-11)$$

onde  $x_{t-i}$  e  $y_{t-j}$  são variáveis com defasagens  $i$  e  $j$ , respectivamente. Comparando-se com a Eq. 2-10, é possível observar uma similaridade entre a formulação da correlação de Pearson e a autocorrelação. De fato, o método de Pearson possui a mesma fundamentação discutida no tópico anterior, inclusive na avaliação de dependência linear entre as variáveis.

Na literatura, existem outras abordagens que visam a extrair a interdependência entre variáveis. Uma técnica conhecida é a correlação de Spearman [134], responsável por medir a dependência estatística entre duas variáveis. Em termos gerais, a correlação de Spearman avalia a relação monotônica entre variáveis, seja linear ou não-linear.

As técnicas que buscam quantificar as relações entre variáveis são essenciais para uma análise do problema e pré-processamento de dados. Os valores de correlação podem servir como critério para selecionar as variáveis que possivelmente possuem uma contribuição positiva na previsão da série temporal, reduzindo assim a dimensionalidade do problema. Vale ressaltar que estas técnicas possuem limitações e geralmente são utilizadas em conjunto com outras metodologias, sejam elas estatísticas ou não.

## 2.3

### Revisão bibliográfica de modelos para previsão de séries temporais

Inferir de forma consistente os valores futuros é uma tarefa que envolve, além de compreender o comportamento de séries temporais, selecionar os modelos de previsão mais adequados para o problema. A escolha nem sempre é tão trivial quanto analisar a dinâmica de séries, e por não existir um tipo de modelo que seja capaz de dominar esta área, as abordagens de diferentes naturezas são importantes para o enriquecimento do trabalho.

O advento de algoritmos complexos e potencialmente mais precisos impactaram a área de ciência de dados, inclusive na tarefa de previsão de séries temporais. Mais recentemente, arquiteturas de redes neurais profundas



tornaram-se as protagonistas em diversas categorias de problemas, sendo consideradas o estado da arte em previsão de séries temporais. Deste modo, é natural que a revisão bibliográfica seja dividida em três momentos: abordagem estatística, ML e *Deep Learning* (DL). Esta revisão bibliográfica tem o interesse em citar as técnicas de previsão mais relevantes e desenvolver um panorama atual desta área, com exemplos práticos encontrados na literatura.

### 2.3.1

#### Primeira fase - Modelos estatísticos

Normalmente, o processo de previsão envolve o ajuste de uma função para estabelecer a relação entre valores passados e o previsto. Por conta disso, a modelagem estatística é uma vertente clássica no processo de previsão, com uma gama de abordagens para diferentes classes de problema. Um modelo muito conhecido é o modelo autorregressivo (AR), expresso pela Eq. 2-12. Como é possível observar pela equação, o modelo AR é útil para casos univariados e admite uma relação linear entre as defasagens e o valor previsto. Os parâmetros do modelo AR ( $z$  e  $a_i, i \in 1, 2, \dots, n$ ) podem ser ajustados usando técnicas como *Expectation-Minimization* e Mínimos Quadrados Restrito [135]. Além disso, o parâmetro  $n$  usualmente é escolhido intuitivamente ou por tentativa e erro.

$$\hat{y}(t) = z + \sum_{i=1}^n a_i y(t-i) \quad (2-12)$$

Além do modelo AR, existem diferentes processos estatísticos que podem ser combinados para proporcionar uma modelagem mais completa de uma série temporal. É baseado neste conceito que o modelo autorregressivo de médias móveis (ARMA) funciona, incorporando a técnica de média móvel para a estimação do termo de erro do processo [35, 36]. Contudo, tanto o método AR quanto o ARMA apresentam limitações, visto que não podem ser aplicados em séries não-estacionárias diretamente. Para tornar a utilização de processos estatísticos aplicáveis neste tipo de problema, foi desenvolvido o modelo autorregressivo integrado de médias móveis (ARIMA) [136].

O modelo ARIMA trata a não-estacionariedade do problema através da parte integrada, que é responsável por realizar diferenciação das séries (explicado anteriormente), de modo a tornar a série estacionária e simples para o ajuste dos parâmetros do modelo [37]. Este modelo é responsável por introduzir uma nova classe de modelos estatísticos de previsão, podendo estender para problemas que incluem sazonalidade (SARIMA) [38] ou que possuem uma longa dependência temporal (FARIMA)[39].

Os processos estatísticos também podem ser aplicados em problemas multivariados. Desta forma, modelos univariados como o AR e ARIMA possuem suas versões aplicadas a múltiplas séries, como o VAR e VARIMA [137], respectivamente. Para exemplificar esta mudança, a Eq. 2-13 mostra a expansão do método AR para o VAR.

$$\hat{y}_s(t) = v + \sum_{s=1}^S \sum_{i=1}^n a_{i,s} y_s(t-i) \quad (2-13)$$

onde  $a_{i,s}$  é o parâmetro do modelo da variável  $y_s$  com defasagem  $i$ ;  $S$  é a quantidade de variáveis; e  $n$  é o total de defasagens usadas no modelo. O estudo sobre modelos estatísticos é bastante extenso e bastante relevante hoje em dia, mesmo com o surgimento de técnicas mais avançadas de Inteligência Computacional. Aplicações recentes podem ser encontradas na literatura, como previsão de taxa de câmbio [40], consumo de energia [41] e, mais recentemente, o nível de infecção na pandemia COVID-19 [42, 43]. Por se tratar de modelos consolidados na área de previsão de séries temporais, processos estatísticos também são populares em estudos comparativos, sendo geralmente uma linha de base para avaliar novas metodologias [44, 45, 46]. Por fim, estes modelos autorregressivos também são utilizados em conjunto com outras técnicas para a formação de modelos híbridos, como ARIMA-SVM [47, 48], ARIMA-LSTM [49, 50], entre outros [51, 52, 53].

Conforme discutido até agora, a abordagem estatística é amplamente utilizada em problemas reais e como modelo básico para comparação com novas metodologias propostas. Este fato pode ser explicado tanto por conta da popularidade quanto pela formulação matemática destas técnicas, já que produzem modelos lineares. Contudo, aproximar séries que possuem relações não-lineares usando um modelo linear pode não ser o mais apropriado, visto que a dinâmica temporal deste tipo de problema tende a ser não-linear [7]. Portanto, é esperado que um método não-linear seja mais promissor em modelar o comportamento não linear e, conseqüentemente, gerar previsões mais confiáveis.

### 2.3.2

#### **Segunda fase - Modelos de *Machine Learning***

O segundo momento a ser analisado nesta exploração sobre modelos de previsão de séries temporais refere-se à utilização de algoritmos de ML. É inegável que técnicas de Inteligência Computacional tiveram (e ainda têm) um impacto na área científica, como na área de visão computacional e reconheci-

mento de padrões. Este efeito também é observado no ramo de séries temporais, sendo considerada hoje em dia uma das técnicas mais relevantes por conta da capacidade de extrair relações não-lineares do problema.

Para a tarefa de previsão, os modelos de ML possuem, em geral, características propícias para uma modelagem não-linear do problema. Por mais que não sejam consideradas técnicas convencionais de modelagem de séries temporais [7], as abordagens de Inteligência Computacional são conhecidas por promover bons resultados de previsão. Dentre os algoritmos mais conhecidos, a abordagem via Redes Neurais é amplamente utilizada por conta de uma rede neural ser um aproximador universal de qualquer função contínua [54]. O problema de séries temporais pode ser readequado para a utilização de Redes Neurais, podendo também ser reformulado para o processo de previsão multivariada, tornando-se adequado para uma maior gama de problemas.

Por mais que aplicações de Redes Neurais sejam as mais populares, outras abordagens de Inteligência Computacional também são conhecidas por produzir bons resultados, como, por exemplo, Programação Genética [55]. Geralmente as abordagens de computação evolucionária são utilizadas em conjunto com técnicas mais populares para previsão de séries temporais para a otimização de parâmetros destes modelos.

Os três modelos mais tradicionais (RN, SIF e algoritmos de Computação Evolucionária), mesmo que sejam amplamente aplicados em problemas de previsão, possuem pontos negativos, como limitações quanto à busca de uma melhor arquitetura de RN e SIF. Neste sentido, os modelos híbridos são concebidos para combinar os aspectos positivos de cada uma das estratégias, surgindo assim abordagens neuro-evolucionária [56], fuzzy-evolucionária [57] e neuro-fuzzy [58].

O interesse em abordagens de ML para séries temporais pode ser evidenciado nas publicações científicas sobre estudos conceituais e avaliações práticas. Quanto ao uso de Redes Neurais têm-se, por exemplo, aplicações clássicas de previsão sobre precificação de produtos [59], séries econômicas e financeiras [60]. Contudo, as aplicações baseadas somente nesta abordagem são menos frequentes nos tempos atuais, por conta da existência de redes neurais profundas e modelos híbridos, como neuro-fuzzy.

Como alternativa aos algoritmos de Redes Neurais, os modelos baseados em lógica *fuzzy* constituem uma categoria de previsores com o potencial de interpretabilidade dos resultados. Consequentemente, as aplicações práticas continuam relevantes até nos dias atuais, com diferentes modificações propostas na literatura para uma previsão mais robusta. Neste trabalho, dedica-se um capítulo para tratar do funcionamento e das características importantes de

um SIF.

### 2.3.3

#### Terceira fase - Deep Learning

O aumento da capacidade de processamento para o treinamento de algoritmos de ML proporcionou o desenvolvimento de Redes Neurais mais complexas, abrindo um novo campo de estudo, denominado DL. Os modelos de DL possuem o status de estado da arte em diferentes tarefas, como reconhecimento de padrões visuais e processamento de linguagem natural [61], com desempenhos comparáveis ao de seres humanos e, em alguns casos, até superando-os [62]. Em previsão de séries temporais, os modelos de DL são utilizados e conhecidos por produzir previsões mais acuradas.

Os modelos de Redes Neurais Recorrentes são os mais apropriados para problemas envolvendo séries temporais, graças à capacidade de recorrência temporal inerente ao modelo. Uma das arquiteturas mais populares para esta tarefa é a *Long short-term Memory* (LSTM), cuja estrutura é bastante popular nos dias atuais. Uma das grandes vantagens deste modelo é a capacidade de aprender as relações temporais de longo e curto prazo, característica muito desejada em problemas de previsão de séries temporais.

A arquitetura básica da rede LSTM serviu como base para o desenvolvimento de diferentes variantes do modelo, como o *Gated Recurrent Unit* (GRU) [63]. Especificamente em tarefas de séries temporais, a hibridização de técnicas de DL com modelos estatísticos também está presente na literatura. Em [50] é proposto um modelo ARIMA-LSTM para modelar tanto a componente linear quanto a não-linear de uma série temporal. Além disso, em [64] é desenvolvido um modelo híbrido de AR e LSTM, onde é efetuada uma combinação linear destas duas abordagens para melhorar a previsão a longo e curto prazo.

No outro espectro de técnicas, os modelos de Redes Neurais Convolucionais também pode ser utilizado para a tarefa de previsão. Por mais que sua arquitetura seja voltada para aplicações de visão computacional, as informações de séries temporais podem ser rearranjadas em um formato bidimensional para explicitar uma relação espaço-temporal entre as variáveis [65]. Deste modo, o uso de redes convolucionais torna-se uma alternativa em relação aos modelos recorrentes, com a vantagem de tornar mais simples a relação espacial entre as variáveis, importante em problemas como previsão de velocidade de vento [66, 67]. Em [68], por exemplo, é utilizada uma técnica convolucional (U-Convolutional) para modelar a relação espaço-temporal de velocidade do vento em diferentes pontos de um parque eólico. Além disso, em [69] é desenvolvida uma metodologia que utiliza um modelo convolucional para a seleção

de características que alimentam um modelo LSTM para previsão.

Recentemente, a criação de modelos baseados no Mecanismo de Atenção tem aumentado o interesse de aplicações em previsão de séries temporais. Conforme descrito em [70], o uso do mecanismo de atenção pode dispensar elementos de recorrências, com a vantagem de ser computacionalmente mais simples de ser treinado e superior em qualidade, como em tarefas de tradução de idiomas. Este efeito também é observado em trabalhos atuais para modelagem de relações com longa dependência, superando em alguns casos modelos de redes neurais recorrentes [71]. Em [72], são discutidos os limites de um modelo de DL recente, denominado Transformer [72], como a capacidade de memória do modelo, comparando-se o seu desempenho com os de arquiteturas mais comuns de DL. Outra potencial característica deste modelo é a interpretabilidade, discutida em [73].

### 2.3.4

#### Considerações finais sobre modelos de previsão

Como é possível observar, a modelagem de séries temporais é bastante diversa e possibilita o uso de diferentes metodologias. Mesmo assim, atualmente é perceptível que as técnicas de DL dominam a área de previsão, principalmente devido à facilidade de incorporar e extrair as relações não-lineares com as defasagens das séries. Contudo, esta abordagem não garante o melhor resultado, sendo necessário avaliar qual modelo usar para o problema em questão. Por mais que estas técnicas possuam características desejáveis para modelagem de séries temporais, uma desvantagem nítida é a falta de transparência dos fatores que influenciam no resultado da previsão.

A dimensionalidade da base de dados disponível é um fator muito importante a ser observado durante o processo de escolha do modelo. As técnicas de DL geralmente demandam uma quantidade alta de registros disponíveis para o treinamento, de modo que todos os parâmetros sejam ajustados corretamente. Já em bases de dados com menos registros, como em problemas com valores coletados anualmente, os processos estatísticos podem ser mais adequadas para entender o comportamento temporal. Abordagens não-lineares não devem ser unicamente consideradas, pois modelos estatísticos são capazes de realizar uma aproximação linear do comportamento de séries temporais [7], e em alguns casos obter até um desempenho superior ao de técnicas mais complexas [64].

Todos os modelos de previsão citados nesta seção, por mais simples ou complexos que sejam, possuem aplicações práticas nos dias atuais. A escolha da metodologia a ser utilizada é função das características do problema e

dos objetivos da solução que se pretende obter. Este trabalho concentra-se no desenvolvimento de um modelo interpretável para previsão de séries temporais, de modo a viabilizar uma análise mais direta dos resultados. No próximo capítulo, será abordado de forma detalhada o funcionamento do Sistema de Inferência Fuzzy, unidade fundamental do modelo proposto.

Prever o comportamento de séries temporais é uma tarefa complexa, com diferentes características que viabilizam uma maior diversidade na modelagem do problema. Com o avanço e a popularização de técnicas de Inteligência Computacional, modelos mais complexos tornam-se opções sólidas para resolução de problemas específicos, principalmente para a modelagem de séries temporais com um comportamento não-linear.

O Sistema de Inferência *Fuzzy* é caracterizado pela inserção da Lógica *fuzzy* para modelar, ao menos teoricamente, qualquer problema aproximável por uma função contínua, linear ou não-linear, sendo considerado um aproximador universal [74]. Tipicamente, este modelo é aplicado a diferentes tipos de problemas, como classificação [75], controle [76] e previsão de séries temporais [77].

O SIF é considerado um modelo com maior facilidade de interpretação, se comparado a outras técnicas de ML. Este fato se deve à tentativa do sistema de simular o raciocínio humano, que geralmente é subjetivo e baseado em informações incertas. Utilizando variáveis linguísticas e criação de regras do tipo "se-então", a análise e avaliação do resultado de um SIF tornam-se mais simples para um humano, devido a uma maior similaridade com as tomadas de decisões no dia-a-dia.

Tipicamente, um SIF pode ser projetado de duas formas: manualmente ou através de técnicas automáticas. No método manual, é necessário que um especialista seja capaz de formular as regras e os termos linguísticos. Por mais que pareça cativante transferir todo o conhecimento de um especialista para o SIF, esta tarefa é laboriosa e pode não modelar completamente o problema. Com o aumento da popularidade da lógica *fuzzy*, foram desenvolvidos métodos para extração de conhecimento a partir de dados, tornando o uso de SIF mais prático e sem a necessidade inicial de um especialista.

Neste capítulo, será exposta uma visão geral de um SIF, abordando-se suas principais componentes deste modelo. Em seguida, é apresentado o aspecto da interpretabilidade, onde são definidos os critérios fundamentais para avaliar a explicabilidade em sistemas baseados em lógica *fuzzy*. Posteriormente,

é discutido como o SIF pode ser ajustado para previsão de séries temporais, especialmente em problemas multivariados. Por fim, a última seção aborda os trabalhos recentes de previsão de séries temporais usando modelos baseados em lógica *fuzzy*, explicando com maior detalhe o funcionamento do modelo AutoMFIS [2], de modo a facilitar a compreensão do modelo proposto neste trabalho.

### 3.1

#### Componentes de um SIF

Para uma melhor compreensão do modelo proposto neste trabalho, o funcionamento básico de um SIF é um tópico importante a ser comentado. Certamente o estudo aprofundado sobre SIFs está fora do escopo deste trabalho, visto que é um assunto muito amplo com sua teoria já consolidada em obras de referência [138, 139]. Ainda assim, comentar alguns aspectos sobre a construção de um sistema *fuzzy* pode ser útil para auxiliar no entendimento das técnicas utilizadas. Deste modo, esta seção expõe uma breve descrição sobre os componentes de um SIF.

A Figura 3.1 mostra a arquitetura de um SIF do tipo Mamdani, que é composto por quatro componentes principais: Fuzzificação, Inferência, Base de Regras e Defuzzificação. Cada um destes elementos será caracterizado a seguir.

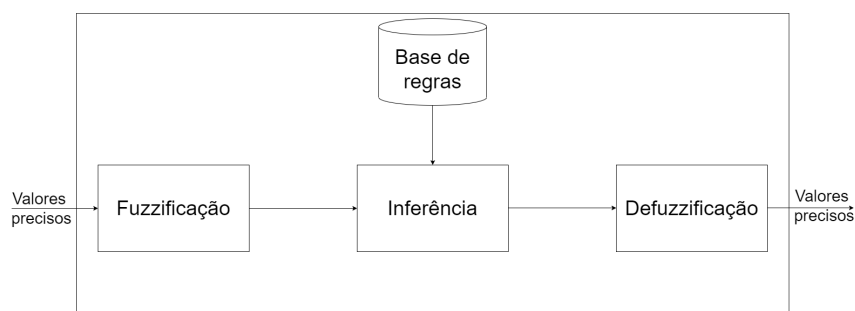


Figura 3.1: Diagrama de blocos de um Sistema de Inferência *fuzzy*

#### 3.1.1

##### Fuzzificação

O estágio de Fuzzificação é responsável por estabelecer a relação entre o valor de entrada do sistema, que é preciso, e a sua respectiva variável linguística. Cada variável linguística possui diferentes valores (chamados de termos linguísticos) que são descritos por conjuntos *fuzzy*. Matematicamente, estes conjuntos são representados por funções de pertinência que podem possuir diferentes formatos, como triangulares, trapezoidais e gaussianas.



Para cada variável linguística, a quantidade de conjuntos *fuzzy* e o definição de suas funções de pertinência são parâmetros que variam de acordo com o problema e que influenciam na representação do mesmo. Estes ajustes podem ser feitos manualmente, a partir do conhecimento de um especialista ou por estimações do usuário. Contudo, é comum que esta etapa seja utilizada uma técnica de ajuste automático por meio de função da distribuição de dados ou algum algoritmo de otimização, como Algoritmos Genéticos [131]. Outras técnicas de geração automática de funções de pertinência serão discutidas no Capítulo 4, quando será discutido o modelo proposto.

Com o dicionário linguístico definido, a relação entre os valores precisos e o domínio *fuzzy* é feita avaliando o grau de pertinência a cada um dos conjuntos. A Figura 3.2 ilustra o processo de fuzzificação de uma variável. Cada valor pode pertencer a mais de um conjunto *fuzzy* com diferentes graus de pertinência, onde a variável  $x$  pertence ao conjunto M e A com diferentes graus (0,3 e 0,7, respectivamente). Esta característica é o cerne da lógica *fuzzy*, onde os valores de  $x$  possuem uma determinada compatibilidade com cada um dos termos linguísticos.

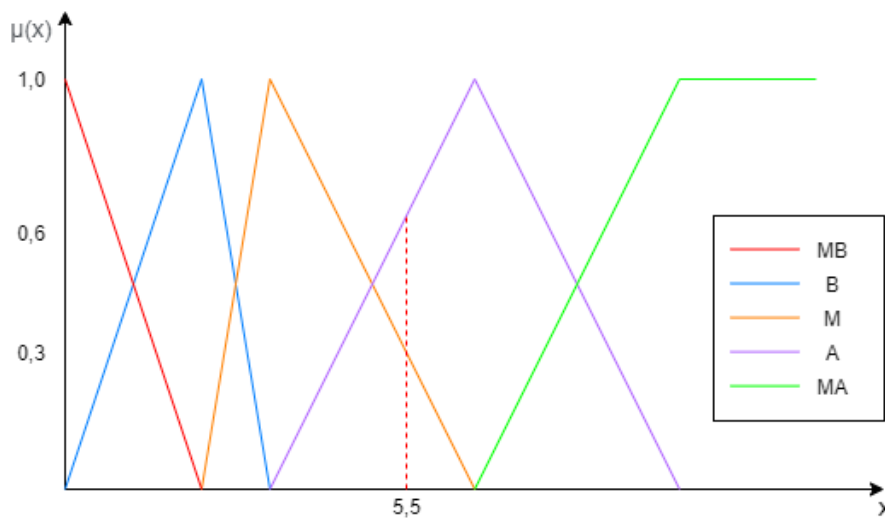


Figura 3.2: Exemplo de Fuzzificação

Por mais que o conceito não seja complexo, é importante notar o papel fundamental da fuzzificação na representação do problema. De uma forma intuitiva, as funções de pertinência caracterizam as variáveis do problema, definidas em um intervalo que seja semanticamente aceitável. Esta definição não apenas influencia na interpretabilidade do sistema, mas também na formulação de regras, e portanto deve-se ter cautela durante a construção do modelo, de modo que a fuzzificação descreva o problema como esperado.

### 3.1.2

#### Base de regras

A Base de Regras é o componente que armazena o conhecimento do modelo em forma de regras de inferência. A base é composta por sentenças do tipo "se-então", que estabelecem uma relação de implicação entre variáveis de entrada e de saída. Por definição, as variáveis de entradas são apresentadas como os antecedentes das regras; a saída do sistema é definida como o termo consequente.

Assim como na definição das variáveis linguísticas, as regras de um sistema *fuzzy* podem ser desenvolvidas manualmente, geralmente fornecidas por um especialista no assunto específico. Contudo, mesmo para um profissional experiente, a extração manual de conhecimento pode ser árdua e imprecisa, dependendo da complexidade do problema. Uma alternativa, muito utilizada em trabalhos utilizando previsão de séries temporais, é a extração automática de conhecimento através dos dados presentes no problema. Um dos métodos mais populares é o de Wang e Mendel [78], mas esta tarefa pode ser realizada por meio de heurísticas [79] ou algoritmos evolutivos [80].

A qualidade da base de regras é um aspecto de extrema importância, haja vista sua influência direta no desempenho do sistema. Em geral, é esperado que o conjunto de regras seja conciso, livre de regras conflitantes ou redundantes e contendo principalmente o conhecimento relevante extraído do problema. Estes fatores influenciam não apenas na acurácia, como também na capacidade de interpretabilidade do modelo, como será discutido ainda neste capítulo.

### 3.1.3

#### Inferência

Com os valores de entrada já mapeados e as regras estabelecidas, a próxima etapa consiste na construção do conjunto *fuzzy* de saída. A inferência é responsável por realizar todas as operações com conjuntos *fuzzy*, como a combinação dos antecedentes, implicação e agregação das regras.

Semanticamente, os antecedentes das regras são geralmente combinados através do conectivo **e**, expresso matematicamente pela norma-t (\*) [139]. Portanto, uma regra será ativada apenas se todos os termos linguísticos que compõem a proposição da regra tiverem um grau de pertinência diferente de 0. O cálculo do grau de pertinência da proposição da regra  $\mu_A(x)$  serve como compatibilidade ao conjunto *fuzzy* de saída, sendo utilizado uma norma-t para o cálculo da operação implicação (Eq. 3-1).

$$\mu_{B^*}(y) = \mu_A(x) * \mu_B(y) \quad (3-1)$$

onde  $\mu_A(x)$  é o grau de ativação do antecedente  $A$  com entrada precisa  $x$ ,  $\mu_B(y)$  é o conjunto fuzzy de saída referente ao consequente  $B$  da regra fuzzy e  $\mu_{B^*}(y)$  é o resultado da inferência. Por fim, faz-se necessária uma agregação dos resultados fornecidos por cada uma das regras ativadas. Isto é realizado, em geral, por uma conorma-t ( $\oplus$ ), que tem no operador máximo o seu representante mais utilizado [139]. Existem outras abordagens possíveis, como será detalhado no próximo capítulo.

### 3.1.4 Defuzzificação

A saída da etapa de inferência ainda é uma variável no domínio *fuzzy*, representado pela composição de conjuntos *fuzzy*. Entretanto, a saída de qualquer modelo deve estar no domínio adequado ao problema. Portanto, a etapa de defuzzificação tem a tarefa de relacionar o resultado *fuzzy* a um valor preciso.

Interpretar o conjunto *fuzzy* como um número preciso pode ser complexo, mas existem diferentes metodologias para representar o resultado [140]. Intuitivamente, o método do centroide traduz com maior fidelidade a forma do conjunto de saída, sendo uma das técnicas de defuzzificação mais populares. Na literatura existem outras abordagens com diferentes características, como a média dos máximos ou o método da altura.

## 3.2 Interpretabilidade

Atualmente, o aspecto de interpretabilidade tem sido cada vez mais investigado. Seja por necessidade, como novas legislações que obrigam modelos de apoio à decisão a serem mais compreensíveis, ou por desejo humano, o paradigma da interpretabilidade tem instigado a pesquisa de novas metodologias para tornar a depuração do resultado mais compreensível. Deste modo, um dos grandes esforços da área de *Explainable AI* está na tarefa de tornar modelos considerados caixa-preta, como LSTM e redes neurais convolucionais, cada vez mais transparentes. Além disso, outra vertente desta área reside no desenvolvimento de novas abordagens que tradicionalmente oferecem o aspecto da explicabilidade.

Antes de prosseguir no tópico, é importante estabelecer o que é considerado interpretabilidade. Por definição, este termo é uma qualidade relativa à capacidade de explicar o significado de um acontecimento, seja uma informação ou ação. Esta ideia também é presente na área de dados, onde o aspecto da interpretabilidade remete à ideia de compreensão do resultado e dos fato-

res que o determinaram. Neste sentido, é possível compreender prontamente a vantagem de um modelo explicável, tornando o resultado mais simples de ser analisado e, conseqüentemente, aumentando a confiabilidade e a aceitabilidade de um modelo.

Ainda assim, a interpretabilidade é um tópico complexo de ser estudado. Por mais que seja fácil definir este termo, analisar o que é considerado interpretável pode ser bastante subjetivo, visto que esta característica também depende da percepção do usuário ao interagir com o modelo. Deste modo, a interpretabilidade como métrica de avaliação tenta traduzir, de forma quantitativa ou qualitativa, os atributos do modelo que influenciam na avaliação do resultado e sua respectiva explicação.

Além da capacidade de ser um aproximador universal, o SIF possui uma característica de grande interesse, a interpretabilidade. Este fator pode ser considerado fundamental na escolha deste modelo quando é desejada uma maior interação humana [81], por conta da proximidade da modelagem de um SIF com o raciocínio humano. Mesmo diante desta grande vantagem, não existe a garantia de que um sistema *fuzzy* seja interpretável em um alto grau, sendo necessário cautela durante o desenvolvimento do modelo.

Em um SIF, a interpretabilidade pode ser avaliada em diferentes níveis de abstração e objetivos. A taxonomia apresentada neste trabalho é definida por Gacto [82], onde as métricas podem ser classificadas quanto aos elementos do SIF (dicionário linguístico e base de regras) e seus aspectos (complexidade e semântica). Esta abordagem ajuda a organizar e identificar as métricas de interpretabilidade, além de contribuir na avaliação localizada de diferentes perspectivas de um modelo *fuzzy*. A Tabela 3.1 lista as métricas relevantes para esta tarefa, agrupadas pela definição de Gacto.

Tabela 3.1: Métricas de interpretabilidade

Nível de abstração	Característica	Métricas
Conjuntos <i>fuzzy</i>	Complexidade	Número de conjuntos <i>fuzzy</i>
		Número de variáveis
	Semântica	Cobertura
		Distinguibilidade
		Complementaridade
Base de regras <i>fuzzy</i>	Complexidade	Tamanho da base de regras
		Comprimento dos antecedentes
	Semântica	Consistência
		Ativação média das regras

Como é possível perceber, a interpretabilidade pode ser avaliada sob diferentes perspectivas de um SIF usando critérios estabelecidos na literatura.

Estas métricas são utilizadas, neste trabalho, para auxiliar na avaliação do tópico de explicabilidade. Os critérios são descritos a seguir:

- **Número de conjuntos *fuzzy*:** Quantidade de conjuntos *fuzzy* necessários, para cada variável linguística, para compor o SIF. Esta métrica também está relacionada com a heurística proposta por Miller [83], em relação ao aspecto de limitação de interpretabilidade humana. Sugerido pelo autor da heurística, a quantidade limite é, em geral, limitado a nove elementos.
- **Número de variáveis:** Quantidade de variáveis utilizadas para modelagem do SIF. Esta característica está relacionada com a dimensionalidade do problema e consequentemente pode afetar a quantidade de regras [84]. Esta métrica não possui uma fórmula quantitativa e portanto a avaliação torna-se subjetiva em relação ao problema. Para a redução de variáveis, geralmente são utilizadas técnicas de pré-processamento de seleção de variáveis.
- **Cobertura:** Avalia, de uma forma qualitativa, se os conjuntos *fuzzy* criados pelo SIF compreendem todo o universo de discurso do problema. De um modo geral, esta métrica analisa se, para qualquer valor de entrada, existe pelo menos um conjunto *fuzzy* correspondente.
- **Distinguibilidade:** Avalia semanticamente os termos linguísticos do SIF, observando características como sobreposição de conjuntos e discernibilidade. A Figura 3.3, por exemplo, mostra exemplo de conjuntos *fuzzy* não interpretáveis (direita) e interpretáveis (esquerda).

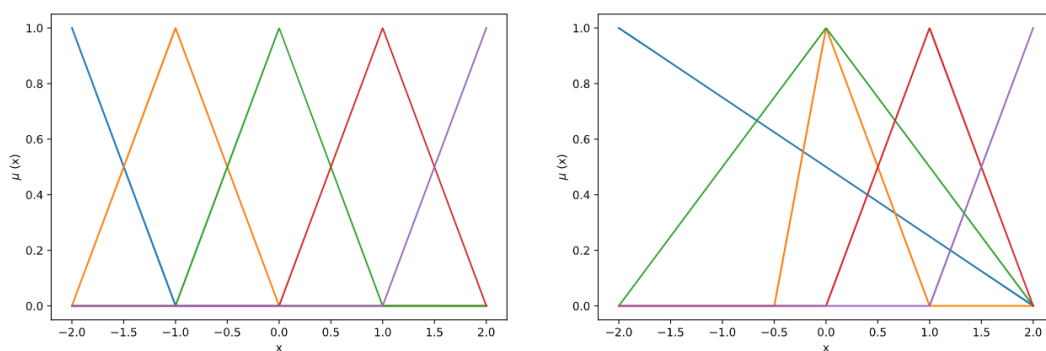


Figura 3.3: Exemplo de conjuntos *fuzzy* interpretáveis e não interpretáveis

- **Complementaridade:** Caracteriza a distribuição dos conjuntos *fuzzy*. No geral, é esperado que, para qualquer valor do universo de discurso, a soma do grau de pertinência de todos os conjuntos ativados deve ter soma igual (ou próximo) a 1.

- **Tamanho da base de regras:** Avalia a capacidade de generalização do modelo, extraindo padrões mais relevantes para a inferência do problema. Quanto menor o tamanho da base de regras, melhor, pois torna-se mais inteligível para ser analisado pelo ser humano. Contudo, bases de regras pequenas podem sofrer por sua simplicidade, prejudicando a acurácia do modelo.
- **Comprimento das regras:** Contabiliza a média da quantidade de antecedentes nas regras do SIF. Esta métrica está relacionada com a heurística proposta por Miller [83], onde é descrito que existe uma limitação da compreensão humana, em termos linguísticos, de uma regra *fuzzy*. Em geral, é estabelecido que a quantidade limite de antecedentes aceitável é igual a nove. Ainda assim, idealmente é esperado que a regra tenha um tamanho bem menor do que o limite máximo estabelecido por esta heurística, visto que um comprimento de regras próximo a nove torna a inteligibilidade muito prejudicada.
- **Consistência:** Determina a coerência da base de regras, analisando se regras com antecedentes semelhantes possuem consequentes diferentes. De um modo geral, este critério pode auxiliar na identificação da presença de regras contraditórias que podem prejudicar a compreensão do problema.
- **Ativação média das regras:** Indica a média das regras ativadas simultaneamente. Esta métrica está diretamente relacionada com o aspecto de interpretabilidade de resultados, e quanto maior a quantidade de ativação média das regras, mais prejudicada é a interpretabilidade. Deste modo, um valor baixo desta métrica indica uma facilidade na avaliação, e torna-se mais simples entender os termos linguísticos que contribuíram para o resultado obtido pelo modelo.

Por mais que as métricas de avaliação de interpretabilidade sejam úteis para a compreensão do modelo, este tópico ainda possui algumas ressalvas que precisam ser discutidas. Alguns critérios, como distinguibilidade e cobertura, podem ter suas condições satisfeitas com uma construção do modelo apropriado e portanto não precisam ser avaliadas para cada problema. Por outro lado, existem métricas que não possuem heurísticas ou critérios absolutos para uma avaliação de interpretabilidade, sendo por vezes subjetiva a sua análise. Por exemplo, o tamanho da base de regras depende de fatores como a complexidade do problema, que impacta diretamente na quantidade necessárias de regras para representá-lo. Idealmente, a análise de critérios de interpretabilidade permite a comparação com outros modelos *fuzzy*, mas na ausência de resultados

de outros SIFs, a avaliação deste tópico é realizada segundo critérios sugeridos na literatura e também por pontos subjetivos do problema.

Conforme será explicado detalhadamente mais adiante, as técnicas de síntese automática tornaram o desenvolvimento de um SIF mais simples, eliminando (pelo menos parcialmente) a necessidade de um especialista para a criação do dicionário linguístico e da base de regras. Contudo, estas técnicas geralmente são concebidas com o objetivo de tornar o modelo mais acurado, podendo gerar conjuntos *fuzzy* que semanticamente não sejam compreensíveis ou até base de regras exageradamente grande, dificultando o entendimento da inferência. Consequentemente, a abordagem puramente voltada à acurácia pode inviabilizar a característica interpretável do SIF, tornando-se semelhante a um modelo caixa-preta [85].

Durante o desenvolvimento de um SIF, a relação entre acurácia e interpretabilidade é um tópico essencial que ajuda a obter um balanço entre estas características, já que esses dois objetivos são geralmente conflitantes. Enquanto a acurácia privilegia sistemas que possuem base de regras mais complexas, a interpretabilidade necessita que o modelo seja simples, de modo que seja mais razoável compreender o resultado. Por conta disso, em diversos trabalhos é possível observar a tentativa de obter o ponto ótimo entre estes dois parâmetros, de modo que estas características não sejam prejudicadas.

A prioridade na relação acurácia-interpretabilidade estabelece a estratégia de otimização do problema, formalmente definida em [86]. Uma abordagem muito frequente consiste na otimização do modelo baseado puramente na acurácia. Uma vez estabelecida a base de regras, são utilizadas técnicas de filtragem para tornar o sistema mais simples. Esta estratégia tenta melhorar o aspecto interpretável do modelo sem comprometer a acurácia obtida, manipulando-se a base de regras para remoção de regras pouco relevantes ou potencialmente redundantes, como em [87] e [88]. De forma contrária a esta primeira abordagem, o SIF pode ser iniciado de forma simplificada (geralmente manualmente desenvolvido) e posteriormente novas regras são adicionadas para aumentar a acurácia. Por fim, as duas características podem ser ajustadas conjuntamente, utilizando principalmente uma abordagem multi-objetiva de otimização [89].

### 3.3

#### Previsão de séries temporais usando SIF

Assim como outros modelos de Inteligência Computacional, o SIF pode ser utilizado para problemas de previsão de séries temporais, usando a estrutura básica citada neste capítulo. Conforme mostra a Figura 3.4, as entradas

do sistema são as defasagens da série e a saída é o valor a ser previsto.

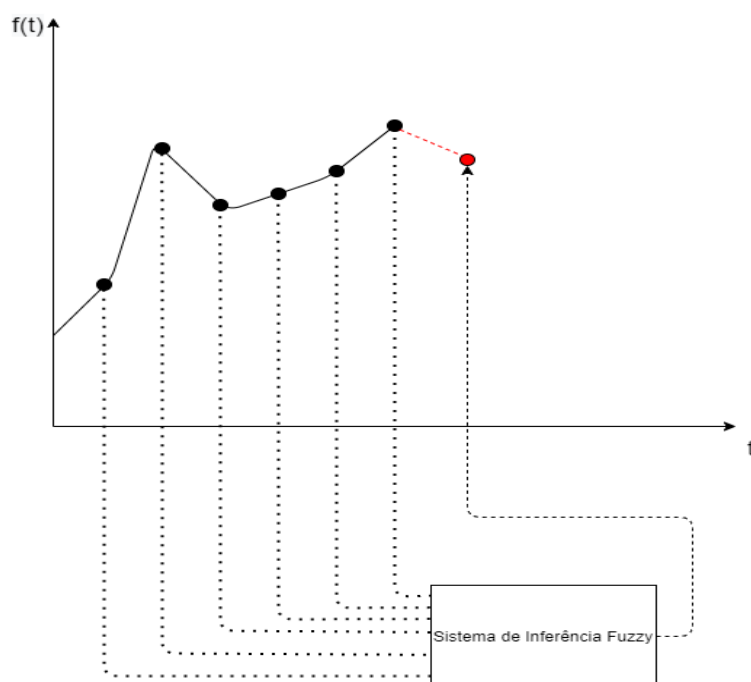


Figura 3.4: Exemplo de configuração de um SIF para previsão de séries

O SIF não possui uma recorrência inerente ao modelo, mas ainda assim é uma técnica amplamente aplicada neste tipo de tarefa. Para realizar o processo de previsão *multi-step*, os valores anteriormente previstos são utilizados como entrada para prever o próximo passo de previsão, e este processo ocorre iterativamente até o horizonte de previsão do problema.

A tarefa de previsão também pode ser generalizada para problemas multivariados, conforme mostra a Figura 3.5. Para realizar este tipo de trabalho, é possível utilizar um sistema único para prever todas as  $n$  variáveis, ou usar  $n$  SIFs para previsão.

### 3.4 Modelos

A ideia de um sistema que tente simular o comportamento do raciocínio humano e descrever o processo de uma forma semanticamente compreensível torna a aplicação de SIFs atrativa para atividades industriais e acadêmicas. Uma das primeiras aplicações deste modelo está presente na área de controle, sendo objeto de pesquisa até os dias de hoje. Ainda assim, o uso de sistemas *fuzzy* possui relevância em outras áreas, como classificação e agrupamento [90]. A previsão de séries temporais também é uma área de atuação importante de sistemas baseados em Lógica *fuzzy*. Além da sua natureza de aproximador



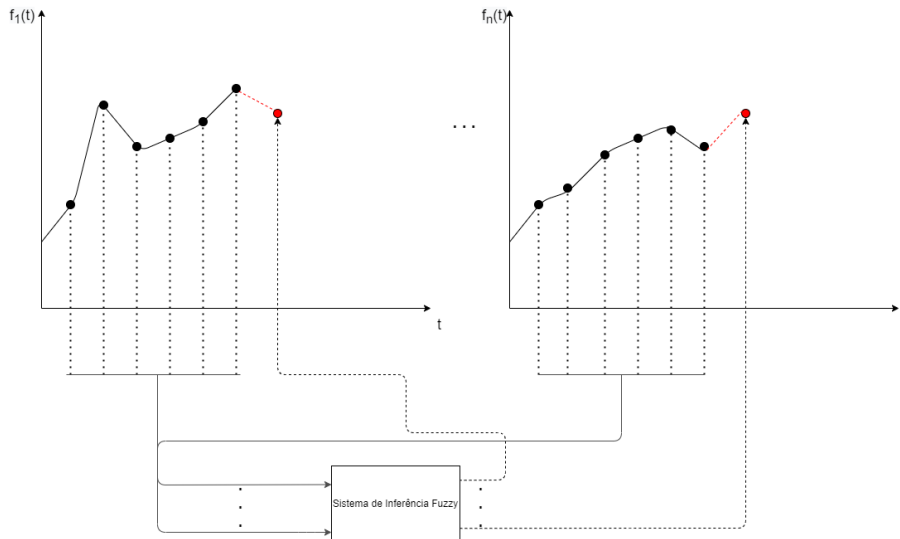


Figura 3.5: Exemplo de previsão multivariada usando SIF

universal [74], a capacidade de interpretabilidade dos resultados torna o uso de SIFs mais atrativos para diferentes aplicações.

Esta seção tem como objetivo reunir as contribuições de metodologias baseadas em sistemas *fuzzy* para previsão de séries temporais e suas principais aplicações. Além disso, no final desta seção é apresentado em linhas gerais o funcionamento do modelo AutoMFIS, considerado a base do modelo proposto. Esta explicação é importante para compreender com maior facilidade o funcionamento do e-AutoMFIS, bem como fundamentar os avanços alcançados nesta nova abordagem proposta.

### 3.4.1

#### Revisão de modelos preditores *fuzzy* usados na literatura

Prever resultados futuros baseados em observações passadas não é uma tarefa fácil, sendo muitas vezes necessário um modelo complexo para captar estas relações. Na área de Inteligência Computacional, por exemplo, existem técnicas que possuem o aspecto de recorrência, essencial para previsão de séries temporais, inerente à sua arquitetura. É o caso de Redes Neurais Recorrentes, como Echo-State Network [91] e LSTM [92], já citadas no capítulo anterior. Por mais que produzam bons resultados, a interpretabilidade é totalmente prejudicada, já que estes modelos são considerados caixa-preta. Quando a explicação do resultado é um fator importante para o problema, é comum a utilização de um SIF para previsão de séries temporais.

O sistema baseado em lógica *fuzzy* preenche o requisito de ser um modelo capaz de obter um bom *trade-off* entre interpretabilidade e acurácia. Com o desenvolvimento de técnicas de síntese automática de SIFs, seu uso em diferen-

tes problemas tornou-se mais simples, substituindo as informações geralmente fornecidas por um especialista pela extração automática de conhecimento obtida a partir da base de dados. Por mais que a automatização do processo de construção do SIF tenha tornado a tarefa menos trabalhosa do ponto de vista humano, estas metodologias necessitam de técnicas de otimização para descobrir a configuração que resulte no desempenho satisfatório. Atualmente, as abordagens de síntese automática de sistemas *fuzzy* podem ser classificadas quanto à estratégia de otimização, usando técnicas de computação evolucionária (*Evolutionary approach*) ou métodos adaptativos e heurísticos (*Evolving approach*).

Os modelos Fuzzy-genéticos utilizam a combinação de sistemas *fuzzy* com Algoritmos Evolutivos, como Algoritmos Genéticos e Programação Genética. Desta forma, é aproveitada a capacidade de otimização do Algoritmo Evolutivo para o desenvolvimento de um SIF, em diferentes níveis. Em [93], é proposto um método no qual o algoritmo de Programação Genética é utilizado para extração de regras de um SIF. Em [94], um Algoritmo Evolucionário multiobjetivo é utilizado para o aprendizado da base de regras de um Sistema Fuzzy, visando à otimização da acurácia e interpretabilidade em problemas de regressão.

Na área de sistemas Fuzzy-Genéticos, existem duas técnicas principais, dependendo do objetivo da otimização da base de regras. O primeiro método, conhecido como abordagem Pittsburgh [95], considera o conjunto de regras *fuzzy* como um indivíduo a ser evoluído, como mostrado no trabalho de [96], onde se utiliza Programação Genética para criação e otimização da base de regras de um SIF. Já a abordagem Michigan é outra técnica conhecida para este tipo de sistema híbrido. Neste método, as regras são avaliadas como indivíduos, gerando então um conjunto de regras contendo aquelas que foram selecionadas pelo algoritmo evolutivo.

A utilização de algoritmos evolutivos para a busca de uma melhor configuração de SIFs é bastante significativa na área científica e é um tópico de pesquisa nos dias atuais. Um grande fator que torna essa abordagem popular e amplamente divulgada é a eficiência de técnicas de computação evolucionária na busca de configurações promissoras em problemas com um grande espaço de busca, como é o caso de otimização de parâmetros de um SIF. Modelos fuzzy-genéticos possuem aplicações em séries temporais, onde um dos problemas mais comuns é definir as melhores defasagens para composição das regras *fuzzy*. Este tópico é discutido em [97], onde o algoritmo evolutivo é usado para seleção de defasagens e otimização de um SIF.

Na literatura, sistemas *fuzzy* e redes neurais são comumente combinados e adaptados para diferentes problemas. Em [98], por exemplo, é proposto um

modelo que utiliza o modelo ANFIS em conjunto com um processo estatístico autorregressivo para a ponderação do resultado previsto. O ANFIS também é aplicado em problemas reais [99, 100] e usado em diferentes abordagens, como em comitê de previsores [101].

Diferentemente de modelos híbridos, os Sistemas Fuzzy auto-adaptativos (ou *Evolving Fuzzy Systems*) não utilizam algoritmos evolutivos para a otimização de sua configuração ou base de regras. O aprendizado deste tipo de técnica é realizada de forma adaptativa, com mudanças graduais em sua estrutura e aprendizado em função do fluxo contínuo de dados [102]. Esta abordagem é bastante diversificada, visto que existem diferentes metodologias para o aprendizado adaptativo de um sistema [103].

O aprendizado *online* permite explorar características que podem ser aproveitadas em problemas de previsão de séries temporais. Conforme discutido no segundo capítulo, as mudanças de comportamento de séries são bastante recorrentes, como efeitos de tendência ou sazonalidade. Este efeito é chamado de deriva de conceito (ou *concept-drift*), onde as características de um problema não são estáveis e mudam com o tempo. Deste modo, os modelos adaptativos podem captar estas variações e modificar a sua estrutura para adequar ao novo comportamento da série.

A capacidade de lidar com problemas inerentes à previsão de séries temporais demonstra o potencial do uso de modelos *fuzzy* adaptativos em tarefas de previsão. Em [104], é desenvolvido um comitê de três modelos adaptativos para agregação e previsão de temperatura usando variáveis climáticas. Em [105], é proposto um modelo *fuzzy* do tipo Takagi-Sugeno para aprendizado online de séries temporais usando a distribuição dos dados apresentados continuamente para o ajuste de sua arquitetura. Outro modelo *fuzzy* adaptativo pode ser visto em [106], onde é utilizado o conceito de granularidade para acomodar e adaptar as representações de termos linguísticos e regras.

Independentemente da estratégia de desenvolvimento utilizada, a modelagem *fuzzy* tem a sua função bem consolidada em previsão de séries temporais. Na literatura, é possível observar aplicações em previsão da velocidade do vento [107, 108], problemas financeiros [109, 110].

### 3.4.2

#### Modelo AutoMFIS

Até o momento foram apresentados algoritmos mais populares de síntese automática de SIF que implementam meta-heurísticas evolutivas para otimização dos modelos ou que utilizam conceitos de aprendizado online dos dados. Nesta seção, será apresentado o funcionamento do modelo *Automatic Multiva-*

riate Fuzzy Inference System (AutoMFIS), voltado para a previsão multivariada de séries temporais [2]. Os conceitos apresentados neste modelo constituem a base para o modelo proposto, que será apresentado no capítulo seguinte.

O AutoMFIS pode ser dividido em três módulos principais, ilustrados na Figura 3.6, a saber: Definição do Dicionário Linguístico, Geração da Base de Regras e Geração dos Valores *crisp* da saída.

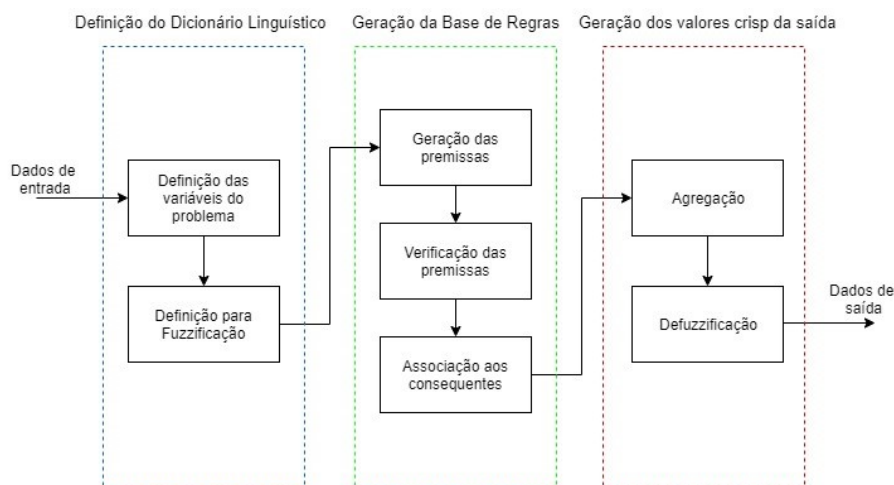


Figura 3.6: Arquitetura modular do AutoMFIS

O módulo de Definição do Dicionário Linguístico é caracterizado como a etapa de definição dos parâmetros para a fuzzificação dos dados. Nesta etapa são estabelecidas quantas variáveis serão consideradas para a composição do problema, assim como a quantidade de conjuntos *fuzzy* a serem utilizados para cada variável.

Devido à característica multivariada do problema, são necessárias técnicas de pré-processamento para selecionar as séries que são mais correlacionadas e, conseqüentemente, contribuir na tarefa de previsão, reduzindo a dimensionalidade do problema. Diferentes metodologias podem ser utilizadas neste processo, sendo o agrupamento hierárquico um dos mais conhecidos.

Outro aspecto avaliado neste módulo é a determinação dos conjuntos *fuzzy* para cada uma das variáveis, tanto na distribuição dos conjuntos quanto na quantidade de conjuntos necessários. Com o objetivo de preservar a interpretabilidade, a distribuição dos conjuntos *fuzzy* no universo de discurso deve ser ajustada de modo que não haja uma grande quantidade de sobreposição entre eles. Na literatura, dois métodos de posicionamentos conhecidos são a distribuição uniforme dos conjuntos e a distribuição usando informação de percentis, que estão disponíveis para utilização no sistema AutoMFIS. Como estas distribuições também são utilizadas no modelo e-AutoMFIS, o Capítulo 4 descreverá de forma mais detalhada as suas características.

A Geração de Base de Regras é o módulo responsável pela criação de todas as regras que compõem e descrevem o problema. De uma forma geral, este módulo é dividido em três etapas: geração dos antecedentes das regras (premissas), verificação de ativação das regras, e associação ao consequente da regra.

A geração dos antecedentes das regras é feita pelo método de geração semi-exaustiva. Esta técnica realiza, de forma iterativa, todas as combinações possíveis dos antecedentes, limitada a um número máximo de antecedentes por regra, o qual aumenta a cada ciclo, excluindo as combinações redundantes, conforme pode ser visto na Figura 3.7. Este processo continua até o tamanho máximo de antecedentes por regras estabelecido. É nesta etapa que o sistema seleciona as defasagens mais adequadas para a formação do termo antecedente da regra. Portanto, o sistema detecta de forma automática quais as defasagens, dentre aquelas consideradas na definição do problema, são as melhores para a geração da regra *fuzzy*.

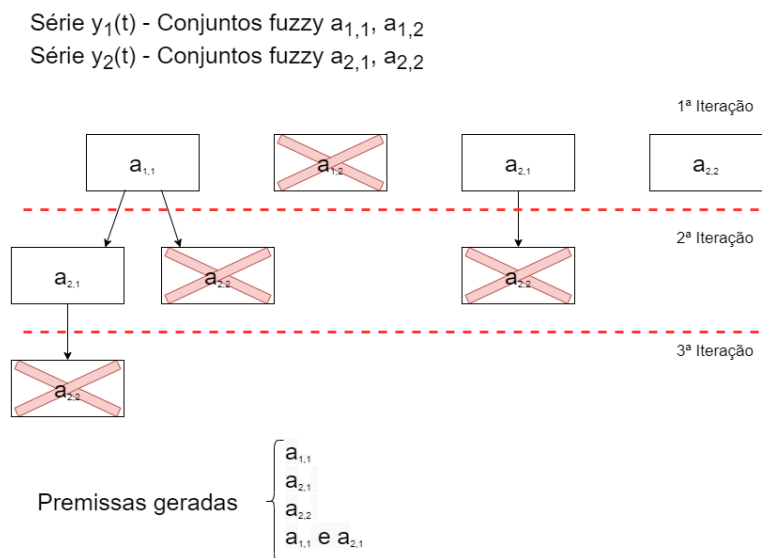


Figura 3.7: Exemplo de geração semi-exaustiva

No exemplo da Figura 3.7, a primeira iteração avalia a ativação média das premissas de acordo com uma métrica de avaliação (que será detalhada na Seção 4.2.2.1). Caso esta métrica seja menor do que o valor mínimo (parâmetro definido pelo usuário), a premissa é descartada e, conseqüentemente. Nas seguintes iterações do processo de geração semi-exaustiva, os termos que não foram eliminados são recombinados para avaliar o crescimento das premissas. Caso esta nova premissa tenha uma métrica de ativação menor do que o limite estabelecido, ela é descartada e o crescimento é interrompido.

A geração semi-exaustiva pode produzir uma base de regras muito grande, com algumas regras pouco (ou nunca) ativadas. Para possibilitar a

redução do conjunto de regras, foi estabelecido um critério de corte de regras baseada na métrica de cardinalidade (que será detalhada no Capítulo 4), que avalia o grau de ativação médio de cada uma das premissas. Caso a premissa da regra tenha uma cardinalidade menor do que a cardinalidade mínima (parâmetro do modelo AutoMFIS), ela é eliminada da base de regras. Deste modo, as regras mais expressivas continuam presentes, com uma base de menor tamanho.

A última etapa do processo de Geração de Base de Regras é a associação, composta da escolha do consequente mais apropriado à premissa para a composição das regras. Neste passo, são avaliadas as premissas das regras e os consequentes possíveis através da métrica de similaridade, visando obter uma maior compatibilidade entre o par premissa-consequente. Como estas métricas são consideradas para o modelo e-AutoMFIS, o próximo Capítulo descreverá em detalhes a sua formulação matemática.

O último módulo do AutoMFIS é a Geração dos valores precisos de saída. Essencialmente, o módulo é dividido em duas etapas: agregação e defuzzificação. A etapa de agregação é responsável por agregar as ativações das regras que possuem o mesmo consequente e premissas diferentes, utilizando o método de agregação ponderada. Este método é importante para que sejam dados pesos distintos às ativações das regras, de modo a inibir regras com premissas que não contribuam com maior acurácia a previsão desta série. A defuzzificação é realizada pelo método da altura, modificado para que o valor defuzzificado de cada série também considere os conjuntos *fuzzy* de todas as outras séries do problema.

Por se tratar de um modelo *fuzzy*, uma das grandes vantagens é a capacidade de interpretabilidade do modelo. A base de regras, e as regras ativadas para gerar um valor de saída, podem ser visualizadas a qualquer instante. Por se tratar de um modelo que contempla a utilização de múltiplas séries para previsão, outra característica importante é a possibilidade de melhorar a previsão pela inserção de características provenientes de outras séries que tenham alguma correlação com a previsão.

Nos dias atuais, é possível observar uma tendência no aumento do volume de bases de dados, característica bastante proveitosa para o treinamento de modelos de ML e DL. Para a tarefa de previsão de séries temporais, uma base de dados maior representa um maior horizonte de observação e, conseqüentemente, a possibilidade de ter uma visão ampla de sua dinâmica, o que contribui para que o modelo aprenda as relações temporais existentes. Contudo, nem todos os modelos se beneficiam com um aumento da quantidade de registros e variáveis, o que pode inclusive inviabilizar computacionalmente o seu uso.

No estado atual, o AutoMFIS possui dificuldades em lidar com bases de dados de alta dimensionalidade, conforme comentado anteriormente. O aumento na quantidade de variáveis, de defasagens e de seus correspondentes conjuntos *fuzzy* influenciam diretamente na dimensionalidade do problema de geração semi-exaustiva de regras. Conseqüentemente, a exploração de relações temporais mais longas acaba comprometida em virtude da viabilidade computacional, como é relatado em [2]. Além disso, o aprendizado deste modelo é global (i.e. as regras são selecionadas observando a ativação na totalidade do conjunto de treinamento), o que dificulta a extração de conhecimento local.

As limitações encontradas no modelo AutoMFIS motivaram o desenvolvimento de uma nova arquitetura para previsão de séries temporais usando sistemas de inferência *fuzzy*. Neste capítulo, é apresentado o modelo e-AutoMFIS, extensão do modelo de previsão multivariada AutoMFIS para problemas com alta dimensionalidade. O capítulo está organizado em três seções principais. A primeira seção detalha as motivações que serviram de base para a idealização do modelo proposto. Em seguida, é exibida detalhadamente a arquitetura do e-AutoMFIS e o seu funcionamento. Por fim, a terceira seção reúne as considerações finais sobre o modelo, além de reunir informações sobre parâmetros e métodos do sistema de inferência *fuzzy*.

## 4.1

### Motivação para a extensão do modelo AutoMFIS

Com a maior capacidade de obtenção e armazenamento de dados, é possível se dispor de mais dados e variáveis que possam ser relevantes para a solução de um problema. Por mais que o modelo AutoMFIS seja um modelo relevante para a previsão de séries multivariadas, algumas características do modelo tornam a sua aplicação limitada em certos problemas que exigem maior poder computacional. Estes fatores motivaram a investigação de modificações na arquitetura, de modo a tornar possível a utilização de um Sistema de Inferência *fuzzy* em problemas de previsão de séries temporais de maior complexidade, seja na quantidade de registros ou no número de variáveis.

A solução para estas limitações foi a implementação de um comitê de previsores AutoMFIS. Esta abordagem é muito conhecida na literatura, e potencialmente supre as necessidades do seu modelo original. Nesta seção, serão explicadas as características que motivaram a implementação da estratégia de aprendizado distribuído.

#### 4.1.1

##### Necessidade e vantagem de um modelo *ensemble*

Quando se deseja tomar uma ação de grande impacto, é comum que a decisão seja discutida por um comitê de especialistas, onde diferentes visões são compartilhadas e, geralmente, se chega a um consenso sobre a ação a ser tomada. Em geral, estas decisões são mais confiáveis e menos enviesadas, se comparadas com uma decisão monocrática. Esta ideia é uma analogia que serve de inspiração para a implementação da abordagem *Ensemble*.

Muitos problemas reais são complexos demais para que apenas um único modelo seja capaz de resolvê-los de forma precisa e robusta. Uma alternativa é considerar um conjunto de modelos especialistas. A técnica de *Ensemble* consiste na criação e no aprendizado de um conjunto de modelos, que são combinados para gerar uma resposta única. Dependendo do tipo do problema, existem diferentes abordagens para a composição do resultado. Em problemas de classificação, por exemplo, é comum o uso de votação dos membros do comitê, onde a classe escolhida é aquela que recebeu mais votos [111]. Em problemas de regressão, a composição por média aritmética é a mais comum das abordagens.

Especificamente em SIF, existem outras estratégias de aprendizado distribuído, de modo a tornar a extração de conhecimento menos complexa. O modelo e-AutoMFIS utiliza a abordagem *Ensemble* dividindo a tarefa de modelagem em diferentes subproblemas para facilitar a seleção de regras mais



relevantes. Desta forma, a essência do comitê de previsores do e-AutoMFIS não está presente na combinação de resultados, mas na agregação das regras selecionadas durante a fase de treinamento.

#### 4.1.2

##### Usabilidade em grandes bases de dados

Com o desenvolvimento de um modelo baseado em comitê de previsores, o e-AutoMFIS torna-se adaptável para diferentes tipos de problemas, inclusive aqueles com um grande volume de dados. Conforme será explicado posteriormente, a subamostragem aleatória reduz a dimensionalidade da base de dados a ser utilizada pelo SIF e, conseqüentemente, o treinamento de cada um dos membros do comitê é facilitado. Em outras palavras, o *Ensemble* transforma um problema de maior dimensionalidade em diversos problemas menores.

A redução da dimensionalidade é um fator importante a ser considerado no modelo e-AutoMFIS. Por incorporar algumas características originais do modelo AutoMFIS, que representa os membros do comitê de previsores, a geração de regras *fuzzy* continua a ser realizada pelo método semi-exaustivo. Simplificando esta abordagem ao considerar que todas as antecedentes possíveis não sejam excluídos durante o processo, a quantidade de conjunções geradas pode ser calculada por uma combinação simples.

É possível observar que o aumento da dimensionalidade pode prejudicar o aprendizado do modelo, tornando-o em alguns casos até impraticável. Para que isto não seja um problema, tipicamente devem ser feitos ajustes nos parâmetros do modelo AutoMFIS, de modo que a geração de regras não inviabilize computacionalmente o seu uso. Com o e-AutoMFIS, este problema é resolvido ao se realizarem sucessivas subamostragens do problema original, tornando a quantidade de variáveis menor para o treinamento do AutoMFIS.

Outra vantagem do modelo e-AutoMFIS, se comparado ao modelo original, é a maior capacidade de lidar com bases de dados contendo uma maior quantidade de valores históricos. Em um problema com uma quantidade alta de registros, é difícil para um modelo único captar todas as relações que auxiliem na previsão de forma satisfatória. Ao dividir uma série extensa em diferentes partes, o sistema tem uma maior facilidade de compreender os comportamentos da dinâmica das séries temporais. Na próxima seção, é descrita de forma detalhada a estrutura do e-AutoMFIS, mostrando-se por meio de diagrama de blocos, o funcionamento do modelo.

## 4.2

### Arquitetura do modelo e-AutoMFIS

O e-AutoMFIS é um modelo de previsão de séries temporais baseado na técnica de criação de múltiplos previsores. O modelo proposto incorpora o conceito básico do AutoMFIS e a técnica de subamostragem aleatória de base de dados para o desenvolvimento de sua arquitetura, conforme mostrado na Figura 4.1.

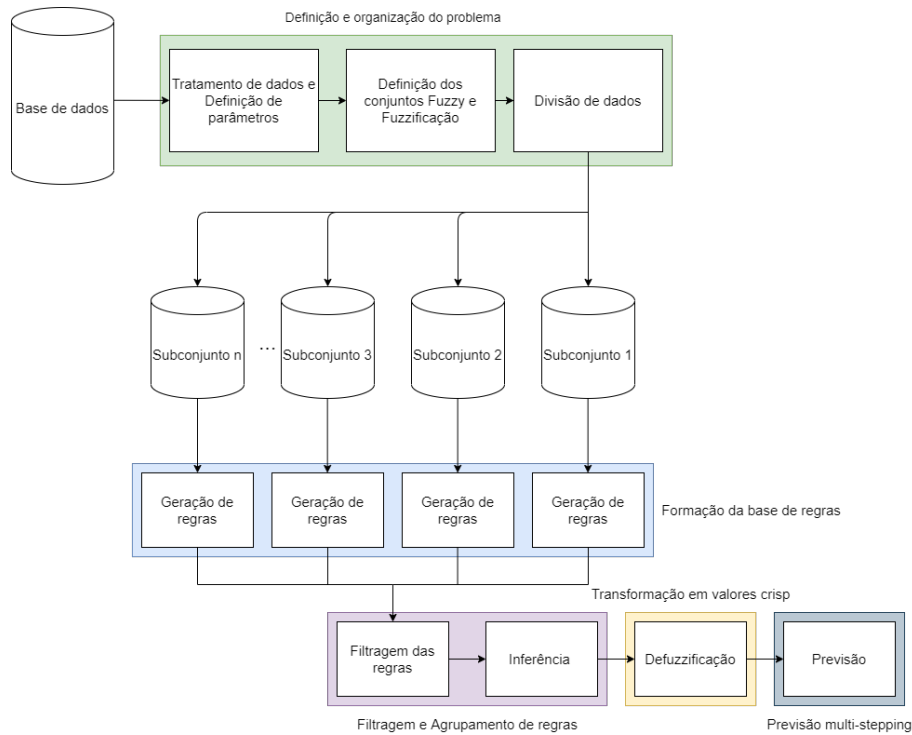


Figura 4.1: Arquitetura simplificada do modelo e-AutoMFIS

Para um melhor entendimento do funcionamento do e-AutoMFIS, consideram-se quatro módulos principais, a saber:

1. **Definição e organização do problema:** responsável pela subamostragem da base de dados a ser usada no treinamento de um membro do comitê de previsores;
2. **Formação da base de regras:** responsável por extrair a informação do subconjunto de dados e fornecer regras linguísticas;
3. **Filtragem e agregação de regras:** encarregado de selecionar as regras mais relevantes do problema e realizar a inferência *fuzzy*;
4. **Defuzzificação:** relação do conjunto *fuzzy* de saída em valores pontuais, retornando ao domínio real do problema;

5. **Previsão *multi-step***: designado para o processo de previsão de séries temporais multivariadas.

Cada um destes tópicos possui uma subseção específica, mostrando com maior detalhes a sua importância, objetivos e diferentes métodos existentes. A arquitetura apresentada na Figura 4.1 é uma simplificação do funcionamento real do e-AutoMFIS, de modo a tornar a explicação modular mais compreensível. No final do capítulo é retratado o funcionamento completo do modelo proposto.

A seguir, os módulos básicos do e-AutoMFIS serão caracterizados, detalhando-se o seu funcionamento e se discutindo a importância dos métodos escolhidos. Ao final de cada subseção, também serão abordados os aspectos da interpretabilidade que podem ser afetados com o ajuste da configuração.

#### 4.2.1

##### Definição e organização do problema

A definição e organização do problema é o ponto inicial do modelo e-AutoMFIS, cujo objetivo maior é a preparação dos dados para uma extração apropriada de regras. Também é nesta etapa que ocorre a geração do dicionário linguístico do modelo, com o ajuste da quantidade de termos linguísticos e suas respectivas funções de pertinência.

Em problemas de previsão de séries temporais é usual que a base de dados seja dividida em um conjunto de treinamento, para o ajuste do modelo (*in-sample*), e um conjunto de teste (*out-sample*), utilizado para avaliar a capacidade de generalização. No geral, o conjunto *in-sample* pode ser segmentado em conjunto de treinamento e validação, prática comum quando se deseja avaliar o desempenho do modelo durante o seu ajuste. Para o modelo proposto, como cada membro do comitê precisa ser avaliado para a seleção de regras, a divisão entre conjunto de treinamento, validação e teste é a mais apropriada. Esta repartição é feita de acordo com a ordem cronológica da série, conforme mostra a Figura 4.2, onde os registros mais antigos são utilizados para a modelagem do sistema e os últimos valores, para a avaliação do processo de previsão. Além disso, todos os processos de construção do modelo (como geração dos termos linguísticos e formulação das regras) são feitos com base no conjunto de treinamento, com exceção da etapa de filtragem que utiliza o resultado do conjunto de validação para a seleção das regras. Portanto, a menos que seja explicitado no texto, os procedimentos do modelo são realizados com base no conjunto de treinamento.

Antes da implementação de qualquer modelo existente, a primeira etapa de um projeto de ML consiste na etapa de pré-processamento de dados. Em-

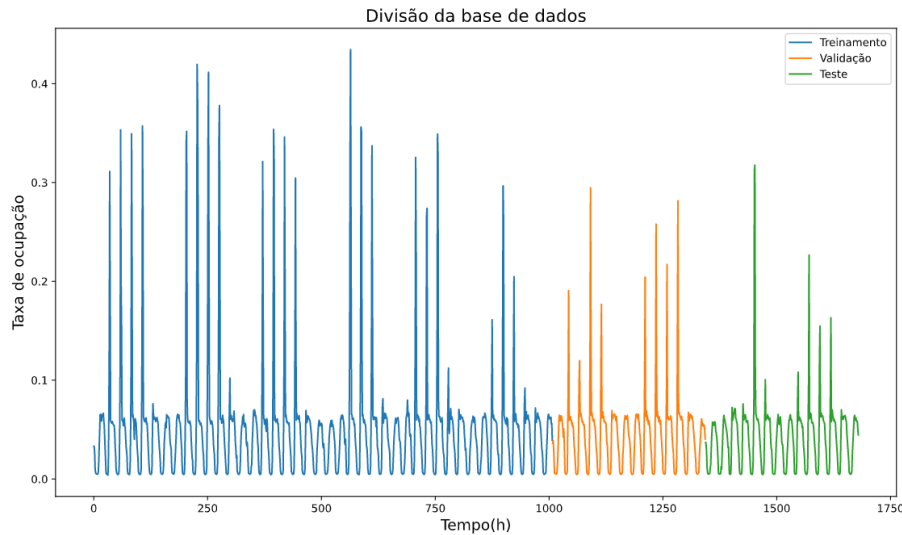


Figura 4.2: Exemplo de divisão de dados entre treinamento, validação e teste em uma aplicação de previsão do volume de tráfego

bora não seja uma atribuição direta do e-AutoMFIS, é um pré-requisito fundamental para que o aprendizado, que é feito de forma automática, tenha um desempenho satisfatório. Especificamente em previsão de séries temporais, a seleção de variáveis e a definição das defasagens são importantes e influenciam diretamente na dimensionalidade do problema e, consequentemente, na interpretabilidade do modelo.

Em previsão de séries temporais, a seleção de variáveis visa quantificar as similaridades entre séries e suas correlações, sejam lineares ou não-lineares. Esta metodologia ajuda a eliminar séries não relacionadas, diminuindo a complexidade do problema e facilitando o aprendizado do modelo e a extração de conhecimento. Na literatura, existe uma grande variedade de técnicas, como correlação de Pearson [132] e seleção de variáveis (ou defasagens) usando métodos iterativos, como RFE (*Recursive Feature Elimination*) [112].

Outro fator importante nesta etapa é a definição da defasagem máxima. Este parâmetro define a quantidade de variáveis defasadas para cada uma das séries selecionadas. Em geral, as técnicas de autocorrelação e correlação cruzada, discutidas no Capítulo 2, auxiliam na escolha inicial de defasagens, com base na captação de relações lineares. Apesar da escolha manual da defasagem máxima, as defasagens efetivamente usadas pelo modelo para cada série temporal são escolhidas automaticamente durante o processo de aprendizado.

O emprego das técnicas citadas para a remoção de características não-estacionárias da série também é importante para o bom uso do e-AutoMFIS. Como os conjuntos *fuzzy* são estáticos (i.e. não mudam de formato em função

do tempo), a presença do termo de tendência pode danificar o significado do conjunto, além de prejudicar o processo de previsão. Portanto, as estratégias citadas no Capítulo 2 para tornar as séries estacionárias são fundamentais na preparação dos dados.

A próxima etapa do e-AutoMFIS é a criação de um dicionário de termos linguísticos. Esta etapa é responsável por estabelecer o universo de discurso, a quantidade de conjuntos *fuzzy* e suas funções de pertinência. A interpretabilidade também é um aspecto importante de ser notado, pois a escolha desses parâmetros pode influenciar diretamente em quão explicável o modelo é. Deste modo, os valores possíveis e métodos foram escolhidos para atender ao máximo de critérios de interpretabilidade, sem prejudicar a acurácia do modelo.

No modelo proposto, as funções de pertinência são geradas automaticamente, e informações como o universo de discurso e as formas dos conjuntos *fuzzy* são obtidos com base nos dados. Esta etapa é fundamental quando são utilizados métodos de síntese automática, visto que as funções de pertinência criadas podem ser especificadas para uma melhor acurácia, mas pouco interpretáveis. Dependendo de como é formulado o problema de criação do dicionário linguístico, é possível que os conjuntos *fuzzy* gerados sejam mais confusos, prejudicando as condições de distinguibilidade e complementaridade. A Figura 4.3 mostra um exemplo sintético de conjuntos *fuzzy* para uma variável qualquer em que tanto o suporte quanto a posição do valor de pico das funções de pertinência são modificados. Neste caso, é possível perceber que a utilização destes parâmetros para a otimização das funções de pertinência podem resultar em modelos que não satisfazem os critérios de interpretabilidade citados anteriormente.

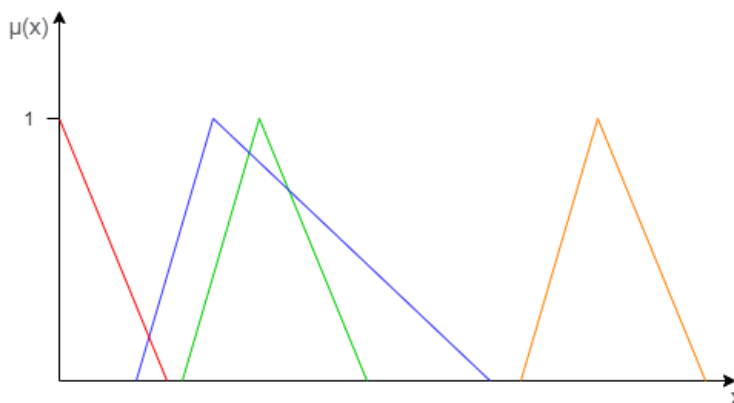


Figura 4.3: Conjuntos *fuzzy* com baixa interpretabilidade no domínio de  $X$

Uma alternativa para a geração de conjuntos *fuzzy* é a partição *fuzzy* forte (*Strong fuzzy Partition*). Matematicamente, dada uma coleção de  $n$

conjuntos *fuzzy* com pertinência  $\mu_{A_n}(x)$ , para qualquer valor numérico  $x$  dentro do universo de discurso  $X$ , a soma das pertinências de todos os conjuntos deve ser igual a 1 (Eq. 4-1).

$$\sum_{i=1}^n \mu_{A_n}(x) = 1, \forall x \in X \quad (4-1)$$

Esta metodologia é muito útil quando se deseja obter um SIF interpretável, visto que critérios de interpretabilidade como cobertura e distinguibilidade são satisfeitos por definição [113]. Além disso, a partição *fuzzy* forte é mais simples de ser ajustada devido à quantidade menor de parâmetros.

Para o e-AutoMFIS, que considera o aspecto de interpretabilidade, foi adotada a ideia de partição *fuzzy* forte para o ajuste das funções de pertinência. Por simplicidade, todas as funções de pertinência possuem um formato triangular, com exceção dos conjuntos *fuzzy* mais ao extremo, que possuem um formato trapezoidal, para que valores fora dos limites estabelecidos no problema sejam representados por algum termo linguístico (i.e. possuir sempre algum conjunto ativado).

No total, o modelo possui três diferentes abordagens. De qualquer forma, os modelos produzidos compartilham de uma mesma representação do problema, exemplificado pela Figura 4.4. Os parâmetros  $c_i$  a ser determinados pelo modelo representam os valores que possuem a máxima pertinência; os parâmetros de base da função de pertinência triangular definidos pelos valores máximos dos conjuntos adjacentes  $c_{i-1}$  e  $c_{i+1}$ . Para aqueles conjuntos *fuzzy* com o formato trapezoidal, as extremidades são abertas e, portanto, não há parâmetros adicionais a serem especificados.

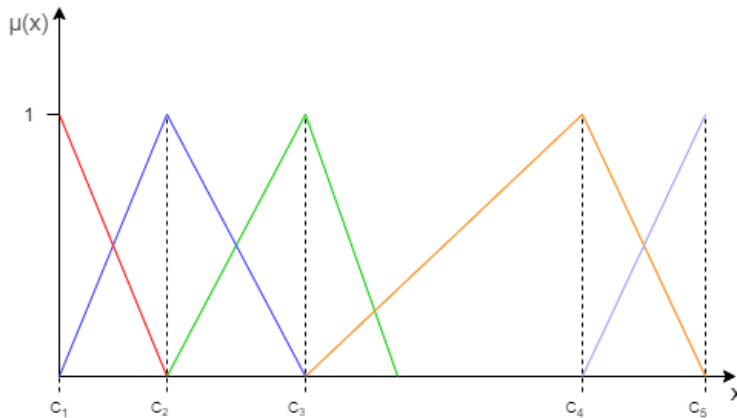


Figura 4.4: Representação da partição *fuzzy* forte

A forma mais simples de particionamento consiste na ordenação uniforme dos conjuntos *fuzzy* em suportes  $\Delta$  ao longo do universo de discurso, conforme

mostra a Figura 4.5. Esta técnica não faz uso de todos os dados disponíveis na base, mas apenas dos valores extremos para determinar os limites de cada variável. Por mais que a implementação seja rápida, uma desvantagem nítida é a incapacidade de ajuste dos conjuntos em função da distribuição dos dados, o que poderia resultar em conjuntos pouco representativos ou que agrupem valores que poderiam ser divididos em diferentes termos linguísticos.

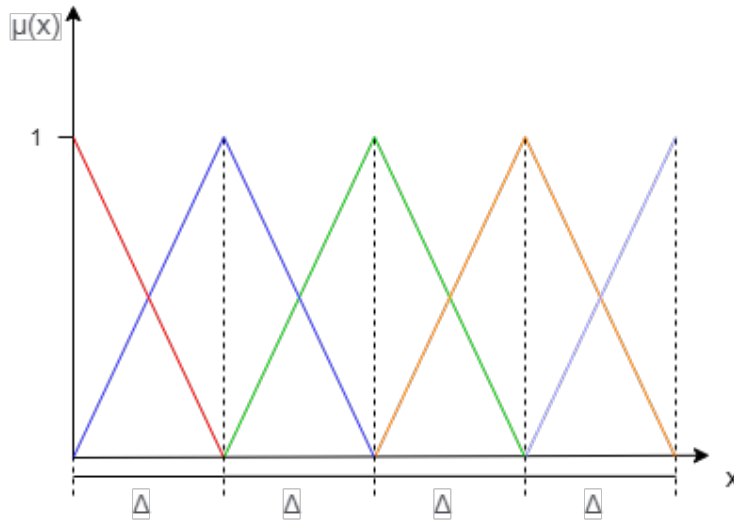


Figura 4.5: Conjuntos *fuzzy* gerados via particionamento uniforme

Em alguns casos, os dados não estão distribuídos de forma homogênea ao longo do universo de discurso e a partição uniforme se torna inadequada. Neste caso, é necessário utilizar outras técnicas que captem a distribuição de dados e conciliem a representação dos conjuntos *fuzzy* com o histograma das variáveis. Um método clássico e muito utilizado na síntese de sistemas *fuzzy* é adequar os parâmetros  $c_i$  em função dos percentis da distribuição dos valores, conforme mostra a Figura 4.6. Este método, conhecido como Tukey, é muito útil quando as variáveis têm uma distribuição normal.

Existem casos em que os dados não obedecem a uma distribuição uniforme ou normal, e a utilização dos métodos citados anteriormente pode não ser a melhor opção. Para contornar este problema, neste trabalho foi proposto um método de agrupamento para determinar os parâmetros a ser ajustados. Este método de particionamento encontra, de forma iterativa, a divisão entre grupos que possuem máxima similaridade intra-cluster e menor variância inter-cluster com a quantidade de agrupamentos previamente definida pelo usuário. Os valores centrais destes  $n$  grupos representam os valores de pico  $c_i$  de cada um dos  $n$  conjuntos *fuzzy*, sendo bastante útil nos problemas em geral. Para exemplificar a finalidade da partição pela técnica de agrupamento, a Figura

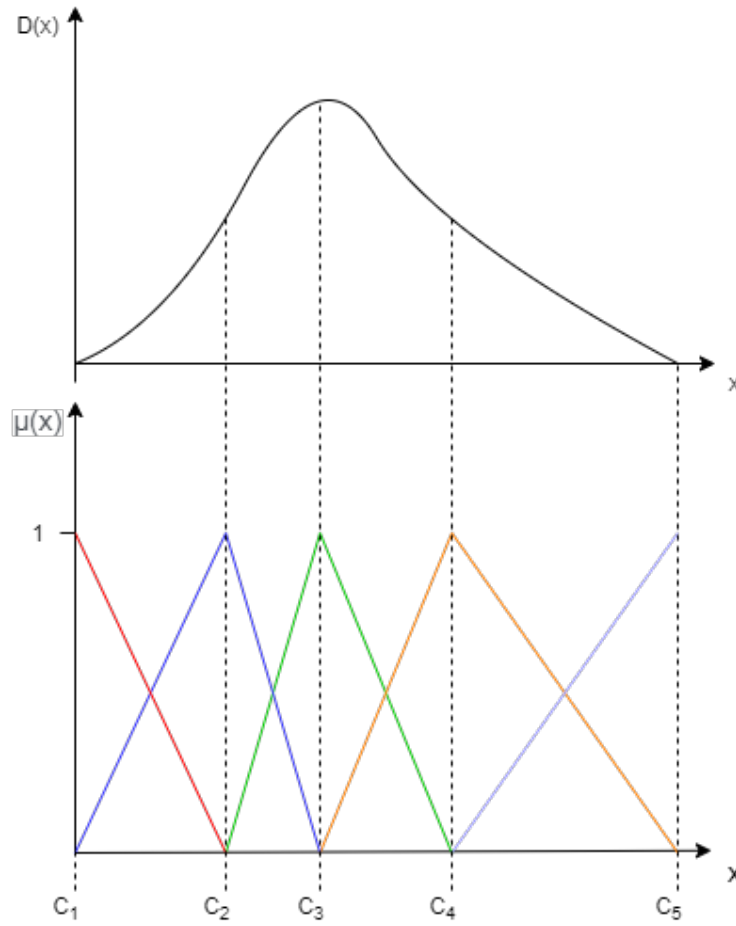


Figura 4.6: Conjuntos *fuzzy* gerados pelo método baseado em percentis

4.7 ilustra uma distribuição sintética de dados e os conjuntos gerados por cada um dos três métodos de partição *fuzzy*.

Cada uma das três abordagens de geração de conjuntos *fuzzy* possui sua relevância, e portanto a escolha do método é baseada principalmente na distribuição dos dados do problema. Esta seleção reforça a necessidade de exploração e análise de dados, com o objetivo de selecionar os métodos e parâmetros mais adequados.

Neste módulo, a interpretabilidade é importante de ser analisada, visto que todos os critérios com o nível de abstração de conjuntos *fuzzy* dependem da construção do dicionário linguístico. A escolha de estratégias nesta primeira etapa, portanto, é norteadada para obter uma melhor relação entre acurácia e interpretabilidade. As seguintes considerações podem ser feitas quanto a esses critérios:

- **Número de conjuntos *fuzzy*:** A quantidade de termos linguísticos é um parâmetro definido pelo usuário, e portanto a interpretabilidade é controlada de acordo com o problema. Geralmente o ajuste do número



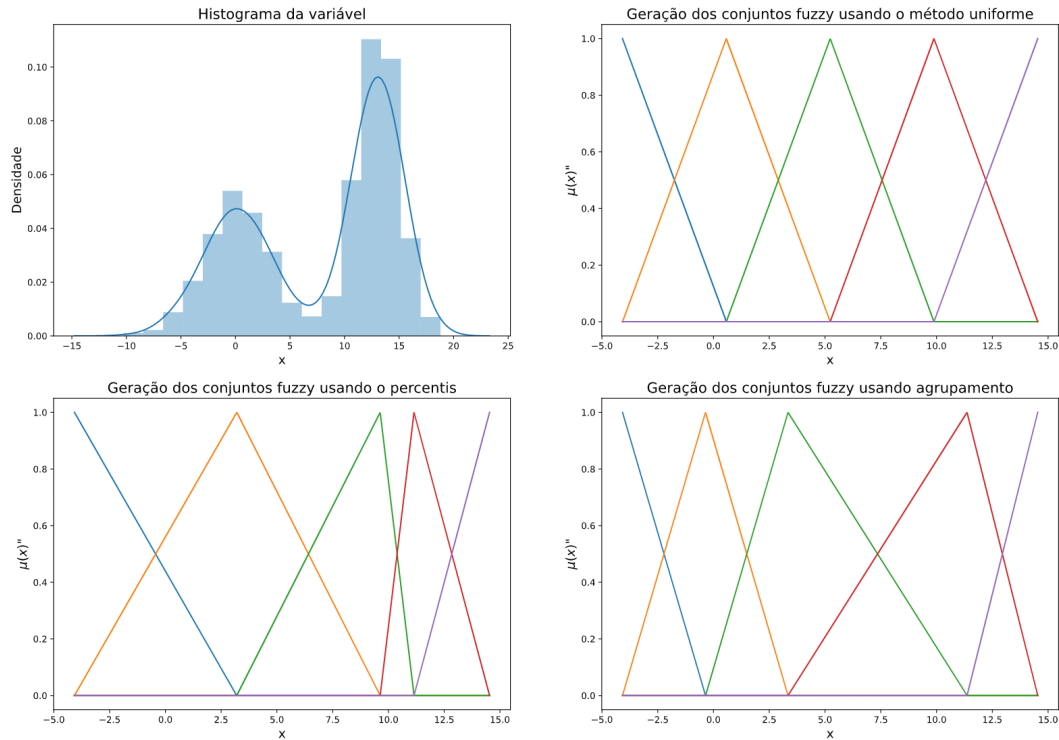


Figura 4.7: Comparação de geração dos conjuntos *fuzzy* usando as metodologias disponíveis para o e-AutoMFIS

de conjuntos *fuzzy* é limitado a nove, satisfazendo a heurística de Miller [83];

- **Número de variáveis:** A quantidade de variáveis define as possíveis relações que podem ser extraídas na etapa de formulação das regras. Este critério é controlado através de técnicas de pré-processamento, definindo as variáveis potencialmente mais importantes para o problema;
- **Cobertura:** A cobertura é garantida para todos os problemas por conta da modelagem de determinação das funções de pertinência via partição forte *fuzzy*. Com o uso de funções trapezoidais, até os valores que estejam fora do escopo inicial do universo de discurso (o que pode ser comum em problemas de previsão de séries temporais) são contemplados com algum nível de ativação;
- **Distinguibilidade:** A distinguibilidade também é um critério que é satisfeito para qualquer problema usando o modelo e-AutoMFIS. Com a partição *fuzzy* utilizada, os conjuntos *fuzzy* possuem formatos bem definidos. Isto inviabiliza a formação de conjuntos que prejudiquem a interpretabilidade, como uma sobreposição muito grande de funções de pertinência ou total separação em relação a outros conjuntos;
- **Complementaridade:** Por fim, a complementaridade também é satisfeita com o uso da partição *fuzzy* forte. Ao definir o problema de espe-

cificação das funções de pertinência com formatos triangulares e trapezoidais, é garantido que para qualquer valor no universo de discurso, a soma dos graus de ativação é sempre igual a 1.

#### 4.2.2

##### Formação da base de regras

Com os termos linguísticos gerados, o próximo passo do e-AutoMFIS é a formação da base de regras. Esta fase é o âmago do modelo, pois contém o conhecimento extraído do problema. Além disso, a geração de regras é o principal aspecto a ser observado sob a ótica da interpretabilidade e um fator importante para determinar a acurácia do modelo.

A formação automática de regras consiste na ligação dos termos antecedentes e respectivos consequentes, e é tópico de grande interesse em SIFs. Existem diferentes abordagens conhecidas, como a geração de regras completas, onde tanto o antecedente quanto o consequente da regra são ajustados simultaneamente. Outra possível metodologia utiliza a geração de termos antecedentes das regras e, numa etapa seguinte, assimila os termos consequentes mais adequados.

O e-AutoMFIS incorpora a mesma metodologia usada no modelo antecessor para a etapa de formação de regras. Esta abordagem consiste em duas etapas principais, a saber: a formulação das premissas e a associação dos consequentes. Estes dois estágios são descritos a seguir, enfatizando-se as condições e critérios em termos de interpretabilidade, seus parâmetros e diferentes métodos.

##### 4.2.2.1

##### Formulação

A formulação é a etapa responsável por elaborar as premissas das regras geradas. A premissa é definida pela combinação de funções de pertinência (mostrada na Eq. 4-2), traduzindo matematicamente a expressão "se  $x_1$  é  $A_{j,1}$  e  $x_2$  é  $A_{j,2}$  e ... e  $x_i$  é  $A_{j,3}$ ".

$$\mu_{A_p} = \mu_{A_{j,1}}(x_1) * \mu_{A_{j,2}}(x_2) * \dots * \mu_{A_{j,3}}(x_i) \quad (4-2)$$

onde  $\mu_{A_p}$  é o grau de pertinência da premissa  $p$  e  $\mu_{A_{j,k}}(x_i)$  é o grau de ativação do termo linguístico associado ao conjunto fuzzy  $k$  da  $i$ -ésima variável e da série  $j \in 1, 2, \dots, s$  a ser prevista. Cada premissa  $p$  tem sua pertinência  $\mu_{A_p}$  calculada através da aplicação de um operador t-norma  $*$  entre os antecedentes

ativados pela regra. É possível observar que a mudança do operador t-norma não tem grande influência no resultado final. Portanto, neste trabalho emprega-se exclusivamente o produto algébrico.

Como o e-AutoMFIS consiste na construção de um comitê de previsores, cada membro do comitê de previsores é responsável por criar a base de regras a partir de seu no seu subconjunto de dados. Além disso, por se tratar de um modelo de previsão de séries multivariadas, cada membro do comitê gera regras para cada uma das  $s$  séries envolvidas.

Para a etapa de formulação utiliza-se a técnica de geração semi-exaustiva de regras, também empregada no modelo AutoMFIS [2]. Este processo consiste na procura iterativa de combinação entre os termos linguísticos. A cada iteração, o método de geração semi-exaustiva avalia as possíveis premissas com a mesma quantidade de antecedentes, começando com regras que possuem apenas um antecedente até chegar ao tamanho máximo de antecedentes, que é um parâmetro escolhido pelo usuário.

O processo semi-exaustivo começa com a avaliação de todos os termos linguísticos das variáveis selecionadas pelo subconjunto, compondo inicialmente regras com apenas um antecedente. Estas premissas são avaliadas segundo um critério de ativação mínima e aquelas que não forem satisfeitas são eliminadas da próxima iteração.

Na próxima iteração, as premissas selecionadas na etapa anterior são combinadas com todos os termos linguísticos de variáveis diferentes das já existentes na regra, de modo a evitar inconsistência de regras (e.g. "Se **salário** é **alto** e **salário** é **baixo**"). Novamente, as premissas são avaliadas e aquelas que tiverem ativação menor do que um valor mínimo têm seu crescimento da premissa finalizado. Já as que tiverem as condições atendidas são inseridas no conjunto de premissas selecionadas e continuam no processo iterativo. Este ciclo continua até alcançar o tamanho máximo permitido por premissa, que é um parâmetro ajustável do modelo.

Conforme citado anteriormente, para que as premissas sejam escolhidas é necessário que o grau de ativação destas seja maior do que um valor mínimo, parâmetro escolhido pelo usuário. Neste trabalho, foram utilizados dois métodos para o cálculo da ativação:

- **Cardinalidade:** média do cálculo do grau de pertinência da premissa para todos os registros do conjunto de treinamento;
- **Cardinalidade não-nula:** média do cálculo do grau de pertinência da premissa para os registros cuja ativação seja não-nula.

Estas duas métricas possuem o mesmo objetivo, que é selecionar as premissas mais relevantes para o conjunto de treinamento. Contudo, cada uma tem sua utilidade em diferentes tipos de problema. A cardinalidade pressupõe que as premissas possuem uma representatividade tanto em frequência quanto em grau de ativação, beneficiando mais a constância de ativação do que apenas ativações altas e menos frequentes. Portanto, esta métrica acaba privilegiando a seleção de regras genéricas (i.e. muito frequentes mas com ativação razoavelmente baixa) e excluindo regras relevantes com ocorrências esporádicas, que podem ajudar a compreender melhor o comportamento de uma série temporal. Nesta métrica, as premissas menos extensas são favorecidas, visto que o aumento da quantidade de antecedentes geralmente implica em mais ativações nulas ao longo do conjunto de treinamento e, consequentemente, mais difíceis de serem selecionadas.

A cardinalidade não-nula é uma métrica que calcula a média do grau de pertinência das premissas das amostras que possuem ativação não-nula. Este cálculo de ativação, utilizado no modelo e-AutoMFIS, é capaz de selecionar regras menos frequentes, mas ainda assim importantes, servindo como uma métrica alternativa à cardinalidade. Em contrapartida, a cardinalidade não-nula pode beneficiar regras muito extensas e com uma frequência extremamente baixa (e.g. premissas que são ativadas uma única vez, mas com um grau de pertinência razoável, podem ser selecionadas). Para contornar este provável efeito colateral, também é utilizado um parâmetro de frequência mínima para que a etapa de formulação aceite regras menos frequentes, mas não aquelas que sejam ativadas pouquíssimas vezes.

#### 4.2.2.2 Associação

Com as premissas já selecionadas, é necessário encontrar o consequente que melhor se ajusta ao problema para a geração de regras consistentes. Este é o objetivo da etapa de associação, a última etapa da formação da base de regras.

No processo de associação, cada premissa irá gerar  $s$  regras referentes a  $s$  séries. Para cada variável  $j \in \{1, 2, \dots, s\}$  a ser prevista, é calculado o grau de pertinência  $\mu_{A_{j,i}}(y)$  da saída  $y$  do conjunto de treinamento para cada conjunto fuzzy  $i$ . Em seguida, é calculada a métrica de associação entre os dados de entrada e saída do conjunto de treinamento, utilizando-se seus respectivos graus de ativação. Por fim, o conjunto *fuzzy* que obtiver a maior compatibilidade é associado à premissa, formando uma regra completa e inserida à base de regras.

A maior complexidade na etapa de associação consiste na escolha da métrica apropriada para o problema. Dependendo de como é calculada a similaridade, podem ser geradas regras cujos consequentes não são pertinentes, prejudicando a acurácia do modelo. Para o modelo e-AutoMFIS, serão utilizadas duas métricas distintas: Grau de Confiança *fuzzy* (GCF) e Compatibilidade Média (CM).

A métrica GCF calcula a similaridade entre dois vetores, definida pela Eq. 4-3. Neste caso, é possível perceber que a equação é, por definição, o valor do cosseno entre estes vetores e portanto a análise do resultado é análoga à interpretação de ortogonalidade. Caso a premissa e o consequente sejam ortogonais, o GCF atinge o valor igual a 0 e, se forem totalmente compatíveis, a métrica tem o valor igual a 1.

$$GCF = \frac{\mu_{A_p}(X) \cdot \mu_{s,n}(Y_s)}{\|\mu_{A_p}(X)\| \times \|\mu_{s,n}(Y_s)\|} \quad (4-3)$$

onde  $\mu_{A_p}(x)$  é o grau de pertinência do antecedente  $p$  para todo o conjunto de treinamento  $x$ ;  $\|\mu_{A_p}(x)\|$  é a norma do vetor  $\mu_{A_p}(x)$ ; e  $\mu_{s,n}$  é o grau de pertinência da série  $s$  ao conjunto  $n$  para o conjunto de saída  $y$  desta série temporal. Em termos práticos, o cosseno da similaridade entre os graus de pertinência da premissa e o candidato a consequente avalia a compatibilidade existente por conta da ativação simultânea destes termos.

Alternativamente ao GCF, a compatibilidade média é uma métrica que busca encontrar o consequente mais adequado para cada premissa, calculada conforme mostra a Eq. 4-4. Para cada premissa obtida na etapa de formulação, são avaliadas as entradas  $X$  no conjunto de treinamento que resultem no grau de ativação  $\mu_{A_p}(X)$  diferente de 0. Também é computado o grau de pertinência  $\mu_{s,n}(Y_s)$  para cada conjunto *fuzzy* existente para a variável de saída. No fim, o termo linguístico escolhido como consequente será aquele que possuir a maior compatibilidade média.

$$CM = \frac{1}{N} \sum_{i=1}^n \mu_{A_p}(x_i) \mu_{s,n}(y_{i,s}), \quad (4-4)$$

onde  $\mu_{A_p}(x_i)$  é o grau de pertinência da  $i$ -ésima entrada para o antecedente  $P$ ;  $y_{i,s} \in Y$  | tal que  $\mu_{A_p}(X) \neq 0$ ; e  $\mu_{s,n}(y_{i,s})$  é o grau de pertinência de  $y_{i,s}$  ao conjunto *fuzzy*  $n$  da série  $s$ . Esta métrica é adequada para a maioria dos casos, visto que o maior interesse da etapa de associação é selecionar o melhor termo consequente para as premissas quando ativadas, e não necessariamente uma similaridade entre os vetores  $\mu_{A_p}(X)$  e  $\mu_{s,n}(Y_s)$ . Além disso, em problemas contendo premissas pouco ativadas (mas importantes), o CM tem uma vantagem em relação ao GCF. Este fato ocorre pois o CM desconsidera, em seu cálculo, a maioria dos registros que são irrelevantes para

a premissa, concentrando-se apenas nos registros que geram alguma ativação.

#### 4.2.2.3

##### Considerações sobre o módulo de formação de regras

A formulação e associação representam as etapas principais da construção da base de regras. Por se tratar de uma estratégia *Ensemble*, a geração das regras faz uso de uma subamostragem da base de dados, como mostrado na Figura 4.8. Portanto, este processo é realizado para cada um dos  $n$  subconjuntos, servindo como entrada para a etapa de filtragem e seleção de regras.

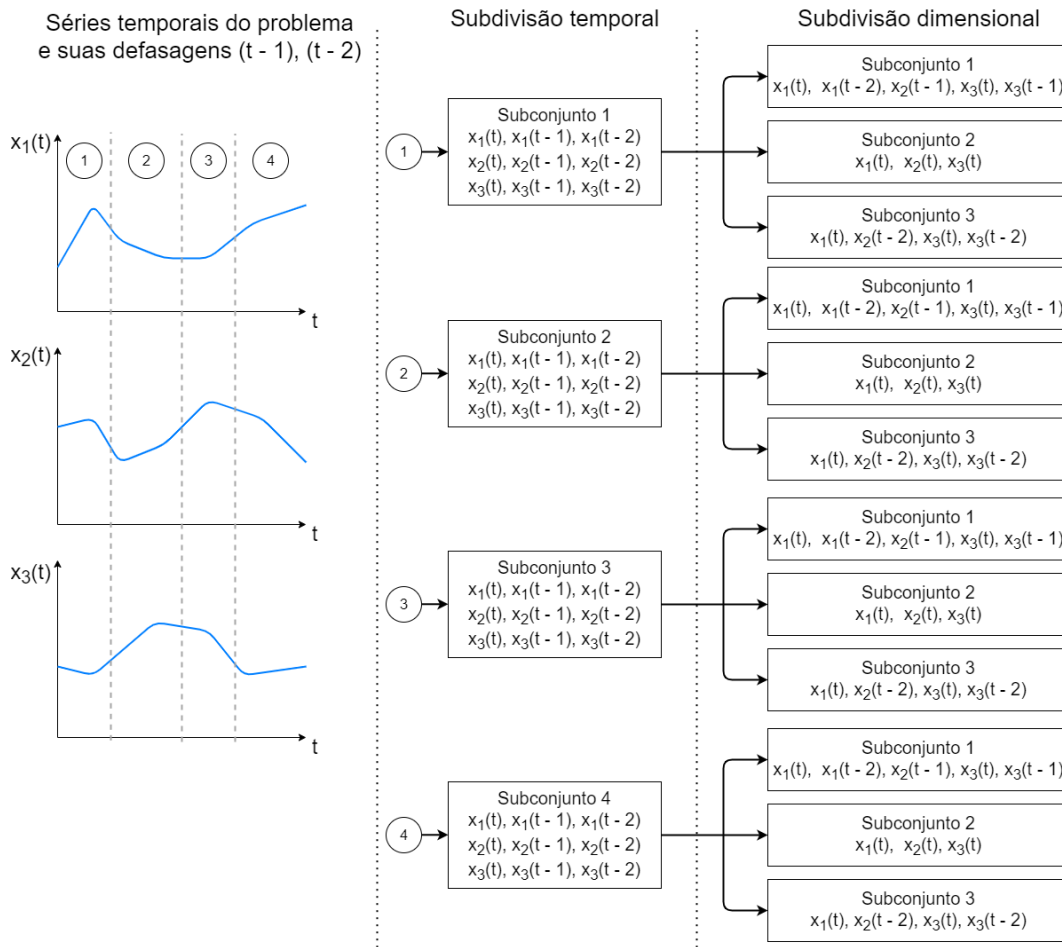


Figura 4.8: Exemplo de subamostragem proposta

Nesta técnica, a subamostragem é feita em duas etapas: temporal e dimensional. A subamostragem temporal divide o histórico da série em  $n$  intervalos de tempo. Em seguida, para cada amostra obtida, que possui variáveis dos problema (i.e. as séries e suas defasagens), é realizada uma subdivisão dimensional, que será utilizado para o treinamento de cada membro do comitê de previsão.

O uso desta técnica promove um aprendizado distribuído que potencialmente pode gerar resultados mais consistentes do que os proporcionados por um modelo único. Além das vantagens já citadas anteriormente, como a facilidade de aprendizado e a diversidade que é proporcionada, a técnica de *Emsemble* também possui benefícios práticos. A subamostragem de variáveis reduz a dimensionalidade do problema e consequentemente reduz o custo computacional do modelo, permitindo o aumento da exploração de parâmetros, como o aumento do número de defasagens. Esta abordagem distribuída também possibilita paralelizar o aprendizado do comitê de previsores, reduzindo o tempo necessário para o treinamento do modelo.

Outro aspecto importante a ser considerado neste módulo é o ajuste de parâmetros e a variedade de métodos existentes. Os parâmetros ajustáveis nesta etapa são:

- **Ativação mínima:** É o limiar de corte para a seleção da premissa gerada. Caso a ativação média da premissa seja maior do que este parâmetro, a regra é adicionada à base de potenciais regras e na próxima etapa da geração semi-exaustiva são avaliadas as inserções de mais um novo antecedente. Este critério é essencial para controlar a admissão de regras e o volume de regras aceitáveis, atuando para balancear a relação entre acurácia e interpretabilidade;
- **Premissa máxima:** Quantidade máxima de antecedentes por premissa. Este parâmetro também controla a extensão da base de regras, visto que o aumento do processo de geração de regras aumenta potencialmente a quantidade de regras selecionadas;
- **Método de formulação:** Parâmetro que define como deve ser avaliada a formulação de regras. Geralmente a escolha deste método é feita de acordo com o tipo de problema, de modo a melhorar a qualidade do conhecimento extraído. Os métodos possíveis são Cardinalidade e Cardinalidade não-nula
- **Método de associação:** Parâmetro que define o critério para a escolha do consequente apropriado para cada premissa. Os métodos possíveis são GCF e CM.

Outro aspecto importante a ser considerado é a avaliação dos critérios de interpretabilidade. Como este módulo está diretamente relacionado com o volume da base de regras, a interpretabilidade é analisada segundo as seguintes considerações:

- **Tamanho da base de regras:** O tamanho da base de regras é inicialmente determinado pela formulação das regras. Portanto, este aspecto deve ser levado em consideração para desenvolver um sistema que seja também interpretável. Para controlar este critério, é necessário ajustar o parâmetro de ativação mínima;
- **Comprimento dos antecedentes:** O comprimento dos antecedentes também é outro critério importante para análise nesta etapa, e pode ser controlado com a escolha de parâmetros apropriados. A quantidade máxima de antecedentes por regra, por exemplo, geralmente é limitado a no máximo 9 termos linguísticos, satisfazendo a heurística de Miller [83].

Em síntese, a interpretabilidade em relação à base de regras pode ser ajustada por meio de uma escolha apropriada dos parâmetros. É difícil encontrar empiricamente o ponto ótimo entre a complexidade e a interpretabilidade com base unicamente nas configurações especificadas até o momento. Para contornar este problema, geralmente o modelo é especificado com parâmetros que possam gerar uma base de regras maior, delegando à próxima etapa a função de redução de complexidade.

### 4.2.3

#### Filtragem e agregação de regras

O processo de geração de regras é realizado especificamente por cada um dos membros do comitê de previsores, que utilizam uma sub-amostragem da base de dados. Dependendo das variáveis selecionadas, um membro do comitê de previsores pode gerar regras que sejam boas para prever uma série, mas ruins para prever outra série. Além disso, pelo método de formulação e associação, existe a possibilidade de que regras geradas sejam sempre ativadas ao mesmo tempo ou que sejam conflitantes. Consequentemente, estes casos podem prejudicar a interpretabilidade do modelo ou até prejudicar a acurácia. Neste caso, é função da etapa de filtragem avaliar e reduzir a base de regras, tornando-a mais concisa.

O módulo de filtragem do e-AutoMFIS é dividido em duas etapas, conforme esquematizado na Figura 4.9. Para uma melhor compreensão, esta primeira fase de filtragem será denominada filtragem interna e a etapa posterior, filtragem externa.

#### 4.2.3.1

##### Filtragem Interna

Na filtragem interna, a atuação das técnicas é feita no nível da base de regras do membro do comitê de previsores e, portanto, a base de conhecimento



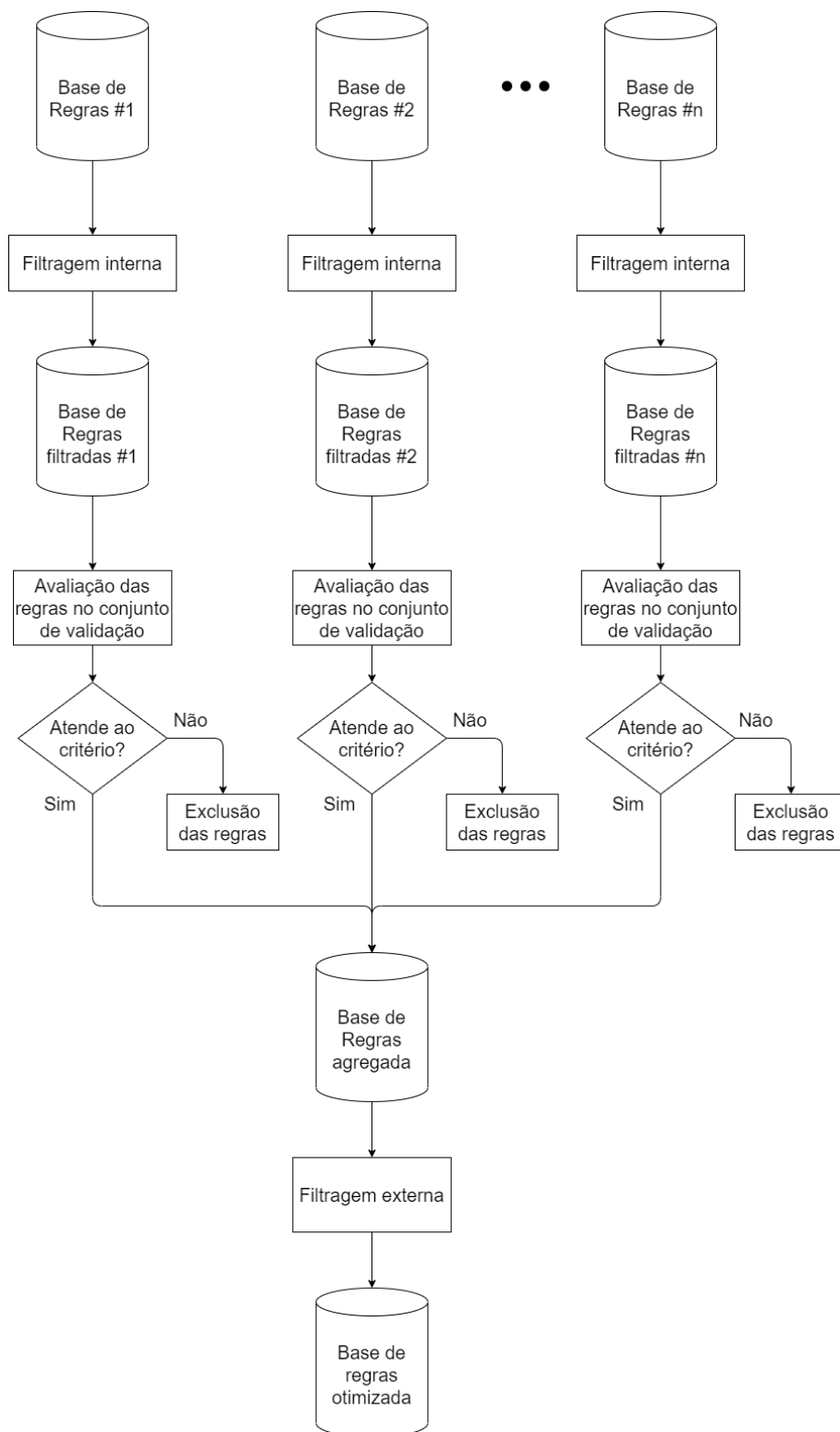


Figura 4.9: Diagrama de blocos da etapa de filtragem

não é ainda otimizada. Para realizar esta tarefa, a filtragem interna possui três objetivos principais:

1. **Remoção de regras semelhantes e conflitantes:** esta etapa localiza as regras potencialmente redundantes ou conflitantes e remove aquelas menos importantes. Para isso, são usadas técnicas para avaliar similaridade entre premissas e seus respectivos consequentes;
2. **Seleção das regras mais relevantes:** separa as regras mais importantes do problema. Para executar esta etapa, é implementada a otimização via mínimos quadrados restrita;
3. **Seleção de regras para composição da base de regras otimizada:** Após a otimização da base de regras, é necessário descobrir se as regras extraídas a partir do subconjunto resultam em uma previsão satisfatória, de forma a incorporá-las ao modelo final. Para isso, é averiguado, no conjunto de validação, o resultado da previsão multivariada e selecionado o conjunto de regras segundo a métrica de avaliação do problema (e.g. métricas de erro na previsão).

A remoção de regras redundantes e conflitantes utiliza um mesmo procedimento inicial, que é a busca de premissas ativadas simultaneamente. Para identificar as regras que são ativadas ao mesmo tempo utiliza-se uma métrica de similaridade entre os graus de pertinência, conforme mostra a Eq. 4-5. Para cada par de premissas, P e Q, calcula-se o seu respectivo grau de ativação  $\mu_{A_P}(x_i)$  e  $\mu_{A_Q}(x_i)$ , para cada vetor de entrada  $x_i$  presente no conjunto de treinamento, de forma a determinar a métrica de similaridade  $S_{P,Q}$ .

$$S_{P,Q} = \sum_{i=1}^n \frac{\min[\mu_{A_P}(x_i), \mu_{A_Q}(x_i)]}{\max[\mu_{A_P}(x_i), \mu_{A_Q}(x_i)]} \quad (4-5)$$

Caso  $S_{P,Q}$  seja maior do que uma constante  $S_{max}$  (parâmetro do modelo e-AutoMFIS), as premissas destas regras são consideradas similares e a regra com menor ativação média é descartada da base de regras original. Para diferenciar a métrica e determinar regras como conflitantes e similares, pode-se ainda determinar dois valores limítrofes  $S_{cmax}$  (conflitantes) e  $S_{smax}$  (similares), avaliando-se as ativações das premissas com o respectivo consequente e fazendo uso do o valor limite adequado. É difícil, empiricamente determinar o melhor valor para estas constantes. Assim, por simplicidade, utiliza-se  $S_{smax} = S_{cmax}$ .

O segundo procedimento de filtragem interna consiste na seleção das regras mais relevantes. Existem algumas abordagens úteis para a implementação

desta otimização, como o uso de Algoritmos Genéticos. Neste trabalho, é utilizada a mesma abordagem apresentada no modelo AutoMFIS [2], que é a otimização por Mínimos Quadrados Restrito (MQR). Este problema pode ser definido como o ajuste de pesos das premissas em função do grau de ativação das mesmas, com sua formulação apresentada na Eq. 4-6.

$$\min \left( \sum_{t=1}^T (\mu_{C_{s,n}}(y_{s,t}) - \sum_{p=1}^P w_p \mu_{A_p}(x_t)) \right) \quad (4-6)$$

$$s.a. \sum_{p=1}^P w_p = 1 \quad (4-7)$$

onde  $\mu_{C_{s,n}}(y_{s,t})$  representa o grau de ativação do consequente C da série  $s$  ao conjunto *fuzzy*  $n$ ;  $\mu_{A_p}(x_t)$  é o grau de pertinência da premissa  $p$  gerado pela entrada  $x_t$ ; e  $w_p$  é o seu respectivo peso, cujo valor varia entre 0 e 1. Utilizando-se um método de solução de problemas MQR, é possível obter os pesos  $w_p$  para cada um dos termos linguísticos das premissas das regras. Intuitivamente, estes pesos podem ser interpretados como o grau de importância das regras para a geração do resultado, onde regras irrelevantes possuem peso próximo de 0. Deste modo, o processo de otimização MQR é utilizado na filtragem interna como um método de seleção de regras, diminuindo a complexidade da base de conhecimento.

Por fim, a última etapa da filtragem interna é a seleção das regras para a composição da base de regras final. Este processo tem como objetivo avaliar se as regras geradas e otimizadas são capazes de generalizar, com uma certa razoabilidade, e realizar a previsão *multi-step*. Esta verificação é feita realizando-se a previsão (com o horizonte definido pelo problema) no conjunto de validação e avaliando-se o resultado segundo as métricas estabelecidas. O critério para seleção da base de regras de cada subsistema consiste na comparação das métricas em função de um limiar de erro  $\epsilon_{min}$ , parâmetro que define o erro mínimo para compor a base de conhecimento final.

#### 4.2.3.2

##### Filtragem Externa

A filtragem externa é a etapa de filtragem realizada após o treinamento de todos os subsistemas, sendo a última etapa de ajustes da base de conhecimento do modelo. Nesta fase, as regras geradas podem ser duplicadas, redundantes e conflitantes. De forma semelhante à filtragem interna, a complexidade da base de regras é reduzida mediante o uso da técnica de filtro de similaridade, que extrai apenas as regras mais importantes.

A segunda parte da filtragem externa tem como objetivo estimar as regras mais relevantes da base de dados, que mais contribuem para uma previsão mais acurada. Novamente, é utilizada a estratégia de MQR para o ajuste de pesos das regras, mas com uma estratégia diferente. Neste caso, regras que possuem um peso menor que um limite definido são eliminadas, para diminuir a complexidade da base de regras, mas as regras selecionadas continuam com seus respectivos parâmetros calculados. Intuitivamente, a ponderação estabelecida tem como objetivo indicar as regras com maior credibilidade para o processo de agregação.

Após a etapa de filtragem, é necessário determinar como as regras ativadas podem ser combinadas para gerar o conjunto *fuzzy* de saída. Esta é a função do estágio de agregação, responsável por unificar em uma ativação as regras que possuem um mesmo consequente. Dadas duas regras ativadas com graus de ativação  $\mu_{A_P}(x)$  e  $\mu_{A_Q}(x)$  e um mesmo consequente  $B_{i,j}$  da série  $i$  e conjunto *fuzzy*  $j$  com grau de pertinência  $\mu_{B_{i,j}}(y)$ , o conjunto *fuzzy* de saída  $\mu_{B_{i,j}^*}(y)$  é definido pela função de agregação, genericamente definida pela Eq. 4-8.

$$\mu_{B_{i,j}^*}(y) = f(\mu_{A_P}(x), \mu_{A_Q}(x), \mu_{B_{i,j}}(y)) \quad (4-8)$$

Na literatura, o *Modus Ponens* Generalizado (MPG) é a técnica de inferência utilizada em SIFs do tipo Mamdani. Neste trabalho, o MPG é utilizado para a geração do conjunto de saída, utilizando o operador máximo para a agregação das regras. Ao invés de usar puramente o grau de pertinência  $\mu_{A_P}(x^*)$  para a composição do resultado  $\mu_{B_{s,n}^*}(y)$ , o modelo e-AutoMFIS faz uso dos pesos  $w_P$  obtidos na etapa de filtragem para ponderar as ativações das regras, conforme a Eq. 4-9.

$$\mu_{B_{s,n}^*}(y) = \min\{\max\{w_P \cdot \mu_{A_P}(x), \dots\}, \mu_{B_{s,n}}(y)\} \quad (4-9)$$

Desta forma, os pesos usados na etapa de filtragem contam como uma métrica de confiabilidade da regra, visando a aumentar a acurácia sem um alto comprometimento na interpretabilidade. A Figura 4.10 mostra a diferença entre a técnica utilizada no e-AutoMFIS e a abordagem clássica, sem o uso dos pesos, sendo perceptível a diferença na saída gerada.

Neste módulo, a interpretabilidade tem o papel de nortear as escolhas dos parâmetros e a construção de seus métodos. As seguintes métricas são relevantes nesta etapa, com suas respectivas soluções:

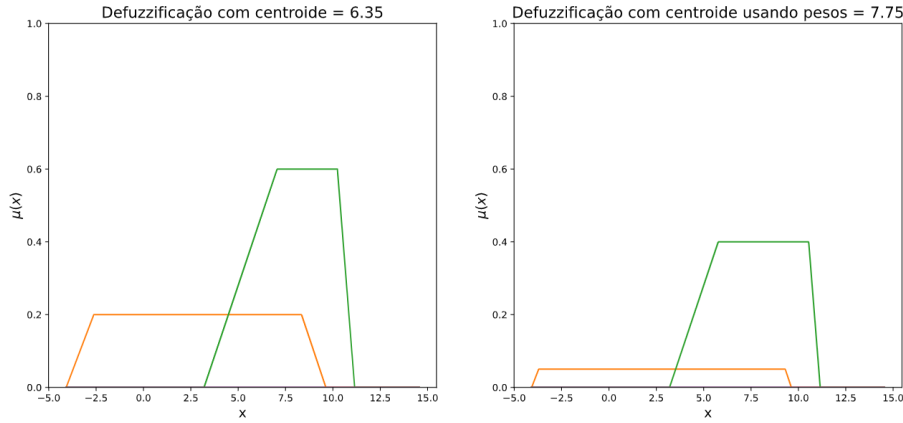


Figura 4.10: Efeito da ponderação na defuzzificação do e-AutoMFIS

- **Tamanho da base de regras:** remoção de regras menos significantes, reduzindo a complexidade do problema;
- **Ativação simultânea:** semelhante ao critério anterior, utilização do filtro de similaridade e do método MQR para identificar regras relevantes e eliminar as menos relevantes;
- **Discernibilidade:** eliminação de regras frequentemente ativadas simultaneamente, e que apontam para consequentes distintos.

#### 4.2.4 Defuzzificação

Com a base de regras já consolidada, o próximo passo de um SIF é a transformação do resultado da inferência, que ainda está no domínio *fuzzy*, para um valor real, que é o domínio estabelecido pelo problema. A defuzzificação é um processo clássico de SIFs baseado em técnicas bem consolidadas, como os métodos do centroide e média dos máximos. Neste trabalho, foram consideradas as técnicas de centroide, altura e altura modificada, que são definidas a seguir. Esta escolha foi feita de modo a contemplar também a forma do conjunto *fuzzy* durante o processo de defuzzificação e não apenas o consequente de maior ativação, como é o método de defuzzificação pelo máximo.

A defuzzificação pelo centroide calcula o valor de saída conforme mostra a Eq. 4-10. Este método interpreta o resultado gerado como sendo um valor central dos conjuntos *fuzzy* ativados.

$$y_c = \frac{\int y_i \mu_B(y_i) dy}{\int \mu_B(y_i) dy} \quad (4-10)$$

Por fim, o modelo e-AutoMFIS permite o uso de uma terceira abordagem para a etapa de defuzzificação, chamado de altura modificada. O cálculo do valor de saída é efetuado conforme a Eq. 4-11.

$$y_c = \frac{\sum_{i \in S} y_i \mu_B(y_i) / \delta}{\sum_{i \in S} \mu_B(y_i) / \delta} \quad (4-11)$$

Este método de defuzzificação foi criado para casos em que os conjuntos *fuzzy* possuem suportes com dimensões diferentes e que podem levar a resultados indesejados, como mostrado na Figura 4.11. No exemplo mostrado, é possível ver que usando a técnica de centroide, conjuntos com maior suporte podem ter uma grande influência para modificar o valor de saída, e mesmo com outro conjunto de maior ativação, o resultado pode ser mais deslocado do que o desejado. Deste modo, a defuzzificação pela altura modificada tenta amenizar este problema, ponderando as ativações pelos seus respectivos suportes.

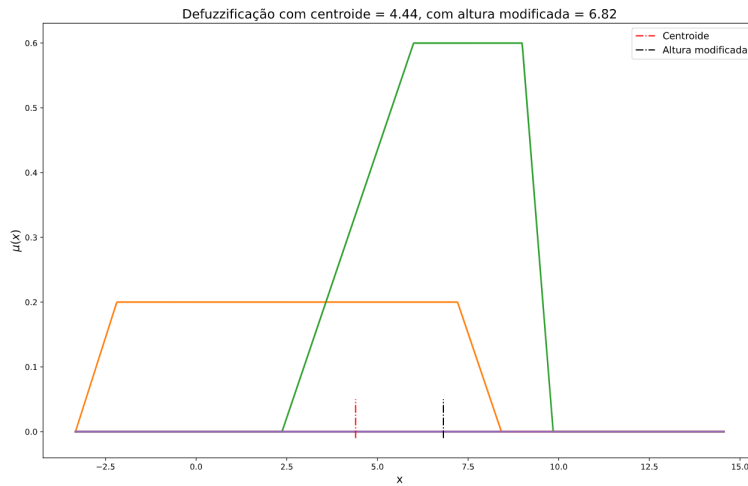


Figura 4.11: Comparação de métodos de defuzzificação

#### 4.2.5

##### Previsão *multi-step*

A última etapa do e-AutoMFIS consiste na previsão *multi-step*. Por se tratar de um modelo multivariado, todas as variáveis selecionadas devem ser previstas, por mais que não tenham suas métricas avaliadas posteriormente. Estes valores previstos são usados como entrada para o próximo passo de previsão, repetindo este processo iterativo até a previsão de  $T$  passos a frente.

No processo de previsão, é nítido que a seleção de variáveis tem um papel fundamental para que o modelo tenha um bom desempenho. Além dos problemas anteriormente citados, como uma maior dificuldade de formulação de

premissas, cada variável precisa ter seu resultado previsto, o que implica em um erro inserido em relação ao valor objetivo. Estes resultados ligeiramente errôneos são usados como entrada para prever os valores das séries no próximo instante de tempo, e o erro acumulado tende a aumentar. Portanto, aumentar a quantidade de séries, além de aumentar a complexidade do problema, também pode incorrer em um maior erro inserido no sistema e, conseqüentemente, tornar o modelo menos eficiente, tanto em acurácia quanto em interpretabilidade.

### 4.3

#### Resumo do modelo e-AutoMFIS

O e-AutoMFIS é uma extensão do AutoMFIS, com o objetivo de previsão de séries multivariadas. A abordagem *Ensemble* é extremamente proveitosa para a proposta do modelo desenvolvido. Além de suprir as deficiências observadas pelo AutoMFIS, esta técnica também auxilia no aprendizado do sistema.

Por mais que seu esquema tenha sido dividido em cinco módulos, a sua estrutura detalhada pode ser vista na Figura 4.12. Neste novo diagrama, é possível observar com maior nitidez a separação entre o treinamento e o teste que avalia a generalização do modelo. Além disso, são explicitadas as características finais que são relevantes para o processo de previsão.

Esta outra visão sobre o funcionamento do modelo e-AutoMFIS mostra uma faceta mais interna do sistema, onde é evidenciada a relação entre os módulos, de forma mais detalhada do que o esquema simplificado da Figura 4.1. Por mais que esta nova representação seja levemente mais complexa, é mais simples entender o ajuste do modelo e posteriormente a utilização do e-AutoMFIS no processo de previsão de séries.

Um tópico abordado neste capítulo, mas que não foi formalmente listado, é a definição de todos os parâmetros e métodos disponíveis pelo modelo. A Tabela 4.1 apresenta estes parâmetros.

Tabela 4.1: Lista de parâmetros do modelo e-AutoMFIS

Nome da variável	Descrição
Lag	Valor da defasagem máxima da série
HPrev	Horizonte de previsão
NumInput	Quantidade de séries para compor o subconjunto
FuzzySets	Quantidade de conjuntos <i>fuzzy</i> por variável
NumPred	Número de subsistemas para o comitê de previsores
MinAct	Ativação mínima das premissas
MaxRule	Quantidade máxima de premissas em uma regra

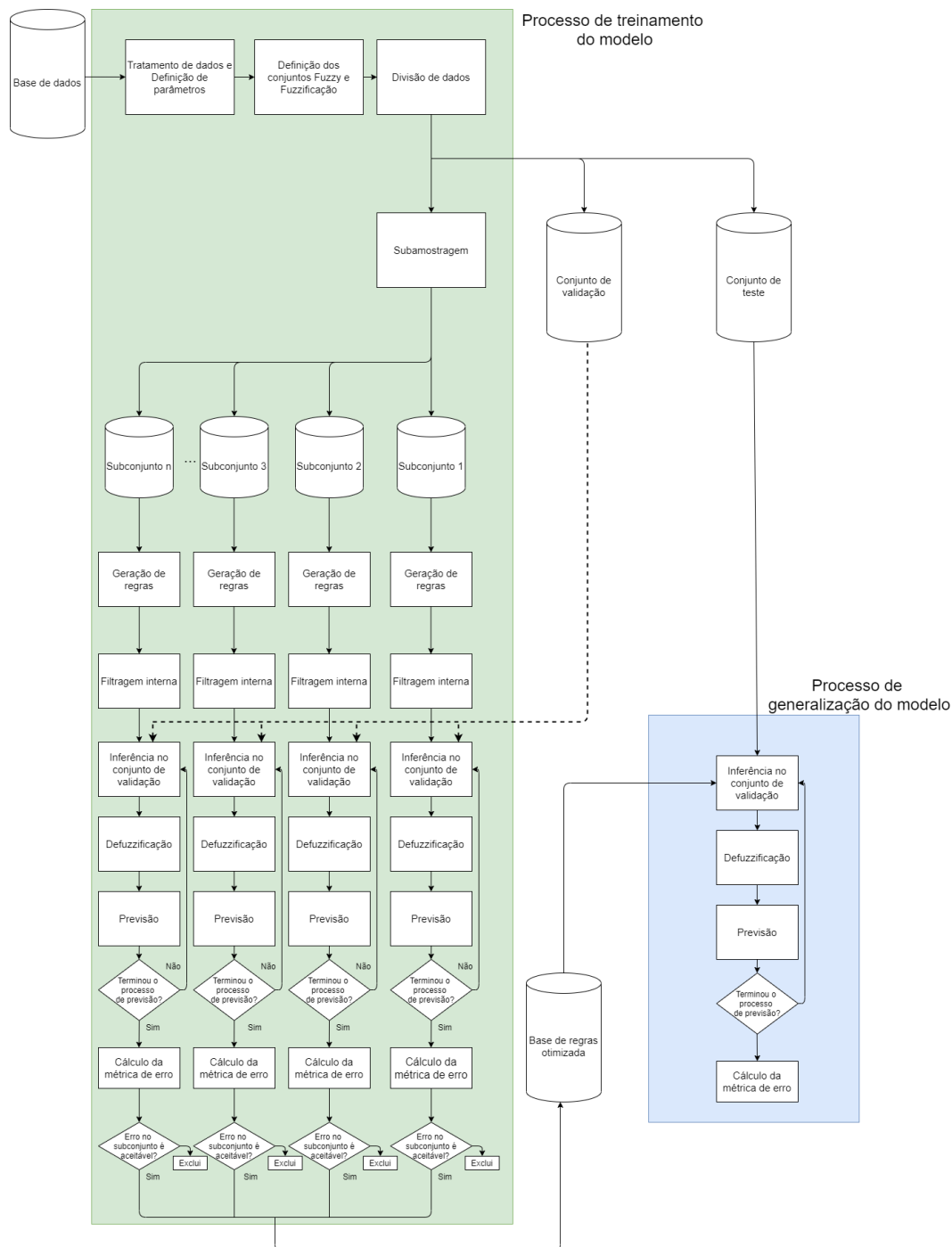


Figura 4.12: Arquitetura completa do modelo e-AutoMFIS



Além disso, a Tabela 4.2 apresenta os métodos existentes no modelo, agrupados pela funcionalidade. Conforme será visto no próximo capítulo, estas tabelas serão utilizadas para organizar os parâmetros testados e para se obter a melhor configuração.

Tabela 4.2: Lista de métodos do modelo e-AutoMFIS

Nome da variável	Descrição	Métodos possíveis
FuzzyMethod	Método para determinar como serão gerados os conjuntos <i>fuzzy</i>	Geração uniforme
		Geração usando percentis
		Geração usando agrupamento
ActivationMethod	Método para calcular a ativação da premissa	Ativação média
		Ativação média não-nula
DefuzzMethod	Método para determinar a defuzzificação	Centroide
		Altura
		Altura modificada

## 5

### Estudos de casos

O modelo proposto neste trabalho, e-AutoMFIS, foi desenvolvido com a proposta de viabilizar a utilização em bases de dados mais volumosas para o problema de previsão de séries multivariadas. Portanto, a busca por bases de dados *benchmark* foi refinada de modo a contemplar trabalhos que usem modelos de previsão multivariadas. Na literatura, é possível encontrar bases de dados, como qualidade do ar [114], consumo ou geração de energia [115] e previsão de velocidade do vento [116].

Este capítulo apresenta a aplicação do e-AutoMFIS em alguns estudos de casos, com o objetivo de evidenciar o aperfeiçoamento em relação ao seu modelo antecessor, além de discutir suas características e potenciais aplicabilidades práticas. Para isso, foram selecionadas cinco bases de dados, cujas informações básicas estão dispostas na Tabela 5.1. Um aspecto importante para a escolha dos estudos de casos foi a presença de outros trabalhos na literatura que também as avaliaram. Deste modo, a comparação de resultados torna-se mais justa.

Tabela 5.1: Informações sobre as bases de dados dos estudos de casos

Base de dados	Registros por série	Séries
Competição M3	134	5
Qualidade do ar	9357	12
Tráfego terrestre	17544	862
Taxa de câmbio	7588	8
Competição CIS	17521	3274

Nas próximas seções serão apresentados os estudos de casos selecionados para este trabalho, mostrando detalhadamente as particularidades das bases de dados, a metodologia utilizada e seus respectivos resultados. Em todos os experimentos realizados são feitas duas análises:

1. **Avaliação da acurácia:** análise em função das métricas de avaliação do problema, sendo possível a comparação com resultados produzidos por outros modelos;

2. **Avaliação da interpretabilidade:** observação qualitativa e quantitativa do e-AutoMFIS por meio dos critérios de interpretabilidade citados na Seção 3.2.

Outro ponto relevante na avaliação dos estudos de casos é a reprodutibilidade dos experimentos, de modo que os resultados obtidos neste trabalho sejam facilmente reproduzidos por futuros usuários que desejem utilizar o modelo. Para isso, além da metodologia utilizada em cada teste, também será disponibilizada uma tabela contendo os parâmetros com a melhor configuração do e-AutoMFIS.

## 5.1

### Competição M3

Competições são eventos bastante úteis para avaliar diferentes modelos e técnicas. Além da disponibilidade da base de dados, em competições os modelos geralmente são testados exaustivamente para conquistar os melhores resultados possíveis, proporcionando um teste mais desafiador para novos modelos.

Na área de previsão de séries temporais, as competições mais conhecidas foram organizadas por Spyros Makridakis, popularmente denominadas de Competições M [117]. Originalmente, estas competições eram conhecidas por uma coletânea de séries temporais de naturezas distintas (como de áreas da indústria, economia e finança) e diferentes intervalos entre amostras. Recentemente, a Competição M5 [118] tem como objetivo a previsão de vendas do mercado de varejo da empresa Walmart, mantendo a tradição de disponibilização de um grande volume de dados.

É natural, portanto, que um dos estudos de casos deste trabalho esteja relacionado com uma das competições organizadas por Makridakis. Neste caso, o primeiro experimento para avaliação do e-AutoMFIS utiliza séries disponíveis pela Competição M3 [119]. Esta escolha foi feita principalmente pela presença de resultados do AutoMFIS, o que possibilita uma comparação direta e mais justa entre modelos da mesma família de previsores, podendo avaliar as melhorias observadas.

Conforme citado anteriormente, este tipo de competição é conhecido pela quantidade elevada de séries temporais. A Competição M3 define o horizonte de previsão com o tamanho igual a 12, e a métrica de avaliação é feita pelo SMAPE (*Symmetric Mean Absolute Percentage Error*), matematicamente definida pela Eq. 5-1.

$$SMAPE = \frac{100\%}{n} \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{(|y_i| + |\hat{y}_i|)/2} \quad (5-1)$$

onde  $\hat{y}_i$  é o  $i$ -ésimo valor previsto pelo modelo,  $y_i$  é o  $i$ -ésimo valor atual e  $N$  é a quantidade total de elementos.

No modelo AutoMFIS, os testes realizados concentraram-se nas séries mensais sobre finanças. Estas foram posteriormente agrupadas usando um método de *clustering* hierárquico. Foram selecionados 3 *clusters* para avaliação, conforme mostrado em [2]. Desta forma, foram mantidos os grupos formados para a avaliação do e-AutoMFIS, tendo o experimento priorizado unicamente o conjunto no qual o AutoMFIS obteve um baixo desempenho, com o objetivo de salientar as vantagens do modelo proposto. A Figura 5.1 mostra os registros históricos deste grupo.

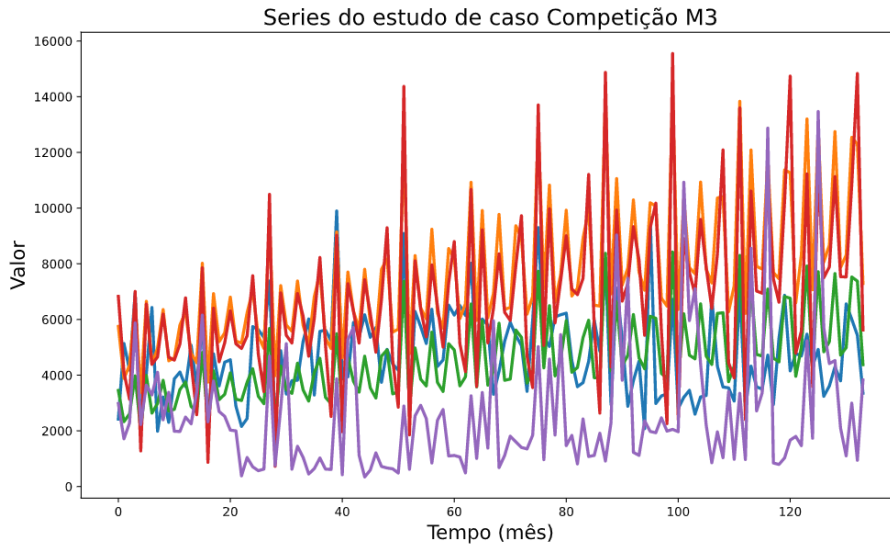


Figura 5.1: Registro histórico de 5 séries financeiras da base de dados Competição M3

Apenas observando graficamente, é possível perceber que as séries possuem uma tendência de crescimento e, conforme citado anteriormente, esta característica é indesejada para uso do modelo e-AutoMFIS. Para resolver este problema, foi realizada a remoção de tendência usando o método de regressão linear. Por mais que semanticamente estes novos valores tenham uma interpretabilidade menor que os originais, este pré-processamento é necessário para garantir que as séries se tornem estacionárias e possibilitem um melhor uso do e-AutoMFIS.

Além disso, por mais que todas as variáveis sejam necessárias para a previsão, é interessante analisar qual a relação existente entre as séries temporais. Para isso, foi utilizada a correlação de Pearson para testar a

dependência linear entre as variáveis. Isto é mostrado na Figura 5.2. Neste caso, é possível perceber que existem séries que não possuem uma correlação linear entre si, e potencialmente ineficazes para a construção da base de regras. Esta informação é utilizada para orientar a subamostragem do problema.

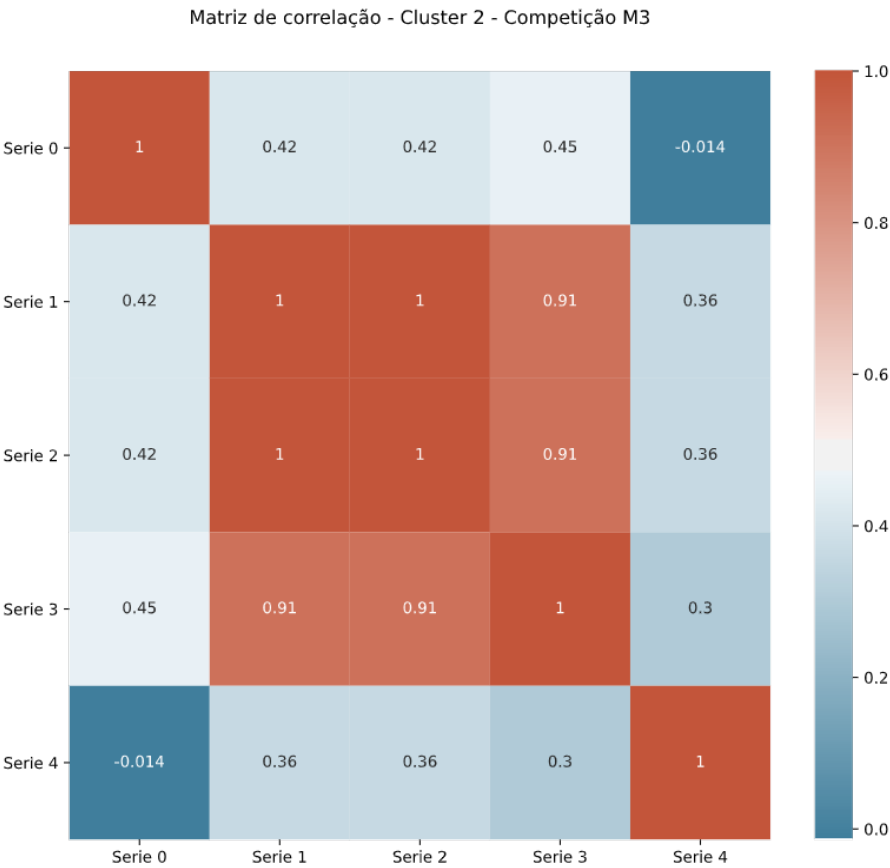


Figura 5.2: Matriz de correlação para o estudo de caso Competição M3

O modelo possui uma quantidade substancial de parâmetros, o que torna o teste exaustivo para avaliar todas as combinações possíveis uma tarefa impraticável. Assim, uma opção viável e empregada neste trabalho é a diminuição do espaço de busca das configurações possíveis, utilizando faixas de valores que, empiricamente, não resultem em uma complexidade elevada (e conseqüentemente uma alta demanda computacional), nem no desenvolvimento de sistemas com baixa acurácia. Em alguns parâmetros, como a quantidade de conjuntos *fuzzy*, a faixa de possíveis valores é limitada por heurísticas que favorecem a interpretabilidade do modelo.

A Tabela 5.2 exhibe os parâmetros testados e a configuração que resultou no melhor resultado obtido. Por se tratar de um modelo estocástico, decorrente da subamostragem da base de dados, cada combinação de parâmetros foi testada três vezes para assegurar uma maior robustez nesta escolha. Vale ressaltar que o desempenho do modelo é avaliado segundo a métrica SMAPE.

Tabela 5.2: Parâmetros testados no estudo de caso Competição M3

Parâmetro	Variáveis testadas	Melhor configuração
Lag	16 a 24	<b>24</b>
NumInput	10 a 20	<b>12</b>
FuzzySets	5, 7, 9	<b>7</b>
NumPred	10 a 15	<b>10</b>
MinAct	0,5 a 0,75	<b>0,65</b>
MaxRule	5 a 7	<b>5</b>
FuzzyMethod	Uniforme Agrupamento	<b>Agrupamento</b>
ActivationMethod	Ativação média Ativação média não-nula	<b>Ativação média não-nula</b>
DefuzzMethod	Centroide Altura ponderada	<b>Centroide</b>

Em relação aos diferentes modelos, a análise comparativa mais direta é em relação à acurácia obtida no processo de previsão. Além do resultado obtido pelo modelo e-AutoMFIS, a Tabela 5.3 dispõe o resultado mediano dos modelos utilizados e o melhor desempenho na competição M3, juntamente com o melhor resultado do modelo AutoMFIS. Para o modelo proposto, também é disponibilizado um resultado médio das dez simulações realizadas com a mesma configuração descrita na Tabela 5.3.

Tabela 5.3: Resultado do Estudo de Caso 1

Método	Desempenho (SMAPE)
Desempenho mediano (Comp. M3)	21,77 %
Melhor desempenho (Comp. M3)	17,02 %
AutoMFIS	19,77 %
e-AutoMFIS	17,12 $\pm$ 0.5 %
e-AutoMFIS (melhor)	<b>16,8%</b>

Observando o aspecto de acurácia, o modelo e-AutoMFIS foi capaz de obter um desempenho superior aos métodos estatísticos utilizados na competição M3. Em relação ao seu modelo anterior (AutoMFIS), a mudança na arquitetura alterou positivamente o resultado, tendo um erro SMAPE calculado 15% menor. Mesmo comparando em relação ao desempenho médio do e-AutoMFIS, o resultado permanece melhor se comparado tanto com o AutoMFIS, quanto com o valor médio dos modelos da competição.

Diversos motivos podem ser responsáveis pela melhoria observada neste estudo de caso. O primeiro aspecto, essencial para qualquer projeto que almeje um bom resultado, é a visualização e análise de dados. As técnicas utilizadas nesta etapa ajudaram a nortear a busca pela otimização de parâmetros,

como a quantidade da defasagem máxima para cada série e a orientação da subamostragem para o treinamento do modelo, avaliando as métricas de autocorrelação e correlação cruzada, respectivamente.

Outro ponto que possivelmente teve uma contribuição positiva na melhoria do desempenho é a utilização da técnica de *Ensemble*. Neste caso, o modelo e-AutoMFIS fornece uma maior diversidade na busca, por conta dos diversos treinamentos usando diferentes subamostragens do conjunto total de dados, tornando mais simples o aprendizado e a seleção das regras. Aliado a isto, a diminuição da complexidade do problema durante a fase de treinamento possibilita a utilização mais ampla de parâmetros e, conseqüentemente, uma potencial melhora nos resultados. Neste experimento, por exemplo, é perceptível que a escolha de defasagens maiores foram benéficas para a inferência do e-AutoMFIS, fato que não pode ser completamente explorado no seu modelo antecessor, devido ao aumento considerável do custo computacional [2]. Para este problema, é relatado que o AutoMFIS foi capaz de utilizar no máximo quatro defasagens, sem que houvesse um alto comprometimento com o desempenho computacional.

O resultado obtido pelo modelo proposto exemplifica a evolução esperada para a previsão de séries temporais, tanto em termos de acurácia quanto na capacidade de analisar combinações mais extensas. Ainda assim, a interpretabilidade é um tópico de alta importância e sua análise é necessária para que o e-AutoMFIS seja considerado um modelo interpretável. A Tabela 5.4 apresenta o conjunto de métricas de interpretabilidade, segundo os critérios apresentados na Seção 3.2.

Tabela 5.4: Critérios de interpretabilidade do estudo de caso Competição M3

Critério	Resultado
Tamanho da base de regras	23,5 regras / série
Comprimento dos antecedentes	2,1 antecedentes
Ativação média das regras	5,2 regras

Uma das melhores formas de avaliar a interpretabilidade é por meio de uma análise comparativa com outros SIFs. Contudo, na ausência de resultados destas métricas, o estudo da interpretabilidade tem um caráter mais qualitativo, sendo observado seus valores e analisados segundo uma razoabilidade humana. Levando em consideração a complexidade geralmente existente na dinâmica de séries financeiras, o tamanho da base de regras não é elevada e é simples de ser entendida.

A complexidade no nível de regra linguística também é aceitável, dentro do que é esperado no quesito de interpretabilidade. A média de antecedentes

de regras representa um bom indicativo das regras geradas, com uma distribuição interessante de ser analisada, apresentada na Figura 5.3. É possível perceber, por exemplo, que a base de conhecimento possui um equilíbrio entre a simplicidade e a complexidade das regras.

Distribuição de antecedentes da base de regras

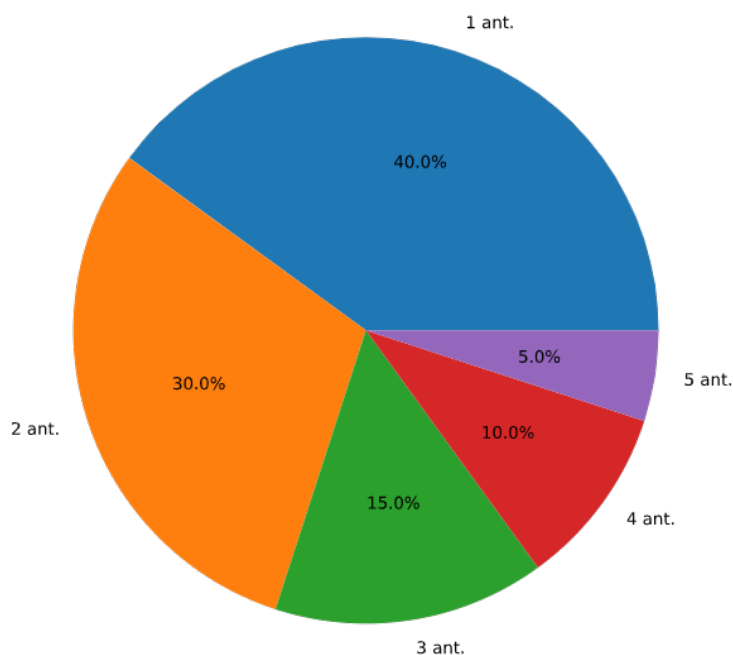


Figura 5.3: Distribuição da quantidade de antecedentes para a base de regras do estudo de caso Competição M3

A ativação simultânea de regras também é um bom indicador de interpretabilidade, e é possível perceber que seu resultado (Tabela 5.4) corrobora a explicabilidade do modelo. Na prática, isto significa que a cada passo de inferência do e-AutoMFIS, cerca de cinco regras são ativadas ao mesmo tempo. Isto aponta que uma quantidade reduzida de regras é disparada ao mesmo tempo. Esta métrica com um valor baixo está relacionada com a eficiência da etapa de filtragem do modelo, sendo bem-sucedido em escolher apenas as regras mais relevantes.

Com base no resultado obtido neste estudo de caso, é aceitável concluir que o e-AutoMFIS possui um desempenho competitivo em relação aos modelos da Competição M3. Também é perceptível a melhoria dos resultados do modelo AutoMFIS, tornando mais prática e precisa a extração de regras, além de possibilitar uma maior exploração dos parâmetros. Por fim, é possível perceber que, dentro do que é possível analisar, a interpretabilidade é mantida para este problema, tendo uma boa consonância com relação a acurácia.



## 5.2

### Air Quality

Com o início do processo de industrialização, países começaram a receber um fluxo migratório para os centros urbanos, hoje conhecidos por possuírem uma alta densidade populacional. Contudo, este crescimento populacional trouxe consigo problemas indesejados, como o aumento de gases poluentes. Hoje em dia, o monitoramento da poluição urbana é essencial, podendo servir como auxílio para políticas de controle de qualidade do ar e controle de tráfego [120].

A difusão de dispositivos para sensoramento de gases, devido ao seu baixo custo, foi um fator importante para a implementação de sistemas de monitoramento de centros urbanos. Consequentemente, dados históricos de concentração de gases estão cada vez mais disponíveis publicamente, fomentando a pesquisa nesta área de qualidade do ar. Na literatura, alguns dos tópicos mais relevantes neste tema são a calibração de sensores [121] e a previsão de concentração de gases [122].

O segundo estudo de caso do modelo e-AutoMFIS aborda a temática da previsão de concentração de gases. Este tópico, além da sua relevância social, também é de bastante interesse no aspecto técnico, pois é esperado que este tipo de problema tenha uma natureza multivariada, no qual outras variáveis podem contribuir para uma previsão mais robusta. A dinâmica da concentração de gases também não é simples de ser modelada, possuindo diversos fatores de influência, como é mostrado na Figura 5.4. Deste modo, o uso de técnicas de ML, como é o caso do e-AutoMFIS, torna-se apropriado para este tipo de problema. Adicionalmente, o modelo proposto neste trabalho insere o caráter explicável de um SIF, podendo ser extremamente útil na análise dos fatores que determinaram a previsão.

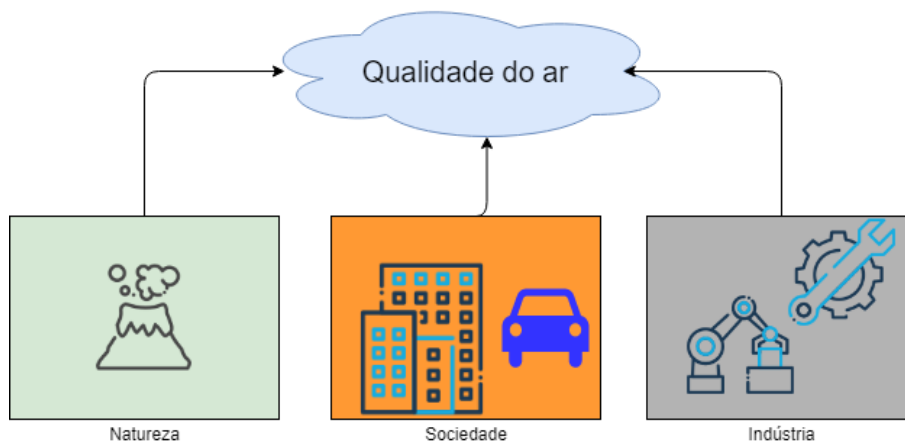


Figura 5.4: Exemplo de fatores de influência na qualidade do ar

A base de dados utilizada neste experimento foi elaborada por [123] e disponibilizada publicamente <sup>1</sup>, contendo no total 12 variáveis, entre concentração real de gases, valores medidos por sensores e variáveis de interesse, como temperatura e umidade relativa. Para uma visão mais ampla destas informações, a Tabela 5.5 fornece, de uma forma sucinta, a descrição das variáveis e suas unidades de medida. Diferentes tarefas podem ser aplicadas nesta base de dados, como calibração de sensores e previsão de compostos, como benzeno e monóxido de carbono.

Tabela 5.5: Variáveis da base de dados Qualidade do ar

Variável	Descrição	Unidade
CO (GT)	Concentração média de CO	mg/m <sup>3</sup>
PT08.S1 (CO)	Resposta horária média do sensor de CO	Óx. Est.
NMHC (GT)	Concentração média de hidrocarbonetos não-metais	µg/m <sup>3</sup>
C6H6 (GT)	Concentração média de Benzeno	µg/m <sup>3</sup>
PT08.S2 (NMHC)	Resposta horária média do sensor de NMHC	Titania
NOx (GT)	Concentração média de compostos nitrogenados	ppb.
PT08.S3 (NOx)	Resposta horária média do sensor de NOx	Óx. Tungs.
NO <sub>2</sub> (GT)	Concentração média de NO <sub>2</sub>	µg/m <sup>3</sup>
PT08.S4 (NO <sub>2</sub> )	Resposta horária média do sensor NO <sub>2</sub>	Óx. Tungs.
PT08.S5 (O <sub>3</sub> )	Resposta horária média do sensor O <sub>3</sub>	Óx. Índio
T	Temperatura	°C
RH	Umidade relativa	%
AH	Umidade absoluta	-

Os registros históricos destas séries foram coletados de hora em hora entre 03/10/2004 e 04/04/2005, totalizando uma base com 9357 registros e 12 variáveis. Quanto à definição do problema, o horizonte de previsão é definido para 24 passos, totalizando a previsão de um dia, com os dados divididos entre treinamento (70%), validação (10%) e teste (20%). Além disso, o desempenho do modelo é avaliado, em termos de acurácia, usando as métricas de Erro Médio Absoluto (MAE, em inglês) e Erro Médio Quadrático (RMSE, em inglês), definidos pelas Eqs. 5-2 e 5-3, respectivamente.

<sup>1</sup>Base de dados disponível em <https://archive.ics.uci.edu/ml/datasets/Air+Quality>

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (5-2)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (5-3)$$

onde  $\hat{y}_i$  é o  $i$ -ésimo valor previsto pelo modelo,  $y_i$  é o  $i$ -ésimo valor atual e  $N$  é a quantidade total de elementos. Por conta da variedade de compostos gasosos, é necessário definir quais são as variáveis desejadas para a avaliação do resultado, de modo que as etapas de pré-processamento e seleção de parâmetros do modelo privilegiem as séries que serão avaliadas. Desta forma, foi escolhido o composto CO para analisar o resultado, uma vez que existe uma gama maior de resultados publicados.

Conforme comentado anteriormente, a base de dados possui informações reais e medidas por sensores, como as variáveis CO(GT) e PT08.S1 (CO). Como o propósito deste experimento é a previsão de séries temporais e as variáveis medidas por sensores podem ser redundantes ou até desnecessárias em relação a seus valores reais de concentração, todas as variáveis relacionadas a informações sensoriais foram removidas da base de dados, reduzindo a dimensionalidade do problema.

Além da análise superficial das variáveis, é importante avaliar a qualidade dos dados. Conforme pode ser visto na Figura 5.5, a base de dados possui valores faltantes para todas as variáveis do problema. Para a maioria das variáveis (que possuem até 20% de dados faltantes), ainda é possível realizar um tratamento de dados para preencher esses valores faltantes. Especificamente neste estudo de caso, foi selecionado um subconjunto de treinamento que tenha uma maior quantidade de dados presentes, e os valores ainda faltantes foram preenchidos pelo método *Backward* (i.e. preencher as lacunas com os valores imediatamente a frente). A única exceção a essa metodologia foi a variável NMHC, já que a quantidade alta de valores faltantes impossibilitou o uso de qualquer técnica apropriada. Portanto, esta variável também foi removida da base de dados.

Outra etapa importante que deve ser realizada durante a etapa de pré-processamento de dados é a avaliação da relação existente entre as variáveis para selecionar apenas as mais relevantes. Para avaliar a dependência linear entre as variáveis, a correlação de Pearson pode servir como indicador para auxiliar na seleção de variáveis. A Figura 5.6 dispõe a matriz de correlação para auxiliar na análise.

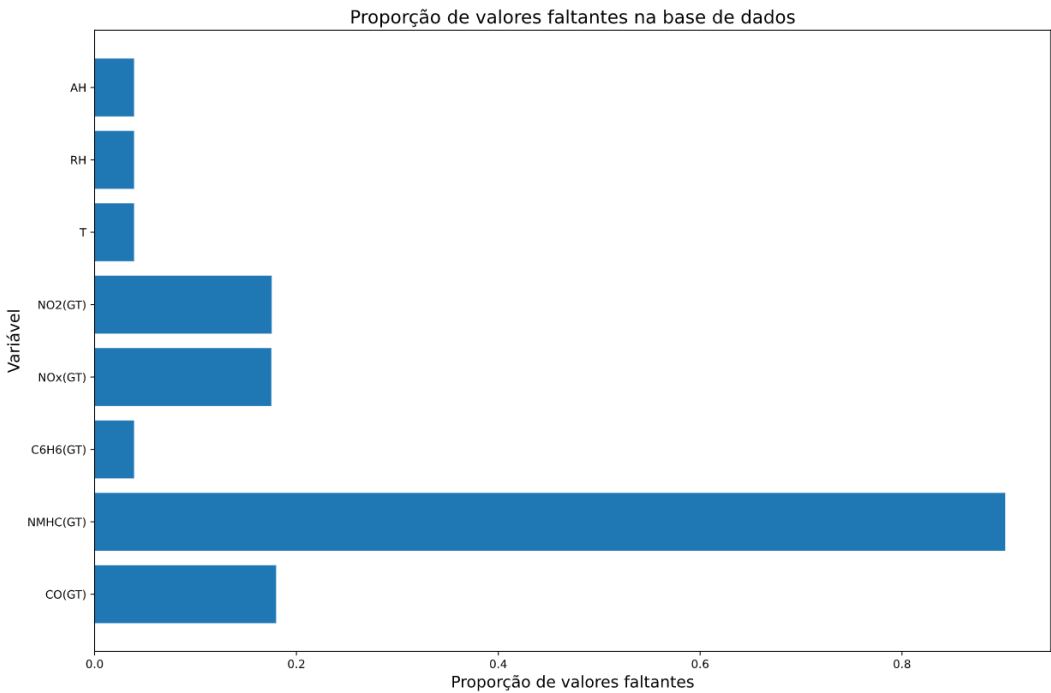


Figura 5.5: Taxa de valores faltantes na base de dados Qualidade do Ar



Figura 5.6: Matriz de correlação de Pearson para a base de dados Qualidade do Ar

Como é possível perceber, as variáveis que expressam a umidade do ar (RH e AH) possuem uma dependência linear desprezível em relação às variáveis que representam as informações de gases. Por mais que não exista

uma associação linear, em um primeiro momento estas variáveis não foram eliminadas, uma vez que pode existir uma relação não-linear. Outro fato interessante a se notar é o comportamento linear entre as quatro variáveis sobre compostos gasosos. Aparentemente, os efeitos sazonais e os efeitos que ocasionam a mudança de concentração de gases têm influência sobre as 4 variáveis analisadas.

Continuando a análise das variáveis, uma técnica extremamente útil para compreender a relevância dessas variáveis é o RFE (*Recursive Feature Elimination*). Esta técnica estima a importância de cada uma das variáveis do problema através da análise do resultado de um modelo ao remover recursivamente uma variável do problema. Após a avaliação desta primeira etapa, a variável considerada mais fraca é eliminada e o processo iterativo continua até que alcance a quantidade de variáveis selecionadas desejada. A Tabela 5.6 mostra o ranqueamento entre as variáveis selecionadas, usando o método *Random Forest* como estimador para o RFE. Novamente, é possível perceber que as variáveis de umidade foram consideradas as menos relevantes para o problema de previsão proposto. Portanto, tanto a variável RH quanto AH foram removidas da base de dados.

Tabela 5.6: Importância das variáveis usando o método RFE

Variável	Ranqueamento
C <sub>6</sub> H <sub>6</sub> (GT)	2 <sup>o</sup>
NO <sub>x</sub> (GT)	1 <sup>o</sup>
NO <sub>2</sub> (GT)	3 <sup>o</sup>
T	4 <sup>o</sup>
RH	6 <sup>o</sup>
AH	5 <sup>o</sup>

Após esta primeira etapa de pré-processamento, 5 variáveis foram selecionadas: CO, C<sub>6</sub>H<sub>6</sub>, NO<sub>x</sub>, NO<sub>2</sub> e T. A Figura 5.7 mostra o comportamento de cada uma das séries. Diferente do primeiro estudo de caso, todas as séries não possuem tendências e, portanto, não é necessário aplicar a técnica de remoção de tendências. Ainda assim, é possível observar um comportamento cíclico nos compostos gasosos, com o período de 24 horas. Intuitivamente isto é esperado, visto que fatores sociais e naturais também possuem ciclos diários, como o transporte para o trabalho e a exposição ao sol. Este comportamento cíclico não é prejudicial para o aprendizado do modelo e, portanto, não é necessário aplicar um método para remoção de ciclos.

A investigação das defasagens também é uma tarefa muito importante neste experimento. De forma a auxiliar na busca deste parâmetro do modelo, efetuou-se uma análise de autocorrelação, cujo resultado pode ser visto na

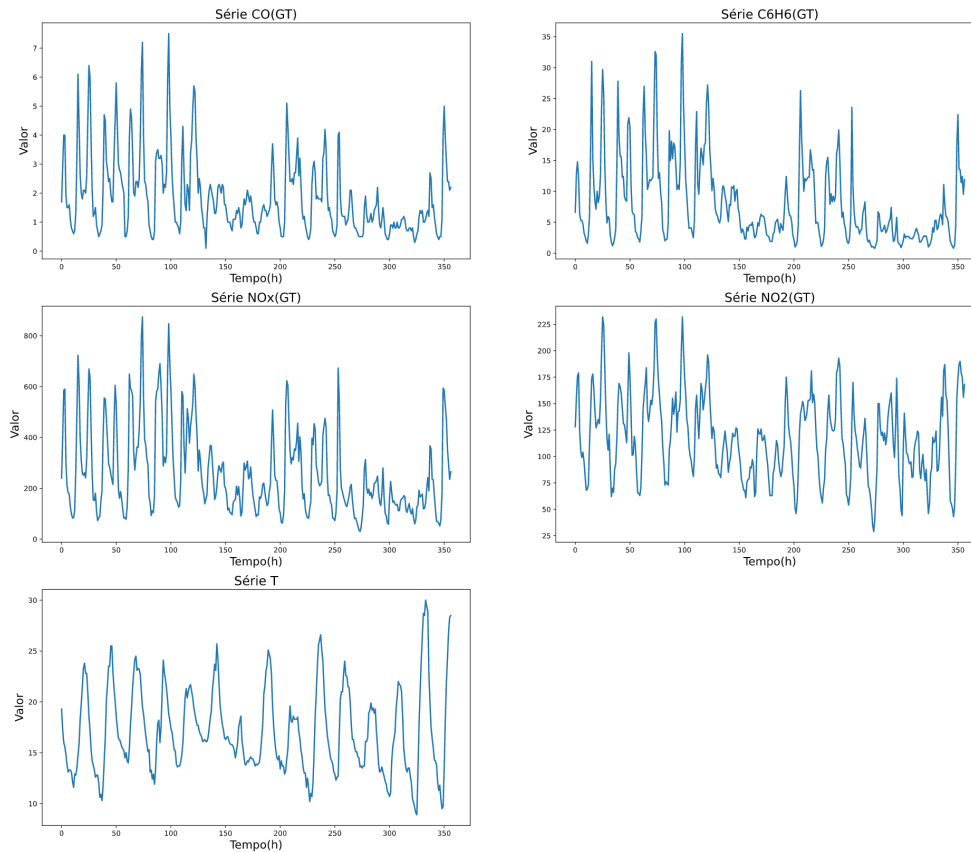


Figura 5.7: Amostra das séries do estudo de caso 'Qualidade do ar'

Figura 5.8. Como é esperado, as séries com 24 passos atrás apresentam o maior valor de correlação, indicando um bom ponto inicial para o parâmetro de defasagem máxima do modelo.

O ajuste dos parâmetros do e-AutoMFIS é feito de forma iterativa, buscando a melhor combinação com base no RMSE do conjunto de validação, com valor médio em 5 execuções. A Tabela 5.7 lista os parâmetros testados e a configuração utilizada para a geração dos resultados finais.

Neste estudo de caso, é importante citar os métodos encontrados na literatura para efeito de comparação. Em [127] é reunido um conjunto de modelos com resultados consolidados, a saber: *Decision Tree* (DT), *Random Forest* (RF), Adaboost, *Support Vector Machine* (SVM), *Gated Recurrent Unit* (GRU), *Recurrent Neural Network* (RNN) e *Long Short-term Memory* (LSTM). A Tabela 5.8 mostra o resultado destes modelos, além do resultado do e-AutoMFIS.

Em termos de acurácia, o e-AutoMFIS possui um resultado competitivo em relação a todos os métodos comparados neste experimento. Observando a métrica RMSE, é possível concluir que o e-AutoMFIS obteve o melhor desempenho, enquanto que analisando-se somente o MAE o modelo proposto é o terceiro melhor avaliado. Este fato pode ser justificado por conta da diferença

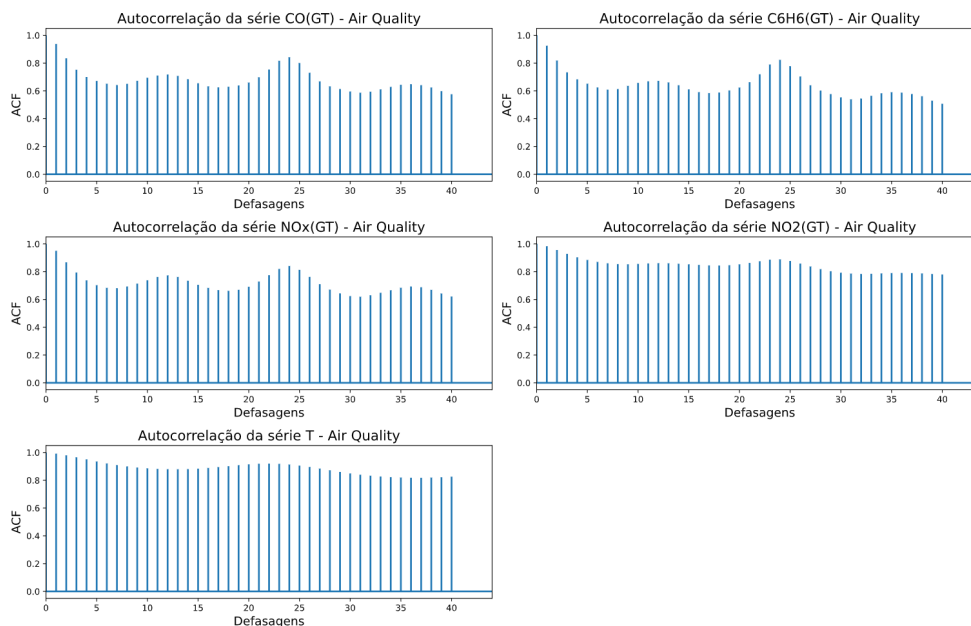


Figura 5.8: Função de autocorrelação para as variáveis selecionadas na base de dados Qualidade do Ar

Tabela 5.7: Parâmetros testados no estudo de caso Qualidade do Ar

Parâmetro	Variáveis testadas	Melhor configuração
Lag	16 a 24	<b>24</b>
NumInput	10 a 20	<b>14</b>
FuzzySets	5, 7, 9	<b>7</b>
NumPred	10 a 15	<b>12</b>
MinAct	0,5 a 0,75	<b>0,70</b>
MaxRule	5 a 7	<b>5</b>
FuzzyMethod	Uniforme Agrupamento	<b>Agrupamento</b>
ActivationMethod	Ativação média Ativação média não-nula	<b>Ativação média não-nula</b>
DefuzzMethod	Centroide Altura ponderada	<b>Altura ponderada</b>

Tabela 5.8: Resultado do estudo de caso Qualidade de Ar

Método	Desempenho (MAE)	Desempenho (RMSE)
DT	0.68	1.00
RF	0.65	0.94
Adaboost	0.53	0.70
SVM	1.37	1.60
GRU	0.42	0.636
LSTM	0.42	0.635
e-AutoMFIS	0.50 ± 0.04	0.62 ± 0.2

de interpretação das métricas. Como o RMSE eleva ao quadrado a diferença entre os valores previsto e observado, antes de efetuar a radiciação, quanto maior o erro calculado, maior é o peso atribuído. Já no caso do MAE, realiza-se a média do módulo da diferença dos valores previstos e observados. Deste modo, é possível concluir que, comparado com estes outros modelos, o e-AutoMFIS possui menos erros destoantes ao longo da previsão, e um erro médio comparável.

É importante ressaltar o bom desempenho do e-AutoMFIS em comparação aos métodos de DL, como GRU e LSTM. Este fato notável indica que é possível que modelos interpretáveis consigam resultados semelhantes a modelos caixa-preta mais complexos e até considerados estado-da-arte, tornando-se vantajoso o uso de modelos como o e-AutoMFIS. Ainda assim, o resultado obtido neste experimento é decorrente de vários passos de pré-processamento de dados, indicando a necessidade do uso de técnicas de seleção de variáveis e tratamento da informação, para que o modelo tenha uma melhor capacidade de aprendizado.

Exposta a análise comparativa de acurácia neste estudo de caso, é indispensável avaliar as métricas de interpretabilidade para afirmar se o e-AutoMFIS pode ser considerado um modelo explicável. A Tabela 5.9 lista os resultados obtidos em termos de interpretabilidade, considerando apenas a base de regras da série de monóxido de carbono (CO).

Tabela 5.9: Critérios de interpretabilidade do estudo de caso Qualidade do Ar

<b>Critério</b>	<b>Resultado</b>
Tamanho da base de regras	30
Comprimento dos antecedentes	3,2
Ativação média das regras	3,8

Como pode ser observado na Tabela 5.8, os modelos usados para comparação não possuem métricas de interpretabilidade, o que torna a comparação inviável neste quesito. Além disso, não foi possível utilizar o modelo AutoMFIS de forma apropriada para gerar resultados, devido ao alto custo computacional. Portanto, para evitar a comparação equivocada dos resultados de seu modelo antecessor, não foi incluído o desempenho do AutoMFIS, tanto em termos de acurácia quanto de interpretabilidade.

Em termos de interpretabilidade, o e-AutoMFIS possui resultados compatíveis com o que é esperado de um modelo interpretável. O tamanho da base de regras, por exemplo, é relativamente pequeno e apropriado para o problema, por conta da quantidade de variáveis existentes no problema. Além disso, a ati-



vação simultânea das regras é, em média, baixa o suficiente para tornar simples o entendimento humano do resultado.

As regras geradas também são importantes de serem avaliadas no aspecto de interpretabilidade. O tamanho médio de antecedentes por regra está dentro do critério de Miller, pois é limitado pelo parâmetro do modelo e-AutoMFIS. Esta quantidade baixa de termos linguísticos facilita o entendimento das informações extraídas, sendo possível entender melhor os termos que influenciam nos resultados.

Neste experimento, o e-AutoMFIS foi capaz de obter um desempenho satisfatório, com valores de acurácia melhores do que alcançados por métodos tradicionais e avançados de ML. Entretanto, o modelo proposto fica levemente atrás em relação a arquiteturas modificadas de DL para problemas específicos, como proposto em [127] (que utiliza uma técnica de inserção adaptativa de entradas em um modelo LSTM), mas ainda possui um resultado competitivo com modelos de DL sem adaptações. Ainda assim, o saldo é positivo para o e-AutoMFIS, visto que é o único método que apresenta o aspecto de interpretabilidade.

### 5.3

#### Traffic-Rate

O modelo e-AutoMFIS foi desenvolvido para abranger problemas com um grande volume de dados, através da extração de conhecimento de subamostras do conjunto total. Para avaliar o desempenho neste tipo de tarefa, foram buscados trabalhos de previsão de séries temporais usando DL para obtenção de base de dados e métodos para comparação. Em [64], uma coleção de 5 problemas são disponibilizados publicamente, tendo 4 deles um comportamento de padrões a curto e longo prazo. Deste modo, foram selecionados dois temas distintos para averiguação do modelo proposto, que serão apresentados no terceiro e quarto estudo de caso.

O terceiro experimento para avaliação do e-AutoMFIS consiste na previsão da taxa de ocupação de ruas em San Diego. Os dados foram coletados durante 48 meses (entre 2015 e 2016) pelo Departamento de Transportes da Califórnia, totalizando 17544 registros. Cada uma das 862 variáveis representam, neste caso, pontos das ruas com os sensores de medição. Para que no decorrer deste estudo de caso seja possível comparar os resultados com outros métodos, foram definidos os mesmos parâmetros do problema, com a divisão de dados entre treinamento (60%), validação (20%) e teste (20%). O horizonte de previsão estabelecido é igual a 24, significando a previsão multi-stepping de um dia a frente da taxa de ocupação das ruas. A métrica de avaliação para

este problema é o RRSE (*Root Relative Squared Error*), definido pela Eq. 5-4

$$RRSE = \sqrt{\sum_{(i,t) \in \Omega_{test}} \frac{y_{i,t} - \hat{y}_{i,t}}{y_{i,t} - \bar{Y}}} \quad (5-4)$$

onde  $\hat{y}_i$  é o  $i$ -ésimo valor previsto pelo modelo,  $y_i$  é o  $i$ -ésimo valor atual e  $N$  é a quantidade total de elementos. Diferentemente dos outros estudos de casos mostrados até o momento, este problema possui um grande desafio quanto à sua dimensionalidade. Como todas as variáveis precisam ser previstas, não é possível excluí-las da base de dados. Uma metodologia muito útil para este tipo de problema é a divisão em subconjuntos para o treinamento do modelo. Existem diferentes técnicas para realizar o agrupamento de séries, como técnicas usando agrupamento hierárquico [124] ou outras métricas de similaridade.

O modelo proposto neste trabalho realiza, dentro da arquitetura desenvolvida, a divisão de dados para treinamento via subamostragem de conjuntos. Contudo, esse procedimento é feito de forma aleatória e potencialmente não deve gerar o melhor resultado, se comparado com alguma técnica prévia de agrupamento. Para contornar este problema, a primeira etapa de pré-processamento consiste em um processo recursivo de agrupamento de séries utilizando métrica de similaridade. Iniciando-se pela primeira variável, são selecionadas 5 outras variáveis que possuam os maiores valores de similaridade para a formação de um grupo. Estas 6 séries são removidas do conjunto de séries, e o processo iterativo de agrupamento continua até que todos os valores pertençam a um grupo. Esta técnica foi utilizada com o objetivo de gerar conjuntos correlatos (e eventualmente próximos fisicamente), podendo ocasionalmente ser útil para formar relações causais entre as séries selecionadas.

Feita a divisão de variáveis dos conjuntos da base de dados, a próxima etapa consiste na análise das componentes das séries para aplicação de técnicas apropriadas para facilitar o treinamento do modelo. Para facilitar a explicação desta etapa, a Figura 5.9 mostra um exemplo do comportamento de uma das variáveis da base de dados. Nenhuma das séries possui o termo de tendência, mas todas compartilham de componentes cíclicas. Como é esperado, este problema possui dois padrões em períodos de 24 horas e 168 horas, representando repetições diárias e semanais.

A existência de padrões temporais com períodos fixos, como ocorre neste experimento, facilita uma nova modificação na arquitetura original do modelo proposto para o aprendizado de regras mais focadas em cada um dos 7 prováveis padrões. Neste problema, a base de dados é dividida manualmente, visto

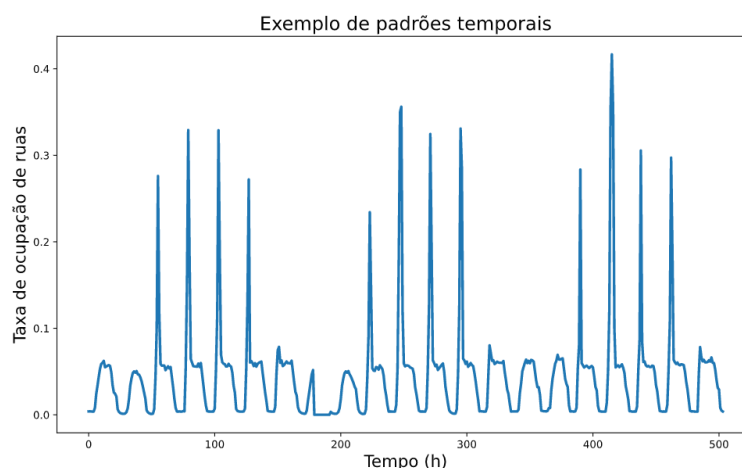


Figura 5.9: Exemplo de padrões temporais no estudo de caso

que neste caso estamos interessados em obter sistemas especializados para diferentes dias da semana. O ajuste é necessário para que o modelo consiga captar, de um modo mais simples, padrões em períodos mais longos.

A agregação também sofreu uma pequena modificação para tornar a base de regras única e mais interpretável. Conforme citado anteriormente, cada membro do comitê de previsores recebe uma subamostragem aleatória de cada um dos 7 padrões que representam os dias da semana, e é esperado que no processo de previsão estas regras sejam ativadas somente para o dia da semana designado. Para assegurar que isto ocorra, em cada antecedente de regras selecionadas é inserido um termo linguístico, usando conjuntos *singleton* para representar o dia da semana. Desta forma, uma regra hipotética "Se  $s_1(t-5)$  é **MA** e  $s_3(t-1)$  é **B** então  $s_2(t+1)$  é **A**" transforma-se em "Se  $s_1(t-5)$  é **MA** e  $s_3(t-1)$  é **B** e Dia da semana é **Domingo** então  $s_2(t+1)$  é **A**" caso a base de regras seja para o padrão de domingo. Esta estratégia, além de reunir as regras extraídas em uma única base de conhecimento, também torna a seleção das regras semanticamente interpretável para o usuário.

Com o pré-processamento de dados estabelecido e a modificação proposta realizada, o próximo passo da metodologia desenvolvida neste estudo de caso é o ajuste de parâmetros do e-AutoMFIS. Por conta da grande dimensionalidade do problema e do número exagerado de grupos gerados (144), a otimização da configuração para cada um desses 144 grupos é basicamente impraticável e provavelmente com um ganho de acurácia que não justifica o custo computacional necessário. Para simplificar o processo de otimização, foi selecionado aleatoriamente um grupo e realizado o mesmo processo de busca pela melhor configuração citado no estudo de caso anterior. A Tabela 5.10 mostra as diferentes configurações testadas, e a melhor obtida. Vale ressaltar que os parâmetros escolhidos para este grupo foram utilizados para todos os

outros 143 grupos restantes.

Tabela 5.10: Parâmetros testados no estudo de caso Taxa de ocupação de ruas

Parâmetro	Variáveis testadas	Melhor configuração
Lag	16 a 24	<b>24</b>
NumInput	10 a 20	<b>14</b>
FuzzySets	5, 7, 9	<b>7</b>
NumPred	10 a 15	<b>12</b>
MinAct	0,5 a 0,75	<b>0,70</b>
MaxRule	5 a 7	<b>5</b>
FuzzyMethod	Uniforme Agrupamento	<b>Agrupamento</b>
ActivationMethod	Ativação média Ativação média não-nula	<b>Ativação média não-nula</b>
DefuzzMethod	Centroide Altura ponderada	<b>Altura ponderada</b>

Para averiguar se a abordagem de agrupamento de séries é apropriada, foram concebidas duas abordagens ingênuas que representam espectros extremos: um treinamento usando todas as 862 variáveis e outro usando uma variável por vez, resultando em um problema de previsão univariada. A Tabela 5.11 reúne o resultado obtido para cada um dos três casos. Como era de se esperar, a utilização da técnica de agrupamento obteve um melhor desempenho em relação aos outros dois, reforçando a vantagem de um modelo multivariado neste problema e a necessidade de uma etapa de seleção de variáveis para tornar o aprendizado mais fácil.

Tabela 5.11: Abordagens testadas no e-AutoMFIS para a base de dados Taxa de ocupação de ruas

Método	Resultado (RRSE)
Univariada	0,62
Multivariada (por agrupamento)	0,52
Multivariada (todas as variáveis)	0,89

A avaliação do desempenho do e-AutoMFIS é feita através da comparação de diferentes métodos apresentados em [64], com técnicas da área estatística, modelos tradicionais de ML e DL. Para elucidar estas abordagens, a Tabela 5.12 define as metodologias utilizadas para uma análise comparativa dos resultados.

Com os métodos já definidos, é possível consolidar os resultados de todos os métodos citados, inclusive aqueles do e-AutoMFIS. Para auxiliar a análise comparativa, a Tabela 5.13 reúne as informações importantes sob a perspectiva da acurácia do modelo.

Tabela 5.12: Métodos usados para a base de dados Taxa de ocupação de ruas

Método	Descrição
AR	Modelo autorregressivo
LRidge	Modelo VAR ( <i>Vector autoregression</i> ) com regularização L2
LSVR	Modelo VAR com função de objetivo SVM, proposto em [128]
TRMF	Modelo autorregressivo usando <i>Temporal regularized matrix factorization</i> , proposto em [129]
VAR-MLP	Modelo híbrido usando Redes Neurais MLP ( <i>Multilayer Perceptron</i> ) e método ARIMA, proposto em [130]
RNN-GRU	Modelo RNN usando GRU ( <i>Gated Recurrent Units</i> )
LSTNet-Skip	Modelo proposto por [64] usando camadas skip-RNN
LSTNet-Attn	Modelo proposto por [64] usando camadas de <i>Temporal Attention</i>

Tabela 5.13: Resultado obtido para a base de dados Taxa de ocupação de ruas

Método	Resultado (RRSE)
AR	0,62
LRidge	0,60
LSVR	0,59
TRMF	0,64
VAR-MLP	0,61
RNN-GRU	0,56
LSTNet-Skip	0,49
LSTNet-Attn	0,53
e-AutoMFIS	0,52 $\pm$ 0,07

Em termos de acurácia, o modelo proposto desempenhou bem a tarefa de previsão das 862 séries de taxa de ocupação de ruas. Esta constatação é corroborada pela análise da métrica de avaliação, indicando o e-AutoMFIS como um sistema competitivo na previsão de séries temporais. Dentre os 8 métodos comparados, o SIF desenvolvido neste trabalho obteve, em média, um desempenho melhor que 7 destes, mostrando uma superioridade com relação a técnicas padrões de previsão de séries. Ainda assim, o e-AutoMFIS possui um resultado comparável com o LSTM-Attention, modelo proposto por [64] para a tarefa de modelagem de padrões temporais a curto e longo prazo.

Um dos fatores que certamente contribuiu para o aprendizado adequado e a generalização no processo de previsão é a divisão temporal de dados para o treinamento do método proposto, de modo que os padrões diários e semanais sejam captados corretamente. A Figura 5.10 ilustra o resultado de previsão *multi-step* de 24 horas, para o período de uma semana. A existência de padrões

em intervalos temporais fixos, como períodos diários, semanais e mensais, é bastante recorrente em problemas reais. Portanto, a estratégia usada neste estudo de caso pode ser generalizada para qualquer outro problema em que o aprendizado local de padrões temporais tem um impacto positivo. O único efeito negativo neste procedimento é que a divisão é feita manualmente e, portanto, necessita de uma análise e manipulação do usuário.

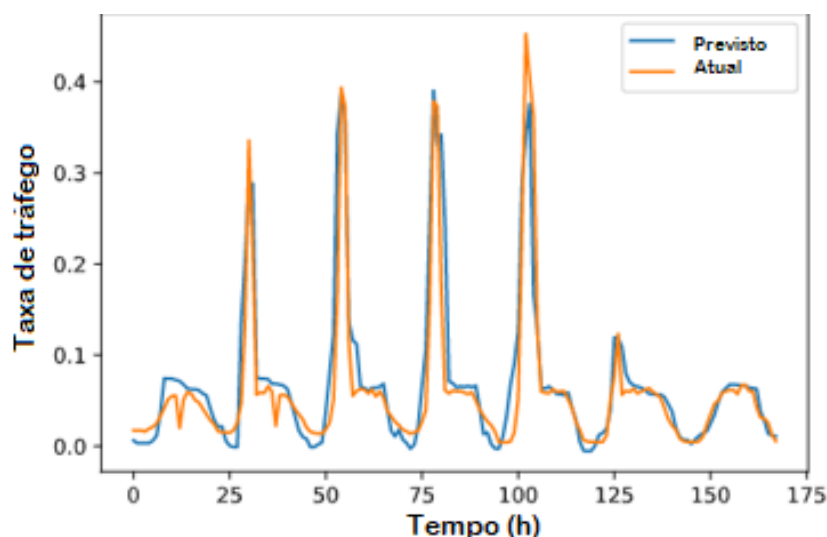


Figura 5.10: Exemplo de previsão para o estudo de caso Taxa de ocupação de ruas

Além da acurácia, é importante considerar o aspecto da interpretabilidade do SIF desenvolvido, de modo a validar o aspecto de explicabilidade do modelo. Para isso, a Tabela 5.14 auxilia com a disposição das métricas de interpretabilidade, obtidas pela média de cada um dos 862 conjuntos de regras do problema. Em um primeiro momento, é possível perceber uma diferença no tamanho da base de regras, bem maior do que mostrado nos estudos de casos anteriores. Este aumento é um efeito adverso da estratégia utilizada para o treinamento, fazendo a concatenação das regras geradas para cada um dos sete dias da semana. Por conta disso, a filtragem definida no escopo deste trabalho não é capaz de eliminar regras que sejam redundantes em dias diferentes, por conta da adição do novo termo linguístico "dia da semana". Ainda assim, este aspecto não influencia diretamente na interpretabilidade humana, pois as regras podem ser facilmente observadas selecionando o dia da semana apropriado.

Por mais que o tamanho da base de regras possa indicar, de forma ingênua, uma dificuldade de análise do conhecimento extraído, as métricas de ativação simultânea de regras e quantidade média de antecedentes indicam uma facilidade na interpretação dos resultados. A ativação média baixa aponta uma boa etapa de filtragem para o mesmo tipo de padrão, aliado a um estágio

Tabela 5.14: Critérios de interpretabilidade do estudo de caso Taxa de ocupação de ruas

Critério	Resultado
Tamanho da base de regras	210
Comprimento dos antecedentes	2,7
Ativação média das regras	3,5

de seleção das regras para a inferência. Além disso, as regras geradas possuem um tamanho médio adequado de antecedentes e sugere que as informações extraídas não são simples demais (com apenas um antecedente) nem muito complexas.

Este experimento é importante para mostrar a capacidade do modelo e-AutoMFIS em lidar com problemas de alta dimensionalidade e modelar diferentes padrões de séries temporais. O resultado em termo de acurácia demonstra a capacidade do SIF desenvolvido de fornecer respostas tão boas quanto aquelas de métodos mais avançados e tradicionais de ML. A interpretabilidade do e-AutoMFIS também foi uma propriedade explorada neste estudo de caso, sendo possível inferir uma potencial facilidade humana em compreender os resultados gerados. Ainda assim, é possível explorar outras metodologias para melhorar a interpretabilidade, como, por exemplo, a fusão de regras.

## 5.4

### Taxa de câmbio

A economia é uma área que possui grande afinidade com previsão de séries temporais. Quer seja no auxílio na definição de políticas públicas do governo ou simplesmente maximizar o lucro de operadores do mercado financeiro, o desenvolvimento de modelos de previsão mais exatos proporciona impactos no cotidiano, como mudanças na estratégia econômica de um país. Ainda que não seja geralmente considerado um objetivo primário, a interpretabilidade em problemas financeiros tem um efeito positivo na compreensão do resultado obtido, inclusive aumentando a credibilidade do ser humano no valor gerado por um modelo matemático.

Com o objetivo de avaliar o comportamento do modelo proposto em um problema mais concreto da área econômica, o quarto estudo de caso contempla a previsão multivariada da taxa de câmbio de países. Assim como o terceiro experimento, a base de dados é organizada por [64]. As informações representam os valores diários da taxa de câmbio de oito países, a saber: Austrália, Grã-Bretanha, Canadá, Suíça, China, Japão, Nova Zelândia e Cingapura. Os dados foram coletados entre 1990 e 2016, totalizando 7588 registros e 8 variá-

veis. Além disso, o horizonte de previsão é igual a 24 (representando a previsão de 24 dias de taxa de câmbio), com divisão de dados entre treinamento (60%), validação (20%) e teste (20%). A métrica de avaliação é o RRSE. Para compreender o comportamento das séries por inspeção visual neste primeiro momento, a Figura 5.11 exibe as séries temporais para o conjunto de treinamento.

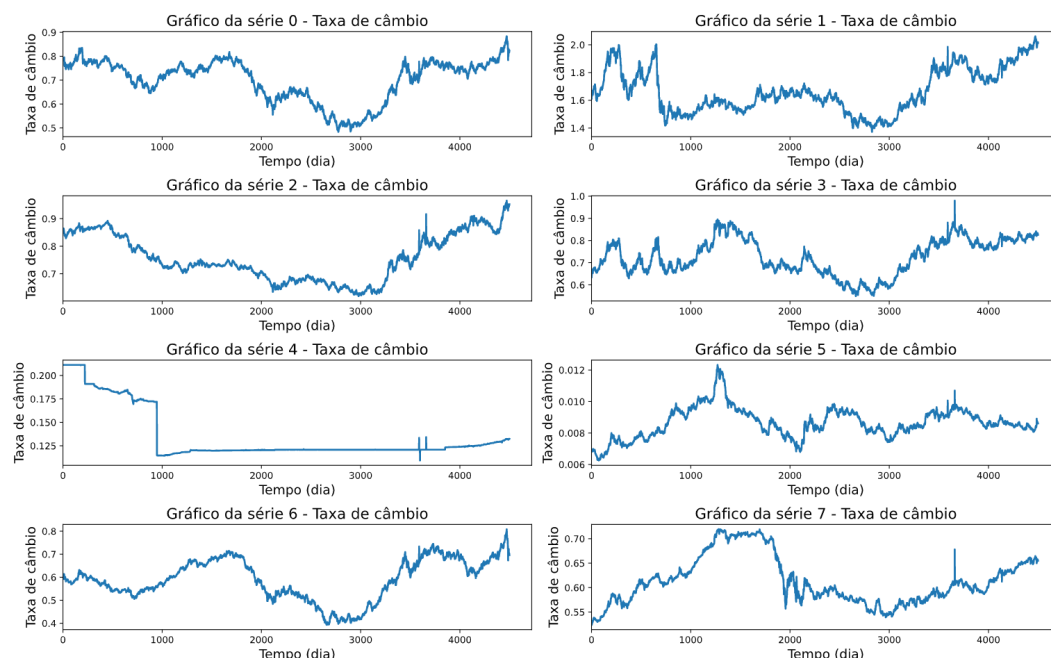


Figura 5.11: Comportamento das séries do estudo de caso - Taxa de Câmbio

Visualmente é possível perceber eventos atípicos que representam uma mudança abrupta da taxa de câmbio. Estes valores, caso não sejam tratados, podem prejudicar a formação dos conjuntos *fuzzy* e, conseqüentemente, a inferência do modelo. As séries também possuem um comportamento mais ruidoso, e portanto o uso da técnica de média móvel torna-se apropriada para suavizar as curvas. Além disso, também foi utilizada a técnica de diferenciação, com o objetivo de assegurar que as séries sejam estacionárias.

Neste experimento, a quantidade de séries disponíveis é bem menor do que as 862 variáveis do terceiro estudo de caso (Seção 5.3), mas ainda assim é importante extrair informações de relações lineares e não-lineares entre as séries de taxa de câmbio para os diferentes países. Como todas as variáveis devem ser previstas, os métodos de pré-processamento podem ser úteis para subdivisão do problema e facilitar a extração de regras. Para mostrar a dependência linear do problema, a Figura 5.12 mostra a matriz de correlação de Pearson.

Como é possível perceber pela matriz de correlação, nem todas as variáveis possuem uma forte dependência linear e, portanto, este resultado serve como um ponto inicial para orientar a subamostragem para o aprendizado da base de regras. Para contribuir no estudo sobre as variáveis, outra técnica



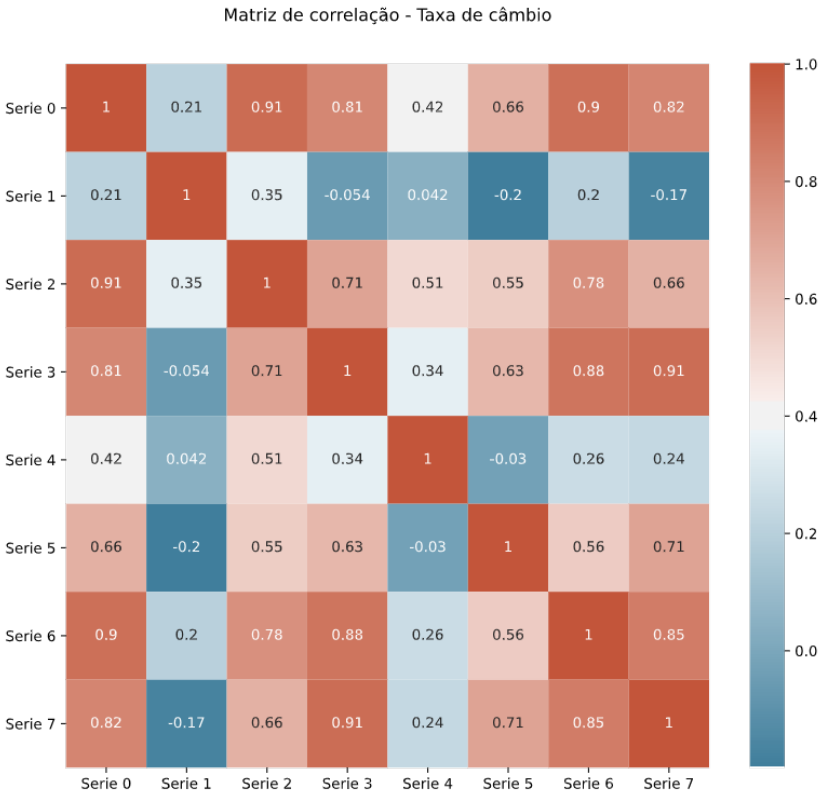


Figura 5.12: Matriz de correlação de Pearson - Taxa de Câmbio

utilizada para ajudar nesta tarefa de pré-processamento é o RFE, método já utilizado no segundo estudo de caso. Como neste estudo de caso não há uma única série alvo, este método é utilizado iterativamente para cada uma das 8 séries.

Os dois métodos de análise de variáveis servem de apoio para agrupar de forma mais adequada as variáveis do grupo e consequentemente reajustar o problema para que o aprendizado seja mais simples e preciso na tarefa de previsão. Este agrupamento será utilizado na etapa de divisão de dados, com a subamostragem das variáveis selecionando apenas as séries e defasagens de um dos dois grupos.

Outra etapa importante para a preparação de dados é a diferenciação de séries. Visualmente, é possível perceber que a dinâmica do comportamento das variáveis pode ser difícil de extrair em sua forma original. Além disso, intuitivamente é mais apropriado redefinir o problema para a previsão da variação diária da taxa de câmbio, sendo uma abordagem amplamente aplicada na área financeira. Por fim, a diferenciação também contribui para reduzir os efeitos não-estacionários das séries, sem comprometer a interpretabilidade semântica do problema.

Com a diferenciação das séries (mostrado na Figura 5.13), é possível evidenciar eventos que geram uma mudança mais acentuada na taxa de

câmbio. Em alguns casos mais raros, estas mudanças são muito acentuadas e podem ser consideradas valores extremos, sendo pouco frequentes neste problema. Especificamente para o modelo e-AutoMFIS, os valores extremos podem prejudicar a formação do dicionário linguístico e, como efeito, dificultar a formação das regras e a tarefa de previsão. Para contornar este problema, os valores considerados extremos (i.e. diferença entre o valor e a média da variável maior que o dobro do desvio padrão) foram substituídos pela média da série, de modo a não prejudicar o treinamento do modelo.

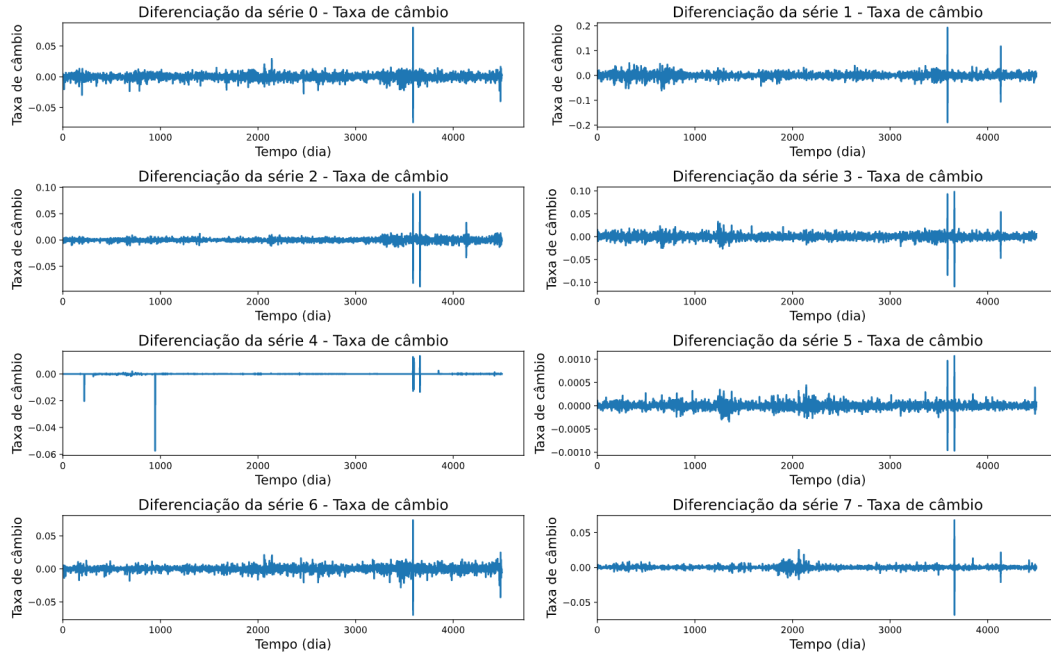


Figura 5.13: Differencição das séries do estudo de caso Taxa de Câmbio

Mesmo após a manipulação de dados propostos na metodologia deste experimento, a distribuição dos dados ainda é bastante estreita, e a extração de conhecimento pode ser enviesada para privilegiar apenas o conjunto com maior densidade de valores como consequente. Para contornar este problema foi utilizada a técnica de normalização z-score, cuja formulação é dada pela Eq. 5-5, reescalando as variáveis para média nula e desvio padrão unitário. É importante notar que, por mais que seja necessário para o aprendizado do e-AutoMFIS, esta normalização dificulta o entendimento imediato do modelo SIF, visto que cada variável agora possui valores diferentes do original. Ainda assim, estes valores podem ser desnormalizados, voltando ao seu domínio de origem e, consequentemente, restaurando o aspecto da interpretabilidade.

$$z_i = \frac{x_i - \bar{x}}{\sigma} \quad (5-5)$$

onde  $z_i$  é o valor normalizado,  $\bar{x}$  é o valor médio da variável  $x$ , e  $\sigma$  é o desvio padrão. Esta técnica atenua o problema exposto quanto à distribuição de dados, ao reescalar a média e desvio padrão.

Realizada a análise e o pré-processamento de dados, a etapa seguinte consiste no ajuste de parâmetros do modelo. A Tabela 5.15 consolida todos os parâmetros com seus respectivos valores. A melhor configuração escolhida é aquela que obteve o menor erro obtido no conjunto de validação na média de 5 inicializações.

Tabela 5.15: Parâmetros testados no estudo de caso Taxa de Câmbio

Parâmetro	Variáveis testadas	Melhor configuração
Lag	16 a 24	<b>24</b>
NumInput	10 a 20	<b>14</b>
FuzzySets	5, 7, 9	<b>7</b>
NumPred	10 a 15	<b>12</b>
MinAct	0,5 a 0,75	<b>0,70</b>
MaxRule	5 a 7	<b>5</b>
ActivationMethod	Ativação média Ativação média não-nula	<b>Ativação média não-nula</b>
DefuzzMethod	Centroide Altura ponderada	<b>Altura ponderada</b>

Por se tratar de um problema originalmente presente no mesmo trabalho que o exposto no terceiro estudo de caso, os métodos selecionados para a análise comparativa estão presentes na Tabela 5.12. Além disso, para facilitar a análise dos resultados, a Tabela 5.16 registra o desempenho de todos os métodos citados, inclusive o desempenho do e-AutoMFIS.

Tabela 5.16: Métodos usados na base de dados Taxa de Câmbio

Método	Resultado (RRSE)
AR	0,0445
LRidge	0,0675
LSVR	0,0662
TRMF	0,0563
VAR-MLP	0,0578
RNN-GRU	0,0626
LSTNet-Skip	0,0449
LSTNet-Attn	0,0590
e-AutoMFIS	0,0613 $\pm$ 0,0081

Em termos de acurácia, é possível concluir que o modelo proposto neste trabalho não obteve um resultado próximo aos melhores resultados obtidos para esta base de dados. Analisando a base de conhecimento gerada, é possível

perceber que a maioria das regras possuíam consequentes iguais (que representa o valor próximo à diferença nula). Consequentemente, os valores previstos na etapa de previsão *multi-stepping* estavam sempre próximos a 0. Por mais que graficamente este fato possa ser observado na Figura 5.13, sem uma diversidade de regras, o modelo não é capaz de captar os momentos em que a variação deveria ser maior, e com isso o erro calculado torna-se maior do que o esperado.

Outro aspecto importante para a avaliação do modelo neste estudo de caso é a interpretabilidade. A Tabela 5.17 mostra o resultado quantitativo dos critérios de interpretabilidade. Algumas considerações podem ser feitas ao analisar esta característica do modelo. Corroborando com o resultado obtido na acurácia, é possível perceber que o modelo e-AutoMFIS possui uma certa limitação para captar a variação neste tipo de problema. As regras geradas neste problema permaneceram genéricas, e mesmo com a etapa de filtragem tentando forçar que as regras tenham uma maior diversidade, este fato não pode ser observado empiricamente neste estudo de caso. Consequentemente, as regras mais extensas tiveram um grau de ativação maior que a ativação mínima, e com isso o critério sobre o comprimento dos antecedentes é maior do que o mesmo para os outros estudos de casos.

Tabela 5.17: Critérios de interpretabilidade do estudo de caso Taxa de Câmbio

Critério	Resultado
Tamanho da base de regras	58 regras/série
Comprimento dos antecedentes	3,5
Ativação média das regras	4,1

Avaliando o desempenho do e-AutoMFIS neste estudo de caso é possível compreender as limitações existentes do modelo. Especificamente em problemas em que a variação a curto prazo não é muito alta, como também ocorre em problemas relacionados à velocidade do vento, o modelo proposto tem uma maior deficiência em extrair eficientemente as regras que sejam úteis para a previsão. Além de investigar novas técnicas de pré-processamento para ajustar este tipo de base de dados (e.g. normalização por partes), também podem ser estudadas novas funcionalidades para o processo de associação e filtragem, visando a aumentar a diversidade dos consequentes das regras.

## 5.5 CIS-Challenge

A área de previsão de séries temporais possui uma gama de funcionalidades no setor energético, como previsão anual de demanda energética ou consumo residencial. Graças à popularização da automação residencial e ao

barateamento de dispositivos como medidores inteligentes, aplicações práticas de análise e previsão do consumo energético residencial ganham popularidade na sociedade e também na área científica [125].

Seguindo esta tendência, o quinto estudo de caso avalia o desempenho do modelo proposto na tarefa de previsão de demanda energética residencial, através da competição hospedada pela *Computational Intelligence Society* do IEEE (CIS-IEEE) [126]. A competição nomeada "*Fuzz-IEEE Competition on Explainable Energy Prediction*"<sup>2</sup> tem como objetivo realizar a previsão mensal, durante um ano, do consumo de energia de 3248 usuários da concessionária britânica. Para isso, são disponibilizados os dados coletados a cada 30 minutos, durante 1 ano, para cada uma das residências, totalizando 17521 registros. A competição também fornece informações auxiliares que podem auxiliar no agrupamento de perfis de consumidores e na previsão de consumo, como tipo de residência, quantidade de quartos e temperatura diária.

O desempenho do modelo é computado como a média entre duas métricas de avaliação. A primeira métrica é o erro RAE (*Relative Absolute Error*) anual, conforme mostra a Eq. 5-6. Além disso, também é calculada a média do erro RAE para cada uma das 3248 residências, matematicamente definida pela Eq. 5-7.

$$year_{RAE} = \frac{\sum_{k=1}^N |y_k - t_k|}{\sum_{k=1}^N |t_k - \bar{t}|} \quad (5-6)$$

$$month_{RAE} = \frac{1}{N} \sum_{k=1}^N \frac{\sum_{i=1}^{12} |y_i - t_i|}{\sum_{i=1}^{12} |t_i - \bar{t}|} \quad (5-7)$$

onde  $y_k$  é o  $k$ -ésimo valor previsto pelo modelo,  $t_k$  é o  $k$ -ésimo valor atual e  $N$  é a quantidade total de elementos. Um dos grandes problemas desta competição é a ausência de valores para diferentes séries na base de dados de consumo energético e de informações complementares. Especificamente na base de dados sobre informações adicionais, apenas as variáveis Tipo de habitação e Número de quartos possuem pelo menos 40% de dados preenchidos (1702 e 1859, respectivamente), enquanto todas as outras informações possuem no máximo 10% de dados presentes. Por conta da ausência elevada desta base de dados, não é possível estabelecer uma metodologia para a definição de um perfil de consumidor baseado em informações socio-econômicas.

<sup>2</sup><https://iee-dataport.org/competitions/fuzz-ieee-competition-explainable-energy-prediction>

A quantidade de dados faltantes sobre o consumo elétrico também é um fator determinante para auxiliar na estratégia de resolução do problema. Como pode ser visto na Figura 5.14, todas as séries possuem valores ausentes em diferentes proporções. As séries que possuem menos do que 2500 valores faltantes possuem lacunas em um pequeno intervalo de tempo (não ultrapassando o período mensal), e portanto estes dados são preenchidos por valores no mesmo dia da semana e horário da semana anterior. Esta etapa de processamento foi implementada em 247 séries temporais.

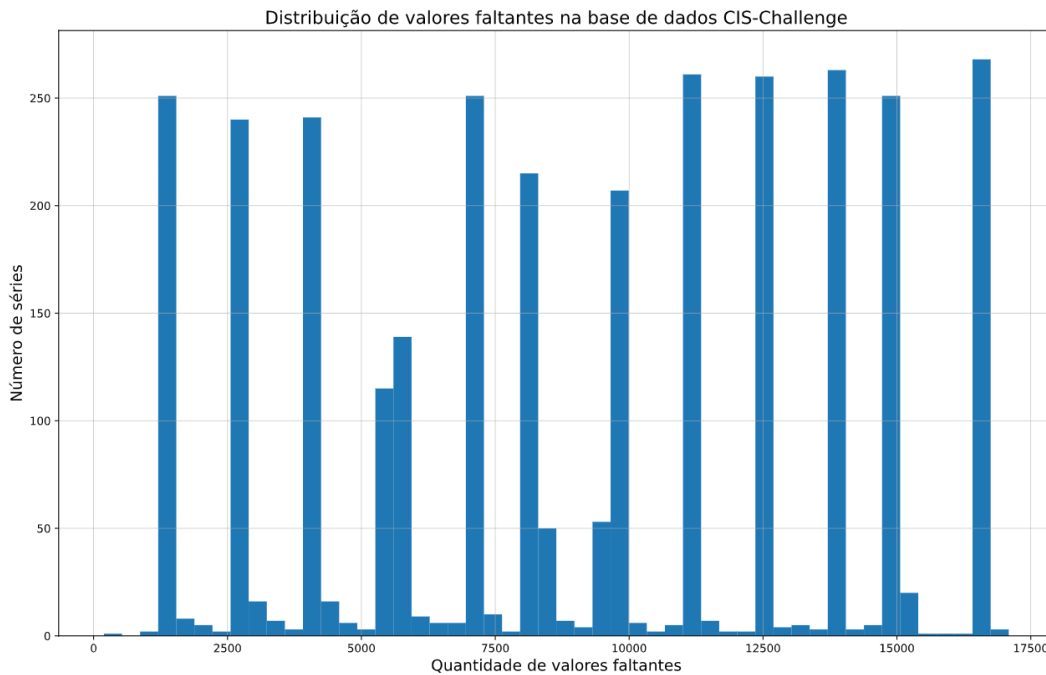


Figura 5.14: Valores faltantes na base de dados CIS-Challenge

O problema de valores ausentes persiste para as 3000 séries restantes. Como as lacunas nestes casos possuem intervalos superiores a um mês, a estratégia comentada anteriormente não é aplicável para essa situação. Um contraponto adicional é que, como o problema disponibiliza apenas o consumo durante um ano, a falta de valores prejudica a análise do consumo mensal da residência. Os valores referentes ao consumo destas 3000 residências, portanto, podem não ser úteis para o treinamento e validação do modelo proposto, visto que podem ser insuficientes para explicar e extrair a dinâmica temporal da série.

Além da inserção de novos valores na base de dados, é perceptível a incompatibilidade entre o intervalo de aquisição dos dados (coletados a cada 30 minutos) e o horizonte de previsão. Caso não seja readequado o intervalo de tempo do problema, seriam necessários 17521 passos de previsão *multi-step* para fornecer o horizonte desejado de um ano, o que é um absurdo sob

o ponto de vista teórico. Entretanto, transformar a base de dados para o intervalo mensal também não é o mais adequado, uma vez que a quantidade limitada para o treinamento (apenas 12 registros por série) pode prejudicar a generalização do modelo. Para contornar este dilema, os dados foram transformados para a representação diária do consumo de energia, servindo como um meio-termo entre o horizonte de previsão e a quantidade de dados disponíveis para ajuste do modelo. Com isso, foi estabelecido que o horizonte de previsão para este problema é de 365, de modo a gerar o resultado para um ano de horizonte.

Observando a característica da base de dados do problema, algumas etapas adicionais de análise e processamento de séries temporais são necessárias para viabilizar o uso do modelo e-AutoMFIS nesta competição. Conforme citado anteriormente, após a imputação de valores faltantes, apenas 247 residências possuem todos os valores disponíveis para o treinamento, de 1/01/2017 até 31/12/2017. Por conta dessa maior disponibilidade de dados, estas séries (classificadas como séries pivôs neste estudo de caso) foram escolhidas para a extração do comportamento do consumo energético. Mesmo assim, uma dúvida ainda permanece: como inferir o consumo mensal do ano de 2018 para as 3000 outras residências que possuem dados faltantes? A solução implementada neste trabalho foi dividida em duas etapas, agrupamento do perfil de consumo energético e identificação das séries mais semelhantes.

A primeira parte da metodologia proposta consiste na seleção das séries com a maior quantidade de registros para realizar a previsão de um ano à frente com maior confiabilidade, visto que talvez seja possível captar as variações de cada mês durante o treinamento. Por conta disso, as 247 séries que possuem todos os valores diários disponíveis durante um ano foram selecionadas. Elas são caracterizadas como "séries pivôs" neste estudo.

Em um primeiro momento de análise do problema, como existe uma quantidade exagerada de séries (cerca de 92%) que não devem ser utilizadas diretamente pelo modelo e-AutoMFIS para o treinamento, é preciso pensar em uma forma de extrapolar as previsões das séries pivôs para todas as 3000 variáveis do problema. Uma proposta que pode ser interessante para este trabalho é realizar uma técnica de agrupamento para tentar extrair diferentes perfis de consumidores. Para isso, foi usado o método de agrupamento K-Means, com distância euclidiana como métrica de avaliação, usando a série inteira com os valores diários de consumo de energia como vetor de entrada. Além disso, para possibilitar que todas as séries estejam em uma mesma escala, todos os valores das séries pivôs foram normalizados usando a Eq. 5-5. Para escolher a quantidade de grupos mais apropriada, foi utilizada a heurística do

cotovelo (*Elbow Method*) [133], iterando de 1 até 12 *clusters*. A Figura 5.15 mostra o resultado das séries que representam os grupos para a quantidade de *clusters* escolhido (6).

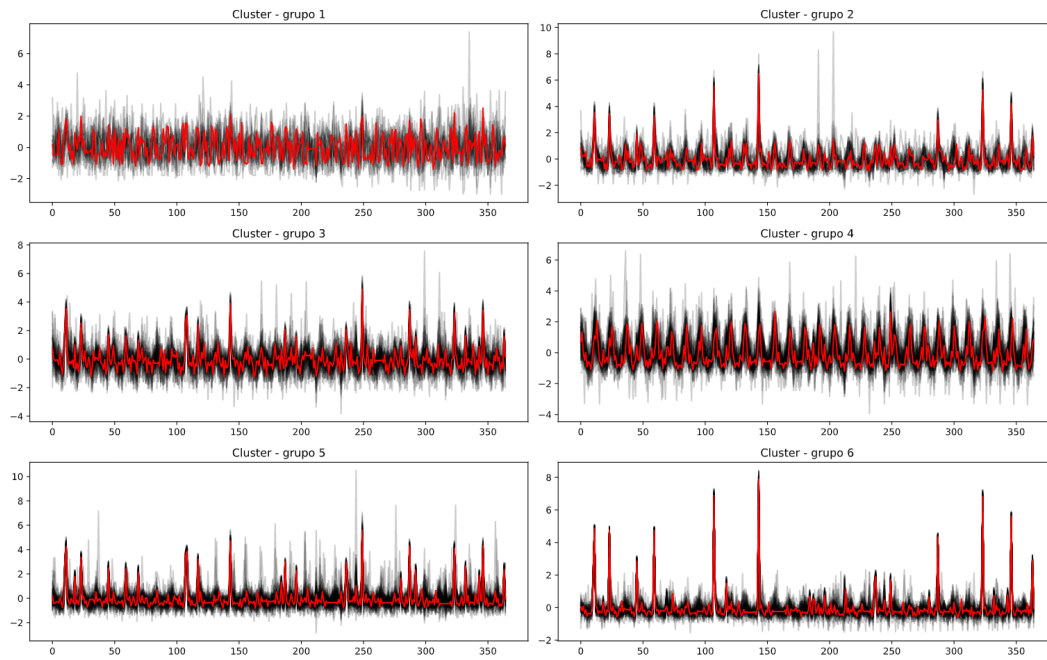


Figura 5.15: Grupos gerados na base de dados CIS-Challenge

Para auxiliar na leitura dos grupos gerados pelo método K-Means, além do gráfico da série que representa o grupo (em vermelho), também foram inseridos os gráficos das séries pivôs em um tom acinzentado. Deste modo, quanto mais séries forem sobrepostas, mais forte é a tonalidade aparente no gráfico. Com base na análise dos gráficos, é possível perceber alguns comportamentos que são recorrentes aos consumidores. O grupo 1, por exemplo, é aparentemente o mais heterogêneo, pois as séries não parecem seguir o mesmo comportamento, com exceção de esporadicamente terem uma pequena variação em relação ao consumo médio. Outros *clusters*, como o grupo 2, 5 e 6, possuem comportamentos próximos, no qual o consumo está geralmente bem próximo do consumo médio, apesar de que em momentos esporádicos, o valor do consumo energético é sujeito a um aumento considerável. As diferenças entre estes três grupos está tanto na amplitude da variação do consumo quanto na frequência em que isto ocorre. O grupo 4 também possui um comportamento bastante interessante a ser analisado. Estas séries possuem variações periódicas bem semelhantes ao longo do tempo, e poucas séries possuem valores muito destoantes do que é representado pelo *cluster* (i.e. existem poucas linhas cinzas diferentes do que é representado pela linha vermelha).

Feita esta primeira análise do problema, já é possível realizar a previsão das séries pivôs usando o modelo e-AutoMFIS. Foram utilizadas duas



abordagens para este tipo de previsão: multivariada usando as informações do agrupamento, e univariada. Como o passo de previsão é diário e o objetivo da competição é avaliar mensalmente os resultados, foi necessário agregar o resultado da previsão (365 valores, que corresponde a um ano de previsão) na escala de tempo mensal. Na ausência de uma base de dados com um período maior para estabelecer um conjunto de validação, a alternativa mais factível desenvolvida neste trabalho foi utilizar uma série pertencente ao mesmo agrupamento (escolhido aleatoriamente). Por mais que não seja a forma mais adequada para a validação do modelo, esta alternativa apresentou-se como uma alternativa razoável. Além disso, por se tratar de um estudo de caso relacionado a competição, as informações sobre o consumo de 2018, que é o horizonte estipulado para a previsão deste trabalho, não são disponibilizadas publicamente, evitando assim que as estratégias utilizadas não sejam sobre-ajustadas para o resultado real do conjunto de teste.

O horizonte de previsão, neste problema, é muito longo e o erro acumulado no processo *multi-step* certamente deve prejudicar o resultado obtido. De modo a tentar suavizar este problema, a cada 30 passos de previsão, a entrada do modelo é substituída pela média entre o vetor de entrada e os valores na base de treinamento referente ao mesmo intervalo de tempo. Nos testes realizados, a abordagem de previsão multivariada gera bons resultados, mas também em alguns casos o resultado é muito destoante do que é apresentado no ano anterior. Em contrapartida, a abordagem univariada gerou resultados mais consistentes, e por conta disso essa abordagem foi escolhida para a análise do resultado. A Figura 5.16 mostra o resultado da previsão de uma série do segundo *cluster* do problema.

Como é possível perceber, o modelo e-AutoMFIS foi capaz de aprender o comportamento básico da série, mas não conseguiu captar quando esse aumento do consumo ocorre. Ainda assim, mensalmente este resultado indica que os valores mensais previstos poderão ser ligeiramente menores do que o que é esperado. Para melhorar ligeiramente a previsão das séries pivôs, o resultado final da previsão de cada mês é calculado como a média entre o valor previsto e a média do ano anterior. Consequentemente, esta estratégia tenta amortecer a variação existente entre os valores mensais de anos anteriores, caso a diferença do resultado seja muito alta. A Tabela 5.18 mostra os testes de configurações realizados, assim como os melhores parâmetros escolhidos.

Após esta primeira etapa de previsão, é preciso definir a estratégia para adequar a inferência do consumo das três mil residências em função dos resultados obtidos pelas séries pivôs. Uma das opções, inclusive escolhida por gerar uma previsão satisfatória, foi utilizar as informações do agrupamento

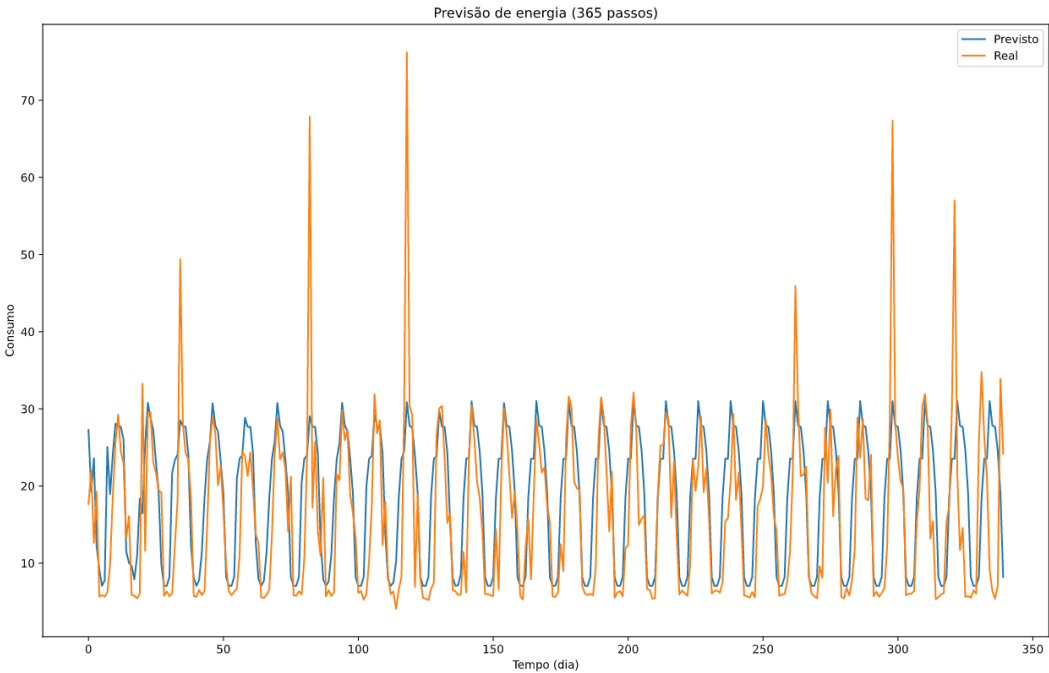


Figura 5.16: Resultado de uma série no estudo de caso CIS-Challenge

Tabela 5.18: Parâmetros testados no estudo de caso Competição CIS

Parâmetro	Variáveis testadas	Melhor configuração
Lag	24 a 30	<b>30</b>
NumInput	5 a 20	<b>15</b>
FuzzySets	5, 7, 9	<b>7</b>
NumPred	10 a 15	<b>15</b>
MinAct	0,65 a 0,75	<b>0,70</b>
MaxRule	5 a 7	<b>5</b>
ActivationMethod	Ativação média Ativação média não-nula	<b>Ativação média não-nula</b>
DefuzzMethod	Centroide Altura ponderada	<b>Altura ponderada</b>

gerado para buscar as séries mais próximas. Como as séries possuem diferentes quantidades de valores faltantes, não é possível aplicar diretamente o resultado gerado pelo K-Means para verificar a qual grupo a série pertence. Ao invés de tentar substituir os valores faltantes usando alguma estratégia, como o preenchimento pela média da série ou repetindo os valores do primeiro mês disponível, a identificação das variáveis mais próximas é feita calculando o erro MAE entre as séries pivôs e cada uma das três mil residências com registros ausentes, computando apenas para os valores presentes. Após o cálculo do erro, são selecionadas as três variáveis que possuem o menor erro calculado e são observados os grupos aos quais pertencem. Caso exista uma maioria, o resultado é inferido como a média da previsão do grupo majoritário e, caso

contrário, é utilizada a previsão da série que tenha o menor erro calculado.

Conforme citado anteriormente, os resultados da previsão estão normalizados e precisam retornar às suas escalas originais. No caso das séries pivôs o procedimento é imediato, visto que tanto a média quanto o desvio padrão já foram calculados. Para as variáveis restantes, é necessário calcular estas medidas estatísticas para reverter o processo de normalização. Deste modo, mesmo que a previsão não utilize diretamente as suas informações temporais, o resultado inferido tem uma escala próxima ao espaço amostral da série.

Exposta a metodologia desenvolvida para este estudo de caso, é possível avaliar o resultado obtido e comparar com outros modelos da literatura. Como a competição ainda está em andamento, poucas entradas de outros competidores foram realizadas, e por conta disso a análise comparativa é bastante reduzida. A Tabela 5.19 reúne os resultados disponibilizados até o presente momento.

Tabela 5.19: Modelos usados para a base de dados Competição CIS

Nome na competição	RAE anual	RAE mensal	RAE total
Benchmark	0.3886	1.6926	1.0406
Sumit Srivastava	0.6461	6.0923	3.3692
e-AutoMFIS	0.4177	1.8654	1.1416

Na ausência de uma diversidade de resultados disponíveis pela competição, o modelo e-AutoMFIS pode ser comparado basicamente com o resultado *Benchmark*, visto que a outra técnica utilizada obteve um resultado muito aquém da abordagem proposta neste estudo de caso. O modelo usado pelo *Benchmark* não é citado, mas o resultado é observado como um valor constante para a previsão mensal de cada uma das séries da competição (que não é o valor médio da série). Como as métricas de erro estão bem próximas, é possível perceber que, no geral, a abordagem utilizada neste estudo de caso é coerente, especialmente na identificação da previsão das três mil séries com valores faltantes. Ainda assim, pode ser pensada futuramente uma nova metodologia para minimização do erro obtido pelo e-AutoMFIS.

## 5.6

### Considerações sobre os estudos de casos

Com a discussão sobre as metodologias utilizadas nos estudos de casos, algumas considerações podem ser feitas para sumarizar as características do modelo proposto. Nos cinco experimentos descritos neste capítulo é possível perceber a competitividade do e-AutoMFIS no processo de previsão, já que ele apresenta métricas de erro próximas aos melhores modelos em cada problema.

Ainda assim, o e-AutoMFIS introduz uma característica muito importante, que é a capacidade de interpretabilidade dos resultados.

Os cinco estudos de casos mostraram de forma explícita que apenas o uso direto do modelo pode não ser o mais apropriado para explorar todo o potencial deste SIF no problema de previsão. De fato, como estas séries temporais possuem particularidades que não foram consideradas inicialmente na concepção da arquitetura do e-AutoMFIS, é imprescindível que o usuário tenha um domínio sobre o assunto para ajustar o modelo, que tem uma proposta de previsão generalizada, para o domínio específico do problema. Como as funcionalidades do sistema são bem definidas, é mais simples realizar qualquer modificação necessária, como foi discutido no terceiro estudo de caso.

Por mais que o modelo trate da síntese automática de SIFs, este fato não exclui a necessidade do usuário preparar previamente os dados e analisá-los, além de realizar qualquer tipo de intervenção manual no treinamento do modelo, caso seja preciso. Mesmo assim, já é pensada uma estratégia para absorver alguns métodos de pré-processamento para o modelo, de modo a automatizar também as etapas mais sensíveis do e-AutoMFIS, como a subamostragem de variáveis.

## 6

### Conclusão e trabalhos futuros

Com o aumento de problemas de alta dimensionalidade, é vantajoso utilizar um modelo que consiga usufruir as características que este tipo de problema fornece. Tendo em mente esta ideia, o presente trabalho teve como proposta desenvolver uma nova metodologia para previsão de séries multivariadas baseada no AutoMFIS, tornando a ideia originalmente concebida apta para tarefas que demandem uma maior complexidade para modelar o problema.

Para validar a nova abordagem desenvolvida neste trabalho, foi investigada na literatura a existência de bases de dados que fossem adequadas para a composição do estudo de caso. No total, cinco problemas, com diferentes naturezas e características, foram selecionados para avaliar o modelo quanto à acurácia e à interpretabilidade. No geral, o e-AutoMFIS apresentou um resultado competitivo em relação a modelos tradicionais de previsão de séries temporais, inclusive baseados em DL. Este resultado é muito importante, pois corrobora a ideia de que nem sempre é necessário um modelo altamente complexo para produzir bons resultados.

Diferentemente da avaliação em termos da acurácia, a interpretabilidade é um aspecto bastante discutido nos estudos de caso, mas sua análise ainda é um tópico aberto. Por mais que fosse discutido sobre os resultados dos critérios de interpretabilidade, a análise comparativa tem um papel fundamental de compreender a complexidade do problema e auxiliar na avaliação quantitativa. Portanto, este tópico ainda necessita de uma investigação mais profunda sob esta perspectiva para confirmar a análise realizada.

A adaptação proposta na arquitetura para tornar o aprendizado deste SIF distribuído por meio de uma abordagem *Ensemble* também se mostrou bastante vantajosa, se comparado com o modelo originalmente proposto em [2]. A mudança no aprendizado provoca uma diversidade na extração das regras, visto que a subamostragem e a divisão temporal da base de dados possibilitam explorar localmente as características mais importantes. Esta nova metodologia soluciona as restrições inerentes ao modelo original quanto ao limite de variáveis e defasagens, além do custo computacional necessário para o treinamento do sistema. Este fato pode ser observado pelo primeiro estudo

de caso, onde foi possível observar uma melhor exploração dos parâmetros e consequentemente um melhor desempenho, se comparado com o autoMFIS.

Outro ponto relevante observado neste trabalho é a dificuldade de otimização do modelo proposto. Uma grande quantidade de parâmetros inviabiliza a busca exaustiva por todas as combinações possíveis, sendo necessário um maior estudo prévio sobre o problema para guiar o alcance das configurações a serem testadas. Além disso, alguns parâmetros são muito sensíveis em função da estabilidade entre a interpretabilidade e complexidade. Um ajuste pequeno em Ativação Mínima (empiricamente, mudanças de 0,05 ou 0,1) podem mudar completamente a dimensão da base de regras, tornando o problema simplificado demais ou demasiadamente volumoso, diminuindo a capacidade de interpretabilidade. Com isso, é essencial avaliar a cada mudança na configuração do e-autoMFIS o efeito gerado em termos de acurácia e complexidade, de modo a buscar um equilíbrio entre estes termos.

O modelo proposto neste trabalho é um passo inicial para tornar esta técnica de síntese automática de SIFs mais robusta e computacionalmente mais rentável para ser utilizada. Ainda assim, o e-autoMFIS possui uma gama de tópicos abertos que podem contribuir para o desenvolvimento de características que podem facilitar o seu uso em diferentes bases de dados. Para tornar esta explicação mais concisa, os trabalhos futuros podem ser divididos em duas linhas principais: análise das características do modelo e modificações no *Ensemble*.

A análise feita neste trabalho visou a compreender o comportamento do modelo e-autoMFIS em diferentes bases de dados, avaliando não somente a acurácia e interpretabilidade nos estudos de caso, mas também como os parâmetros podem influenciar no resultado final. Tanto o aspecto da interpretabilidade quanto a análise de sensibilidade dos parâmetros ainda são tópicos a serem discutidos, de modo a enriquecer o entendimento do modelo proposto. Para isso, o modelo e-autoMFIS poderia ser testado em problemas de previsão de séries temporais que tenham sido modelados através de SIFs, tornando a comparação mais completa em termos de interpretabilidade.

Como trabalho futuro, também é proposto um novo ajuste na funcionalidade do e-autoMFIS, especialmente na abordagem *Ensemble*. No geral, a divisão temporal e dimensional da base de dados é feita de forma aleatória, com exceção de casos em que estas etapas foram ajustadas manualmente para melhorar o resultado obtido. Para tornar esta característica mais proveitosa e eliminar a intervenção manual nesta etapa, a subamostragem da base de dados pode ser orientada por diferentes métricas existentes na literatura. Para a subamostragem de variáveis, por exemplo, as métricas de correlação cruzada e

autocorrelação podem ser usadas como ponderação para a composição do subconjunto, privilegiando as variáveis com maior dependência linear e ao mesmo tempo possibilitando a interação com outras.

Finalmente, a busca por uma configuração mais adequada pode ser implementada usando algoritmos mais sofisticados de otimização, como Algoritmo Genético. Além de ser uma estratégia mais inteligente do que uma busca exaustiva, a seleção dos parâmetros também pode ser extremamente útil para determinar o desejado equilíbrio entre acurácia e interpretabilidade, desde que seja usada uma função de aptidão adequada para o problema.

## Referências bibliográficas

- [1] RAO, M.. An imagined reality: Malthusianism, neo-malthusianism and population myth. *Economic and Political Weekly*, p. PE40–PE52, 1994.
- [2] COUTINHO, J. R.; TANSCHKEIT, R.; VELLASCO, M. ; KOSHIYAMA, A.. **Automfis: Fuzzy inference system for multivariate time series forecasting**. In: 2016 IEEE INTERNATIONAL CONFERENCE ON FUZZY SYSTEMS (FUZZ-IEEE), p. 2120–2127. IEEE, 2016.
- [3] LEI, M.; SHIYAN, L.; CHUANWEN, J.; HONGLING, L. ; YAN, Z.. A review on the forecasting of wind speed and generated power. *Renewable and Sustainable Energy Reviews*, 13(4):915–920, 2009.
- [4] PEARRE, N. S.; SWAN, L. G.. Statistical approach for improved wind speed forecasting for wind power production. *Sustainable Energy Technologies and Assessments*, 27:180–191, 2018.
- [5] PARMEZAN, A. R. S.; SOUZA, V. M. ; BATISTA, G. E.. Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model. *Information sciences*, 484:302–337, 2019.
- [6] DODGE, Y.. **The concise encyclopedia of statistics**. Springer Science & Business Media, 2008.
- [7] CHATFIELD, C.. **Time-series forecasting**. CRC press, 2000.
- [8] TASKAYA-TEMIZEL, T.; CASEY, M. C.. A comparative study of autoregressive neural network hybrids. *Neural Networks*, 18(5-6):781–789, 2005.
- [9] QI, M.; ZHANG, G. P.. Trend time-series modeling and forecasting with neural networks. *IEEE Transactions on Neural Networks*, 19(5):808–816, 2008.
- [10] HARVEY, A. C.. Trends and cycles in macroeconomic time series. *Journal of Business & Economic Statistics*, 3(3):216–227, 1985.



- [11] ADHIKARI, R.; AGRAWAL, R. K.. **An introductory study on time series modeling and forecasting.** arXiv preprint arXiv:1302.6613, 2013.
- [12] FUNG, W.; LAM, K. S.; HUNG, W.; PANG, S. ; LEE, Y.. **Impact of urban temperature on energy consumption of hong kong.** Energy, 31(14):2623–2637, 2006.
- [13] GRANGER, C. W. J.; NEWBOLD, P.. **Forecasting economic time series.** Academic Press, 2014.
- [14] SWANSON, N. R.; WHITE, H.. **A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks.** Review of Economics and Statistics, 79(4):540–550, 1997.
- [15] HARVEY, N.. **Why are judgments less consistent in less predictable task situations?** Organizational Behavior and Human Decision Processes, 63(3):247–263, 1995.
- [16] HOGARTH, R. M.; MAKRIDAKIS, S.. **Forecasting and planning: An evaluation.** Management science, 27(2):115–138, 1981.
- [17] LAWRENCE, M.; GOODWIN, P.; O'CONNOR, M. ; ÖNKAL, D.. **Judgmental forecasting: A review of progress over the last 25 years.** International Journal of forecasting, 22(3):493–518, 2006.
- [18] FILDES, R.; GOODWIN, P.; LAWRENCE, M. ; NIKOLOPOULOS, K.. **Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning.** International Journal of Forecasting, 25(1):3–23, 2009.
- [19] GRANGER, C. W.; JEON, Y.. **Long-term forecasting and evaluation.** International Journal of Forecasting, 23(4):539–551, 2007.
- [20] LORENC, A. C.. **Analysis methods for numerical weather prediction.** Quarterly Journal of the Royal Meteorological Society, 112(474):1177–1194, 1986.
- [21] SINGH, S.; MOHAPATRA, A. ; OTHERS. **Repeated wavelet transform based arima model for very short-term wind speed forecasting.** Renewable energy, 136:758–768, 2019.
- [22] LEE, D.; PARK, Y.-G.; PARK, J.-B. ; ROH, J. H.. **Very short-term wind power ensemble forecasting without numerical weather**

- prediction through the predictor design.** *Journal of Electrical Engineering and Technology*, 12(6):2177–2186, 2017.
- [23] POTTER, C. W.; NEGNEVITSKY, M.. **Very short-term wind forecasting for tasmanian power generation.** *IEEE Transactions on Power Systems*, 21(2):965–972, 2006.
- [24] NASON, G. P.. **Stationary and non-stationary time series.** *Statistics in Volcanology*, 60, 2006.
- [25] MADDALA, G. S.; LAHIRI, K.. **Introduction to econometrics**, volumen 4. Macmillan New York, 2009.
- [26] R ROSCA, E.. **Stationary and non-stationary time series.** *The USV Annals of Economics and Public Administration*, 10(1):177–186, 2011.
- [27] FRISTON, K.; JOSEPHS, O.; ZARAHN, E.; HOLMES, A.; ROUQUETTE, S. ; POLINE, J.-B.. **To smooth or not to smooth?: Bias and efficiency in fmri time-series analysis.** *Neuroimage*, 12(2):196–208, 2000.
- [28] CORTÉS-IBÁÑEZ, J. A.; GONZÁLEZ, S.; VALLE-ALONSO, J. J.; LU-ENGO, J.; GARCÍA, S. ; HERRERA, F.. **Preprocessing methodology for time series: An industrial world application case study.** *Information Sciences*, 514:385–401, 2020.
- [29] BRYCE, R.; SPRAGUE, K.. **Revisiting detrended fluctuation analysis.** *Scientific reports*, 2(1):1–6, 2012.
- [30] BASHAN, A.; BARTSCH, R.; KANTELHARDT, J. W. ; HAVLIN, S.. **Comparison of detrending methods for fluctuation analysis.** *Physica A: Statistical Mechanics and its Applications*, 387(21):5080–5090, 2008.
- [31] PODOBNIK, B.; STANLEY, H. E.. **Detrended cross-correlation analysis: a new method for analyzing two nonstationary time series.** *Physical Review Letters*, 100(8):084102, 2008.
- [32] DERPANIS, K. G.. **Overview of the ransac algorithm.** *Image Rochester NY*, 4(1):2–3, 2010.
- [33] RAY, B. K.. **Long-range forecasting of ibm product revenues using a seasonal fractionally differenced arma model.** *International Journal of Forecasting*, 9(2):255–269, 1993.

- [34] PAL, A.; PRAKASH, P.. **Practical time series analysis: master time series data processing, visualization, and modeling using Python**. Packt Publishing Ltd, 2017.
- [35] GUJARATI, D. N.; PORTER, D. C.. **Econometria básica-5**. Amgh Editora, 2011.
- [36] WHITTLE, P.. **Hypothesis testing in time series analysis**, volumen 4. Almqvist & Wiksells boktr., 1951.
- [37] BOX, G. E.; JENKINS, G. M.; REINSEL, G. C. ; LJUNG, G. M.. **Time series analysis: forecasting and control**. John Wiley & Sons, 2015.
- [38] MARTINEZ, E. Z.; SILVA, E. A. S. D. ; FABBRO, A. L. D.. **A sarima forecasting model to predict the number of cases of dengue in campinas, state of são paulo, brazil**. *Revista da Sociedade Brasileira de Medicina Tropical*, 44(4):436–440, 2011.
- [39] SHU, Y.; JIN, Z.; ZHANG, L.; WANG, L. ; YANG, O. W.. **Traffic prediction using farima models**. In: 1999 IEEE INTERNATIONAL CONFERENCE ON COMMUNICATIONS (CAT. NO. 99CH36311), volumen 2, p. 891–895. IEEE, 1999.
- [40] NYONI, T.. **Modeling and forecasting naira/usd exchange rate in nigeria: a box-jenkins arima approach**, 2018.
- [41] DE OLIVEIRA, E. M.; OLIVEIRA, F. L. C.. **Forecasting mid-long term electric energy consumption through bagging arima and exponential smoothing methods**. *Energy*, 144:776–788, 2018.
- [42] BENVENUTO, D.; GIOVANETTI, M.; VASSALLO, L.; ANGELETTI, S. ; CICCIOZZI, M.. **Application of the arima model on the covid-2019 epidemic dataset**. *Data in brief*, 29:105340, 2020.
- [43] PERONE, G.. **An arima model to forecast the spread of covid-2019 epidemic in italy**. *arXiv preprint arXiv:2004.00382*, 2020.
- [44] REBANE, J.; KARLSSON, I.; PAPAPETROU, P. ; DENIC, S.. **Seq2seq rnns and arima models for cryptocurrency prediction: A comparative study**. In: SIGKDD FINTECH'18, LONDON, UK, AUGUST 19-23, 2018, 2018.
- [45] SIAMI-NAMINI, S.; NAMIN, A. S.. **Forecasting economics and financial time series: Arima vs. lstm**. *arXiv preprint arXiv:1803.06386*, 2018.

- [46] SINGH, S.; PARMAR, K. S.; MAKKHAN, S. J. S.; KAUR, J.; PESHORIA, S. ; KUMAR, J.. **Study of arima and least square support vector machine (ls-svm) models for the prediction of sars-cov-2 confirmed cases in the most affected countries.** *Chaos, Solitons & Fractals*, 139:110086, 2020.
- [47] ORDÓÑEZ, C.; LASHERAS, F. S.; ROCA-PARDINAS, J. ; DE COS JUEZ, F. J.. **A hybrid arima-svm model for the study of the remaining useful life of aircraft engines.** *Journal of Computational and Applied Mathematics*, 346:184–191, 2019.
- [48] WANG, P.; ZHANG, H.; QIN, Z. ; ZHANG, G.. **A novel hybrid-garch model based on arima and svm for pm2. 5 concentrations forecasting.** *Atmospheric Pollution Research*, 8(5):850–860, 2017.
- [49] LIU, J.; TAN, X. ; WANG, Y.. **Cssap: software aging prediction for cloud services based on arima-lstm hybrid model.** In: 2019 IEEE INTERNATIONAL CONFERENCE ON WEB SERVICES (ICWS), p. 283–290. IEEE, 2019.
- [50] CHOI, H. K.. **Stock price correlation coefficient prediction with arima-lstm hybrid model.** *arXiv preprint arXiv:1808.01560*, 2018.
- [51] KHASHEI, M.; HAJIRAHIMI, Z.. **A comparative study of series arima/mlp hybrid models for stock price forecasting.** *Communications in Statistics-Simulation and Computation*, 48(9):2625–2640, 2019.
- [52] BÜYÜKŞAHİN, Ü. Ç.; ERTEKİN, Ş.. **Improving forecasting accuracy of time series data using a new arima-ann hybrid method and empirical mode decomposition.** *Neurocomputing*, 361:151–163, 2019.
- [53] FAULINA, R.. **Hybrid arima-anfis for rainfall prediction in indonesia**, 2017.
- [54] CYBENKO, G.. **Approximation by superpositions of a sigmoidal function.** *Mathematics of Control, Signals and Systems*, 2(4):303–314, 1989.
- [55] SANTINI, M.; TETTAMANZI, A.. **Genetic programming for financial time series prediction.** In: EUROPEAN CONFERENCE ON GENETIC PROGRAMMING, p. 361–370. Springer, 2001.
- [56] ABDULHAI, B.; PORWAL, H. ; RECKER, W.. **Short-term traffic flow prediction using neuro-genetic algorithms.** *ITS Journal-Intelligent Transportation Systems Journal*, 7(1):3–41, 2002.

- [57] ADNAN, R. M.; YUAN, X.; KISI, O.; ADNAN, M. ; MEHMOOD, A.. Stream flow forecasting of poorly gauged mountainous watershed by least square support vector machine, fuzzy genetic algorithm and m5 model tree using climatic data from nearby station. *Water Resources Management*, 32(14):4469–4486, 2018.
- [58] NAYAK, P. C.; SUDHEER, K.; RANGAN, D. ; RAMASASTRI, K.. A neuro-fuzzy computing technique for modeling hydrological time series. *Journal of Hydrology*, 291(1-2):52–66, 2004.
- [59] CHAKRABORTY, K.; MEHROTRA, K.; MOHAN, C. K. ; RANKA, S.. Forecasting the behavior of multivariate time series using neural networks. *Neural networks*, 5(6):961–970, 1992.
- [60] KAASTRA, I.; BOYD, M.. Designing a neural network for forecasting financial and economic time series. *Neurocomputing*, 10(3):215–236, 1996.
- [61] KHURANA, D.; KOLI, A.; KHATTER, K. ; SINGH, S.. Natural language processing: State of the art, current trends and challenges. *arXiv preprint arXiv:1708.05148*, 2017.
- [62] POPEL, M.; TOMKOVA, M.; TOMEK, J.; KAISER, Ł.; USZKOREIT, J.; BOJAR, O. ; ŽABOKRTSKÝ, Z.. Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals. *Nature communications*, 11(1):1–15, 2020.
- [63] WANG, Y.; LIAO, W. ; CHANG, Y.. Gated recurrent unit network-based short-term photovoltaic forecasting. *Energies*, 11(8):2163, 2018.
- [64] LAI, G.; CHANG, W.-C.; YANG, Y. ; LIU, H.. Modeling long-and short-term temporal patterns with deep neural networks. In: *THE 41ST INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH & DEVELOPMENT IN INFORMATION RETRIEVAL*, p. 95–104, 2018.
- [65] TASCIKARAOGLU, A.; SANANDAJI, B. M.; CHICCO, G.; COCINA, V.; SPERTINO, F.; ERDINC, O.; PATERAKIS, N. G. ; CATALÃO, J. P.. Compressive spatio-temporal forecasting of meteorological quantities and photovoltaic power. *IEEE Transactions on Sustainable Energy*, 7(3):1295–1305, 2016.

- [66] DOWELL, J.; PINSON, P.. **Very-short-term probabilistic wind power forecasts by sparse vector autoregression**. *IEEE Transactions on Smart Grid*, 7(2):763–770, 2015.
- [67] PERSSON, C.; BACHER, P.; SHIGA, T. ; MADSEN, H.. **Multi-site solar power forecasting using gradient boosted regression trees**. *Solar Energy*, 150:423–436, 2017.
- [68] BASTOS, B. Q.; OLIVEIRA, F. L. C. ; MILIDIÚ, R. L.. **U-convolutional model for spatio-temporal wind speed forecasting**. *International Journal of Forecasting*, 2020.
- [69] LIVIERIS, I. E.; PINTELAS, E. ; PINTELAS, P.. **A cnn-lstm model for gold price time-series forecasting**. *Neural computing and applications*, 32(23):17351–17360, 2020.
- [70] VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER, L. ; POLOSUKHIN, I.. **Attention is all you need**. *arXiv preprint arXiv:1706.03762*, 2017.
- [71] WU, N.; GREEN, B.; BEN, X. ; O'BANION, S.. **Deep transformer models for time series forecasting: The influenza prevalence case**. *arXiv preprint arXiv:2001.08317*, 2020.
- [72] LI, S.; JIN, X.; XUAN, Y.; ZHOU, X.; CHEN, W.; WANG, Y.-X. ; YAN, X.. **Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting**. *arXiv preprint arXiv:1907.00235*, 2019.
- [73] LIM, B.; ARIK, S. O.; LOEFF, N. ; PFISTER, T.. **Temporal fusion transformers for interpretable multi-horizon time series forecasting**. *arXiv preprint arXiv:1912.09363*, 2019.
- [74] KOSKO, B.. **Fuzzy systems as universal approximators**. *IEEE transactions on computers*, 43(11):1329–1333, 1994.
- [75] ROUBOS, J. A.; SETNES, M. ; ABONYI, J.. **Learning fuzzy classification rules from labeled data**. *Information sciences*, 150(1-2):77–93, 2003.
- [76] NGUYEN, A.-T.; TANIGUCHI, T.; ECIOLAZA, L.; CAMPOS, V.; PALHARES, R. ; SUGENO, M.. **Fuzzy control systems: Past, present and future**. *IEEE Computational Intelligence Magazine*, 14(1):56–68, 2019.

- [77] BOSE, M.; MALI, K.. Designing fuzzy time series forecasting models: A survey. *International Journal of Approximate Reasoning*, 111:78–99, 2019.
- [78] WANG, L.-X.; MENDEL, J. M.. Generating fuzzy rules by learning from examples. *IEEE Transactions on systems, man, and cybernetics*, 22(6):1414–1427, 1992.
- [79] NOZAKI, K.; ISHIBUCHI, H. ; TANAKA, H.. A simple but powerful heuristic method for generating fuzzy rules from numerical data. *Fuzzy sets and systems*, 86(3):251–270, 1997.
- [80] HERRERA, F.; LOZANO, M. ; VERDEGAY, J. L.. Generating fuzzy rules from examples using genetic algorithms. In: *FUZZY LOGIC AND SOFT COMPUTING*, p. 11–20. World Scientific, 1995.
- [81] ALCALÁ-FDEZ, J.; ALONSO, J. M.. A survey of fuzzy systems software: Taxonomy, current research trends, and prospects. *IEEE Transactions on Fuzzy Systems*, 24(1):40–56, 2015.
- [82] GACTO, M. J.; ALCALÁ, R. ; HERRERA, F.. Interpretability of linguistic fuzzy rule-based systems: An overview of interpretability measures. *Information Sciences*, 181(20):4340–4360, 2011.
- [83] MILLER, G. A.. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, 63:81–97, 1956.
- [84] ALCALÁ, R.; NOJIMA, Y.; HERRERA, F. ; ISHIBUCHI, H.. Multiobjective genetic fuzzy rule selection of single granularity-based fuzzy classification rules and its interaction with the lateral tuning of membership functions. *Soft Computing*, 15(12):2303–2318, 2011.
- [85] ALONSO, J. M.; MAGDALENA, L.. Special issue on interpretable fuzzy systems, 2011.
- [86] CASILLAS, J.; CORDÓN, O.; TRIGUERO, F. H. ; MAGDALENA, L.. Interpretability issues in fuzzy modeling, volumen 128. Springer, 2013.
- [87] DOS PRAZERES TEIXEIRA, T.. Sistema de Inferência Fuzzy com Extração Automática de Regras para Diagnóstico de Desempenho de Turbinas a Gás Aeronáuticas. PhD thesis, PUC-Rio, 2016.

- [88] KETATA, R.; BELLAJ, H.; CHTOUROU, M. ; AMER, M. B.. **Adjustment of membership functions, generation and reduction of fuzzy rule base from numerical data.** Malaysian Journal of Computer Science, 20(2):147–169, 2017.
- [89] ALCALÁ, R.; GACTO, M. J.; HERRERA, F. ; ALCALÁ-FDEZ, J.. **A multi-objective genetic algorithm for tuning and rule selection to obtain accurate and compact linguistic fuzzy rule-based systems.** International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 15(05):539–557, 2007.
- [90] YANG, M.-S.. **A survey of fuzzy clustering.** Mathematical and Computer modelling, 18(11):1–16, 1993.
- [91] SHEN, L.; CHEN, J.; ZENG, Z.; YANG, J. ; JIN, J.. **A novel echo state network for multivariate and nonlinear time series prediction.** Applied Soft Computing, 62:524–535, 2018.
- [92] SAGHEER, A.; KOTB, M.. **Time series forecasting of petroleum production using deep lstm recurrent networks.** Neurocomputing, 323:203–213, 2019.
- [93] MENDES, R. R.; DE VOZNIKA, F. B.; FREITAS, A. A. ; NIEVOLA, J. C.. **Discovering fuzzy classification rules with genetic programming and co-evolution.** In: EUROPEAN CONFERENCE ON PRINCIPLES OF DATA MINING AND KNOWLEDGE DISCOVERY, p. 314–325. Springer, 2001.
- [94] GACTO, M. J.; ALCALÁ, R. ; HERRERA, F.. **A multiobjective evolutionary algorithm for tuning fuzzy rule based systems with measures for preserving interpretability.** In: IFSA/EUSFLAT CONF., p. 1146–1151, 2009.
- [95] SMITH, S.. **A learning system based on genetic adaptive algorithms [phd dissertation].** University of Pittsburgh, 1980.
- [96] KOSHIYAMA, A. S.; VELLASCO, M. M. ; TANSCHKEIT, R.. **Gpfis-control: a genetic fuzzy system for control tasks.** Journal of Artificial Intelligence and Soft Computing Research, 4(3):167–179, 2014.
- [97] LUKOSEVICIUTE, K.; RAGULSKIS, M.. **Evolutionary algorithms for the selection of time lags for time series forecasting by fuzzy inference systems.** Neurocomputing, 73(10-12):2077–2088, 2010.



- [98] SARICA, B.; EĞRIOĞLU, E. ; AŞIKGİL, B.. **A new hybrid method for time series forecasting: Ar-anfis**. *Neural Computing and Applications*, 29(3):749–760, 2018.
- [99] WEI, L.-Y.. **A hybrid anfis model based on empirical mode decomposition for stock time series forecasting**. *Applied Soft Computing*, 42:368–376, 2016.
- [100] CHENG, C.-H.; WEI, L.-Y.. **One step-ahead anfis time series model for forecasting electricity loads**. *Optimization and Engineering*, 11(2):303–317, 2010.
- [101] MELIN, P.; SOTO, J.; CASTILLO, O. ; SORIA, J.. **A new approach for time series prediction using ensembles of anfis models**. *Expert Systems with Applications*, 39(3):3494–3506, 2012.
- [102] GOMIDE, F.. **Evolving Granular Neural Networks From Data Streams**, p. 1–9. American Cancer Society, 2017.
- [103] ŠKRJANC, I.; IGLESIAS, J. A.; SANCHIS, A.; LEITE, D.; LUGHOFFER, E. ; GOMIDE, F.. **Evolving fuzzy and neuro-fuzzy approaches in clustering, regression, identification, and classification: a survey**. *Information Sciences*, 490:344–368, 2019.
- [104] SOARES, E.; COSTA JR, P.; COSTA, B. ; LEITE, D.. **Ensemble of evolving data clouds and fuzzy models for weather time series prediction**. *Applied Soft Computing*, 64:445–453, 2018.
- [105] ANGELOV, P.; FILEV, D.. **Simpl\_ets: A simplified method for learning evolving takagi-sugeno fuzzy models**. In: THE 14TH IEEE INTERNATIONAL CONFERENCE ON FUZZY SYSTEMS, 2005. FUZZ'05., p. 1068–1073. IEEE, 2005.
- [106] LEITE, D.; GOMIDE, F.; BALLINI, R. ; COSTA, P.. **Fuzzy granular evolving modeling for time series prediction**. In: 2011 IEEE INTERNATIONAL CONFERENCE ON FUZZY SYSTEMS (FUZZ-IEEE 2011), p. 2794–2801. IEEE, 2011.
- [107] JIANG, P.; YANG, H. ; HENG, J.. **A hybrid forecasting system based on fuzzy time series and multi-objective optimization for wind speed forecasting**. *Applied energy*, 235:786–801, 2019.
- [108] DAMOUSIS, I. G.; ALEXIADIS, M. C.; THEOCHARIS, J. B. ; DOKOPOULOS, P. S.. **A fuzzy model for wind speed prediction and power**

- generation in wind parks using spatial correlation. *IEEE Transactions on Energy Conversion*, 19(2):352–361, 2004.
- [109] CHEN, M.-Y.. **A high-order fuzzy time series forecasting model for internet stock trading.** *Future Generation Computer Systems*, 37:461–467, 2014.
- [110] SINGH, P.; BORAH, B.. **Forecasting stock index price based on m-factors fuzzy time series and particle swarm optimization.** *International Journal of Approximate Reasoning*, 55(3):812–833, 2014.
- [111] KIM, H.; KIM, H.; MOON, H. ; AHN, H.. **A weight-adjusted voting algorithm for ensembles of classifiers.** *Journal of the Korean Statistical Society*, 40:437–449, 2011.
- [112] YAN, K.; ZHANG, D.. **Feature selection and analysis on correlated gas sensor data with recursive feature elimination.** *Sensors and Actuators B: Chemical*, 212:353–363, 2015.
- [113] MENCAR, C.; LUCARELLI, M.; CASTIELLO, C. ; MARIA, F. A.. **Design of strong fuzzy partitions from cuts.** In: 8TH CONFERENCE OF THE EUROPEAN SOCIETY FOR FUZZY LOGIC AND TECHNOLOGY (EUSFLAT-13). Atlantis Press, 2013.
- [114] LIU, Z.; HU, B.; WANG, L.; WU, F.; GAO, W. ; WANG, Y.. **Seasonal and diurnal variation in particulate matter (pm 10 and pm 2.5) at an urban site of beijing: analyses from a 9-year study.** *Environmental Science and Pollution Research*, 22(1):627–642, 2015.
- [115] JALLAL, M. A.; GONZALEZ-VIDAL, A.; SKARMETA, A. F.; CHABAA, S. ; ZEROUAL, A.. **A hybrid neuro-fuzzy inference system-based algorithm for time series forecasting applied to energy consumption prediction.** *Applied Energy*, 268:114977, 2020.
- [116] DE JONG, P.; DARGAVILLE, R.; SILVER, J.; UTEMBE, S.; KIPERSTOK, A. ; TORRES, E. A.. **Forecasting high proportions of wind energy supplying the brazilian northeast electricity grid.** *Applied Energy*, 195:538–555, 2017.
- [117] MAKRIDAKIS, S.; ANDERSEN, A.; CARBONE, R.; FILDES, R.; HIBON, M.; LEWANDOWSKI, R.; NEWTON, J.; PARZEN, E. ; WINKLER, R.. **The accuracy of extrapolation (time series) methods: Results of a forecasting competition.** *Journal of forecasting*, 1(2):111–153, 1982.

- [118] MAKRIDAKIS, S.; SPILIOTIS, E. ; OTHERS. **The m5 competition and the future of human expertise in forecasting**. *Foresight: The International Journal of Applied Forecasting*, p. 33–37, 2021.
- [119] MAKRIDAKIS, S.; HIBON, M.. **The m3-competition: results, conclusions and implications**. *International journal of forecasting*, 16(4):451–476, 2000.
- [120] CROXFORD, B.; PENN, A. ; HILLIER, B.. **Spatial distribution of urban pollution: civilizing urban traffic**. *Science of the total environment*, 189:3–9, 1996.
- [121] SPINELLE, L.; GERBOLES, M.; VILLANI, M. G.; ALEIXANDRE, M. ; BONAVIDACOLA, F.. **Field calibration of a cluster of low-cost commercially available sensors for air quality monitoring. part b: No, co and co2**. *Sensors and Actuators B: Chemical*, 238:706–715, 2017.
- [122] CINAR, Y. G.; MIRISAE, H.; GOSWAMI, P.; GAUSSIER, E.; AIT-BACHIR, A. ; STRIJOV, V.. **Time series forecasting using rnns: an extended attention mechanism to model periods and handle missing values**. *arXiv preprint arXiv:1703.10089*, 2017.
- [123] DE VITO, S.; PIGA, M.; MARTINOTTO, L. ; DI FRANCIA, G.. **Co, no2 and nox urban pollution monitoring with on-field calibrated electronic nose by automatic bayesian regularization**. *Sensors and Actuators B: Chemical*, 143(1):182–191, 2009.
- [124] MURTAGH, F.; CONTRERAS, P.. **Algorithms for hierarchical clustering: an overview**. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(1):86–97, 2012.
- [125] MEHDIPOUR PIRBAZARI, A.; FARMANBAR, M.; CHAKRAVORTY, A. ; RONG, C.. **Short-term load forecasting using smart meter data: A generalization analysis**. *Processes*, 8(4):484, 2020.
- [126] TRIGUERO, I.. **Fuzz-ieee competition on explainable energy prediction**, 2020.
- [127] MUNKHDALAI, L.; MUNKHDALAI, T.; PARK, K. H.; AMARBAYASGALAN, T.; ERDENEBAATAR, E.; PARK, H. W. ; RYU, K. H.. **An end-to-end adaptive input selection with dynamic weights for forecasting multivariate time series**. *IEEE Access*, 7:99099–99114, 2019.

- [128] VAPNIK, V.; GOLOWICH, S. E. ; SMOLA, A. J.. **Support vector method for function approximation, regression estimation and signal processing.** In: ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS, p. 281–287, 1997.
- [129] YU, H.-F.; RAO, N. ; DHILLON, I. S.. **Temporal regularized matrix factorization for high-dimensional time series prediction.** Advances in neural information processing systems, 29:847–855, 2016.
- [130] ZHANG, G. P.. **Time series forecasting using a hybrid arima and neural network model.** Neurocomputing, 50:159–175, 2003.
- [131] BUCKLEY, J. J.; HAYASHI, Y.. **Fuzzy genetic algorithm and applications.** Fuzzy sets and systems, 61(2):129–136, 1994.
- [132] LIAO, T. W.. **Clustering of time series data—a survey.** Pattern recognition, 38(11):1857–1874, 2005.
- [133] KODINARIYA, T. M.; MAKWANA, P. R.. **Review on determining number of cluster in k-means clustering.** International Journal, 1(6):90–95, 2013.
- [134] DE WINTER, J. C.; GOSLING, S. D. ; POTTER, J.. **Comparing the pearson and spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data.** Psychological methods, 21(3):273, 2016.
- [135] HAMILTON, J. D.. **Time series analysis.** Princeton university press, 1994.
- [136] HARVEY, A. C.. **Time Series and Statistics,** p. 22–24. Palgrave Macmillan UK, 1990.
- [137] VU, K. M.. **The ARIMA and VARIMA time series: their modelings, Analyses and Applications.** AuLac Technologies Inc., 2007.
- [138] PEDRYCZ, W.; GOMIDE, F.. **Fuzzy systems engineering: toward human-centric computing.** John Wiley & Sons, 2007.
- [139] KLIR, G.; YUAN, B.. **Fuzzy sets and fuzzy logic,** volumen 4. Prentice hall New Jersey, 1995.
- [140] ROYCHOWDHURY, S.; PEDRYCZ, W.. **A survey of defuzzification strategies.** International Journal of intelligent systems, 16(6):679–695, 2001.