



EDUARDO ALMEIDA SOARES

**SISTEMAS EVOLUTIVOS BASEADOS EM REGRAS PARA
PREVISÃO DE SÉRIES TEMPORAIS METEOROLÓGICAS**

LAVRAS – MG

2017

EDUARDO ALMEIDA SOARES

**SISTEMAS EVOLUTIVOS BASEADOS EM REGRAS PARA PREVISÃO DE SÉRIES
TEMPORAIS METEOROLÓGICAS**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Engenharia de Sistemas e Automação, área de concentração em Inteligência Computacional, para a obtenção do título de Mestre.

Prof. DSc. Daniel Furtado Leite
Orientador

LAVRAS – MG

2017

**Ficha catalográfica elaborada pela Coordenadoria de Processos Técnicos
da Biblioteca Universitária da UFLA**

Soares, Eduardo Almeida

Sistemas Evolutivos Baseados em Regras para Previsão de
Séries Temporais Meteorológicas / Eduardo Almeida Soares.
– Lavras : UFLA, 2017.

75 p. : il.

Dissertação(Mestrado acadêmico)–Universidade Federal
de Lavras, 2017.

Orientador: Prof. DSc. Daniel Furtado Leite.

Bibliografia.

1. Aprendizado de Máquina. 2. Sistemas Inteligentes
Evolutivos. 3. Previsão de Séries Temporais. I. Leite, Daniel
Furtado. . II. Título.

EDUARDO ALMEIDA SOARES

**SISTEMAS EVOLUTIVOS BASEADOS EM REGRAS PARA PREVISÃO DE SÉRIES
TEMPORAIS METEOROLÓGICAS**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Engenharia de Sistemas e Automação, área de concentração em Inteligência Computacional, para a obtenção do título de Mestre.

APROVADA em 25 de agosto de 2017.

Prof. DSc. Pyramo Costa Jr PUC–MG/PPGEE
Prof. DSc. Bruno Sielly Jales Costa Ford Research & Innovation Center/EUA

Prof. DSc. Daniel Furtado Leite
Orientador

**LAVRAS – MG
2017**

*Dedico este trabalho à minha família, minha mãe Andréa, minha vó Cida, meu pai Zé Paulino
(in memoriam) e meus irmãos Gustavo e Zé Maria.*

AGRADECIMENTOS

À Deus, pela força nos momentos de grande dificuldade durante esta caminhada.

Ao Prof. Daniel F. Leite, por sua amizade, conselhos e dedicação durante este período. Sua orientação foi de extrema importância.

Aos membros da banca, pela disponibilidade e contribuição.

Aos amigos do programa de Pós-Graduação em Engenharia de Sistemas e Automação. Com certeza, o apoio destes foi fundamental neste trabalho.

À Universidade Federal de Lavras, ao Departamento de Engenharia, ao Programa de Pós-Graduação em Engenharia de Sistemas e Automação, aos professores e colaboradores. Por terem proporcionado esta oportunidade.

À CAPES, pelo auxílio financeiro.

À minha mãe Andréa, meu pai Zé Paulino (*in memoriam*), minha vó Cida e meus irmãos Gustavo e Zé Maria. Que são minha base.

Aos meus amigos e todos aqueles que de alguma forma contribuíram.

Meus sinceros agradecimentos.

Eu sou dono e senhor de meu destino; Eu sou o comandante de minha alma.
(William E Henley)

RESUMO

Este trabalho considera métodos inteligentes evolutivos para previsão de séries temporais meteorológicas. São avaliados os métodos *evolving Takagi-Sugeno* (eTS), *eXtended Takagi-Sugeno* (xTS), *Dynamic Evolving Neural Fuzzy Inference System* (DENFIS), *Fuzzy-Set Based evolving Modeling* (FBeM), e uma variação de um método inteligente baseado em nuvens de dados chamado *Typicality and Eccentricity Data Analytics* (TEDA). O objetivo é desenvolver modelos centrados em dados, não lineares e variantes no tempo para previsão de temperatura média mensal. Valores passados de temperatura mínima, máxima e média, assim como valores passados de variáveis exógenas como o nível de nebulosidade, precipitação de chuva e umidade do ar são levados em conta na previsão. É proposto um método não-paramétrico baseado na correlação de Spearman para ordenação e seleção das características e atrasos de tempo mais importantes para a previsão. As bases de dados foram obtidas de estações meteorológicas localizadas nas cidades de São Paulo, Manaus, Porto Alegre, e Natal. Estas cidades têm características climáticas particulares. Ademais, um *ensemble* de nuvens de dados, modelos fuzzy e operadores de agregação fuzzy é proposto para a obtenção de previsões singulares e granulares das séries temporais. Previsões granulares envolvem uma faixa de valores possíveis de temperatura e, portanto, contribuem com uma noção sobre o erro e a incerteza associados às previsões singulares.

Palavras-chave: Aprendizado de Máquina. Nuvens de Dados. Sistemas Inteligentes Evolutivos. Previsão de Séries Temporais. Fluxo de Dados Online.

ABSTRACT

This work considers evolving intelligent methods for weather time series prediction. We evaluate the methods evolving Takagi-Sugeno (eTS), eXtended Takagi-Sugeno (xTS), Dynamic Evolving Neural Fuzzy Inference System (DENFIS), Fuzzy-Set Based evolving Modeling (FBEM), and a variation of cloud-based intelligent method known as typicality-and-eccentricity-based method for data analysis (TEDA). The objective is to develop data-centric nonlinear and time-varying models to predict mean monthly temperature. Past values of minimum, maximum and mean monthly temperature, as well as previous values of exogenous variables such as cloudiness, rainfall and humidity are considered in the analysis. A non-parametric Spearman correlation based method is proposed to rank and select the most relevant features and time delays for a more accurate prediction. The datasets were obtained from weather stations located in main Brazilian cities such as Sao Paulo, Manaus, Porto Alegre, and Natal. These cities are known to have particular weather characteristics. Additionally, an ensemble of cloud and fuzzy models and fuzzy aggregation operators is developed to give single-valued and granular predictions of the time series. Granular predictions convey a range of possible temperature values. Therefore, it provides a notion about the error and uncertainty associated with the single-valued predictions.

Keywords: Machine Learning. Data Clouds. Evolving Intelligent Systems. Weather Time Series Prediction. Online Data Stream.

LISTA DE FIGURAS

| | |
|---|----|
| Figura 2.1 – Série temporal da temperatura média mensal em Lavras-MG | 17 |
| Figura 2.2 – Série temporal não estacionária | 19 |
| Figura 2.3 – Série temporal estacionária | 19 |
| Figura 2.4 – Exemplo de nerônio artificial | 22 |
| Figura 2.5 – Representação de uma rede ANFIS | 25 |
| Figura 3.1 – Representação de uma função potencial | 31 |
| Figura 3.2 – Algoritmo de aprendizagem: Takagi-Sugeno Evolutivo | 32 |
| Figura 3.3 – Algoritmo de aprendizagem: DENFIS | 34 |
| Figura 3.4 – Algoritmo de aprendizagem: FBeM | 38 |
| Figura 3.5 – Algoritmo de aprendizagem: TEDA Predict | 43 |
| Figura 4.1 – Séries temporais de temperatura das estações meteorológicas: (a) Sao Paulo, (b) Manaus, (c) Porto Alegre, e (d) Natal | 48 |
| Figura 4.2 – <i>Ensemble</i> de modelos fuzzy e de nuvens de dados | 49 |
| Figura 5.1 – Comparação de <i>RMSE</i> para diferentes quantidades de características - Es- tação de São Paulo | 52 |
| Figura 5.2 – Comparação de <i>Tempo(s)</i> para diferentes quantidades de características - Estação meteorológica de São Paulo | 52 |
| Figura 5.3 – Estimção FBeM de um passo para a estação meteorológica de São Paulo . | 53 |
| Figura 5.4 – Comparação entre <i>RMSE</i> / Características - Estação meteorológica de Manaus | 55 |
| Figura 5.5 – Comparação de <i>Tempo(s)</i> para diferentes quantidades de características - Estação meteorológica de Manaus | 56 |
| Figura 5.6 – Estimção TEDA de um passo para a estação de Manaus | 57 |
| Figura 5.7 – Comparação de <i>RMSE</i> para diferentes quantidades de características - Es- tação meteorológica de Porto Alegre | 59 |
| Figura 5.8 – Comparação de <i>Tempo(s)</i> para diferentes quantidades de características - Estação meteorológica de Porto Alegre | 59 |
| Figura 5.9 – Estimção TEDA Predict de um passo para estação de Porto Alegre | 60 |
| Figura 5.10 – Comparação de <i>RMSE</i> para diferentes quantidades de características - Es- tação meteorológica de Natal | 62 |
| Figura 5.11 – Comparação de <i>Tempo(s)</i> para diferentes quantidades de características - Estação meteorológica de Natal | 62 |

| | |
|--|----|
| Figura 5.12 – Estimação TEDA Predict de um passo para estação de Natal | 63 |
| Figura 5.13 – Previsão granular e singular de temperatura utilizando um <i>ensemble</i> de modelos fuzzy e de nuvens de dados considerando a estação de Manaus | 64 |

LISTA DE TABELAS

| | |
|---|----|
| Tabela 3.1 – Comparação entre nuvens de dados e <i>clusters</i> | 39 |
| Tabela 3.2 – Comparação entre métodos de inferência fuzzy | 39 |
| Tabela 5.1 – Previsão: Estação meteorológica de São Paulo | 51 |
| Tabela 5.2 – Previsão: Estação meteorológica de Manaus | 54 |
| Tabela 5.3 – Previsão: Estação meteorológica de Porto Alegre | 58 |
| Tabela 5.4 – Previsão: Estação meteorológica de Natal | 61 |
| Tabela 5.5 – Sumário da previsão de temperatura média mensal fornecido pelo <i>ensemble</i> utilizando o operador de agregação média ponderada pelo erro | 63 |

SUMÁRIO

| | | |
|----------------|--|-----------|
| 1 | INTRODUÇÃO | 12 |
| 1.1 | Objetivos | 14 |
| 1.2 | Contribuições | 14 |
| 1.2.1 | Produtos da Dissertação | 15 |
| 1.2.2 | Autoria e Coautoria de Artigos Relacionados | 15 |
| 1.3 | Organização do trabalho | 16 |
| 2 | MODELAGEM DE SÉRIES TEMPORAIS METEREOLÓGICAS | 17 |
| 2.1 | Conceitos Elementares | 17 |
| 2.1.1 | Componentes de Séries Temporais | 18 |
| 2.2 | Métodos Estatísticos | 19 |
| 2.2.1 | Método Médias Móveis Simples | 20 |
| 2.2.2 | Método ARIMA | 20 |
| 2.2.3 | Método Autorregressivo Não-Linear com entradas Exógenas | 21 |
| 2.3 | Sistemas Inteligentes | 21 |
| 2.3.1 | Redes Neurais Artificiais | 22 |
| 2.3.2 | Sistemas Fuzzy | 23 |
| 2.3.3 | Sistemas Neuro-Fuzzy | 24 |
| 3 | SISTEMAS INTELIGENTES EVOLUTIVOS | 27 |
| 3.1 | Modelos Evolutivos | 27 |
| 3.2 | Sistemas Fuzzy Evolutivos | 28 |
| 3.2.1 | Modelo Takagi-Sugeno Evolutivo | 29 |
| 3.2.1.1 | Modelo Takagi-Sugeno Estendido | 32 |
| 3.3 | Redes Neuro fuzzy evolutivas | 33 |
| 3.3.1 | Sistema de Inferência Neuro-Fuzzy Evolutivo Dinâmico | 33 |
| 3.4 | Fluxo de Dados Granulares | 34 |
| 3.5 | Sistemas Granulares Evolutivos | 35 |
| 3.5.1 | Modelagem Evolutiva Granular Fuzzy | 35 |
| 3.6 | Inferência Fuzzy Baseada em Nuvens de Dados | 38 |
| 3.6.1 | Tipicidade e Excentricidade | 40 |
| 3.6.1.1 | TEDA Predict | 42 |
| 4 | METODOLOGIA | 44 |

| | | |
|-----|--|----|
| 4.1 | Seleção de Características Baseadas na Correlação de Spearman | 44 |
| 4.2 | Dados Utilizados | 45 |
| 4.3 | <i>Ensemble</i> de Nuvens de Dados e Modelos Fuzzy | 47 |
| 5 | RESULTADOS E DISCUSSÕES | 50 |
| 5.1 | Resultados para a Estação Meteorológica de São Paulo | 50 |
| 5.2 | Resultados para a Estação Meteorológica de Manaus | 53 |
| 5.3 | Resultados para a Estação Meteorológica de Porto Alegre | 56 |
| 5.4 | Resultados para a Estação Meteorológica de Natal | 60 |
| 5.5 | Resultados do <i>Ensemble</i> | 62 |
| 6 | CONCLUSÃO | 65 |
| | REFERÊNCIAS | 67 |
| | APENDICE A – Operadores de Agregação | 73 |
| .1 | Agregação T-norma | 73 |
| .2 | Agregação S-Norma | 73 |
| .3 | Agregação de média | 74 |
| | APENDICE B – Método de Mínimos Quadrados Recursivos | 75 |

1 INTRODUÇÃO

O desenvolvimento de modelos capazes de prever valores futuros de séries temporais é uma tarefa importante, e ao mesmo tempo complexa em diversas áreas de conhecimento. No caso de séries temporais meteorológicas, as características de não-linearidade e não estacionariedade induzidas por diversos fatores atmosféricos tornam este problema ainda mais difícil de lidar (LEITE et al., 2011).

Previsões de séries temporais meteorológicas são úteis em diversas aplicações e podem servir de base para muitos empreendimentos humanos. Por exemplo, a agricultura é extremamente sensível a mudança climática. Previsões meteorológicas são importantes para que produtores tomem decisões sobre certas e para a proteção da propriedade. Áreas, como energia, transporte, aviação, e planejamento de inventário, claramente podem se beneficiar de previsões meteorológicas para tomada de decisões. Em geral, qualquer sistema que seja sensível ao estado da atmosfera pode se beneficiar de previsões (SOARES et al., 2017).

Desde os anos 70, diversos métodos e modelos têm sido propostos para previsão de séries temporais (FALK et al., 2012) (BREIDT, 2005). Métodos tradicionais de estatística se baseiam na observação de padrões em gráficos, médias móveis ponderadas, análise de tendências. Estes métodos são lineares e estacionários de uma maneira geral, e.g., regressão linear, filtro de Kalman, e o método de *Box & Jenkins*. Ademais, transformada discreta de Fourier, transformadas *wavelet* (FALOUTSOS; RANGANATHAN; MANOLOPOULOS, 1994), e polinômios de Chebyshev (CAI; NG, 2004) são métodos de representação de séries temporais não adaptativos. O mesmo é válido para métodos de Markov, métodos limiares, e métodos paramétricos não-lineares baseados em autorregressão (HAMILTON, 1990) (TONG, 1990). Estes métodos assumem que dados históricos estão disponíveis. Logo, eles apresentam desvantagens quando há mudanças no ambiente em que os dados estão inseridos (HAYKIN, 2008) (IZAKIAN; PEDRYCZ; JAMAL, 2013).

Previsores estatísticos e inteligentes não-lineares têm sido propostos especificamente para a tarefa de análise de dados meteorológicos, como previsores baseados em redes neurais (HUNG et al., 2009) (GHAZALI et al., 2014), regras fuzzy (DEB et al., 2017), *deep learning* (HOSSAIN et al., 2015), e *support vector machine* (SARHADI et al., 2016). Devido as características intrínsecas de não-linearidade, estes métodos geralmente aproximam comportamentos dinâmicos complexos mais facilmente. No entanto, estes métodos não são capazes de capturar certas não-estacionariedades como aquelas provenientes condições atmosféricas e oceânicas.

Os modelos gerados não são equipados com algoritmos incrementais *online* capazes de adaptar seus parâmetros e estrutura.

Com o aumento do número de dispositivos capazes de coletar dados de modo contínuo — incluindo dados/variáveis que afetam o clima — um grande número de dados têm sido produzido. Um modo de lidar com grandes volumes de dados é através do uso de uma classe de métodos não-paramétricos denominado de sistemas inteligentes evolutivos (ANGELOV et al., 2006), (KASABOV, 2007). A abordagem evolutiva inteligente é um modo efetivo e eficiente de lidar com dados de séries temporais devido à habilidade de adaptação dos estrutural de modelos às diferentes situações, providenciando respostas rápidas às mudanças, à i.e, *concept drifts* e *concept shifts* (LUGHOFER; ANGELOV, 2011). Estes dois conceitos se referem â mudanças graduais e abruptas, respectivamente, da fonte geradora de dados.

Este trabalho considera e compara os métodos *evolving Takagi-Sugeno* (eTS), *eXtended Takagi-Sugeno* (xTS), *Dynamic Evolving Neural-Fuzzy Inference System* (DENFIS), *Fuzzy-Set Based evolving Modeling* (FBeM), e uma variação do método *Typicality and Eccentricity Data Analytics* (TEDA). O objetivo dos métodos evolutivos supracitados é prever a temperatura média mensal (previsão de um passo) em cidades com diferentes padrões climáticos. São considerados nas análises valores passados de temperaturas mensais médias, mínimas e máximas, assim como as variáveis exógenas: nebulosidade, precipitação e humidade. Um método não-paramétrico baseado na correlação de Spearman é proposto para ordenação e seleção das características mais importantes e número de atrasos de tempo a serem considerados pelos métodos evolutivos.

A previsão de temperatura mensal é utilizada em diversas áreas. Na agricultura, por exemplo, utiliza-se à previsão de temperatura média mensal para o planejamento do plantio, e todas as atividades referentes à este. No setor de energia, à previsão de temperatura média mensal é utilizada por exemplo na tomada de decisão sobre a abertura de comportas em hidrelétricas. Diversos outros setores, também se beneficiam de previsões mensais de temperatura em suas atividades.

Os modelos inteligentes desenvolvidos desempenham o papel de um modelos autorregressivos não-lineares com entradas exógenas, *NARX*, sem assumir uma forma funcional específica. Nuvens de dados e regras fuzzy são criadas e adaptadas sem restrições espaciais. Adicionalmente, um *ensemble* de modelo de nuvens e modelos fuzzy é proposto utilizando diferentes operadores de agregação fuzzy para fusão da contribuição de modelos individuais. O *ensemble*

é desenvolvido para obtenção de valores de previsão pontuais e granulares das séries temporais. Previsões realizadas por *ensemble* podem exibir uma melhor capacidade de generalização, uma redução na variância de modelos individuais, e uma grande tolerância à incertezas nos dados (LEITE et al., 2016), (BUENO et al., 2015). As previsões granulares fornecem uma faixa de possíveis valores de temperatura e, portanto, uma ideia sobre o erro e a incerteza associados aos dados. Experimentos usando dados reais das estações meteorológicas de São Paulo, Manaus, Porto Alegre, e Natal evidenciam a utilidade do método de seleção de características baseado em correlação de *Spearman* proposto e também dos vários modelos evolutivos para a previsão de temperatura média mensal um passo à frente.

1.1 Objetivos

O objetivo deste trabalho é realizar a previsão de temperatura média mensal das estações meteorológicas de São Paulo, Manaus, Porto Alegre e Natal à partir de fluxo de dados numéricos incertos. Os dados mensais utilizados nesta pesquisa são os dados de temperatura média, mínima e máxima, além dos dados de nebulosidade, precipitação e humidade que funcionam como variáveis exógenas para os métodos evolutivos utilizados. Os dados desta pesquisa correspondem entre o período de Janeiro de 1990 à Dezembro de 2015. Tem-se como objetivos gerais desta pesquisa:

- Propor um método para ordenação e seleção de características baseado na correlação de Spearman.
- Determinar o modelo evolutivo mais preciso, mais parcimonioso, de fácil interpretação, e mais rápido computacionalmente para a tarefa de previsão de séries temporais com mudanças graduais de comportamento.
- Propor um *ensemble* de nuvem de dados e modelos fuzzy utilizando operadores de agregação fuzzy, para obtenção de valores de previsão singulares e granulares.

1.2 Contribuições

As principais contribuições deste trabalho são:

- Avaliação de métodos evolutivos para lidar com séries temporais meteorológicas. Evidencia-se as características que tornam os modelos evolutivos resultantes mais apropriados para este problema de previsão.

- Proposta de método não-paramétrico para ordenação e seleção de características baseado na correlação de Spearman. Este método auxilia a compactação dos modelos, tornando-os mais interpretáveis e menos custosos computacionalmente. A correlação de Spearman não assume linearidade.

- Proposta de um *ensemble* de nuvens de dados e modelos fuzzy utilizando operadores de agregação fuzzy na camada de saída. Incrementa-se a informação de previsão numérica oferecida por métodos inteligentes quaisquer com a ideia de previsão granular. Um invólucro sobre previsões singulares auxilia a tomada de decisão em diferentes contextos.

Alguns trabalhos, produtos diretos da dissertação e relacionados, já foram publicados. Sendo sete publicações em congressos. Além disso, um artigo foi submetido para periódico, e outro para congresso. A lista dos trabalhos é apresentada a seguir:

1.2.1 Produtos da Dissertação

1. **SOARES, E. A** ; MOTA, V ; POUÇAS, R ; LEITE, D. F. . Cloud-Based Evolving Intelligent Method for Weather Time Series Prediction. In: 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2017, Naples. 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2017.

2. **SOARES, E. A.** ; COSTA JR, P. ; COSTA, B. ; LEITE, D. F. . Ensemble of Evolving Data Cloud and Fuzzy-Based Models for Weather Time Series Prediction. Applied Soft Computing. (submetido).

3. MOTA, V ; **SOARES, E. A.** ; LEITE, D. F. . Modelagem Fuzzy Incremental para Previsão Climática. In: 46º Congresso Brasileiro de Engenharia Agrícola - CONBEA, 2017, Maceió. Anais do 46º Congresso Brasileiro de Engenharia Agrícola - CONBEA, 2017.

1.2.2 Autoria e Coautoria de Artigos Relacionados

1. **SOARES, E. A.** ; POUÇAS, R. ; PRADO, J. W. ; LAMOUNIER, S. ; CATALANO, M. D. . Sistemas Fuzzy para Classificação de Vendas de Telemarketing Bancário. In: 1º Encontro Internacional de Lógica Fuzzy aplicada em Administração - ELFA, 2016, Rio de Janeiro. Anais do 1º Encontro Internacional de Lógica Fuzzy aplicada em Administração - ELFA. Rio de Janeiro: COPPEAD / UFRJ, 2016. p. 1-15.

2. MOTA, V ; DAMASCENO, F. ; **SOARES, E. A.** ; LEITE, D. F. . Fuzzy Clustering Methods Applied to the Evaluation of Compost Bedded Pack Barns. In: 2017 IEEE Interna-

tional Conference on Fuzzy Systems (FUZZ-IEEE), 2017, Naples. 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2017.

3. PRADO, J. W. ; CHAIN, C. P. ; CARVALHO, F. M. ; POUÇAS, R. ; **SOARES, E. A.** . Aplicação de sistemas Neuro-Fuzzy para previsão de Séries Temporais Financeiras. In: 1º Encontro Internacional de Lógica Fuzzy aplicada em Administração - ELFA, 2016, Rio de Janeiro. Anais do 1º Encontro Internacional de Lógica Fuzzy aplicada em Administração - ELFA. Rio de Janeiro: COPPEAD / UFRj, 2016. p. 1-16.

4. POUÇAS, R ; PRADO, J. W. ; **SOARES, E. A.** ; LAMOUNIER, S. ; CATALANO, M. D. . Algoritmos de agrupamento Fuzzy C-Means e Gustafson-Kessel para detecção de Spam. In: 1º Encontro Internacional de Lógica Fuzzy aplicada em Administração - ELFA, 2016, Rio de Janeiro. Anais do 1º Encontro Internacional de Lógica Fuzzy aplicada em Administração - ELFA. Rio de Janeiro: COPPEAD / UFRJ, 2016. p. 1-14.

5. MOTA, V ; CAMPOS, A. T. ; DAMASCENO, F. ; **SOARES, E. A.** ; LEITE, D. F. . Métodos de Agrupamento Fuzzy Aplicados à Sistemas de Galpão de Compostagem. In: 46º Congresso Brasileiro de Engenharia Agrícola - CONBEA, 2017, Maceió. Anais do 46º Congresso Brasileiro de Engenharia Agrícola - CONBEA, 2017.

6. GARCIA, C. M. ; CATALANO, M. D. ; **SOARES, E. A.** ; BARBOSA, B. H. G. . Teaching-Learning-Based Optimization no Treinamento de Redes Neurais Artificiais para Problemas de Classificação. Congresso Brasileiro de Inteligência Computacional - CBIC. (submetido).

1.3 Organização do trabalho

Este trabalho é dividido em 6 capítulos que são descritos abaixo:

Este capítulo introduz o problema a ser tratado e menciona de maneira geral quais métodos serão utilizados para construção de modelos previsores fuzzy. O capítulo 2 apresenta uma revisão sistemática sobre o tema de modelagem e previsão de séries temporais meteorológicas, e discute as características gerais que tais séries podem ter. O capítulo 3 introduz o tema sistemas inteligentes evolutivos e apresenta os principais métodos utilizados neste trabalho. A metodologia utilizada neste projeto de pesquisa é apresentada no capítulo 4. O capítulo 5 apresenta os resultados obtidos com as análises realizadas e discute os resultados encontrados neste trabalho de dissertação. O capítulo 6 trata-se das conclusões obtidos, e indica possíveis trabalhos.

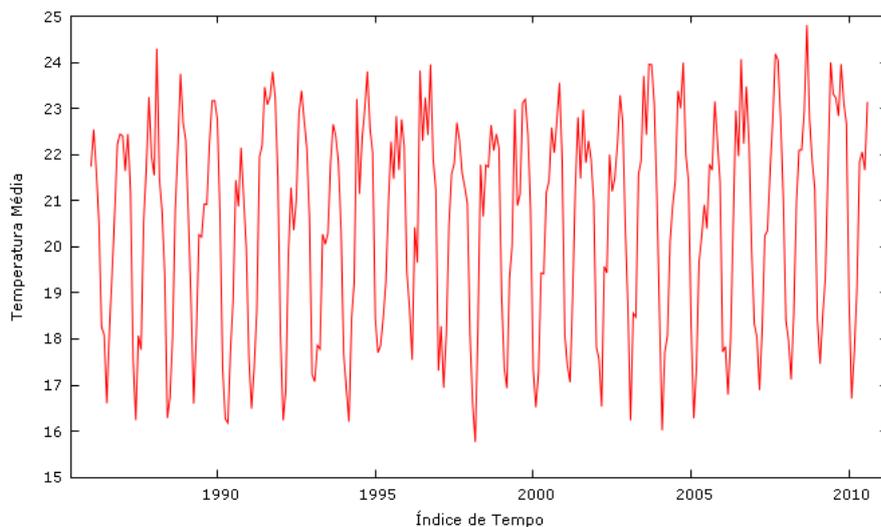
2 MODELAGEM DE SÉRIES TEMPORAIS METEOROLÓGICAS

Neste capítulo são apresentadas características e conceitos relacionados à modelagem e previsão de séries temporais através de técnicas convencionais estatísticas e de inteligência computacional. O capítulo visa tornar o texto auto-contido.

2.1 Conceitos Elementares

Segundo (MORETTIN; TOLOI, 2006), uma série temporal pode ser compreendida como a representação matemática do comportamento de um processo ou sistema através de um conjunto de observações ordenadas no tempo. Séries temporais refletem o comportamento de sistemas físicos ou fenômenos a partir de dados coletados periodicamente. A Figura 2.1, mostra uma série temporal típica. Trata-se da série temporal de temperatura média mensal entre os anos de 1986 e 2017 da cidade de Lavras, Minas Gerais.

Figura 2.1 – Série temporal da temperatura média mensal em Lavras-MG



Fonte: Adaptado de <<http://www.inmet.gov.br/>>

Também são exemplos de séries temporais:

- Valores de diários de precipitação.
- Índice horário de temperatura.
- Valores mensais de nebulosidade.

(MORETTIN; TOLOI, 2006) define que pode-se obter uma série temporal de um evento contínuo através da determinação de um intervalo de tempo de amostragem, Δt . Logo, para analisar uma série Z_t , é necessário defini-la num espaço de tempo correspondente ao intervalo

$[0, t]$. A forma mais simples de se interpretar uma série temporal é através da conceituação da série como um conjunto de valores discretos que apresentam dependência temporal. (BOX et al., 2015), lista os objetivos da análise de séries temporais:

- Entender o mecanismo que gera os valores da série. Isto possibilita explicar o comportamento da série em um determinado intervalo de tempo.
- Identificar componentes sazonais, cíclicos e de tendência acontecendo na sequência de dados.
- Realizar previsão de comportamentos futuros, em curto, médio ou longo prazo.

Segundo (CRYER; KELLET, 1986), séries temporais podem não possuir amostras completamente independentes entre si. Estudos de modelos que identifiquem essa dependência entre amostras são importantes. As relações de dependência em uma série temporal tornam possível a realização de previsão de valores futuros.

2.1.1 Componentes de Séries Temporais

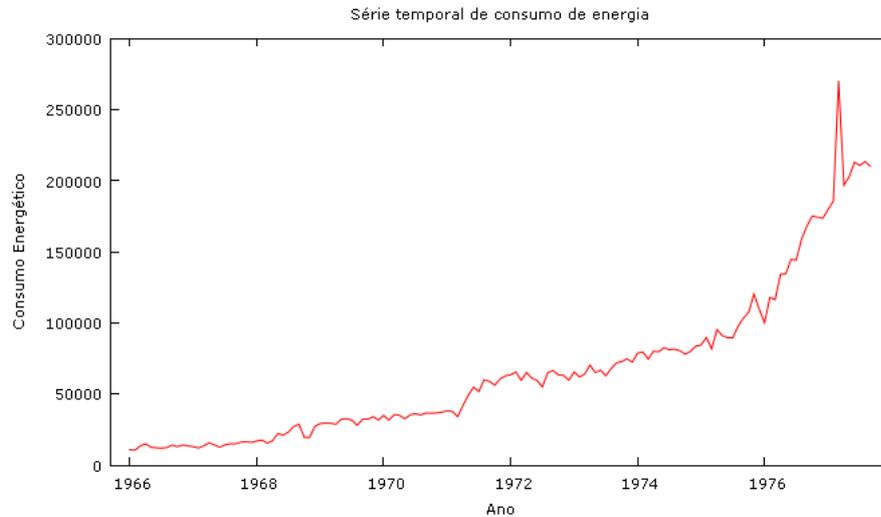
Uma das características mais importantes a serem observadas em séries temporais é a estacionariedade (HAMILTON, 1994). Uma série temporal é considerada estacionária quando se desenvolve aleatoriamente no tempo, em torno de um valor médio μ constante. Em geral, a maioria das séries não apresentam estacionariedade. A Figura 2.2 mostra uma série não-estacionária de consumo anual de energia. Já a Figura 2.3 ilustra uma série estacionária do mesmo conjunto de dados utilizados na Figura 2.2. Entretanto, para a obtenção da série temporal ilustrada na Figura 2.3 fez-se o uso de técnicas que tornam séries não-estacionárias em séries temporais. No caso, usou-se a técnica de diferença logarítmica, vide (MORETTIN; TOLOI, 2006).

Conforme observado nas Figuras 2.2 e 2.3, séries não-estacionárias podem ser transformadas em séries estacionárias através de técnicas que proporcionem isto (MARPLE, 1987). Para que isso ocorra, é necessário que se analise outros componentes como:

Tendência: componente que indica como a série se comporta ao longo do tempo em relação ao seu deslocamento. Pode ser expressa por: $z_t = T_t + a_t$, onde z_t é a série temporal analisada, T_t é a tendência da série, e a_t é o erro;

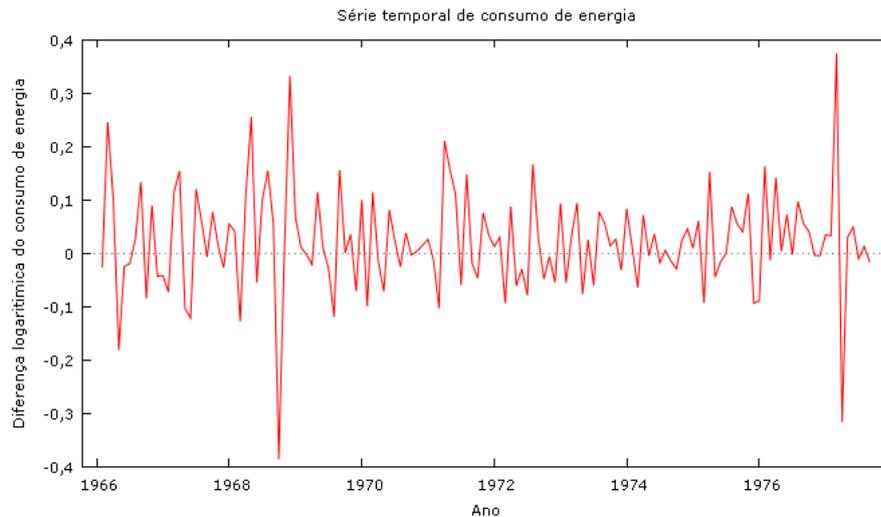
- Ciclicidade: componente que indica a presença de ciclos em um período fixo em torno da tendência;
- Sazonalidade: componente que possibilita identificar regularidades ao longo da série;

Figura 2.2 – Série temporal não estacionária



Fonte: Adaptado de (MORETTIN; TOLOI, 2006)

Figura 2.3 – Série temporal estacionária



Fonte: Adaptado de (MORETTIN; TOLOI, 2006)

- Erro aleatório: componente que não pode ser modelado, pois representa a variabilidade intrínseca aos dados.

2.2 Métodos Estatísticos

Nesta seção são apresentados os principais conceitos e modelos estatísticos utilizados em análise e previsão de séries temporais. Os modelos utilizam elementos estatísticos para compreender o comportamento da série e assim identificar suas características. Os modelos apresentados são baseados nos métodos de médias móveis simples (MA), método auto regres-

sivo integrado médias móveis (ARIMA) e o método autorregressivo não-Linear com entradas exógenas (NARX).

2.2.1 Método Médias Móveis Simples

O método de médias móveis simples, MA, utiliza a média das observações passadas como critério para realizar previsões, podendo essa média ser simples, ponderada ou centrada (WEI, 1994). Sendo a série $z_t, z_{t+1}, z_{t+2}, \dots, z_{t+h}$, em que em que $z_t(j), j = 1, 2, \dots, h$, podemos descrever a seguinte equação.

$$z_t = \mu_t + a_t \quad (2.1)$$

onde a_t é o erro, que é igual a zero; e μ_t é uma variável que possui valor desconhecido e que pode mudar lentamente com tempo. O método de média móveis simples consiste em realizar o cálculo da média aritmética das h observações mais recentes, podendo ser representada por:

$$M_t = \frac{z_t + z_{t+1} + z_{t+2} + \dots + z_{t+h}}{h} \quad (2.2)$$

Segundo (VELICER; FAVA, 2003), quanto maior for a janela de observações, i.e. quanto maior o valor de h , a previsão terá uma suavidade maior. Se a série apresentar mudanças abruptas de comportamento, um menor número de observações deve ser utilizado nos cálculos. Caso a série não se altere drasticamente com o tempo, um número de observações maior pode deve ser utilizado, para que se tenha mais eficiência. Conforme (HUANG et al., 1998), a denominação média móvel acontece porque a cada iteração a média das observações é recalculada, e a observação mais antiga é eliminada.

2.2.2 Método ARIMA

O método auto regressivo integrado de médias móveis (ARIMA), foi proposto em 1970 por *Box & Jenkins* (BOX et al., 2015). Este método visa, através do uso de funções de autocorrelação e autocorrelação parcial, identificar o comportamento da série e realizar previsões futuras. O método ARIMA se baseia em uma equação que representa o comportamento aproximado da série, ou seja:

$$z_t = \alpha_0 + \alpha_1 z_{t-1} + \dots + \alpha_p z_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q} \quad (2.3)$$

onde o termo α_0 representa uma constante, α_1 a α_p são parâmetros que ponderam os valores passados de Z_t do instante imediatamente anterior até o mais distante, representado por p . Os valores de ε representam uma sequência de termos aleatórios e independentes uns dos outros; ε_t é a porção não controlável do modelo, chamada de ruído branco. Os parâmetros β_1 a β_q possibilitam escrever a série em função dos termos passados do ruído branco. Em geral, ε_t é considerado como tendo distribuição normal, média zero e variância constante.

2.2.3 Método Autorregressivo Não-Linear com entradas Exógenas

Em séries temporais, um métodos autorregressivo não-linear com entradas exógenas (*NARX*), é um método autorregressivo não-linear no qual utiliza entradas exógenas em suas operações (LEONTARITIS; BILLINGS, 1985). Logo, o método *NARX* relaciona o valor atual da série temporal com ambos:

- Valores passados da mesma série temporal e;
- Valores correntes e passados de séries temporais exógenas associadas a série temporal principal, ou seja, séries temporais externas que influenciam a série temporal principal.

Adicionalmente um valor de erro é adicionado ao modelo *NARX*. O valor de erro é adicionado pois apenas o conhecimento dos outros valores não possibilita uma precisão correta na previsão dos valores da série temporal (LEONTARITIS; BILLINGS, 1985). O modelo *NARX* pode ser descrito algebricamente por:

$$z_t = F(z_{t-1}, z_{t-2}, z_{t-3}, \dots, u_t, u_{t-1}, u_{t-2}, u_{t-3}, \dots) + \varepsilon_t \quad (2.4)$$

onde, z_t é a série temporal principal, e u_t são as séries exógenas. De modo que, a série exógena u_t ajude a prever os valor da da série principal z_t . O valor de erro é indicado por ε_t . A função F pode assumir qualquer função não-linear, i.e. uma função polinomial, dentre outras. F também pode assumir o valor de uma rede neural artificial, ou de outros sistemas inteligentes.

2.3 Sistemas Inteligentes

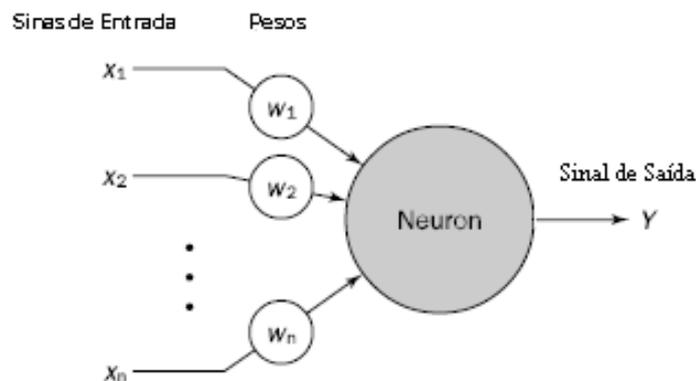
Sistemas inteligentes podem ser entendidos como sistemas que possuem a habilidade de compreender o ambiente no qual estão inseridos e aprender com este (YAGER; ZADEH, 2012). Nesta seção são apresentados métodos de inteligência computacional úteis no contexto

de previsão de séries temporais. São apresentados métodos baseados em redes neurais artificiais, sistemas fuzzy e sistemas neuro-fuzzy.

2.3.1 Redes Neurais Artificiais

Redes neurais são formadas por um número de unidades de processamento interconectadas, denominados neurônios (WANG, 2003), (NILSSON, 2014), (NEGNEVITSKY, 2005). Os neurônios artificiais são conectados uns aos outros através de ligações ponderadas denominadas sinapses. Cada neurônio recebe um número de sinais de entrada e produz um sinal de saída (RUSSELL; NORVIG, 1995). A Figura 2.4 ilustra um neurônio artificial.

Figura 2.4 – Exemplo de neurônio artificial



Fonte: Adaptado de (NEGNEVITSKY, 2005)

Segundo (QIAN et al., 2010), os pesos de conexões pré-sinápticas, $w_i, i = 1, \dots, n$, funcionam como memória de longo prazo em redes neurais. Os pesos expressam a importância de cada variável de entrada. Uma rede neural artificial aprende através do ajuste destes pesos que acontece em cada iteração de um algoritmo. O neurônio computa a soma ponderada dos valores de entrada e compara o resultado com um valor limiar θ , conforme a equação abaixo. Se o valor do somatório for menor que o limiar θ , então o valor de saída é -1. Caso contrário, se o valor for maior que θ , então a saída assume o valor 1, como segue:

$$X = \sum_{i=1}^n x_i w_i \quad (2.5)$$

onde, X é a soma ponderada; x_i é o valor da entrada i ; w_i é o peso associado a entrada i ; n é o número de entradas; e Y é a saída do neurônio. As redes neurais *feed-forward* consistem basicamente de neurônios, conforme os descritos, e dispostos paralelamente e em camadas. Além

disso, diferentes redes neurais usam diferentes funções de ativação para a definição de Y . Funções de ativação mapeiam o valor somatório de um neurônio, X , de forma não-linear, em um valor de saída Y . São exemplos de redes neurais usadas na previsão de séries temporais as redes *MLP*, *RBF*, entre outras (NEGNEVITSKY, 2005). Os métodos tradicionais de previsão de séries temporais utilizando redes neurais se baseiam no ajuste dos pesos de forma recorrente para encontrar uma aproximação de uma função f . Trabalhos como (KAASTRA; BOYD, 1996), (DOUCOURE; AGBOSSOU; CARDENAS, 2016), e (KAUR; KUMAR; SEGAL, 2016), demonstram a eficiência dessa abordagem para tarefa de previsão de séries temporais.

2.3.2 Sistemas Fuzzy

Sistemas fuzzy foram propostos por (ZADEH, 1965). Diferentemente de redes neurais que tentam emular a fisiologia do cérebro humano, sistemas fuzzy se propõe a lidar com a incerteza do raciocínio. Sistemas fuzzy permitem a utilização de variáveis linguísticas para criação de uma base de regras que compõem um modelo.

Um conjunto fuzzy pode ser definido como um conjunto com fronteiras suaves (DUBOIS, 1980). Sendo A um conjunto fuzzy no universo X , a pertinência de um elemento $x \in X$ no conjunto A é:

$$\mu_A(x) : X \rightarrow [0, 1] \quad (2.6)$$

Se $\mu_A(x) = 1$, então x está totalmente em A ; Se $\mu_A(x) = 0$, então x não está em A ; $0 < \mu_A(x) < 1$ indica que x pertence parcialmente em A .

Para cada elemento x do universo X , a função de pertinência $\mu_A(x)$ indica o grau de x em A . Graus de pertinência variam entre 0 e 1 (PEDRYCZ, 1994).

A interseção entre conjuntos fuzzy, pode ser determinada conforme:

$$\mu_{A \cap B}(x) = \min[\mu_A(x), \mu_B(x)] = \mu_A(x) \cap \mu_B(x) \quad (2.7)$$

A união de dois conjuntos fuzzy, A e B é:

$$\mu_{A \cup B}(x) = \max[\mu_A(x), \mu_B(x)] = \mu_A(x) \cup \mu_B(x) \quad (2.8)$$

Os operadores *min* e *max* são casos particulares de T e S normas. Previsão de séries temporais utilizam basicamente o método de inferência fuzzy baseado no modelo de Takagi-

Sugeno (TAKAGI; SUGENO, 1985),(LIU; LEE, 2016). Este método consiste de uma função linear na parte consequente de regras fuzzy. Um sistema MISO de ordem 1 pode ser representada conforme:

$$\text{Se } x_1 \text{ é } A_1 \text{ e } x_2 \text{ é } A_2, \dots, \text{ e } x_n \text{ é } A_n \text{ então } y = a_0 + a_1 * x_1 + a_2 * x_2 + \dots + a_n * x_n \quad (2.9)$$

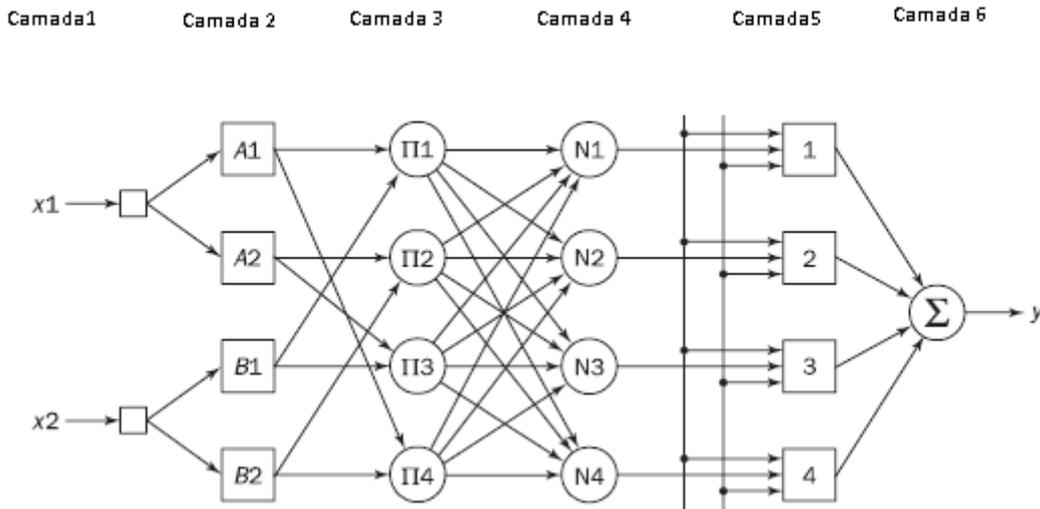
onde y é a saída da regra; f a função de saída; e n é o número de variáveis.

Para previsão de séries temporais utilizando modelos fuzzy, (KHASHEI; HEJAZI; BI-JARI, 2008), (CASTILLO; MELIN, 2002) e (NEGNEVITSKY, 2005) utilizam os seguintes procedimentos: (i) Definição do domínio da variável x ; (ii) definição dos conjuntos fuzzy que compõem esta variável x ; (iii) fuzzificação dos dados provenientes das séries temporais. À partir da série temporal fuzzificada são criadas as regras fuzzy necessárias, ocorre então a fase de defuzzificação e o valor de previsão é retornado. Diversos trabalhos, (SONG; CHISSOM, 1993a), (HUARNG, 2001), (CHEN; HWANG, 2000), (SONG; CHISSOM, 1993b), mostram resultados satisfatórios ao utilizarem modelos fuzzy para previsão de séries temporais.

2.3.3 Sistemas Neuro-Fuzzy

Sistemas Fuzzy e redes neurais são abordagens complementares para a construção de sistemas inteligentes (NEGNEVITSKY, 2005). Sistemas neuro-fuzzy combinam a habilidade de aprendizagem das redes neurais com a capacidade de representação de conhecimento de sistemas fuzzy (SUN; JANG, 1993). A estrutura de um sistema neuro-fuzzy é similar à estrutura de uma rede neural multicamada. Como um sistema neuro-fuzzy é basicamente uma rede neural multicamada, pode-se aplicar então algoritmos de aprendizagem similares aos utilizados pelas redes neurais, como por exemplo o algoritmo *backpropagation* (NEGNEVITSKY, 2005). Diversos trabalhos (ABDULSHAHED; LONGSTAFF; FLETCHER, 2015), (YUN et al., 2008), (MELLIT et al., 2007), (EL-SHAFIE; JAAFER; AKRAMI, 2011), demonstraram sucesso ao utilizar a abordagem neuro-fuzzy para previsão de séries temporais. Estes trabalhos utilizaram um sistema de neuro-fuzzy que utiliza o sistema de inferência fuzzy Takagi-Sugeno, denominado *Adaptative Neuro-fuzzy Inference System* (ANFIS) proposto por (JANG, 1993). O ANFIS possui 6 camadas, conforme demonstra a Figura 2.5.

Figura 2.5 – Representação de uma rede ANFIS



Fonte: Adaptado de (NEGNEVITSKY, 2005)

A primeira camada representa as entradas dos dados. A camada de fuzzificação, ou camada 2, é camada na qual ocorre o processo de fuzzificação. No modelo ANFIS proposto por (JANG, 1993), é utilizada uma função de ativação do tipo sino.

$$y_i^{(2)} = \frac{1}{1 + \left(\frac{x_i^{(2)} - a_i}{c_i}\right)^{2b_i}} \quad (2.10)$$

onde \$x_i^{(2)}\$ é a entrada, \$y_i^{(2)}\$ é a saída do neurônio \$i\$ na camada 2. \$a_i\$, \$b_i\$ e \$c_i\$ são parâmetros que controlam respectivamente o centro, largura e altura da função de ativação sino do neurônio \$i\$.

Na camada 3, cada neurônio corresponde a uma regra do tipo Takagi-Sugeno. E é obtido através da respectiva equação.

$$y_i^{(3)} = \prod_{j=1}^k x_{ji}^{(3)} \quad (2.11)$$

onde \$x_{ji}^{(3)}\$ são as entradas e \$y_i^{(3)}\$ é a saída da regra do neurônio \$i\$ na camada 3. A camada 4, ou camada de normalização, é a camada que recebe todos os valores da camadas de regras anterior e calcula o valor normalizado da determinada regra.

$$y_i^{(4)} = \frac{x_{ii}^{(4)}}{\sum_{j=1}^n x_{ji}^{(4)}} = \frac{\mu_i}{\sum_{j=1}^n \mu_j} = \bar{\mu}_i \quad (2.12)$$

onde $y_i^{(4)}$ é a entrada do neurônio j localizado na camada 3 ao neurônio i na camada 4, n representa o número total de neurônios, μ é o valor normalizado. Na camada de defuzzificação, camada 5, cada neurônio nesta camada é conectado ao respectivo neurônio na camada de normalização. Nesta camada os valores ponderados das partes consequentes das regras são calculados.

$$y_i^{(5)} = x_i^{(5)} [k_{i0} + k_{i1}x_1 + k_{i2}x_2] = \bar{\mu}_i [k_{i0} + k_{i1}x_1 + k_{i2}x_2] \quad (2.13)$$

Em que $x_i^{(5)}$ é a entrada e $y_i^{(5)}$ a saída do neurônio i de defuzzificação na camada 5, k_{i0} , k_{i1} , k_{i2} são um conjunto de parâmetros consequentes da regra i . A sexta camada faz um somatório de todas as saídas da camada de defuzzificação e produz a saída média, y , do modelo ANFIS.

$$y = \sum_{i=1}^n x_i^{(6)} = \sum_{i=1}^n \bar{\mu}_i [k_{i0} + k_{i1}x_1 + k_{i2}x_2] \quad (2.14)$$

O ANFIS utiliza um modelo híbrido de aprendizagem, no qual combina estimador de mínimos quadrados com o método de gradiente descendente (NEGNEVITSKY, 2005). O ANFIS possui habilidade de convergir e generalizar rapidamente, o que é importante quando se trabalha com aprendizado em tempo real (NEGNEVITSKY, 2005), (JANG, 1993), (ÇAYDAŞ; HASÇALIK; EKICI, 2009).

3 SISTEMAS INTELIGENTES EVOLUTIVOS

Neste capítulo são apresentados os principais conceitos sobre os sistemas evolutivos inteligentes utilizados nesta pesquisa.

3.1 Modelos Evolutivos

De acordo com (KASABOV, 2007), um modelo evolutivo pode ser classificado como um modelo que se adapta e se desenvolve de modo contínuo. Este processo de adaptação está relacionado com o ambiente que o envolve. Variações temporais dificultam a previsão do comportamento de um processo. Um modelo evolutivo deve ser capaz de perceber mudanças e melhorar sua capacidade.

O termo evolutivo possui uma amplitude maior quando comparado ao termo evolucionário. Enquanto o termo evolucionário se refere a uma população de indivíduos e suas gerações, o termo evolutivo se refere ao desenvolvimento das funcionalidades e estruturas de modelos (KASABOV, 2007). Conforme (ANGELOV, 2009), sistemas evolutivos aprendem com a experiência, mudanças graduais, geração de resultados provenientes da extração de regras à partir de dados. Estas são características essenciais de um sistema autônomo.

Para análises fluxos de dados *online*, são necessários métodos com maior nível de adaptabilidade quando comparados com os métodos tradicionais de sistemas adaptativos (ÅSTRÖM; WITTENMARK, 2013), (ANGELOV, 2009), de identificação de parâmetros de sistemas proposta (LJUNG, 1988), e também dos métodos apresentados por (TIBSHIRANI; WALTHER; HASTIE, 2001). Uma frente de pesquisas para enfrentar os novos desafios é denominada sistemas inteligentes evolutivos.

Segundo (ANGELOV, 2013), um modelo evolutivo se difere de um modelo adaptativo, pois um modelo adaptativo supõe uma estrutura fixa; em que os parâmetros se ajustam conforme a necessidade. Em um modelo evolutivo não somente os parâmetros são adaptáveis, mas também sua estrutura se adapta conforme necessidade. Um modelo evolutivo pode manifestar comportamentos tais como (KASABOV, 2007):

- Aleatório: Não há regras que definam o modelo, sendo este imprevisível;
- Caótico: O modelo é previsível apenas em curto prazo;
- Quase-periódico: O modelo é previsível, porém está sujeito a erros. As regras se aplicam ao longo do tempo, mas de forma modificada.

- Periódico: o modelo repete os padrões ao longo do tempo, sendo totalmente previsível. As regras do modelo são fixas.

As séries temporais meteorológicas tratadas neste projeto têm caráter caótico e quase-periódico. Em outras palavras, é possível prevê-las em curto prazo com uma certa precisão através de regras e modelos evolutivos. Em ambientes dinâmicos não-lineares e não-estacionários, as equações que definem estes processos são complicadas de serem obtidas. Desta forma, sistemas de regras fuzzy e redes neurais provêm uma alternativa na modelagem.

Segundo (LUGHOFER; SAYED-MOUCHAWEH, 2015), a motivação por sistemas evolutivos provém do aumento da complexidade e do aumento da capacidade de aquisição de dados, requerendo algoritmos de processamento rápido em um contexto de fluxo de dados *online*. Processamento de fluxo de dados é uma das principais razões de se utilizar sistemas evolutivos. Fluxos de dados são sequências infinitas de amostras, não podendo ser processados em uma única vez (LUGHOFER; SAYED-MOUCHAWEH, 2015).

Redes neurais evolutivas foram introduzidas em (KASABOV, 2001), enquanto que sistemas baseados regras fuzzy evolutivas foram introduzidas por (ANGELOV; BUSWELL, 2002). Sistemas evolutivos podem ser divididos em duas categorias:

- Aprendizagem direta (fase única): Trata-se de métodos de aprendizagem supervisionados. Um modelo não-linear é obtido a partir de pares de dados de entrada e saída.
- Aprendizagem indireta (duas fases): Trata-se de métodos que particionam os dados grupos.

O problema tratado neste trabalho requer métodos de aprendizagem direta. Apesar do problema de previsão de série temporal um tempo a frente, em princípio, não disponibilizar supervisão, é possível recorrer a pares de entrada e saída históricos para o desenvolvimento e adaptação do modelo previsor.

3.2 Sistemas Fuzzy Evolutivos

Sistemas fuzzy evolutivos modificam sua estrutura de regras e parâmetros de acordo com os dados. A base de regras que compõe um sistema fuzzy evolutivo pode expandir ou reduzir conforme a necessidade para acomodar novas informações (ANGELOV et al., 2006). Sistemas fuzzy evolutivos, podem ser entendidos como a combinação de regras de inferência fuzzy, métodos recursivos de aprendizagem de máquina e mecanismos para compactação e representação dos dados de entrada.

(ANGELOV; BUSWELL, 2002) propõe os sistemas fuzzy evolutivos do tipo Takagi-Sugeno. O modelo Takagi-Sugeno evolutivo é útil e eficiente para controle e aproximação de funções quando necessita-se de adaptabilidade e flexibilidade. Exemplos de aplicação são: computação e automação industriais, sensores inteligentes, previsão, entre outros.

3.2.1 Modelo Takagi-Sugeno Evolutivo

Sistemas fuzzy baseado em regras do tipo Takagi-Sugeno (TAKAGI; SUGENO, 1985) possuem os antecedentes das regras na forma linguística e os termos consequentes são combinações lineares das variáveis de entrada. Conforme (ANGELOV; BUSWELL, 2002), o sistema Takagi-Sugeno evolutivo é uma versão recursiva do sistema Takagi-Sugeno onde a estrutura do modelo é flexível, e os antecedentes da regras são determinados por um processo de agrupamento por espalhamento (ORDERS; FADLEY, 1983). As regras são do tipo:

$$R^i : \text{Se } x_1 \text{ é } A_{i1} \text{ e } \dots \text{ e } x_n \text{ é } A_{in} \text{ Então } y_i = a_{i0} + a_{i1}x_1 + \dots + a_{in}x_n \quad (3.1)$$

onde, R^i representa a i -ésima regra fuzzy, $i = 1, \dots, g^k$, em que g^k se refere ao número de regras na iteração k . A variável x_j , $j = 1, \dots, n$, possui n variáveis de entrada; A_{ij} são os conjuntos fuzzy antecedentes; y_i é a saída local do modelo; e a_{ij} são coeficientes.

O grau de ativação de uma regra define o quanto o i -ésimo modelo local terá participação na saída. Funções de pertinência Gaussiana definem os antecedentes A_{ij} conforme:

$$\mu_{ij} = \exp\left(-\frac{4}{r^2} \|x_j - x_{ij}^*\|^2\right) \quad (3.2)$$

onde r representa o espalhamento da função de pertinência. O valor de r é sempre positivo; ele define a zona de influência da regra i . A norma Euclidiana é definida por $\|\cdot\|$ e x_{ij}^* , é o centro. O grau de ativação, τ , de cada regra do modelo Takagi-Sugeno evolutivo é definido pela T-norma produto conforme:

$$\tau_i = \mu_{i1}(x_1) \times \mu_{i2}(x_2) \times \dots \times \mu_{in}(x_n) \quad (3.3)$$

A saída do modelo evolutivo Takagi-Sugeno é dado pela média ponderada da saída, y_i , de cada regra i :

$$y = \sum_{i=1}^{g^k} \lambda_i y_i \quad (3.4)$$

onde,

$$\lambda_i = \frac{\tau_i}{\sum_{j=1}^{g^k} \tau_j} \quad (3.5)$$

é o grau de ativação normalizado de cada regra. A base de regras do modelo Takagi-Sugeno evolutivo é atualizada a cada iteração por um algoritmo de agrupamento recursivo não supervisionado, ou seja, a base de regras é atualizada a cada nova amostra apresentada. Este agrupamento é realizado no espaço de entrada. Os grupos são projetados nos eixos das variáveis de entrada para criação dos antecedentes das regras. O algoritmo de agrupamento utilizado pelo modelo Takagi-Sugeno evolutivo é uma versão incremental do algoritmo de agrupamento subtrativo (CHIU, 1994). Este algoritmo baseia-se na noção de que a representividade de cada amostra pode ser mensurada através de uma função que estima seu potencial (YAGER; FILEV, 1994). A função potencial de uma amostra x^k é definida como a medida de proximidade desta amostra dentre todas as outras (ANGELOV; FILEV, 2004):

$$P(x^k) = \frac{1}{k-1} \sum_{i=1}^{k-1} \exp(-r \|x^k - x^i\|^2) \quad (3.6)$$

onde $k = 2, 3, \dots$ representa o índice das amostras.

O objetivo da função potencial é encontrar amostras de dados que possam ser definidas como centros das regiões onde há aglomeração de dados, vide Figura 3.1. A Figura 3.1 ilustra um caso em que o potencial da amostra A é menor que o potencial da amostra B . Logo a amostra B é o centro do grupo. Quanto maior o número de amostras em uma determinada região, maior será o potencial das amostras desta região.

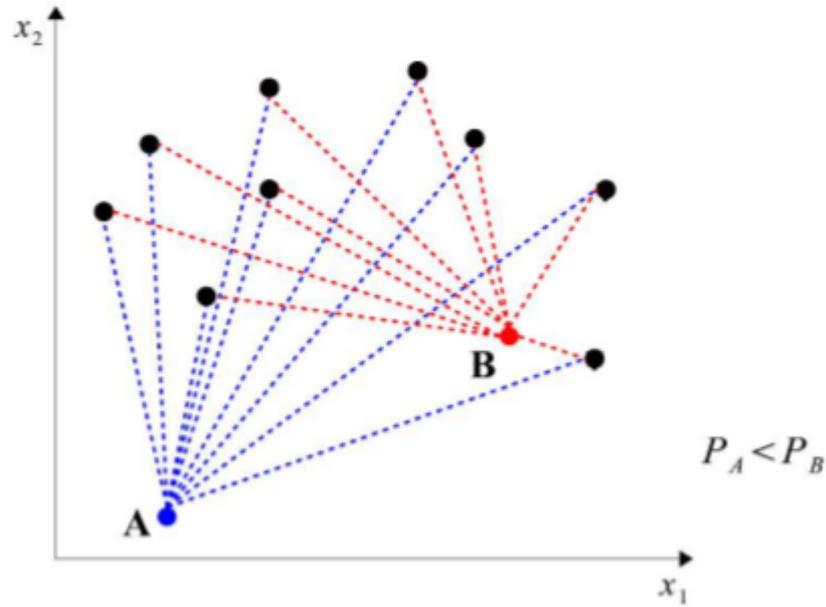
Conforme (ANGELOV; FILEV, 2004), a função potencial de uma nova amostra pode ser calculada recursivamente através de:

$$P^k(z^k) = \frac{k-1}{(k-1)(\vartheta^k + 1)(\sigma^k - 2\nu^k)} \quad (3.7)$$

onde,

$$\vartheta^k = \sum_{j=1}^{m+1} (z_j^k)^2, \quad (3.8)$$

Figura 3.1 – Representação de uma função potencial



Fonte: Adaptado de (ANGELOV et al., 2006)

$$\sigma^k = \sum_{l=1}^{k-1} \sum_{j=1}^{m+1} (z_j^l)^2, \quad (3.9)$$

$$v^k = \sum_{j=1}^{m+1} z_j^k \beta_j^k, \quad (3.10)$$

$$\beta_j^k = \sum_{l=1}^{k-1} z_j^l \quad (3.11)$$

Os parâmetros β_j^k e σ^k são calculados recursivamente, enquanto os valores de ϑ^k e v^k são calculado à partir de z^k . O potencial dos centros dos grupos são atualizados através de:

$$P^k(x^*) = \frac{(k-1)P^{k-1}(x_l^*)}{k-2 + P^{k-1}(x_l^*) + P^{k-1}(x_l^*) \sum_{j=1}^{m+1} d_j^{k(k-1)}} \quad (3.12)$$

onde x^* é o centro do grupo. Este possui dimensão l , em que $l = 1 \times m + 1$; e $d_j^{k(k-1)} = z_j^k - z_j^{k-1}$.

Caso uma nova amostra possua potencial maior que os centros dos grupos existentes, esta amostra se torna centro de um grupo. Caso exista um grupo próximo, esta amostra passa a ser o centro desse grupo; ao contrário, um novo grupo é criado. Caso o potencial desta nova amostra seja menor que o potencial do centro de todos os grupos existentes, os parâmetros da parte consequente da regra que rege o grupo com menor distância Euclidiana para esta amostra

são atualizados através do algoritmo de quadrados mínimos recursivo. O pseudo-código abaixo resume o procedimento de aprendizagem.

Figura 3.2 – Algoritmo de aprendizagem: Takagi-Sugeno Evolutivo

Algoritmo de aprendizagem: Takagi-Sugeno Evolutivo

- 1: Calcular o potencial da nova amostra $P(x^k)$;
 - 2: **PARA** $j = 1, \dots, q^k$ **FAZER**
 - 3: Calcular o potencial do centro c^j ;
 - 4: **FIM PARA**
 - 5: **SE** $P(x^k) > P(c^j) \forall_j$ **ENTÃO**
 - 6: **SE** x^k é próximo do centro de algum grupo j **ENTÃO**
 - 7: O centro do grupo j , c^j , é substituído por x^k ;
 - 8: **SE NÃO**
 - 9: Um novo grupo é criado tendo como centro x^k ;
 - 10: **FIM SENÃO**
 - 11: **SENÃO**
 - 12: Os parâmetros consequentes do grupo mais próximo são atualizados
 - 13: **FIM SE**
 - 14: **FIM**
-

Fonte: Adaptado de (ANGELOV et al., 2006)

3.2.1.1 Modelo Takagi-Sugeno Estendido

Uma versão estendida do modelo evolutivo Takagi-Sugeno foi proposta por (ANGELOV; ZHOU, 2006). A versão denominada *eXtended Takagi-Sugeno*, ou *xTS*, estima o raio de atuação de cada grupo de forma recursiva à partir das amostras de dados. Outra funcionalidade introduzida no *xTS* é a idade dos grupos, utilizada para medir a relevância dos grupos. A idade de um grupo é definida pela diferença entre o número de amostras já processadas e o valor médio dos índices temporais das amostras representadas pelo grupo:

$$\text{idade}^j = k - \frac{2A^j}{k-1} \quad (3.13)$$

onde A^j representa o tempo acumulado de chegada. O índice de idade assume valores entre $(0, k]$. Se os valores da idade estão próximos à 0, isto indica que o grupo é novo. Caso contrário, se o valor do índice está próximo ao valor de k , isto sugere que o grupo é antigo e que seu centro deve ser substituído por novas amostras.

3.3 Redes Neuro fuzzy evolutivas

Uma rede neuro fuzzy evolutiva é um modelo evolutivo neural cuja estrutura pode ser interpretada como um conjunto de regras fuzzy. Um sistema de inferência neuro fuzzy é constituído de regras e um método de inferência representado em uma estrutura evolutiva. Sistemas de inferência neuro fuzzy evolutivos são sistemas em que os mecanismo de aprendizado e inferência evolui conforme a necessidade (KASABOV, 2007).

3.3.1 Sistema de Inferência Neuro-Fuzzy Evolutivo Dinâmico

Proposto por (KASABOV, 2001), o *Dynamic Evolving Neural-Fuzzy Inference System*, (DENFIS), pode ser aprender a partir de dados tanto em modo *online* quanto *offline*. Ele pode ser aplicado em previsão de séries temporais. Este sistema evolui através do aprendizado incremental, supervisionado ou não-supervisionado. Durante sua operação, novas regras fuzzy podem ser criadas, adicionadas, e excluídas da estrutura da rede neural. Um método de agrupamento evolutivo (ECM) é aplicado em combinação com um algoritmo um de mínimos quadrados recursivos (GRANT, 1987) durante o aprendizado. DENFIS utiliza o modelo de inferência fuzzy do tipo Takagi-Sugeno conforme:

$$\left\{ \begin{array}{l} \text{Se } x_1 \text{ é } A_{11} \text{ e } x_2 \text{ é } A_{12} \text{ e } \dots \text{ e } x_q \text{ é } A_{1q} \text{ então } y \text{ é } f_1((x_1, x_2, \dots, x_q)) \\ \text{Se } x_1 \text{ é } A_{21} \text{ e } x_2 \text{ é } A_{22} \text{ e } \dots \text{ e } x_q \text{ é } A_{2q} \text{ então } y \text{ é } f_2((x_1, x_2, \dots, x_q)) \\ \vdots \\ \text{Se } x_1 \text{ é } A_{m1} \text{ e } x_2 \text{ é } A_{m2} \text{ e } \dots \text{ e } x_q \text{ é } A_{mq} \text{ então } y \text{ é } f_m((x_1, x_2, \dots, x_q)) \end{array} \right. \quad (3.14)$$

onde $x_j \text{ é } A_{ij}, i = 1, 2, \dots, m; j = 1, 2, \dots, q$ são $m \times q$ proposições fuzzy que formam q antecedentes para cada regra fuzzy. $x_j, j = 1, 2, \dots, q$, são as variáveis antecedentes definidas sobre seus respectivos universos de discurso X_j . R_{ij} são os conjuntos fuzzy definidos pelas funções de pertinência $\mu_{R_{ij}} : X_j \rightarrow [0, 1], j = 1, 2, \dots, q$. Na parte consequente das regras fuzzy, y é a variável consequente, e as funções $f_i, i = 1, 2, \dots, m$, são utilizadas.

Em ambos modos *online* e *offline*, o modelo DENFIS utiliza funções de pertinência do tipo triangular com três parâmetros, a, b, c , conforme definido abaixo:

$$\mu(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x \leq b \\ \frac{c-x}{c-b} & b \leq x < c \\ 0 & x \geq c \end{cases} \quad (3.15)$$

onde b é a posição do centro de um grupo ECM; e a e c são obtidos a partir do raio de grupos ECM.

No modelo DENFIS online, as regras são criadas e atualizadas ao mesmo tempo em que ocorre o particionamento do espaço de entrada. Essencialmente, os seguintes passos de aprendizado são realizados. Para detalhes do algoritmo ver (KASABOV, 2007).

Figura 3.3 – Algoritmo de aprendizagem: DENFIS

Algoritmo de aprendizagem: DENFIS

- 1: Escolher os parâmetros iniciais;
 - 2: Para $k = 1, \dots$, FAZER
 - 3: Receber $(x, y)^k$;
 - 4: Proceder agrupamento ECM;
 - 5: Determinar centros, e alocar dados de acordo com uma métrica;
 - 6: Obter as regras correspondentes a grupos;
 - 7: Criar os antecedentes utilizando centros dos grupos;
 - 8: Calcular os parâmetros das funções consequentes, via MQR.
 - 9: **FIM**
-

Fonte: Adaptado de (KASABOV, 2001)

3.4 Fluxo de Dados Granulares

Grânulos de informação desempenham um papel importante nas atividades cognitivas e tomadas de decisões (BARGIELA; PEDRYCZ, 2012). No contexto de inteligência computacional, granular é uma forma de resolver problemas através da decomposição deste, em que cada parte resultante pode ser solucionada separadamente (LEITE et al., 2012). O conceito de computação granular é discutido em (YAO et al., 2000), (PEDRYCZ; SKOWRON; KREINOVICH, 2008), (LEITE; COSTA; GOMIDE, 2013). Grânulos são conjuntos de elementos associados por sua similaridade espacial, temporal ou por relacionamento funcional (YAO; VASILAKOS; PEDRYCZ, 2013). O termo computação granular se refere ao processamento de grânulos de informação (PEDRYCZ et al., 2007).

A granulação de uma série temporal é uma abordagem para lidar com dinâmicas complexas (PEDRYCZ et al., 2007). O nível de abstração suportado por um grânulo de informação é relacionado ao número de elementos neste grânulo. Neste caso, a medida de cardinalidade de um grânulo funciona como um descritor deste. Quanto maior a granularidade, maior o nível de abstração e menor o nível de especificidade (LEITE et al., 2011). Segundo (PEDRYCZ; SKOWRON; KREINOVICH, 2008), a granularidade de um grânulo A finito pode ser dada por:

$$\text{card}(A) = \sum_{i=1}^n A(x_i) \quad (3.16)$$

onde $A(x)$ é a descrição formal de um grânulo de informação (PEDRYCZ; SKOWRON; KREINOVICH, 2008). Em sistemas fuzzy, são utilizados graus de pertinência para a realização dos cálculos.

3.5 Sistemas Granulares Evolutivos

Sistemas granulares evolutivos podem ser definidos como um conjunto de abordagens de modelagem à partir de fluxo de dados incertos ou numéricos e adaptam modelos baseados em regras. Nesta seção será apresentada a abordagem Modelagem Evolutiva baseada em Conjuntos Fuzzy (FBeM) (LEITE et al., 2011).

3.5.1 Modelagem Evolutiva Granular Fuzzy

Um grânulo consiste de um conjunto de objetos que possuem características em comum como equivalência, semelhança, proximidade, funcionalidade ou indistinguibilidade (PEDRYCZ; SKOWRON; KREINOVICH, 2008). O modelo FBeM consiste de um conjunto de regras fuzzy. FBeM é útil quando os dados são ilimitados e algoritmos convencionais apresentam problema de escalabilidade. FBeM utiliza aprendizagem incremental e fornece aproximações granulares e singulares de funções não-lineares (LEITE et al., 2011). O método FBeM não necessita de dados *a priori* para iniciar sua fase de aprendizagem, ou seja, estes modelos aprendem do zero. Regras e grânulos são criados dinamicamente e adaptados ao longo do tempo (LEITE et al., 2011). Para cada grânulo de informação existe uma regra correspondente. Segundo (LEITE et al., 2015), a parte antecedente das regras FBeM consiste de hiper retângulos fuzzy, e a parte consequente é composta de termos funcionais e linguísticos. Uma regra FBeM possui a seguinte forma:

$$\begin{aligned}
R^i : & \text{Se}(x_1 \text{ é } A_1^i) \text{ e } \dots \text{ e } (x_j \text{ é } A_j^i) \text{ e } \dots \text{ e } (x_n \text{ é } A_n^i) \\
& \text{Então } (y_1 \text{ é } B_1^i) \text{ e } \bar{y}_1 = p_k^1(x_j \forall_j) \text{ e } \dots \\
& \text{e } (y_m \text{ é } B_m^i) \text{ e } \bar{y}_m = p_m^i(x_j \forall_j)
\end{aligned} \tag{3.17}$$

onde $(y_m \text{ é } B_m^i)$ é a parte linguística da equação e $\bar{y}_m = p_m^i(x_j \forall_j)$ é a parte funcional. x_j e y_j são variáveis provenientes do fluxo de dados $(x, y)^{[h]}$, onde $h = 1, 2, \dots$. A_j^i e B_k^i são funções de pertinência criadas conforme os dados; p_k^i representa uma função afim.

O conseqüente de uma regra FBeM é composto pela união entre um conjunto fuzzy funcional. Estes fornecem uma aproximação da saída e limites de tolerância na aproximação. Grânulos no espaço de saída são formados por conjuntos fuzzy trapezoidais $B_k^i = (u^i, v^i, \gamma^i, U^i)$. A função local p^i é adaptada usando amostras que pertencem ao grânulo γ^i . As funções p^i não são necessariamente lineares (LEITE et al., 2015). A função linear do modelo FBeM é:

$$p^i(x_1, \dots, x_n) = a_0^i + \sum_{j=1}^n a_j^i x_j \tag{3.18}$$

onde o algoritmo de mínimos quadrados recursivos vide Anexo B é utilizado para determinar os coeficientes de p^i . Se funções p^i de ordem superior são utilizadas para aproximar f , então o número de coeficientes aumenta.

Cada regra FBeM contribui para saída do sistema, pois a representação trapezoidal permite que todos grânulos sobreponham. A saída singular do sistema FBeM é dada pelo valor da média ponderada de todas as regras criadas, conforme:

$$p_k = \frac{\sum_{i=1}^c \min(A_1^i, \dots, A_n^i) p_k^i}{\sum_{i=1}^c \min(A_1^i, \dots, A_n^i)} \tag{3.19}$$

A saída granular FBeM dada por B^{i*} pode ser mais relevante que a saída numérica p_k . A saída granular B^{i*} é obtida de:

$$ch(B^1, \dots, B^c) = (\min(u^1, \dots, u^c), \min(u^1, \dots, u^c), \max(Y^1, \dots, Y^c), \max(U^1, \dots, U^c)) \tag{3.20}$$

onde $ch(\cdot)$ é a casca convexa (MOORE; KEARFOTT; CLOUD, 2009).

A máxima largura que um conjunto fuzzy A_j^i pode assumir no FBeM é limitada por ρ . Onde, $largura(A_j^i) = L_j^i - l_j^i \leq \rho$, $j = 1, \dots, n$; $i = 1, \dots, c$. Diferentes valores de ρ produzem diferentes representações dos fluxos de dados em diferentes níveis de granularidade. Sendo a região de expansão de um conjunto A_j^i denotada por:

$$E_j^i = [mp(A_j^i) - \frac{\rho}{2}, mp(A_j^i) + \frac{\rho}{2}] \quad (3.21)$$

onde $mp(A_j^i)$ é ponto central de A_j^i . A região de expansão auxilia o FBeM a determinar quando uma amostra de dados deve ser considerada ou não em um grânulo já existente. Para dados normalizados o valor de ρ assume valores entre $[0, 1]$. Se o valor de ρ é igual à 0 então o grânulo não é capaz de se expandir. Portanto, o processo de aprendizagem no FBeM cria uma regra para cada amostra de dados, o que pode levar a um *overfitting* e uma alta complexidade computacional. Se o ρ assume o valor 1, então o FBeM cria um único grânulo, o que pode não ser ideal em um ambiente em que os dados são não-estacionários. O FBeM utiliza um procedimento rápido para evoluir o ρ durante o tempo, como segue. Sendo r a diferença entre o número de grânulos corrente e o número de grânulos h_r passos anteriores, $r = c^{[h]} \setminus c^{[h-h_r]}$. Se a quantidade de grânulos crescer mais rápido do que uma dada taxa de crescimento η , ou seja, $r > \eta$, então o ρ é incrementado:

$$\rho(novo) = (1 + \frac{r}{h_r})\rho(velho) \quad (3.22)$$

A equação 3.22 controla o valor de ρ de modo que rejeita grandes valores de bases de regras, e portanto evita o aumento de complexidade. Se o valor de ρ cresce à uma taxa menor do que o valor η , ou seja, $r \leq \eta$, então o valor de ρ decresce conforme:

$$\rho(novo) = (1 - \frac{\eta - r}{h_r})\rho(velho) \quad (3.23)$$

Este procedimento mantém o modelo de granularidade variando no tempo de acordo com o fluxo de dados. No FBeM não necessariamente existem regras antes do início do procedimento de aprendizagem. Portanto, o procedimento incremental para criação de regras ocorre sempre que uma entrada (x_1, \dots, x_n) não pertence à região de expansão (E_1^i, \dots, E_n^i) , $i = 1, \dots, c$, ou a saída $y \notin E_k^i$, $i = 1, \dots, c$. Caso contrário, a base de regras não é alterada.

O algoritmo de aprendizagem FBeM é como segue.

Figura 3.4 – Algoritmo de aprendizagem: FBeM

Algoritmo de aprendizagem: FBeM

-
- 1: **DEFINIR** os parâmetros p, h_r, η, ψ, c ;
 - 2: **LER** $(x, y)^{[h]}$, $h = 1$
 - 3: **CRIAR** grânulo y^{c+1} ;
 - 4: **PARA** $h = 2, \dots$ **FAZER**
 - 5: **LER** $(x, y)^{[h]}$
 - 6: **FORNECER** aproximações singulares $p(x^{[h]})$;
 - 7: **FORNECER** aproximações granulares B^{i*} ;
 - 8: **CALCULAR** o erro $\varepsilon^{[h]} = y^{[h]} - p(x^{[h]})$;
 - 9: **SE** $y^{[h]}$ ou $x^{[h]}$ não estão na região de expansão dos grânulos $E^i \forall i$;
 - 10: **CRIAR** grânulo γ^{c+1} ;
 - 11: **SENÃO**
 - 12: **ADAPTAR** o grânulo mais ativo $\gamma^i, i = \max_i(S(x, A^1), \dots, (S(x, A^c)))$;
 - 13: **ADAPTAR** os parâmetros α_i^j utilizando mínimos quadrados recursivos
 - 14: **FIM SE**
 - 15: **SE** $h = \alpha$, $\alpha = 1, 2, \dots$
 - 16: **COMBINAR** os grânulos semelhantes
 - 17: **ATUALIZAR** a granularidade ρ
 - 18: **REMOVER** os grânulos inativos
 - 19: **FIM SE**
 - 20: **FIM**
-

Fonte: Adaptado de (LEITE et al., 2015)

3.6 Inferência Fuzzy Baseada em Nuvens de Dados

(ANGELOV; YAGER, 2011) apresentaram um novo método para definição da parte antecedente de uma regra fuzzy. Nesse caso, não há necessidade de definir funções de pertinência para cada variável. Também não há necessidade da utilização de conectivos lógicos termos. A parte antecedente em AnYa é formada por uma nuvem de dados que é muito semelhante aos *clusters*. Porém, nuvem de dados se difere de *cluster* por não possuir formato específico, fronteiras ou parâmetros. A Tabela 3.2 abaixo expõe as diferenças entre nuvem e *cluster* de dados.

As nuvens de dados utilizam densidade baseada em todas as amostras de dados. Regras fuzzy AnYa não requerem o uso de um tipo particular de distância, e.g. Euclidiana, Mahalono-bis, e não requer a definição de centros de conjuntos fuzzy (ANGELOV; YAGER, 2011).

Uma regra fuzzy do tipo AnYa pode ser definida conforme:

Tabela 3.1 – Comparação entre nuvens de dados e *clusters*

| Características | Nuvem de dados | <i>Clusters</i> |
|------------------------|--|------------------------------|
| Fronteiras | Dependem do <i>threshold</i> escolhido | Definidas |
| Centros | Não possui | Definidos |
| Calculos de distâncias | Entre todos os dados (vetores) | Entre os centros (escalares) |
| Funções de Pertinência | Não parametrizadas | Parametrizadas |

Fonte: Adaptado de (ANGELOV; YAGER, 2011)

Tabela 3.2 – Comparação entre métodos de inferência fuzzy

| Método de inferência | Antecedentes (SE) | Consequentes (ENTÃO) | Método de Defuzzificação |
|----------------------------------|--|---|---|
| Mandami Takagi-Sugeno AnYa | Conjuntos fuzzy Conjuntos fuzzy Nuvem de dados | Váriaveis linguísticas Funções Matemáticas Váriaveis linguísticas e/ou Funções Matemáticas | Centro de gravidade Soma ponderada Centro de gravidade e/ou Soma ponderada |

Fonte: Adaptado de (ANGELOV; YAGER, 2011)

$$R^i = \text{Se } (x \sim X^i) \text{ Então } (u^i) \quad (3.24)$$

onde \sim representa a forma linguística "associado a". $X^i \in R^n$ representa a i -ésima nuvem de dados. $x = [x_1, x_2, \dots, x_n]^T$ é um vetor de entradas, e $u^i = [u_1^i, u_2^i, \dots, u_n^i]$ é um vetor de saídas.

Uma nuvem de dados representa todas as amostras de dados anteriores diretamente, ao contrário das funções de pertinência tradicionais que não representam a distribuição real dos dados (ANGELOV; YAGER, 2011). O grau do conjunto de pertinência de uma nuvem de dados é expresso pela densidade normalizada relativa de uma amostra x_k :

$$\lambda_k^i = \frac{y_k^i}{\sum_{i=1}^n y_k^i} \quad (3.25)$$

onde λ_k^i representa a i -ésima densidade local de uma nuvem associada a uma amostra de dados. A densidade global de uma amostra de dados pode ser definida por:

$$\Gamma_k = K \left(\sum_{j=1}^k d_{kj} \right) \quad (3.26)$$

onde d_{kj} representa uma métrica de distância. A densidade local pode ser definida através de um núcleo do tipo Cauchy conforme:

$$Y_k^i = \frac{1}{1 + \frac{\sum_{j=1}^{m^i} (d_{kj}^i)^z}{M^i}} = \frac{M^i}{M^i + \sum_{j=1}^{m^i} (d_{kj}^i)^z} \quad (3.27)$$

assim como a densidade local a densidade global pode ser calculada por um núcleo do tipo Cauchy:

$$Y_k^i = \frac{1}{1 + \frac{\sum_{j=1}^{K-1} (d_{kj}^i)^z}{K-1}} = \frac{K-1}{K-1 + \sum_{j=1}^{K-1} (d_{kj}^i)^z} \quad (3.28)$$

O grau da função de pertinência de uma nuvem de dados é normalizado, isto é:

$$\sum_{i=1}^n \gamma^i = 1 \quad (3.29)$$

Regras AnYa podem utilizar tanto o método de defuzzificação através de média ponderada, utilizado pelo sistema de inferência Takagi-Sugeno, como o método defuzzificação pelo centro de gravidade, utilizado pelo método Mandami (ANGELOV; YAGER, 2011).

3.6.1 Tipicidade e Excentricidade

TEDA (*Typicality and Eccentricity Data analytics*) é um algoritmo baseado em regras fuzzy para aprendizagem incremental a partir de grandes volumes de dados (ANGELOV, 2014). A aprendizagem no TEDA é baseada nos conceitos de tipicidade e excentricidade (SOARES et al., 2017), conceitos relacionados respectivamente a proximidade e densidade dos dados em um espaço. Mais especificamente, o conceito de excentricidade está ligado ao quanto uma amostra é distinta das demais da base de conhecimento. Ao contrário, o conceito de tipicidade é baseado no quão similar uma amostra de dados é das demais amostras. TEDA não requer informação *a priori* sobre os dados ou parâmetros pré-definidos. Por utilizar uma abordagem recursiva, TEDA é capaz de processar fluxos de modo rápido, *online*, e geralmente em tempo real (KANGIN; ANGELOV, 2015). TEDA lida com fluxo de dados e séries temporais.

Sendo $X = [x_{(1)}, \dots, x_{(K)}]'$, $X \in \mathfrak{R}^{K \times n}$, um conjunto de dados; $k = 1, \dots, K$ um índice de tempo. A soma das distâncias, $\pi(\cdot)$, da k -ésima amostra, $x_{(k)}$, para todas as amostras de X é dada por:

$$\pi(x_{(k)}) = \sum_{j=1}^K d(x_{(k)}, x_{(j)}), \quad (3.30)$$

onde $d(\cdot)$ pode ser qualquer métrica de distância. Recursivamente, $\pi(\cdot)$ no índice de tempo k , i.e., a soma das distâncias entre as k -ésimas amostras de dados e a amostra anterior do fluxo de dados é dada por:

$$\pi(x_{(k)}) = \pi(x_{(k-1)}) + d(x_{(k)}, x_{(k-1)}). \quad (3.31)$$

A excentricidade $\varepsilon(\cdot)$ de $x_{(k)}$ em relação a todas as amostras contidas na base de dados é dada por :

$$\varepsilon(x_{(k)}) = \frac{2\pi(x_{(k)})}{\sum_{j=1}^K \pi(x_{(j)})} \quad (3.32)$$

a tipicidade $\gamma(\cdot)$ é o complemento da excentricidade, i.e.,

$$\gamma(x_{(k)}) = 1 - \varepsilon(x_{(k)}). \quad (3.33)$$

Assumindo $d(\cdot)$ como norma-2, a excentricidade pode ser calculada por:

$$\varepsilon(x_{(k)}) = \frac{1}{2k} + \frac{(\mu_x^k - x_{(k)})^T (\mu_x^k - x_{(k)})}{2k[\sigma_x^k]^2} \quad (3.34)$$

onde a tipicidade pode ser calculada recursivamente através de:

$$\gamma(x_{(k)}) = \frac{k-1}{k} - \frac{(\mu_x^k - x_{(k)})^T (\mu_x^k - x_{(k)})}{k[\sigma_x^k]^2}. \quad (3.35)$$

A média μ_x^k , média dos quadrados $\mu_{x^T x}^k$, e variância $[\sigma_x^k]^2$ das nuvens de dados são atualizadas através de:

$$\mu_x^k = \frac{(k-1)\mu_x^{k-1} + x_{(k)}}{k}, \quad (3.36)$$

$$\mu_{x^T x}^k = \frac{(k-1)\mu_{x^T x}^{k-1} + x_{(k)}^T x_{(k)}}{k}, \quad (3.37)$$

e

$$[\sigma_x^k]^2 = \mu_{x^T x}^k - [\mu_x^k]^T \mu_x^k, \quad (3.38)$$

sendo $\mu_x^0 = \mu_{x^T x}^0 = [\sigma_x^0]^2 = 0$.

Ao se adotar norma 2, as nuvens de dados formadas pelo TEDA assumem forma circular, o que pode trazer desvantagens ao não cobrir todo o espaço de dados.

3.6.1.1 TEDA Predict

O algoritmo TEDA para previsão, difere linguisticamente e funcionalmente na composição dos sistemas de regras fuzzy (SOARES et al., 2017). Uma regra fuzzy TEDA Predict é dada por:

$$R_i(x) : \text{Se } (x \sim x_i^*) \text{ então } y_i = x^T \Theta_i \quad (3.39)$$

onde $i = 1, \dots, N$; Θ_i é uma matriz de coeficientes, e \sim indica o relacionamento de uma amostra x à regra R_i ; e x_i^* é um ponto representativo na nuvem de dados (ponto focal). Em particular, a regra R_i mapeia as amostras $x \in \mathfrak{R}^n$ que são próximas à i -ésima nuvem em um valor de saída $y \in \mathfrak{R}$.

A excentricidade local, ε_i , e a tipicidade local, γ_i , são definidas conforme as amostras são associadas a i -ésima nuvem de dados. Uma nuvem e uma regra correspondente são criadas quando o valor da tipicidade local $\gamma_i \forall i$ é menor que um limiar α , i.e.,

$$\max_{i=1, \dots, N} \gamma_i(x_{(k)}) \leq \alpha. \quad (3.40)$$

caso contrário, a amostra $x_{(k)}$ é atribuída a regra ativa mais relevante. O nível de ativação da i -ésima regra para uma amostra é:

$$w_i(x_{(k)}) = \frac{\gamma_i(x_{(k)})}{\sum_{j=1}^N \gamma_j(x_{(k)})} \quad (3.41)$$

Assim que uma amostra é atribuída a uma regra R_{i^*} , a excentricidade local, ε_{i^*} , e a tipicidade local, γ_{i^*} , são atualizadas baseadas nas equações 3.34 e 3.35. Além disso, os parâmetros consequentes Θ_{i^*} são atualizados através dos Mínimos Quadrados Recursivos (MQR), vide Anexo B. As estimativas de saída são dadas por:

$$\bar{y}_{(k)} = \sum_{i=1}^N w_i \bar{y}_{i(k)}. \quad (3.42)$$

O valor real $y_{(k)}$ se torna disponível após a estimação (estimação de um passo). O algoritmo de aprendizagem e estimação do TEDA Predict é sumarizado abaixo.

Figura 3.5 – Algoritmo de aprendizagem: TEDA Predict

Algoritmo de aprendizagem: TEDA Predict

- 1: **DETERMINAR** α , $N = 0$;
 - 2: **LER** a primeira amostra $x_{(1)}$;
 - 3: **CRIAR** a regra R_N , $N = N + 1$;
 - 4: **PARA** $k = 2, \dots$
 - 5: **LER** $x_{(k)}$;
 - 6: **CALCULAR** $\gamma_i(x_{(k)})$ e $\varepsilon_i(x_{(k)})$, $i = 1, \dots, N$;
 - 7: **CALCULAR** os níveis de ativação $w_i(x_{(k)})$, $i = 1, \dots, N$;
 - 8: **FORNECER** estimaco $\bar{y}_{(k)}$;
 - 9: **SE** equaco 3.40 for verdade
 - 10: **CRIAR** regra R_{N+1} , $N = N + 1$;
 - 11: **SENO**
 - 12: **ATUALIZAR** γ_{i^*} e ε_{i^*} onde $i^* = \operatorname{argmax} w_i(x_{(k)})$;
 - 13: // A sada $y_{(k)}$ se torna disponvel
 - 14: **ATUALIZAR** Θ_{i^*} usando MQR;
 - 15: **FIM**
 - 16: **FIM**
-

Fonte: Adaptado de (SOARES et al., 2017)

4 METODOLOGIA

Neste capítulo é apresentada a metodologia empregada na realização dos experimentos, detalhes das bases de dados, fórmulas para cálculo de erro de previsão, e interconexão de modelos.

4.1 Seleção de Características Baseadas na Correlação de Spearman

A ordenação e seleção de características possibilitam a redução da complexidade de um problema. Um subconjunto de características mais relevantes ou discriminativas facilita a interpretação de um modelo, reduz a chance de *overfitting*, e pode produzir melhores resultados ao eliminarem características que podem confundir o processo descoberta de padrões, tendências e relacionamentos. Seja $[x_1, \dots, x_i, \dots, x_n]$ um conjunto de características, e y a variável dependente. As características podem ser pontuadas através de

$$S_i = \frac{1}{n} \sum_{j=1}^n |\rho(x_i, x_j)| - |\rho(x_i, y)|, \quad (4.1)$$

onde $i = 1, \dots, n$; ρ é o coeficiente de Spearman (SPEARMAN, 1904),

$$\rho(x_i, x_j) = 1 - \frac{6 \sum_{k=1}^K d_{ij}^{(k)2}}{K(K^2 - 1)}, \quad (4.2)$$

onde K é o número de amostras. Além disso,

$$d_{ij}^{(k)} = \text{rank}(x_{i(k)}) - \text{rank}(x_{j(k)}) \quad (4.3)$$

é a diferença entre as classificações das amostras após ordenações destas de modo ascendente. Se os valores consecutivos de uma variável, $x_{i(k)}$ e $x_{i(k+1)}$, são idênticos, então valores fracionários, iguais à média de suas posições na ordem ascendente de valores, são atribuídos. Isto é equivalente a uma média sobre todas as permutações possíveis.

O ρ de Spearman varia de -1 à 1 . Um valor de 1 (ou -1) implica que uma função monotonicamente crescente (decrecente) descreve a relação entre as variáveis perfeitamente. Uma

correlação de 0 indica que não há tendência do valor de uma variável aumentar ou diminuir quando outra variável aumenta. O coeficiente de Spearman aumenta conforme as variáveis tornam-se mais próximas de serem funções monótonas perfeitas umas das outras. Diferentemente do coeficiente de Pearson (PENG; LONG; DING, 2005), o ρ de Spearman não assume linearidade ou normalidade.

Quanto mais próximo o de 0 o valor de $S_i \in [-1, 1]$, mais importante é a i -ésima característica. A ideia é que uma característica interessante possui baixa correlação com outras características, o que implica que esta transmite informações únicas e diferentes. A alta correlação com a variável dependente implica que esta auxilia na previsão. Uma característica é deixada de com base na sua pontuação S_i até que uma redução significativa de desempenho de previsão seja notada. Em outras palavras, quando o erro quadrado médio das previsões é maior que um limiar ϵ , o processo de exclusão de variáveis menos importantes segundo sua pontuação é cessado. Como a velocidade de processamento é essencial no ambiente *online*, o método de seleção de características proposto nesta seção é útil para reduzir o problema de dimensionalidade (PENG; LONG; DING, 2005).

4.2 Dados Utilizados

O propósito deste trabalho é prever a temperatura média mensal das cidades de São Paulo, Manaus, Porto Alegre e Natal utilizando os métodos eTS, xTS, DENFIS, FBeM, e TEDA Predict. Seis variáveis, temperatura média (x_1), mínima (x_2) e máxima (x_3), nebulosidade (x_4), precipitação (x_5), e umidade do ar (x_6) são consideradas. Seus valores correspondem â dados mensais, e correspondem entre o período de Janeiro de 1990 à Dezembro 2015 (312 amostras). Os dados das diferentes estações meterológicas brasileiras estão disponíveis no *site* do Instituto Nacional de Meteorologia conforme <<http://www.inmet.gov.br/>>.

Os dados foram redimensionados na escala $[0, 1]$ para considerar as variáveis na mesma proporção. A normalização dos dados é relaizada de forma recorrente, para favorecer um ambiente online. Ou seja, os valores máximos e mínimos são atualizados conforme os dados vão chegando através do fluxo de dados. A normalização do k -ésimo valor da i -ésima variável é dada por:

$$x_{i(k)}^r = \frac{x_{i(k)} - \min_{\forall j} (x_{i(j)})}{\max_{\forall j} (x_{i(j)}) - \min_{\forall j} (x_{i(j)})} \quad (4.4)$$

em que $j = 1, \dots, k - 1$; e $x_{i(k)}^r \in [0, 1]$.

À princípio, uma janela temporal de tamanho 24, relacionada a um período de 2 anos, foi considerada para a seleção de características. Por tanto, o conjunto original dos dados de entrada é dado por

$$X = [x_{1(k-23)} \dots x_{6(k-23)} \ x_{1(k-22)} \dots \dots x_{6(k-22)} \dots \dots x_{1(k)} \dots x_{6(k)}]. \quad (4.5)$$

A saída dos modelos evolutivos é a temperatura média estimada para o próximo mês ($\bar{y}_{(k+1)} = x_{1(k+1)}$).

O erro de previsão é quantificado através do índice de erro denominado raiz do erro médio quadrático (*RMSE*) conforme

$$RMSE = \frac{1}{k_c} \sum_{k=1}^{k_c} \sqrt{(\bar{y}_{(k+1)} - y_{(k+1)})^2}, \quad (4.6)$$

onde k_c é o número de iterações. O índice de erro Não-Dimensional (NDE),

$$NDE = \frac{RMSE}{std(y_{(k)\forall k})}, \quad (4.7)$$

é útil para comparar a acurácia de diferentes dos previsores para fluxo de dados diferentes, isto é, fluxo de dados com diferentes desvios-padrão.

No tocante ao mesmo tempo de processamento e construção de modelo, foi considerado o *software Matlab R2017a*, com uma CPU de 1.8GHz e 4Gb de memória RAM.

As séries temporais de temperatura das cidades de São Paulo, Manaus, Porto Alegre e Natal são ilustradas na Fig. 4.1. Nota-se que 288 amostras são utilizadas dado o fato da janela temporal de 2 anos (Janeiro/1990 - Dezembro/1991) necessárias para constituir a primeira amostra. Note que as sériessão não-estacionárias, visto variações do valor média. Visualmente,

aas variações são evidentes principalmente nas Figuras 4.1(b) e 4.1(d). As demais séries também são não-estacionárias.

Os parâmetros iniciais para obtenção dos resultados foram definidos da seguinte forma: (i) eTs – $\Omega = 750$; (ii) xTS – $\Omega = 750$; (iii) DENFIS – $dthr = 0.1$; (iv) FBeM – $\rho = 0.7, h_r = 48, \eta = 2$; (v) TEDA – $\alpha = 2$. Apesar do FBeM poder processar dados fuzzy, nesta pesquisa considerou-se dados pontuais.

4.3 Ensemble de Nuvens de Dados e Modelos Fuzzy

Primeiramente, analisou-se os modelos eTS, xTS, DENFIS, FBeM e TEDA Predict individualmente, conforme as descrições em capítulos anteriores.

Um *ensemble* de modelos e operadores de agregação é também proposto para avaliação. eTS, xTS, FBeM e TEDA Predict são construídos de forma incremental à partir do zero usando os dados X . É dado um passo de aprendizagem para cada um dos modelos seguido de um passo de agregação. A ideia do *ensemble* é ponderar e agregar a contribuição dos modelos individuais e fornecer uma previsão singular $y_{(k+1)}$. A ponderação é baseada no valor corrente do *RMSE* de um previsor e em uma agregação média M , vide Anexo A,

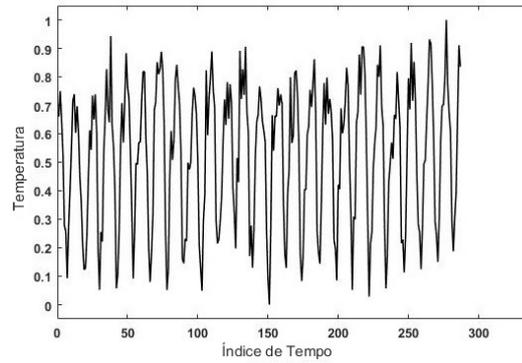
$$\bar{y}_{(k+1)} = \sum_{p=1}^4 w_p \bar{y}_{p(k+1)}, \quad (4.8)$$

onde

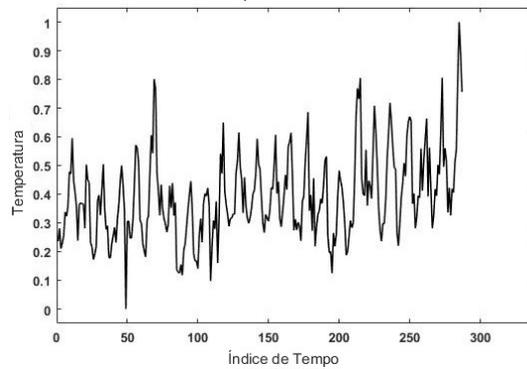
$$w_p = RMSE(modelo_p) \quad (4.9)$$

na k_c -ésima iteração. Além disso, os operadores T_{min} e S_{max} são aplicados em cada saída individual, $\bar{y}_{p(k+1)}$, fornecendo limiares superiores e inferiores, respectivamente, na estimação do *ensemble*. A arquitetura do esquema evolutivo é mostrada na Fig. 4.2. A previsão granular dada pelas normas T e S, envolvem a previsão M e podem auxiliar na tomada de decisões em certos contextos.

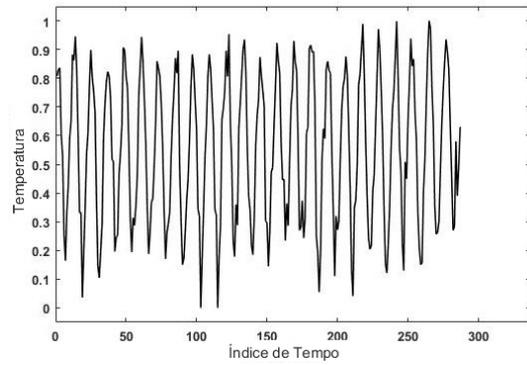
Figura 4.1 – Séries temporais de temperatura das estações meteorológicas: (a) Sao Paulo, (b) Manaus, (c) Porto Alegre, e (d) Natal



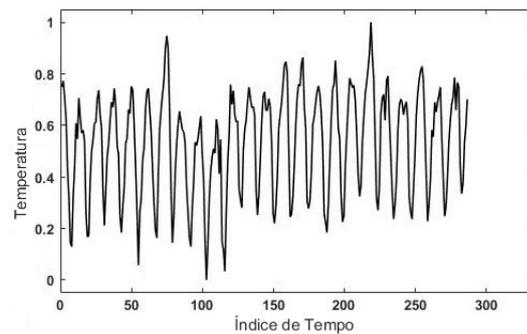
(a)



(b)



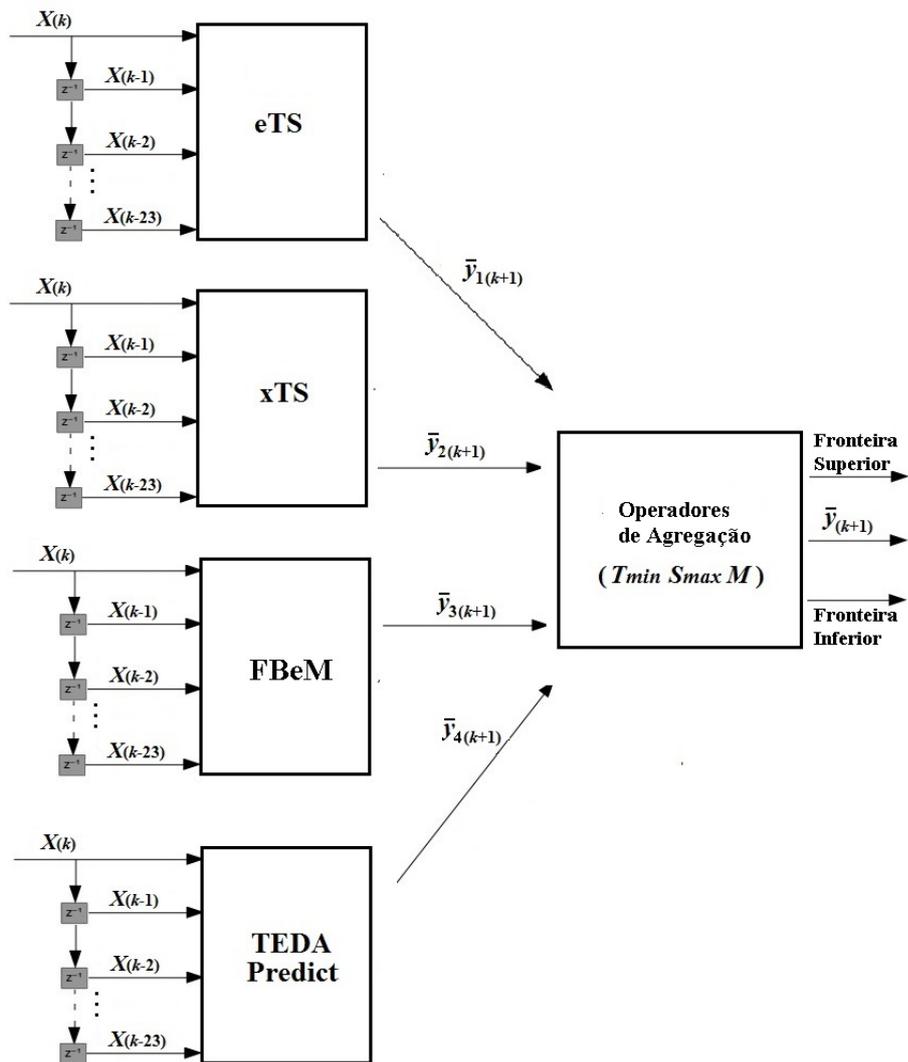
(c)



(d)

Fonte: Arquivo Pessoal

Figura 4.2 – Ensemble de modelos fuzzy e de nuvens de dados



Fonte: Arquivo Pessoal

5 RESULTADOS E DISCUSSÕES

Neste capítulo são apresentados os resultados obtidos partir das simulações computacionais. As simulações foram realizadas utilizando os métodos evolutivos apresentados no capítulo III. A metodologia para obtenção dos resultados apresentados neste capítulo está no capítulo IV.

5.1 Resultados para a Estação Meteorológica de São Paulo

Simulações computacionais foram realizadas para avaliar a acurácia dos métodos evolutivos eTS, xTS, DENFIS, FBeM e TEDA Predict combinados com o método proposto para ordenação e seleção de características baseado na correlação de Spearman. A Tabela 5.1 resume os resultados obtidos pelos métodos evolutivos individualmente para a estação meteorológica de São Paulo considerando diferentes números de características. Os parâmetros iniciais foram ajustados de modo que a estrutura final dos modelos evolutivos não contivessem um número maior que 10 modelos locais.

Nota-se na Tabela 5.1 que os modelos evolutivos são melhores quando todas as 144 características são consideradas. Algumas informações são perdidas quando algumas características são eliminadas do modelo. No entanto, um grande número de características podem ser removidas sem uma perda substancial de acurácia, conforme pode ser observado na Figura 5.1. Note que a rede neuro-fuzzy DENFIS não pode processar um número maior que 72 características ao mesmo tempo, pois sofre de *overfitting*. Portanto, o método para seleção de características proposto neste trabalho é fundamental para este método.

Como indica a Tabela 5.1 e ilustra a Figura 5.1, o método FBeM utilizando 144 características possui o menor valor de *RMSE* (0,0321) quando comparado com os demais métodos utilizados. O método TEDA Predict com 144 características apresentou valor *RMSE* de 0,0329 bem próximo ao valor obtido pelo método FBeM. No entanto, várias de características podem ser deixadas de fora sem uma grande perda de acurácia. Por exemplo, o valor de *RMSE* do TEDA Predict tem um incremento de 0,0329 para 0,0357 utilizando um terço da dimensão original do espaço de entrada. O número de termos antecedentes das regras do TEDA Predict, a compactação do modelo resultante, e o tempo de processamento são melhorados ao custo de uma perda marginal de acurácia na previsão. No caso da rede DENFIS, o método para seleção de características foi imprescindível para que este não sofresse da "maldição da dimensionalidade". Porém, nota-se que conforme as características são desconsideradas, o valor do *RMSE* para DENFIS aumenta de forma polinomial. A Figura 5.2 faz uma comparação entre o tempo

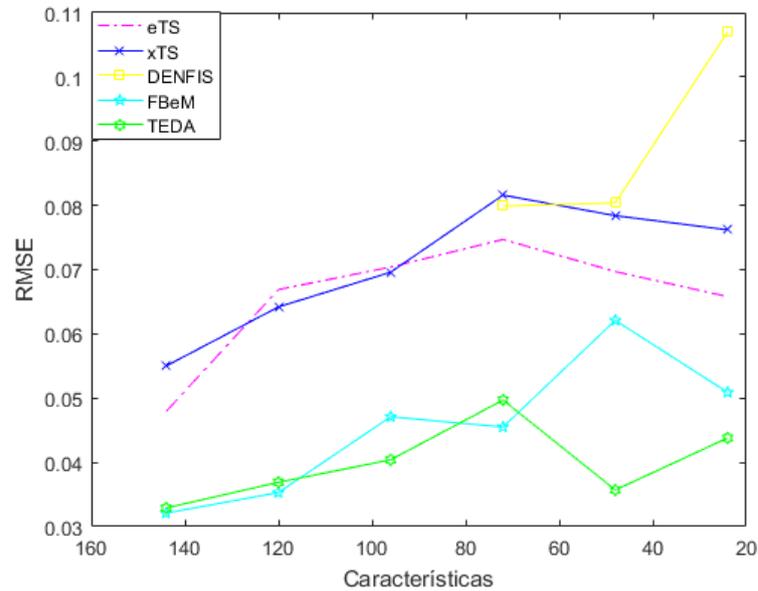
Tabela 5.1 – Previsão: Estação meteorológica de São Paulo

| eTS | | | | |
|-------------------|---------------|---------------|-------------|------------------|
| – | <i>RMSE</i> | <i>NDE</i> | Tempo(s) | # regras |
| # características | | | | |
| 144 | 0,0479 | 0,4154 | 1,95 | 6 |
| 120 | 0,0669 | 0,5802 | 1,76 | 6 |
| 96 | 0,0704 | 0,6106 | 1,69 | 6 |
| 72 | 0,0747 | 0,6479 | 1,72 | 6 |
| 48 | 0,0697 | 0,6045 | 1,82 | 6 |
| 24 | 0,0658 | 0,5707 | 1,21 | 6 |
| xTS | | | | |
| – | <i>RMSE</i> | <i>NDE</i> | Tempo(s) | # regras |
| # características | | | | |
| 144 | 0,0550 | 0,4770 | 1,49 | 5 |
| 120 | 0,0642 | 0,5568 | 1,52 | 5 |
| 96 | 0,0696 | 0,6036 | 1,56 | 5 |
| 72 | 0,0816 | 0,7077 | 1,65 | 5 |
| 48 | 0,0784 | 0,6800 | 1,58 | 5 |
| 24 | 0,0762 | 0,6609 | 1,41 | 5 |
| DENFIS | | | | |
| – | <i>RMSE</i> | <i>NDE</i> | Tempo(s) | # regras |
| # características | | | | |
| 144 | – | – | – | – |
| 120 | – | – | – | – |
| 96 | – | – | – | – |
| 72 | 0,0799 | 0,8360 | 2,90 | 10 |
| 48 | 0,0804 | 0,8400 | 2,27 | 10 |
| 24 | 0,1070 | 1,118 | 2,01 | 10 |
| FBeM | | | | |
| – | <i>RMSE</i> | <i>NDE</i> | Tempo(s) | # regras |
| # características | | | | |
| 144 | 0,0321 | 0,2791 | 1,25 | 5,12 |
| 120 | 0,0353 | 0,3069 | 1,16 | 5,92 |
| 96 | 0,0471 | 0,4095 | 1,12 | 5,46 |
| 72 | 0,0455 | 0,3956 | 1,02 | 5,02 |
| 48 | 0,0621 | 0,5400 | 0,90 | 5,96 |
| 24 | 0,0509 | 0,4426 | 0,79 | 5,80 |
| TEDA Predict | | | | |
| – | <i>RMSE</i> | <i>NDE</i> | Tempo(s) | #nuvens de dados |
| # características | | | | |
| 144 | 0,0329 | 0,2854 | 3,75 | 10 |
| 120 | 0,0369 | 0,3208 | 2,80 | 10 |
| 96 | 0,0404 | 0,3513 | 3,09 | 10 |
| 72 | 0,0497 | 0,4321 | 3,04 | 10 |
| 48 | 0,0357 | 0,3104 | 2,82 | 10 |
| 24 | 0,0438 | 0,3808 | 2,75 | 10 |

de processamento dos métodos utilizados considerando diferentes números de características no espaço de entrada.

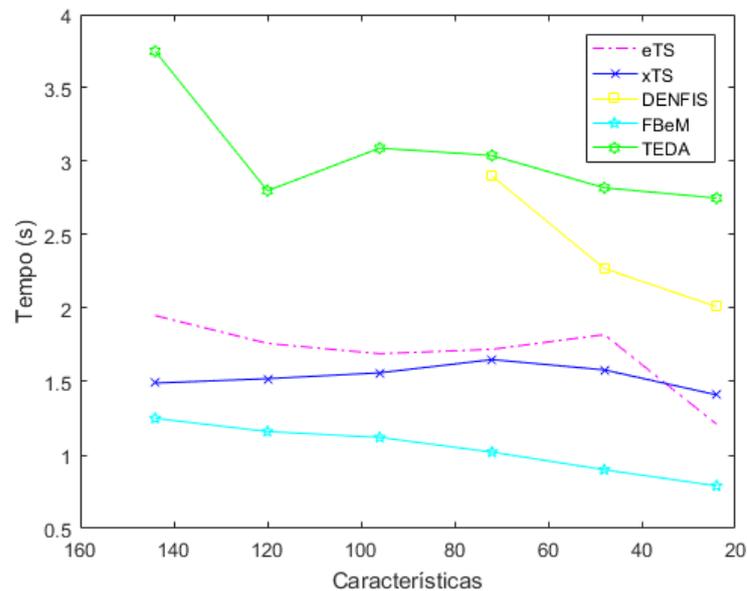
Dentre os métodos mais rápidos destaca-se o FBeM utilizando um sexto da dimensionalidade original do espaço de entradas. Por conta da complexidade algébrica envolvida nos

Figura 5.1 – Comparação de $RMSE$ para diferentes quantidades de características - Estação de São Paulo



Fonte: Arquivo Pessoal

Figura 5.2 – Comparação de $Tempo(s)$ para diferentes quantidades de características - Estação meteorológica de São Paulo

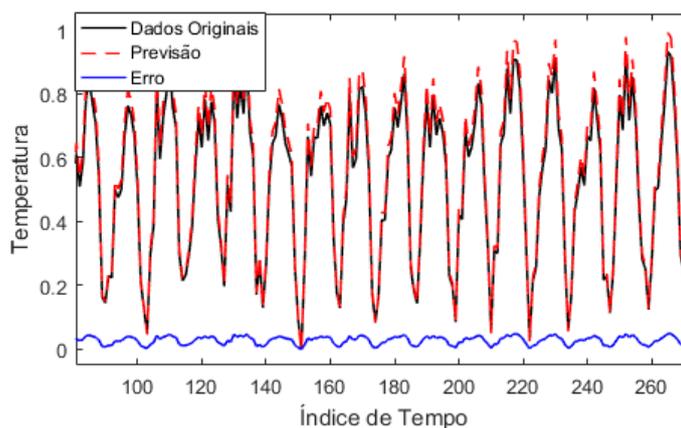


Fonte: Arquivo Pessoal

métodos TEDA Predict e DENFIS, estes destacam-se por gastarem um maior tempo para processar os dados. Em particular o método TEDA Predict é o mais lento visto que vários cálculos de distâncias adicionais entre as amostras, são necessários. Em geral, nota-se que conforme a dimensionalidade do espaço de entradas reduz, os métodos tendem a reduzir o tempo de processamento. Por exemplo, o TEDA Predict com 120 características no espaço de entrada, ou seja,

retirando 24 características da base original, tem-se uma redução de 0,95 segundos no processamento. Os métodos eTS e xTS registraram tempo de processamento similares. A Figura 5.3 ilustra a previsão de um passo para a cidade de São Paulo utilizando o método FBeM (Melhor caso).

Figura 5.3 – Estimação FBeM de um passo para a estação meteorológica de São Paulo



Fonte: Arquivo Pessoal

5.2 Resultados para a Estação Meteorológica de Manaus

O resultado da previsão utilizando os dados da estação de Manaus é mostrado na Tabela 5.2. O resultado foi obtido a partir de simulações computacionais utilizando os métodos eTS, xTS, DENFIS, FBeM e TEDA Predict. Conforme nos resultados apresentados na seção anterior para a estação meteorológica de São Paulo, um seletor de características baseado na correlação de Spearman foi utilizado. A rede neuro-fuzzy DENFIS foi capaz de processar os dados somente com o uso da metade das variáveis originais. O método de seleção proposto é necessário para evitar que seja possível ajustar todos os parâmetros de sua estrutura conexionista.

O previsor TEDA forneceu os melhores índices de erro utilizando todos as características originais para este caso. Aparentemente, nesse caso, quando os primeiros conjuntos de características são deixados de fora, observa-se uma queda súbita de desempenho. No entanto, a crescente tendência das taxas de erro não é monotônica, i.e., utilizando apenas 24 características, TEDA forneceu melhores resultados que quando comparado consigo mesmo utilizando 48 e 120 características de entrada.

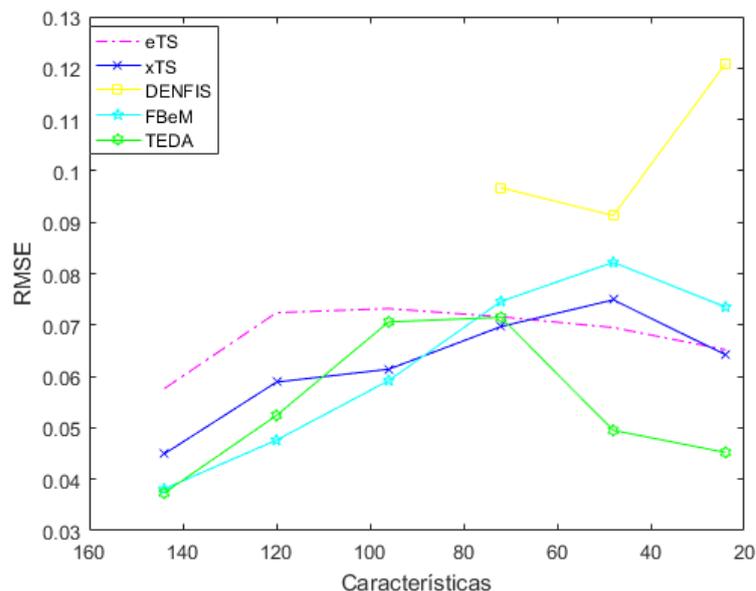
FBeM apresentou resultados similares ao TEDA, viz., um valor *RMSE* de 0,0381 contra 0,0373 do último. No entanto, FBeM apresentou um tempo de processamento de 1,10 segun-

Tabela 5.2 – Previsão: Estação meteorológica de Manaus

| eTS | | | | |
|-------------------|---------------|---------------|-------------|-------------------|
| – | <i>RMSE</i> | <i>NDE</i> | Tempo(s) | # regras |
| # características | | | | |
| 144 | 0,0576 | 0,5009 | 1,97 | 5 |
| 120 | 0,0724 | 0,6296 | 1,83 | 5 |
| 96 | 0,0732 | 0,6365 | 1,80 | 5 |
| 72 | 0,0716 | 0,6226 | 1,76 | 5 |
| 48 | 0,0695 | 0,6043 | 1,75 | 5 |
| 24 | 0,0652 | 0,5669 | 1,41 | 5 |
| xTS | | | | |
| – | <i>RMSE</i> | <i>NDE</i> | Tempo(s) | # regras |
| # características | | | | |
| 144 | 0,0449 | 0,3904 | 1,69 | 7 |
| 120 | 0,0589 | 0,5122 | 1,62 | 7 |
| 96 | 0,0614 | 0,5339 | 1,54 | 7 |
| 72 | 0,0697 | 0,6061 | 1,61 | 7 |
| 48 | 0,0749 | 0,6513 | 1,68 | 7 |
| 24 | 0,0642 | 0,5583 | 1,32 | 7 |
| DENFIS | | | | |
| – | <i>RMSE</i> | <i>NDE</i> | Tempo(s) | # regras |
| # características | | | | |
| 144 | – | – | – | – |
| 120 | – | – | – | – |
| 96 | – | – | – | – |
| 72 | 0,0967 | 0,8356 | 3,07 | 10 |
| 48 | 0,0913 | 0,7939 | 2,12 | 10 |
| 24 | 0,1209 | 1,1510 | 1,78 | 10 |
| FBeM | | | | |
| – | <i>RMSE</i> | <i>NDE</i> | Tempo(s) | # regras |
| # características | | | | |
| 144 | 0,0381 | 0,3313 | 1,10 | 8,03 |
| 120 | 0,0476 | 0,4139 | 1,01 | 8,10 |
| 96 | 0,0592 | 0,5147 | 1,17 | 7,19 |
| 72 | 0,0746 | 0,6486 | 1,94 | 7,91 |
| 48 | 0,0822 | 0,7147 | 1,09 | 6,15 |
| 24 | 0,0735 | 0,6391 | 0,96 | 6,08 |
| TEDA Predict | | | | |
| – | <i>RMSE</i> | <i>NDE</i> | Tempo(s) | # nuvens de dados |
| # características | | | | |
| 144 | 0,0373 | 0,3243 | 3,47 | 10 |
| 120 | 0,0524 | 0,4556 | 3,23 | 10 |
| 96 | 0,0706 | 0,6139 | 3,17 | 10 |
| 72 | 0,0715 | 0,6217 | 2,94 | 10 |
| 48 | 0,0495 | 0,4304 | 2,98 | 10 |
| 24 | 0,0452 | 0,3930 | 2,96 | 10 |

dos, ou seja 2,27 segundos a menos que o TEDA. Este processou os dados em 3,47 segundos. Esta diferença de tempo de processamento pode tornar FBeM mais desejável em certas aplicações, mas não na aplicação discutida aqui. Comparações do índice *RMSE* para diferentes métodos são apresentadas na Figura 5.4.

Figura 5.4 – Comparação entre *RMSE* / Características - Estação meteorológica de Manaus



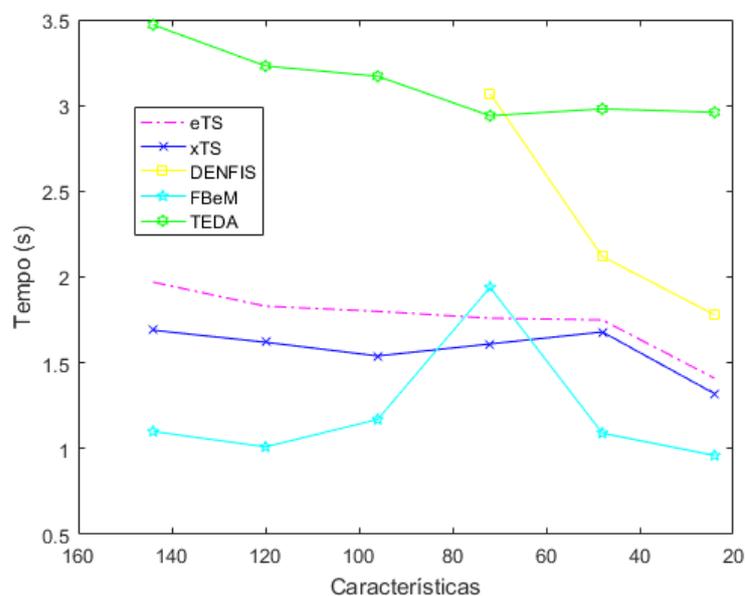
Fonte: Arquivo Pessoal

É interessante notar que o método DENFIS apresenta melhor acurácia utilizando um terço das variáveis originais; sendo este o mínimo global para estes métodos. Novamente, o seletor de características proposto foi benéfico para eliminar características que confundem o preditor. Porém, com a retirada de outras características, o índice de erro volta a aumentar, pois características importantes para a tarefa de previsão são desconsideradas. Nota-se que o FBeM tem um aumento, quase-linear de erro conforme as características são retiradas. Com 24 características, este método apresenta um valor interessante de erro visto a existência de um mínimo local. eTS e xTS apresentam índices de erros similares, sendo que xTS apresenta uma leve piora conforme a retirada das características, esta tendência não se repete no eTS. este apresenta um leve declínio no erro após a retirada de 96 características.

A Figura 5.5 apresenta a comparação entre os tempos de processamento dos métodos utilizados variando o número de características. O método FBeM mostrou ser o mais rápido dos métodos, processando um sexto da dimensionalidade original dos dados em 0,96 segundos.

eTS e xTS apresentam índices de tempo similares, sendo que o xTS apresentou ligeira vantagem sobre o primeiro. É interessante notar que conforme as características são retiradas a rede DENFIS apresenta uma redução quase-linear no tempo de processamento. O método TEDA devido ao maior número de cálculos de distância é o mais lento dos métodos. Porém conforme a retirada das características, o tempo de processamento deste método reduz, o que pode ser útil em aplicações em que há a necessidade de uma resposta mais rápida, mesmo

Figura 5.5 – Comparação de $Tempo(s)$ para diferentes quantidades de características - Estação meteorológica de Manaus



Fonte: Arquivo Pessoal

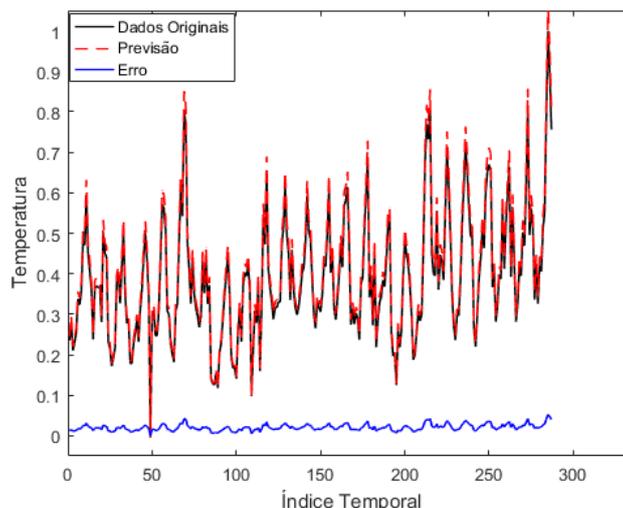
que isto custe uma perda de acurácia. É interessante notar que no método TEDA após às 24 primeiras características serem retiradas, um aumento no tempo de processamento é acusado. Isto pode ter ocorrido graças à diversos fatores, como características que confundiram o modelo preditor, ou outros processos paralelos que poderiam estar em execução no momento em que o teste foi realizado. De qualquer modo, nota que esta diferença no tempo é quase desprezível.

A Figura 5.6 ilustra o melhor caso de previsão de um passo para a cidade de Manaus. A previsão é dada pelo método TEDA utilizando 144 características. É importante notar na figura que uma mudança de conceito é visível após $k \approx 200$; é notável que a temperatura gradualmente torna-se maior. Um método não evolutivo não poderia ser capaz de notar tal mudança. Ademais, nota-se na Figura 5.6 que o erro de previsão representado pela curva azul (curva de menor amplitude) é maior nos picos e nos vales da componente sazonal da série temporal. Em princípio, nenhum dos métodos inteligentes evolutivos considerados neste trabalho possui procedimentos particulares para lidar com essa questão.

5.3 Resultados para a Estação Meteorológica de Porto Alegre

Os resultados obtidos para a cidade de Porto Alegre são mostrados na Tabela 5.3. Os resultados foram obtidos do mesmo modo que aqueles apresentados nas seções anteriores. É interessante notar neste experimento que a utilização de 24 características no desenvolvimento

Figura 5.6 – Estimação TEDA de um passo para a estação de Manaus



Fonte: Arquivo Pessoal

do previsor TEDA foi suficiente para alcançar a melhor acurácia. A interpretação do modelo e a velocidade de processamento são claramente melhoradas com a compactação de regras baseadas em um menor número de nuvens.

Apesar do resultado obtido pelo método TEDA ser ligeiramente melhor em termos dos índices *RMSE* e *NDE*, os índices obtidos pelo método FBeM, o tempo de processamento do FBeM em seu melhor caso, ou seja, com a dimensionalidade original dos dados, foi 1.73 segundos mais rápido que o melhor caso para o TEDA. Esta diferença de processamento pode ser crucial em certas aplicações. A rede DENFIS em seu melhor caso, superou o método eTS em termos dos índices *RMSE* e *NDE*. A Figura 5.7 apresenta a comparação entre os índices *RMSE* para os métodos utilizados de acordo com os diferentes números características.

Nota-se na Figura 5.7 que conforme a dimensionalidade do espaço de entrada é reduzida, a rede DENFIS, tende a ter um aumento no seu índice *RMSE* de modo contínuo, porém isto não é válido para os demais métodos. Por exemplo, o eTS, o FBeM e o TEDA possuem melhores índices de acurácia com 50% da dimensionalidade original do espaço de entradas, do que utilizando dois terços da tamanho original das amostras. A Figura 5.8 ilustra como o tempo de processamento dos métodos utilizados se comportam conforme a redução do espaço de entradas.

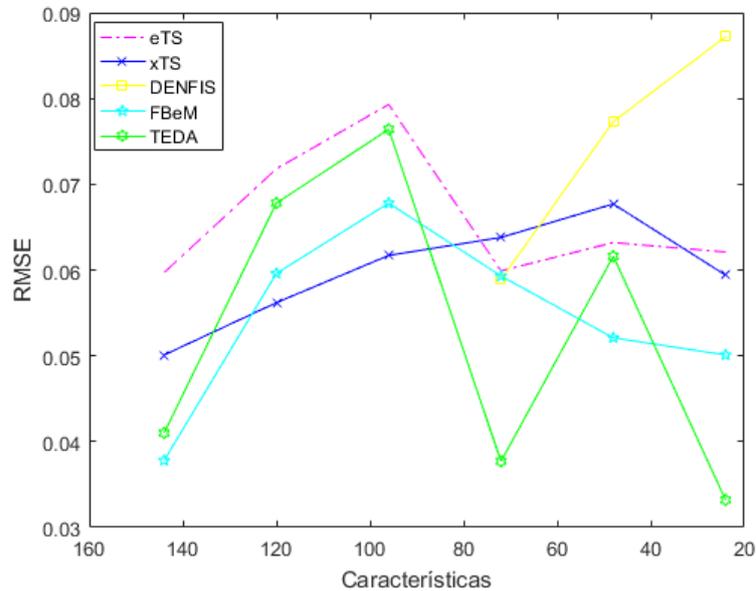
O método FBeM destaca-se como o mais rápido dos métodos utilizados para este tipo de tarefa, gastando 0.96 segundos em seu melhor tempo com 24 características no espaço de entrada. O método TEDA Predict mesmo tendo o melhor índice de acurácia foi 1,73 segun-

Tabela 5.3 – Previsão: Estação meteorológica de Porto Alegre

| eTS | | | | |
|-------------------|---------------|---------------|-------------|-------------------|
| – | <i>RMSE</i> | <i>NDE</i> | Tempo(s) | # regras |
| # características | | | | |
| 144 | 0,0597 | 0,5191 | 1,82 | 6 |
| 120 | 0,0718 | 0,6243 | 1,71 | 6 |
| 96 | 0,0793 | 0,6896 | 1,68 | 6 |
| 72 | 0,0599 | 0,5209 | 1,57 | 6 |
| 48 | 0,0632 | 0,5496 | 1,37 | 6 |
| 24 | 0,0621 | 0,5400 | 1,13 | 6 |
| – | xTS | | | |
| # características | <i>RMSE</i> | <i>NDE</i> | Tempo(s) | # regras |
| 144 | 0,0501 | 0,4357 | 1,62 | 5 |
| 120 | 0,0562 | 0,4887 | 1,57 | 5 |
| 96 | 0,0617 | 0,5365 | 1,58 | 5 |
| 72 | 0,0638 | 0,5548 | 1,64 | 5 |
| 48 | 0,0677 | 0,5887 | 1,81 | 5 |
| 24 | 0,0594 | 0,5165 | 1,54 | 5 |
| – | DENFIS | | | |
| # características | <i>RMSE</i> | <i>NDE</i> | Tempo(s) | # regras |
| 144 | – | – | – | – |
| 120 | – | – | – | – |
| 96 | – | – | – | – |
| 72 | 0,0589 | 0,512 | 3,59 | 10 |
| 48 | 0,0773 | 0,6721 | 2,04 | 10 |
| 24 | 0,0872 | 0,7582 | 1,62 | 10 |
| – | FBeM | | | |
| # características | <i>RMSE</i> | <i>NDE</i> | Tempo(s) | # regras |
| 144 | 0,0378 | 0,3286 | 1,19 | 8,44 |
| 120 | 0,0596 | 0,5182 | 1,04 | 8,14 |
| 96 | 0,0678 | 0,5895 | 1,07 | 8,02 |
| 72 | 0,0593 | 0,5156 | 1,00 | 6,93 |
| 48 | 0,0521 | 0,4530 | 0,97 | 8,93 |
| 24 | 0,0501 | 0,4356 | 0,96 | 6,36 |
| – | TEDA Predict | | | |
| # características | <i>RMSE</i> | <i>NDE</i> | Tempo(s) | # nuvens de dados |
| 144 | 0,0410 | 0,3565 | 3,48 | 10 |
| 120 | 0,0678 | 0,5895 | 3,69 | 10 |
| 96 | 0,0764 | 0,6643 | 2,73 | 10 |
| 72 | 0,0377 | 0,3278 | 3,19 | 10 |
| 48 | 0,0616 | 0,5356 | 3,08 | 10 |
| 24 | 0,0332 | 0,2887 | 2,69 | 10 |

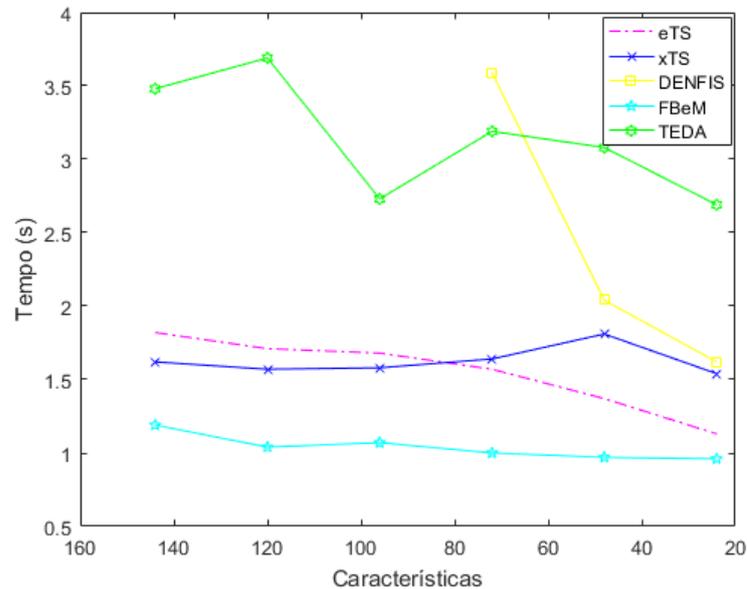
dos mais lento que o FBeM com o mesmo número de amostras no espaço de entrada. Esta diferença pode ser significativa em diversas aplicações que exijam uma maior velocidade de processamento, mesmo que isso signifique uma pequena perda de acurácia. O DENFIS apresenta uma queda quase-linear no tempo de processamento visto a eliminação de características.

Figura 5.7 – Comparação de $RMSE$ para diferentes quantidades de características - Estação meteorológica de Porto Alegre



Fonte: Arquivo Pessoal

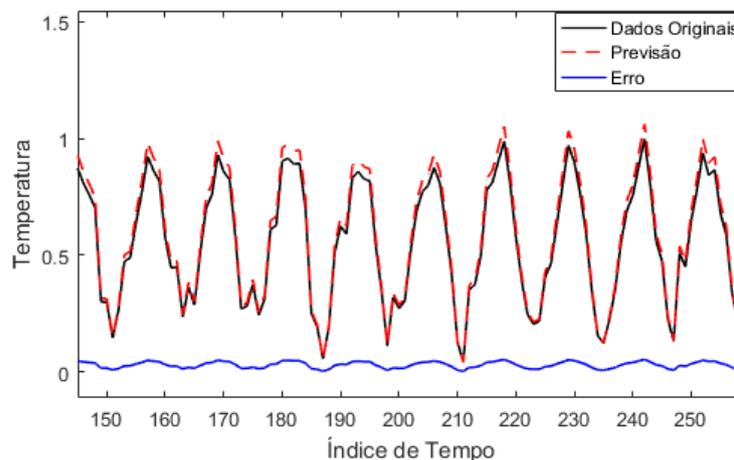
Figura 5.8 – Comparação de $Tempo(s)$ para diferentes quantidades de características - Estação meteorológica de Porto Alegre



Fonte: Arquivo Pessoal

Os métodos eTS e xTS possuem tempo de processamento similares, sendo o primeiro ligeiramente mais rápido conforme decresce o número de características no espaço de entrada. A Figura 5.9 ilustra a previsão da série temporal da cidade de Porto Alegre utilizando o TEDA (melhor caso).

Figura 5.9 – Estimação TEDA Predict de um passo para estação de Porto Alegre



Fonte: Arquivo Pessoal

5.4 Resultados para a Estação Meteorológica de Natal

Os resultados dos experimentos para estação meteorológica de Natal são apresentados na Tabela 5.4.

O TEDA Predict produziu o melhor resultado dentro os métodos utilizados para a cidade de Natal. O TEDA Predict utilizou toda a dimensionalidade original dos dados para atingir o melhor índice de acurácia. Porém conforme demonstrado na Figura 5.10, os índices de erro do TEDA Predict tem uma tendência linear crescente quando a dimensionalidade original do espaço de entrada dos dados é diminuída.

Conforme apresentado na Tabela 5.4, o FBeM é capaz de produzir o melhor índice de *RMSE* e *NDE* utilizando apenas dois terços da dimensionalidade original dos dados. Além de produzir os melhores índices de acurácia o FBeM foi o mais rápido dos métodos utilizados. A compactação do modelo e o tempo de processamento foi claramente beneficiada pela utilização do método para ordenação e seleção de características proposto. O método proposto claramente é benéfico em ambientes que requerem processamento em tempo real ou *online*. Portanto, os resultados obtidos pelo FBeM podem ser mais relevantes que os obtidos pelo TEDA Predict em determinadas aplicações.

A Figura 5.11 evidencia que a compactação dos modelos melhorou o tempo de processamento em todos os casos apresentados. Dentre todos os métodos apresentados, destaca-se o FBeM como o mais rápido, com o tempo de processamento no melhor caso de 1,09 segundos. O TEDA Predict diminui seu tempo de processamento em 1,19 segundos com a compactação

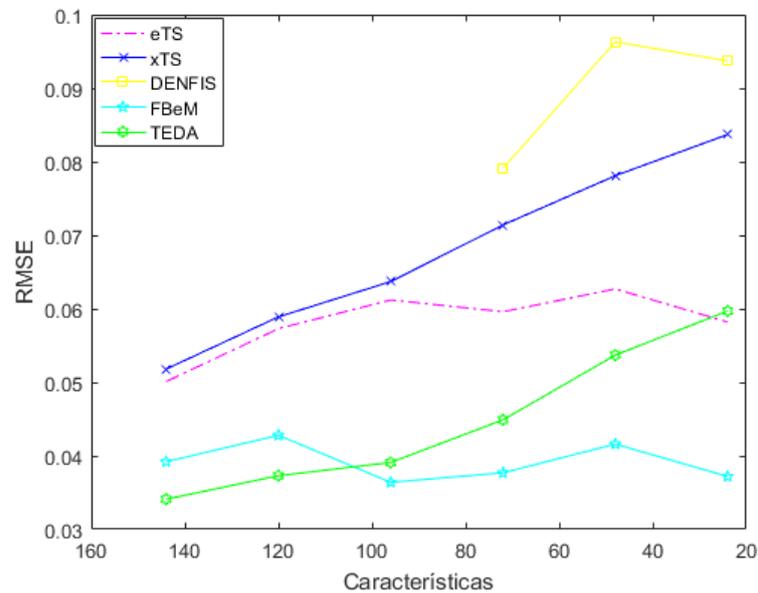
Tabela 5.4 – Previsão: Estação meteorológica de Natal

| eTS | | | | |
|-------------------|---------------|---------------|-------------|-------------------|
| – | <i>RMSE</i> | <i>NDE</i> | Tempo(s) | # regras |
| # características | | | | |
| 144 | 0,0501 | 0,4356 | 1,82 | 3 |
| 120 | 0,0573 | 0,4982 | 1,80 | 3 |
| 96 | 0,0612 | 0,5321 | 1,72 | 3 |
| 72 | 0,0596 | 0,5182 | 1,67 | 3 |
| 48 | 0,0627 | 0,5472 | 1,66 | 3 |
| 24 | 0,0582 | 0,5060 | 1,52 | 3 |
| xTS | | | | |
| – | <i>RMSE</i> | <i>NDE</i> | Tempo(s) | # regras |
| # características | | | | |
| 144 | 0,0518 | 0,4504 | 1,42 | 4 |
| 120 | 0,0589 | 0,5121 | 1,40 | 4 |
| 96 | 0,0637 | 0,5539 | 1,37 | 4 |
| 72 | 0,0714 | 0,6208 | 1,39 | 4 |
| 48 | 0,0781 | 0,6791 | 1,34 | 4 |
| 24 | 0,0837 | 0,7278 | 1,27 | 4 |
| DENFIS | | | | |
| – | <i>RMSE</i> | <i>NDE</i> | Tempo(s) | # regras |
| # características | | | | |
| 144 | – | – | – | – |
| 120 | – | – | – | – |
| 96 | – | – | – | – |
| 72 | 0,0791 | 0,6878 | 3,27 | 10 |
| 48 | 0,0963 | 0,8373 | 2,49 | 10 |
| 24 | 0,0937 | 0,8147 | 2,34 | 10 |
| FBeM | | | | |
| – | <i>RMSE</i> | <i>NDE</i> | Tempo(s) | # regras |
| # características | | | | |
| 144 | 0,0392 | 0,3408 | 1,48 | 8,44 |
| 120 | 0,0428 | 0,3721 | 1,69 | 8,14 |
| 96 | 0,0364 | 0,3165 | 1,73 | 7,12 |
| 72 | 0,0377 | 0,3278 | 1,19 | 9,95 |
| 48 | 0,0416 | 0,43617 | 1,08 | 9,95 |
| 24 | 0,0372 | 0,3234 | 1,09 | 9,52 |
| TEDA Predict | | | | |
| – | <i>RMSE</i> | <i>NDE</i> | Tempo(s) | # nuvens de dados |
| # características | | | | |
| 144 | 0,0341 | 0,2730 | 3,81 | 10 |
| 120 | 0,0373 | 0,3243 | 3,76 | 10 |
| 96 | 0,0391 | 0,3400 | 3,31 | 10 |
| 72 | 0,0449 | 0,3904 | 2,96 | 10 |
| 48 | 0,0537 | 0,4669 | 2,79 | 10 |
| 24 | 0,0597 | 0,5191 | 2,62 | 10 |

realizada, isto ao custo de uma perda marginal na acurácia. Os métodos eTS e xTS apresentaram uma leve diminuição no tempo de processamento, porém considerável em diversas aplicações.

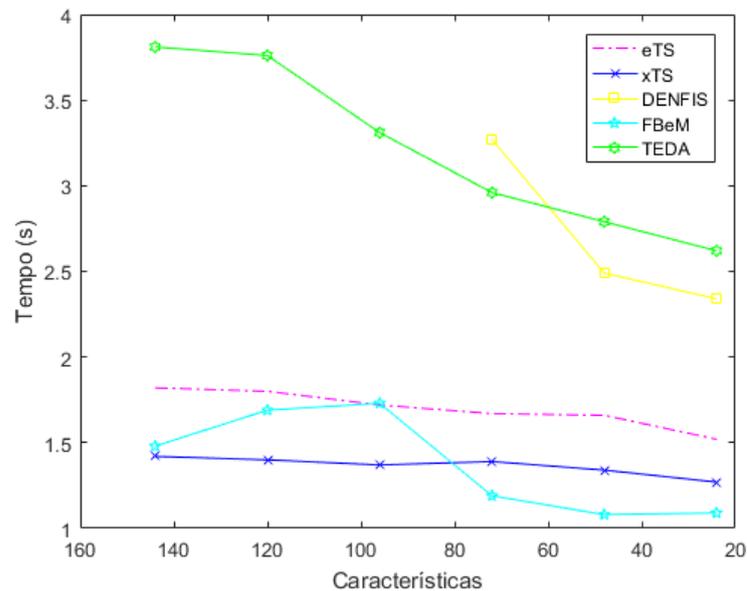
A Figura 5.12 ilustra o melhor caso de previsão obtidos nos experimentos realizados.

Figura 5.10 – Comparação de $RMSE$ para diferentes quantidades de características - Estação meteorológica de Natal



Fonte: Arquivo Pessoal

Figura 5.11 – Comparação de $Tempo(s)$ para diferentes quantidades de características - Estação meteorológica de Natal

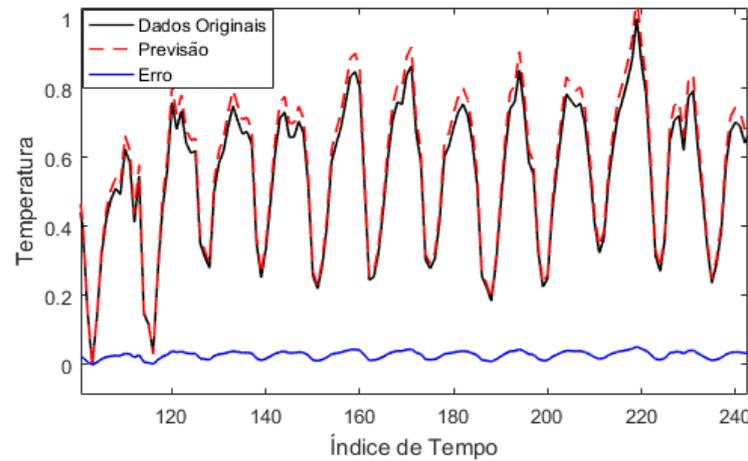


Fonte: Arquivo Pessoal

5.5 Resultados do *Ensemble*

Um *ensemble* de modelos previsores baseados em nuvens de dados e regras fuzzy (Fig. 4.2) é proposto utilizando operadores de agregação fuzzy. Um resumo do custo computacional

Figura 5.12 – Estimação TEDA Predict de um passo para estação de Natal



Fonte: Arquivo Pessoal

e das estimativas de previsões singulares utilizando a agregação de erro ponderado M (conforme Anexo A) para todos os conjuntos de dados são mostrados na Tabela 5.5.

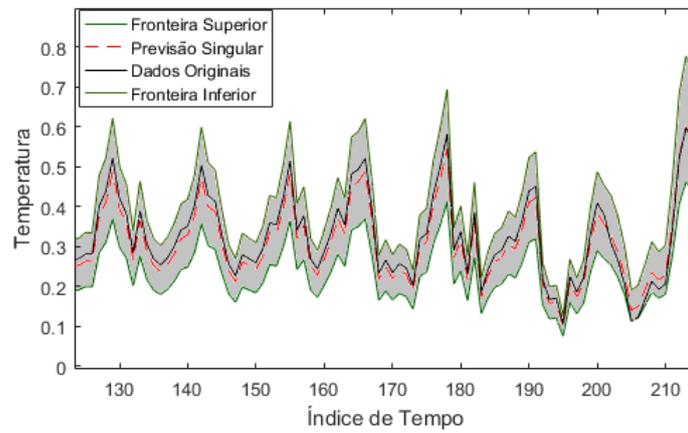
Tabela 5.5 – Sumário da previsão de temperatura média mensal fornecido pelo *ensemble* utilizando o operador de agregação média ponderada pelo erro

| Cidades | $RMSE$ | NDE | Tempo(s) | #Modelos Locais |
|--------------|---------------|---------------|----------|-----------------|
| São Paulo | 0.0362 | 0.3147 | 8.74 | 26.12 |
| Manaus | 0.0361 | 0.3142 | 8.53 | 30.03 |
| Porto Alegre | 0.0389 | 0.3384 | 8.41 | 29.44 |
| Natal | 0.0364 | 0.3170 | 8.70 | 29.44 |

Para os dados da estação de Manaus, o *ensemble* forneceu os melhores resultados globais, enquanto produziu resultados ligeiramente piores em comparação com TEDA e FBeM individualmente para as outras estações meteorológicas. O número de unidades locais mencionados na Tabela 5.5 se referem a soma das nuvens de dados e dos *clusters* desenvolvido nos modelos individuais. A Figura 5.13 ilustra a previsão singular e granular para a estação de Manaus considerando o *ensemble* proposto. O invólucro da série temporal original é criado à partir dos operadores de agregação T_{min} e S_{max} , conforme Anexo A.

Muitas vezes, o invólucro das estimativas pode ser um fator mais importante do que a estimativa singular para a tomada de decisões. Por exemplo, nos setores de água e energia, a fronteira de temperatura superior pode sugerir a abertura de comportas de barragens em usinas hidrelétricas para otimizar redes de distribuição de água e para abastecer depósitos de água de

Figura 5.13 – Previsão granular e singular de temperatura utilizando um *ensemble* de modelos fuzzy e de nuvens de dados considerando a estação de Manaus



Fonte: Arquivo Pessoal

outras usinas hidrelétricas. Na agricultura, a fronteira de temperatura inferior pode sugerir que uma região é inadequada, prejudicial ou arriscada para certas atividades agrícolas e florestais.

6 CONCLUSÃO

Neste trabalho, os métodos *evolving Takagi-Sugeno* (eTS), *eXtended Takagi-Sugeno* (xTS), *Dynamic Neural-Fuzzy Evolving Inference System* (DENFIS), *Fuzzy-Set Based evolving Modeling* (FBeM) e uma variação do método evolutivo baseado em tipicidade e excentricidade dos dados, TEDA, foram aplicados na previsão de séries temporais meteorológicas. Os valores passados de temperatura mensal máxima, mínima e média, bem como os valores históricos de variáveis exógenas, como nebulosidade, precipitação e humidade, foram considerados nas análises. Em princípio, uma janela temporal de tamanho 24, correspondente a 2 anos, foi utilizada.

Foi proposto um método de ranqueamento e seleção de características baseado na correlação de Spearman para identificar o melhor conjunto de características e os atrasos de tempo a serem usados pelos modelos previsores evolutivos. O método proposto é não-paramétrico, e as características foram selecionadas através da pontuação obtida pelo ρ de Spearman. O método apresentou resultados encorajadores no sentido de que reduziu a complexidade computacional global e acelerou as etapas de processamento, propriedade importante no contexto de fluxo de dados.

De um modo geral, os experimentos demonstraram que os métodos evolutivos são eficientes para prever séries temporais meteorológicas em um curto prazo. O modelo baseado em nuvens de dados, tipicidade e excentricidade, apresentou os melhores índices RMSE e NDE nos testes realizados. O número necessário de cálculos algébricos no algoritmo TEDA é maior. Portanto, este método tende a ser mais lento que outros métodos evolutivos. O método FBeM obteve valores de RMSE e NDE bem próximos aos índices obtidos pelo TEDA, superando este em um dos testes. FBeM mostrou-se ser o mais rápido dos métodos utilizados. O tempo de processamento não é uma restrição em problemas de previsão mensais, diários e horários. No entanto, pode ser um ponto crítico em aplicações com fluxo de dados de alta frequência e alta dimensão. Logo, para cada problema um método diferente é aconselhado.

Um *ensemble* baseado em nuvens e modelos fuzzy utilizando operadores de agregação foi construído para fornecer previsões singulares e granulares das séries temporais. O *ensemble* forneceu melhor acurácia global em uma das quatro bases de dados avaliadas. O método TEDA Predict, e FBeM individualmente forneceram melhores resultados as demais bases. O *ensemble* evolutivo proposto é capaz de produzir um invólucro ao redor dos dados originais, e portanto fornecer um ideia sobre o erro e a incerteza associada aos dados. Fronteiras ao redor das previsões podem ter diferentes interpretações diferentes contextos.

Trabalhos futuros irão discutir procedimentos para acumular e manter novas variáveis estatísticas associadas as nuvens de dados e regras fuzzy, como informações sobre especificidade, entropia e cardinalidade. Estas variáveis podem ser úteis na otimização estrutural dos modelos durante o aprendizado *online*. Processamento de dados intervalares e heterogêneos também são tópicos para investigações futuras. Além do mais imagens de satélites serão consideradas para previsões de temperatura mais condizentes com a realidade. Fluxos com diferentes frequências também serão considerados, assim como diferentes variáveis atmosféricas e previsões múltiplos passos à frente.

O cálculo da similaridade entre dados e protótipos neste estudo se baseou essencialmente no inverso da métrica de distância Euclidiana. Uma variedade de medidas de similaridade entre objetos pontuais em espaços precisa ser explorada, já que estas fornecem resultados diversos. Medidas de similaridade entre dados granulares, intervalares e fuzzy, são raras. Em trabalhos futuros, estas devem ser investigadas.

REFERÊNCIAS

- ABDULSHAHED, A. M.; LONGSTAFF, A. P.; FLETCHER, S. The application of anfis prediction models for thermal error compensation on cnc machine tools. **Applied Soft Computing**, Elsevier, v. 27, p. 158–168, 2015.
- ANGELOV, P. Evolving fuzzy systems. **Encyclopedia of Complexity and Systems Science**, Springer, p. 3242–3255, 2009.
- ANGELOV, P. Anomaly detection based on eccentricity analysis. In: IEEE. **Evolving and Autonomous Learning Systems (EALS), 2014 IEEE Symposium on**. [S.l.], 2014. p. 1–8.
- ANGELOV, P.; BUSWELL, R. Identification of evolving fuzzy rule-based models. **IEEE Transactions on Fuzzy Systems**, IEEE, v. 10, n. 5, p. 667–677, 2002.
- ANGELOV, P. et al. **Evolving fuzzy systems**. [S.l.]: IEEE Press, 2006.
- ANGELOV, P.; YAGER, R. Simplified fuzzy rule-based systems using non-parametric antecedents and relative data density. In: IEEE. **Evolving and Adaptive Intelligent Systems (EAIS), 2011 IEEE Workshop on**. [S.l.], 2011. p. 62–69.
- ANGELOV, P.; ZHOU, X. Evolving fuzzy systems from data streams in real-time. In: **2006 International Symposium on Evolving Fuzzy Systems**. [S.l.: s.n.], 2006. p. 29–35.
- ANGELOV, P. P. **Evolving rule-based models: a tool for design of flexible adaptive systems**. [S.l.]: Physica, 2013. v. 92.
- ANGELOV, P. P.; FILEV, D. P. An approach to online identification of takagi-sugeno fuzzy models. **IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)**, IEEE, v. 34, n. 1, p. 484–498, 2004.
- ÅSTRÖM, K. J.; WITTENMARK, B. **Adaptive control**. [S.l.]: Courier Corporation, 2013.
- BARGIELA, A.; PEDRYCZ, W. **Granular computing: an introduction**. [S.l.]: Springer Science & Business Media, 2012. v. 717.
- BELIAKOV, G.; WARREN, J. Appropriate choice of aggregation operators in fuzzy decision support systems. **IEEE Transactions on Fuzzy Systems**, IEEE, v. 9, n. 6, p. 773–784, 2001.
- BOX, G. E. et al. **Time series analysis: forecasting and control**. [S.l.]: John Wiley & Sons, 2015.
- BREIDT, J. **Nonlinear Time Series: Nonparametric and Parametric Methods**. [S.l.]: Taylor & Francis, 2005.
- BUENO, L. et al. Evolving ensemble of fuzzy models for multivariate time series prediction. In: IEEE. **Fuzzy Systems (FUZZ-IEEE), 2015 IEEE International Conference on**. [S.l.], 2015. p. 1–6.
- CAI, Y.; NG, R. Indexing spatio-temporal trajectories with chebyshev polynomials. In: ACM. **Proceedings of the 2004 ACM SIGMOD international conference on Management of data**. [S.l.], 2004. p. 599–610.

- CASTILLO, O.; MELIN, P. Hybrid intelligent systems for time series prediction using neural networks, fuzzy logic, and fractal theory. **IEEE Transactions on Neural Networks**, IEEE, v. 13, n. 6, p. 1395–1408, 2002.
- ÇAYDAŞ, U.; HASÇALIK, A.; EKICI, S. An adaptive neuro-fuzzy inference system (anfis) model for wire-edm. **Expert Systems with Applications**, Elsevier, v. 36, n. 3, p. 6135–6139, 2009.
- CHEN, S.-M.; HWANG, J.-R. Temperature prediction using fuzzy time series. **IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)**, IEEE, v. 30, n. 2, p. 263–275, 2000.
- CHIU, S. L. Fuzzy model identification based on cluster estimation. **Journal of Intelligent & fuzzy systems**, IOS Press, v. 2, n. 3, p. 267–278, 1994.
- CRYER, J. D.; KELLET, N. **Time series analysis**. [S.l.]: Springer, 1986. v. 101.
- DEB, C. et al. A review on time series forecasting techniques for building energy consumption. **Renewable and Sustainable Energy Reviews**, Elsevier, v. 74, p. 902–924, 2017.
- DOUCOURE, B.; AGBOSSOU, K.; CARDENAS, A. Time series prediction using artificial wavelet neural network and multi-resolution analysis: Application to wind speed data. **Renewable Energy**, Elsevier, v. 92, p. 202–211, 2016.
- DUBOIS, D. J. **Fuzzy sets and systems: theory and applications**. [S.l.]: Academic press, 1980. v. 144.
- EL-SHAFIE, A.; JAAFER, O.; AKRAMI, S. A. Adaptive neuro-fuzzy inference system based model for rainfall forecasting in klang river, malaysia. **International Journal of Physical Sciences**, Academic Journals, v. 6, n. 12, p. 2875–2888, 2011.
- FALK, M. et al. A first course on time series analysis: examples with sas. 2012.
- FALOUTSOS, C.; RANGANATHAN, M.; MANOLOPOULOS, Y. **Fast subsequence matching in time-series databases**. [S.l.]: ACM, 1994. v. 23.
- GHAZALI, R. et al. The performance of a recurrent honn for temperature time series prediction. In: IEEE. **Neural Networks (IJCNN), 2014 International Joint Conference on**. [S.l.], 2014. p. 518–524.
- GRANT, I. H. Recursive least squares. **Teaching Statistics**, Wiley Online Library, v. 9, n. 1, p. 15–18, 1987.
- HAMILTON, J. D. Analysis of time series subject to changes in regime. **Journal of econometrics**, Elsevier, v. 45, n. 1-2, p. 39–70, 1990.
- HAMILTON, J. D. **Time series analysis**. [S.l.]: Princeton university press Princeton, 1994. v. 2.
- HAYKIN, S. Neural networks: A comprehensive foundation. **Neural Networks**, v. 2, n. 2004, p. 41, 2004.
- HAYKIN, S. S. **Adaptive filter theory**. [S.l.]: Pearson Education India, 2008.

- HOSSAIN, M. et al. Forecasting the weather of nevada: A deep learning approach. In: **IEEE. Neural Networks (IJCNN), 2015 International Joint Conference on**. [S.l.], 2015. p. 1–6.
- HUANG, N. E. et al. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. In: **THE ROYAL SOCIETY. Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences**. [S.l.], 1998. v. 454, n. 1971, p. 903–995.
- HUARNG, K. Effective lengths of intervals to improve forecasting in fuzzy time series. **Fuzzy sets and systems**, Elsevier, v. 123, n. 3, p. 387–394, 2001.
- HUNG, N. Q. et al. An artificial neural network model for rainfall forecasting in bangkok, thailand. **Hydrology and Earth System Sciences**, Copernicus GmbH, v. 13, n. 8, p. 1413–1425, 2009.
- IZAKIAN, H.; PEDRYCZ, W.; JAMAL, I. Clustering spatiotemporal data: An augmented fuzzy c-means. **IEEE transactions on fuzzy systems**, IEEE, v. 21, n. 5, p. 855–868, 2013.
- JAIN, A. K.; MAO, J.; MOHIUDDIN, K. M. Artificial neural networks: A tutorial. **Computer**, IEEE, v. 29, n. 3, p. 31–44, 1996.
- JANG, J.-S. Anfis: adaptive-network-based fuzzy inference system. **IEEE transactions on systems, man, and cybernetics**, IEEE, v. 23, n. 3, p. 665–685, 1993.
- JOHNSON, C. R. **Lectures & adaptive parameter estimation**. [S.l.]: Prentice-Hall, Inc., 1988.
- KAASTRA, I.; BOYD, M. Designing a neural network for forecasting financial and economic time series. **Neurocomputing**, Elsevier, v. 10, n. 3, p. 215–236, 1996.
- KANGIN, D.; ANGELOV, P. Evolving clustering, classification and regression with teda. In: **IEEE. Neural Networks (IJCNN), 2015 International Joint Conference on**. [S.l.], 2015. p. 1–8.
- KASABOV, N. Evolving fuzzy neural networks for supervised/unsupervised online knowledge-based learning. **IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)**, IEEE, v. 31, n. 6, p. 902–918, 2001.
- KASABOV, N. **Evolving connectionist systems: the knowledge engineering approach**. [S.l.]: Springer Science & Business Media, 2007.
- KAUR, T.; KUMAR, S.; SEGAL, R. Application of artificial neural network for short term wind speed forecasting. In: **IEEE. Power and Energy Systems: Towards Sustainable Energy (PESTSE), 2016 Biennial International Conference on**. [S.l.], 2016. p. 1–5.
- KHASHEI, M.; HEJAZI, S. R.; BIJARI, M. A new hybrid artificial neural networks and fuzzy regression model for time series forecasting. **Fuzzy sets and systems**, Elsevier, v. 159, n. 7, p. 769–786, 2008.
- LEITE, D. et al. Evolving fuzzy granular modeling from nonstationary fuzzy data streams. **Evolving Systems**, Springer, v. 3, n. 2, p. 65–79, 2012.
- LEITE, D.; COSTA, P.; GOMIDE, F. Evolving granular neural networks from fuzzy data streams. **Neural Networks**, Elsevier, v. 38, p. 1–16, 2013.

- LEITE, D. et al. Fuzzy granular evolving modeling for time series prediction. In: IEEE. **Fuzzy Systems (FUZZ), 2011 IEEE International Conference on**. [S.l.], 2011. p. 2794–2801.
- LEITE, D. et al. Evolving granular fuzzy model-based control of nonlinear dynamic systems. **IEEE Transactions on Fuzzy Systems**, IEEE, v. 23, n. 4, p. 923–938, 2015.
- LEITE, D. et al. Fuzzy granular neural network for incremental modeling of nonlinear chaotic systems. In: IEEE. **Fuzzy Systems (FUZZ-IEEE), 2016 IEEE International Conference on**. [S.l.], 2016. p. 64–71.
- LEONTARITIS, I.; BILLINGS, S. A. Input-output parametric models for non-linear systems part i: deterministic non-linear systems. **International journal of control**, Taylor & Francis, v. 41, n. 2, p. 303–328, 1985.
- LIU, Y.; LEE, S.-M. Stability and stabilization of takagi–sugeno fuzzy systems via sampled-data and state quantized controller. **IEEE Transactions on Fuzzy Systems**, IEEE, v. 24, n. 3, p. 635–644, 2016.
- LJUNG, L. System identification toolbox. **The Matlab user's guide**, 1988.
- LUGHOFER, E.; ANGELOV, P. Handling drifts and shifts in on-line data streams with evolving fuzzy systems. **Applied Soft Computing**, Elsevier, v. 11, n. 2, p. 2057–2068, 2011.
- LUGHOFER, E.; SAYED-MOUCHAWEH, M. **Adaptive and on-line learning in non-stationary environments**. [S.l.]: Springer, 2015.
- MARPLE, S. L. **Digital spectral analysis: with applications**. [S.l.]: Prentice-Hall Englewood Cliffs, NJ, 1987. v. 5.
- MELLIT, A. et al. An anfis-based forecasting for solar radiation data from sunshine duration and ambient temperature. In: IEEE. **Power Engineering Society General Meeting, 2007. IEEE**. [S.l.], 2007. p. 1–6.
- MOORE, R. E.; KEARFOTT, R. B.; CLOUD, M. J. **Introduction to interval analysis**. [S.l.]: SIAM, 2009.
- MORETTIN, P. A.; TOLOI, C. **Análise de séries temporais**. [S.l.]: Blucher, 2006.
- NEGNEVITSKY, M. **Artificial intelligence: a guide to intelligent systems**. [S.l.]: Pearson Education, 2005.
- NILSSON, N. J. **Principles of artificial intelligence**. [S.l.]: Morgan Kaufmann, 2014.
- ORDERS, P.; FADLEY, C. Single-scattering cluster calculations and fourier-transform analyses of normal photoelectron diffraction. **Physical Review B**, APS, v. 27, n. 2, p. 781, 1983.
- PEDRYCZ, W. Why triangular membership functions? **Fuzzy sets and Systems**, Elsevier, v. 64, n. 1, p. 21–30, 1994.
- PEDRYCZ, W. et al. Granular computing-the emerging paradigm. **Journal of uncertain systems**, Citeseer, v. 1, n. 1, p. 38–61, 2007.

PEDRYCZ, W.; SKOWRON, A.; KREINOVICH, V. **Handbook of granular computing**. [S.l.]: John Wiley & Sons, 2008.

PENG, H.; LONG, F.; DING, C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. **IEEE Transactions on pattern analysis and machine intelligence**, IEEE, v. 27, n. 8, p. 1226–1238, 2005.

QIAN, Y. et al. Positive approximation: an accelerator for attribute reduction in rough set theory. **Artificial Intelligence**, Elsevier, v. 174, n. 9-10, p. 597–618, 2010.

RUSSELL, S.; NORVIG, P. Artificial intelligence : A modern approach. **Artificial Intelligence**. Prentice-Hall, Egnlewood Cliffs, Citeseer, v. 25, p. 27, 1995.

SARHADI, A. et al. Water resources climate change projections using supervised nonlinear and multivariate soft computing techniques. **Journal of Hydrology**, Elsevier, v. 536, p. 119–132, 2016.

SOARES, E. et al. Cloud-based evolving intelligent method for weather time series prediction. In: IEEE. **Fuzzy Systems (FUZZ), 2017 IEEE International Conference on**. [S.l.], 2017.

SONG, Q.; CHISSOM, B. S. Forecasting enrollments with fuzzy time series—part i. **Fuzzy sets and systems**, Elsevier, v. 54, n. 1, p. 1–9, 1993.

SONG, Q.; CHISSOM, B. S. Fuzzy time series and its models. **Fuzzy sets and systems**, Elsevier, v. 54, n. 3, p. 269–277, 1993.

SPEARMAN, C. The proof and measurement of association between two things. **The American journal of psychology**, JSTOR, v. 15, n. 1, p. 72–101, 1904.

SUN, C.-T.; JANG, J.-S. A neuro-fuzzy classifier and its applications. In: IEEE. **Fuzzy Systems, 1993., Second IEEE International Conference on**. [S.l.], 1993. p. 94–98.

TAKAGI, T.; SUGENO, M. Fuzzy identification of systems and its applications to modeling and control. **IEEE transactions on systems, man, and cybernetics**, IEEE, n. 1, p. 116–132, 1985.

TIBSHIRANI, R.; WALTHER, G.; HASTIE, T. Estimating the number of clusters in a data set via the gap statistic. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, Wiley Online Library, v. 63, n. 2, p. 411–423, 2001.

TONG, H. **Non-linear time series: a dynamical system approach**. [S.l.]: Oxford University Press, 1990.

VELICER, W. F.; FAVA, J. L. Time series analysis. **Handbook of psychology**, Wiley Online Library, 2003.

WANG, S.-C. Artificial neural network. In: **Interdisciplinary computing in java programming**. [S.l.]: Springer, 2003. p. 81–100.

WEI, W. W.-S. **Time series analysis**. [S.l.]: Addison-Wesley publ Reading, 1994.

YAGER, R. R.; FILEV, D. P. Essentials of fuzzy modeling and control. **New York**, 1994.

YAGER, R. R.; ZADEH, L. A. **An introduction to fuzzy logic applications in intelligent systems**. [S.l.]: Springer Science & Business Media, 2012. v. 165.

YAO, J. T.; VASILAKOS, A. V.; PEDRYCZ, W. Granular computing: perspectives and challenges. **IEEE Transactions on Cybernetics**, IEEE, v. 43, n. 6, p. 1977–1989, 2013.

YAO, Y. et al. Granular computing: basic issues and possible solutions. In: **Proceedings of the 5th joint conference on information sciences**. [S.l.: s.n.], 2000. v. 1, p. 186–189.

YOUNG, P. C. **Recursive estimation and time-series analysis: an introduction**. [S.l.]: Springer Science & Business Media, 2012.

YUN, Z. et al. Rbf neural network and anfis-based short-term load forecasting approach in real-time price environment. **IEEE Transactions on power systems**, IEEE, v. 23, n. 3, p. 853–858, 2008.

ZADEH, L. A. Fuzzy sets. **Information and control**, Elsevier, v. 8, n. 3, p. 338–353, 1965.

APÊNDICE A – Operadores de Agregação

Operadores de agregação $A : [0, 1]^n \rightarrow [0, 1], n > 1$ combinam valores de entrada na unidade de hipercubo $[0, 1]^n$ em um valor de saída singular $[0, 1]$. Estes operadores podem ser utilizados para realizar a fusão de informações em *ensembles* de previsores evolutivos.

Um operador de agregação A deve satisfazer duas propriedades fundamentais: (i) condições de fronteiras: $A(0, 0, \dots, 0) = 0$ e $A(1, 1, \dots, 1) = 1$; (ii) monotonicidade em todos os argumentos, i.e., dado $x^1 = (x_1^1, \dots, x_n^1)$ e $x^2 = (x_1^2, \dots, x_n^2)$, se $x_j^1 \leq x_j^2 \forall j$ então $A(x^1) \leq A(x^2)$. Importantes categorias de operadores de agregação são descritas abaixo (BELIAKOV; WARREN, 2001), (LEITE; COSTA; GOMIDE, 2013).

.1 Agregação T-norma

T-normas (T) são operadores associativos, comutativos e monótonos na unidade de hipercubo. $T(\alpha, \alpha, \dots, 0) = 0$, e $T(\alpha, 1, \dots, 1) = \alpha$; $\alpha \in [0, 1]$, são condições de fronteiras das T-normas. O elemento neutro de uma T-norma é $e = 1$. O operador mínimo,

$$T_{min}(x) = \min_{j=1, \dots, n} x_j, \quad (1)$$

é a T-norma mais forte desde que

$$T(x) \leq T_{min}(x), \forall x \in [0, 1]^n. \quad (2)$$

O mínimo também é idempotente, simétrico e Lipschitz contínuo.

.2 Agregação S-Norma

S-normas (S) são operadores comutativos, associativos e monótonos na unidade do hipercubo nos quais as condições de fronteiras são dadas por $S(\alpha, \alpha, \dots, 1) = 1$, e $S(\alpha, 0, \dots, 0) = \alpha$, i.e. $e = 0$ é o elemento neutro de uma S-norma. S-normas são mais fortes que T-normas. O operador máximo:

$$S_{max}(x) = \max_{j=1,\dots,n} x_j, \quad (3)$$

é a mais fraca S-norma. Segue que

$$S(x) \geq S_{max}(x) \geq T(x), \quad \forall x \in [0, 1]^n. \quad (4)$$

.3 Agregação de média

Um operador de agregação A é considerado média se para todo $x \in [0, 1]^n$ ele é delimitado por

$$T_{min}(x) \leq A(x) \leq S_{max}(x). \quad (5)$$

A regra é que o valor de saída não pode ser menor ou maior que qualquer valor de entrada. Um exemplo de operador de média é dado por:

$$M(x) = \frac{1}{n} \sum_{j=1}^n x_j. \quad (6)$$

Os operadores de média são idempotentes, estritamente crescentes, simétricos, homogêneos, e Lipschitz contínuos.

APÊNDICE B – Método de Mínimos Quadrados Recursivos

O método de mínimos quadrados recursivos é utilizado na adaptação do parâmetro a_i^j da função consequente. Sendo $(x, y)^{[h]}$ os pares de amostras disponíveis para treinamento em um dado instante h . O ajuste dos coeficientes a_i^j de p^i podem ser realizados através de:

$$y[h] = a_0^i + \sum_{j=1}^n a_j^i x_j^{[h]}. \quad (7)$$

A equação 7 pode ser reescrita na forma de matriz por:

$$Y = X\Omega^i, \quad (8)$$

onde $Y = [y^{[h]}]$, $X = [1 \ x_1^{[h]} \ \dots \ x_n^{[h]}]$, e $\Omega^i = [a_0^i \ \dots \ a_n^i]^T$ é o vetor de parâmetros desconhecidos. Para estimar os coeficientes a_i^j temos que,

$$Y = X\Omega^i + E \quad (9)$$

Onde E é o erro de aproximação. O algoritmo de mínimos quadrados recursivos escolhem Ω^i para fazer a minimização da função

Ω^i é dado por,

$$\Omega^i = (X^T X)^{-1} X^T Y. \quad (10)$$

Assumindo $P = (X^T X)^{-1}$ e utilizando a matriz de inversão lemma (YOUNG, 2012), é possível evitar a inversão $X^T X$ utilizando,

$$P(novo) = P(velho) \left[I - \frac{X X^T P(velho)}{1 + X^T P(velho) X} \right], \quad (11)$$

Onde I é a matriz de identidade. Geralmente os valores iniciais de entradas da diagonal principal P são valores grandes. Após as transformações matemáticas, o vetor de parâmetros é reajustado recursivamente,

$$\Omega^i(novo) = \Omega^i(velho) + P(novo) X (Y - X^T \Omega^i(velho)) \quad (12)$$

As provas de convergência do algoritmo de mínimos quadrados recursivos podem ser encontradas em (ÅSTRÖM; WITTENMARK, 2013) e (JOHNSON, 1988).