



Julio Ribeiro Coutinho

**AutoMFIS: Um Sistema Fuzzy para Previsão
de Séries Temporais Multivariadas**

Dissertação de Mestrado

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

Orientador : Prof. Ricardo Tanscheit
Co-Orientador: Prof. Marley Maria Bernardes Rebuszi Vellasco

Rio de Janeiro
Setembro de 2015



Julio Ribeiro Coutinho

**AutoMFIS: Um Sistema Fuzzy para Previsão de Séries
Temporais Multivariadas**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-Graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica do Centro Técnico Científico da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Ricardo Tanscheit

Orientador

Departamento de Engenharia Elétrica – PUC-Rio

Profa. Marley Maria Bernardes Rebuzzi Vellasco

Co-orientadora

Departamento de Engenharia Elétrica – PUC-Rio

Prof. Karla Tereza Figueiredo Leite

Departamento de Engenharia Elétrica – PUC-Rio

Prof. Alexandre Gonçalves Evsukoff

UFRJ

Prof. José Eugenio Leal

Coordenador Setorial do Centro

Técnico Científico

Rio de Janeiro, 10 de setembro de 2015

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

Julio Ribeiro Coutinho

Formado em Engenharia Mecânica pela PUC-Rio, teve como principais interesses durante a graduação modelagem matemática, tendo concluído um domínio adicional em matemática, e simulações computacionais, mais especificamente elementos finitos. Em sua experiencial profissional, teve a oportunidade de entrar em contato com ferramentas e métodos de apoio à decisão, o que o levou a buscar um mestrado nesta área. cursou disciplinas em estatística, otimização e inteligência computacional, concentrando-se nesta última. Suas áreas de interesse são ciência de dados e métodos de apoio à decisão.

Ficha Catalográfica

Coutinho, Julio Ribeiro

AutoMFIS: Um Sistema Fuzzy para Previsão de Séries Temporais Multivariadas / Julio Ribeiro Coutinho; orientador: Ricardo Tanscheit; co-orientador: Marley Maria Bernardes Rebuszi Vellasco. — 2015.

108 f. : il. (color.) ; 30 cm

Dissertação (Mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica, 2015.

Inclui bibliografia.

1. Engenharia Elétrica – Teses. 2. Sistema de Inferência Fuzzy. 3. Previsão de Séries Temporais Multivariadas. 4. Interpretabilidade. 5. Modelagem Orientada a Dados. I. Tanscheit, Ricardo. II. Vellasco, Marley Maria Bernardes Rebuszi. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. IV. Título.

CDD: 621.3

Agradecimentos

À minha família – tanto os que estão próximos quanto os que estão, infelizmente, distantes –, pelo infinito apoio e incentivo.

Ao meu orientador Ricardo Tanscheit e co-orientadora Marley Velasco, por serem uma fonte de direcionamento e conhecimento, pela constante disponibilidade para discussões, e por, ao custo de incontáveis revisões, me ajudarem a “curar” meus vícios de escrita – ou pelo menos da maior parte deles –, todos fatores essenciais para a produção desta dissertação. Também agradeço pela paciência e compreensão com a minha ocasional ausência, causada pelos meus compromissos profissionais.

Aos professores Karla Figueiredo, José Mauro, Cristiano Fernandes, Reinaldo Souza, Marco Aurélio Pacheco e Alexandre Street, pelos ensinamentos passados, e aos meus colegas de mestrado, em especial Adriano Koshiyama, sendo nossas colaborações a origem e alicerces deste trabalho.

À PUC-Rio, por oferecer um ambiente e estrutura que me proporcionaram um desenvolvimento inestimável, desde o início da minha graduação até o presente.

À CAPES por oferecer os auxílios necessários para minha participação nesse programa de mestrado.

Finalmente, a todos meus colegas de trabalho, cujo convívio e interação me ajudaram a definir o caminho a ser trilhado na minha vida profissional, que levou-me a iniciar esse mestrado.

Resumo

Coutinho, Julio Ribeiro; Tanscheit, Ricardo; Vellasco, Marley Maria Bernardes Rebuzzi. **AutoMFIS: Um Sistema Fuzzy para Previsão de Séries Temporais Multivariadas**. Rio de Janeiro, 2015. 108p. Dissertação de Mestrado — Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

A série temporal é a representação mais comum para a evolução no tempo de uma variável qualquer. Em um problema de previsão de séries temporais, procura-se ajustar um modelo para obter valores futuros da série, supondo que as informações necessárias para tal se encontram no próprio histórico da série. Como os fenômenos representados pelas séries temporais nem sempre existem de maneira isolada, pode-se enriquecer o modelo com os valores históricos de outras séries temporais relacionadas. A estrutura formada por diversas séries de mesmo intervalo e dimensão ocorrendo paralelamente é denominada série temporal multivariada. Esta dissertação propõe uma metodologia de geração de um Sistema de Inferência Fuzzy (SIF) para previsão de séries temporais multivariadas a partir de dados históricos, com o objetivo de obter bom desempenho tanto em termos de acurácia de previsão como no quesito interpretabilidade da base de regras – com o intuito de extrair conhecimento sobre o relacionamento entre as séries. Para tal, são abordados diversos aspectos relativos ao funcionamento e à construção de um SIF, levando em conta a sua complexidade e clareza semântica. O modelo é avaliado por meio de sua aplicação em séries temporais multivariadas da base completa da competição M3, comparando-se a sua acurácia com as dos métodos participantes. Além disso, através de dois estudos de caso com dados reais públicos, suas possibilidades de extração de conhecimento são exploradas por meio de dois estudos de caso construídos a partir de dados reais. Os resultados confirmam a capacidade do AutoMFIS de modelar de maneira satisfatória séries temporais multivariadas e de extrair conhecimento da base de dados.

Palavras-chave

Sistema de Inferência Fuzzy ; Previsão de Séries Temporais Multivariadas ; Interpretabilidade ; Modelagem Orientada a Dados.

Abstract

Coutinho, Julio Ribeiro; Tanscheit, Ricardo (advisor); Vellasco, Marley Maria Bernardes Rebuszi. **AutoMFIS: A Fuzzy System for Multivariate Time Series Forecast**. Rio de Janeiro, 2015. 108p. Msc. Dissertation — Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

A time series is the most commonly used representation for the evolution of a given variable over time. In a time series forecasting problem, a model aims at predicting the series' future values, assuming that all information needed to do so is contained in the series' past behavior. Since the phenomena described by the time series does not always exist in isolation, it is possible to enhance the model with historical data from other related time series. The structure formed by several different time series occurring in parallel, each featuring the same interval and dimension, is called a multivariate time series. This dissertation proposes a methodology for the generation of a Fuzzy Inference System (FIS) for multivariate time series forecasting from historical data, aiming at good performance in both forecasting accuracy and rule base interpretability – in order to extract knowledge about the relationship between the modeled time series. Several aspects related to the operation and construction of such a FIS are investigated regarding complexity and semantic clarity. The model is evaluated by applying it to multivariate time series obtained from the complete M3 competition database and by comparing it to other methods in terms of accuracy. In addition knowledge extraction possibilities are explored through two case studies built from actual data. Results confirm that AutoMFIS is indeed capable of modeling time series' behaviors in a satisfactory way and of extracting meaningful knowledge from the databases.

Keywords

Fuzzy Inference System ; Multivariate Time Series Forecast ; Interpretability ; Data-driven Modelling.

Sumário

1	Introdução	12
1.1	Inteligência Computacional para Previsão	13
1.2	Objetivo	14
1.3	Metodologia de experimentação	15
1.4	Contribuições	15
1.5	Organização	16
2	Previsão de Séries Temporais	17
2.1	Modelagem da Série Temporal	18
2.2	Correlação entre Séries Temporais	19
2.3	Processo Autoregressivo Vetorial (VAR)	24
2.4	Outros Modelos Multivariados	30
3	Automatic Multivariate Fuzzy Inference System (AutoMFIS)	34
3.1	Sistema de Inferência Fuzzy para Previsão	35
3.2	Modelagem para Extração de Conhecimento	38
3.3	Modelo AutoMFIS	40
4	Experimentos	62
4.1	Comparação com métodos da competição M3	62
4.2	Avaliação da interpretabilidade	73
5	Conclusão	100
	Referências bibliográficas	102

Lista de figuras

2.1	Fluxograma da modelagem de um conjunto de séries temporais.	18
2.2	Exemplo de dendrograma de um Clustering Hierárquico.	22
2.3	Multistep para um SIF.	30
2.4	Esquema de uma rede MLP para previsão univariada.	32
2.5	Esquema de redes MLP para previsão multivariada.	32
2.6	Esquema de um SIF para previsão univariada.	33
2.7	Esquema para previsão multivariada.	33
3.1	Diagrama de um SIF.	35
3.2	Exemplos de conjuntos fuzzy.	35
3.3	Esquema de um SIF para previsão univariada.	37
3.4	Esquema para previsão multivariada.	37
3.5	Multistep para um SIF.	38
3.6	Ajuste de um SIF a partir de dados.	39
3.7	Resumo: Funcionamento do modelo AutoMFIS.	40
3.8	Resumo: Funcionamento do modelo AutoMFIS em 3 Etapas.	41
3.9	Série com tendência.	43
3.10	Série estacionária com out-of-sample extrapolando o domínio das funções de pertinência.	44
3.11	Folga na definição do domínio.	44
3.12	Dicionário de baixa interpretabilidade.	45
3.13	Segmentação uniforme do domínio.	45
3.14	Segmentação pelos percentis da distribuição de valores.	46
3.15	Geração orientada a dados.	47
3.16	Geração de regras do AutoMFIS.	48
3.17	Exemplo de geração iterativa.	50
3.18	Representação dos vetores da premissa e conseqüente no hipercubo $[0, 1]^3$.	52
3.19	Série não explicável por regras de tamanho 1.	57
3.20	Resumo: Funcionamento do modelo AutoMFIS.	60
4.1	62 séries temporais mensais de finanças - Competição M3.	65
4.2	Dendrograma do clustering hierárquico.	66
4.3	Séries do cluster 1.	67
4.4	Séries do cluster 2.	67
4.5	Séries do cluster 3.	67
4.6	Exemplos de relacionamentos entre séries.	76
4.7	Gráfico das séries populacionais Lince-Lebre.	77
4.8	Correlação cruzada Lince X Lebre.	78
4.9	Autocorrelação Lince.	78
4.10	Autocorrelação Lebre.	79
4.11	Previsão fora da amostra pela abordagem acoplada de defuzzificação.	80

4.12	Previsão fora da amostra pela abordagem tradicional de defuzzificação.	81
4.13	Previsões dentro da amostra com agregação não-ponderada.	85
4.14	Previsões dentro da amostra com agregação ponderada.	85
4.15	Previsão fora da amostra com agregação não-ponderada.	86
4.16	Previsão fora da amostra com 7 conjuntos fuzzy.	86
4.17	Previsão fora da amostra com com 9 conjuntos fuzzy.	87
4.18	Séries populacionais Lince-Lebre adicionadas de uma série aleatória.	87
4.19	Correlação cruzada Lince X Ruído.	88
4.20	Correlação cruzada Lebre X Ruído.	88
4.21	Previsão fora da amostra para o modelo com 1 defasagem.	91
4.22	Série de consumo de energia (Indústria, Sudeste).	94
4.23	Série de consumo de energia multivariada.	94
4.24	Autocorrelação do consumo de energia.	95
4.25	Correlação Cruzada do consumo de energia com sua cópia defasada.	95

Lista de tabelas

2.1	Métricas de ligação.	23
3.1	Propriedades de uma t-norma.	36
3.2	T-normas utilizadas.	36
3.3	Parâmetros de interpretabilidade.	38
3.4	Exemplo de conjunto de treinamento univariado	39
4.1	Métodos da Competição M3, divididos em categorias.	64
4.2	Séries temporais por tipo e intervalo.	64
4.3	Comparativo dos resultados: Benchmarks M3 x AutoMFIS.	70
4.4	Comparativo dos resultados: Benchmarks M3 x AutoMFIS modificado.	71
4.5	Testes estatísticos: Cluster 1	72
4.6	Testes estatísticos: Cluster 2	73
4.7	Testes estatísticos: Cluster 3	74
4.8	Base de Regras - M3 Cluster 1.	75
4.9	Base de Regras - M3 Cluster 2.	75
4.10	Base de Regras - M3 Cluster 3.	75
4.11	Tamanho média da base de regras por série.	75
4.12	Erro SMAPE das melhores configurações para Predador-Presa.	80
4.13	Quantidade de regras das melhores configurações para Predador-Presa.	81
4.14	Centros dos conjuntos Fuzzy: Originais X Ajustados.	82
4.15	Quantidade de regras por conjunto fuzzy das variáveis de saída (% do total).	83
4.16	Erro e quantidade de regras por limiar de corte.	84
4.17	Quantidade de regras por método de agregação.	84
4.18	Erro e quantidade de regras com e sem série ruído.	88
4.19	Base de Regras - Lebre.	89
4.20	Base de Regras - Lince.	90
4.21	Erro SMAPE das melhores configurações para Predador-Presa, por quantidade de defasagens.	90
4.22	Quantidade de regras das melhores configurações para Predador-Presa, por quantidade de defasagens.	91
4.23	Base de Regras - Lebre (1 defasagem).	92
4.24	Base de Regras - Lince (1 defasagem).	92
4.25	Erro SMAPE das melhores configurações para Consumo de Energia.	96
4.26	Quantidade de regras das melhores configurações para Consumo de Energia.	96
4.27	Base de Regras - Consumo Atrasado (2 defasagens).	96
4.28	Base de Regras - Consumo Atrasado (3 defasagens).	96
4.29	Base de Regras - Consumo Atrasado (4 defasagens).	97
4.30	Regras adicionadas pela eliminação do critério de corte (2 defasagens).	97

4.31	Presença de regras esperados por método de associação.	98
4.32	Regras não ideais para conjunto fuzzy M (SD)	98

1

Introdução

No contexto atual, em todas as esferas da sociedade, de abundância de informação, tem-se o desafio de entender como estruturá-la, armazená-la e, por fim, utilizá-la para geração de valor. Esse último passo depende da capacidade de se extrair conhecimento significativo da massa de dados disponível e tem impulsionado o desenvolvimento de diversas áreas do conhecimento, como a Inteligência Computacional, que oferece ferramentas para tratar uma grande variedade de problemas [1, 2, 3, 4].

Um grande volume dos dados existentes é quantitativo. Destes, são de especial interesse os que descrevem valores dinâmicos, pois os mesmos são uma fonte de incerteza para qualquer iniciativa que deles dependa. Estes valores mutáveis comumente têm a sua evolução no tempo descrita por uma série temporal: uma sequência de medições sucessivas em intervalos de tempo fixos. Como muitos fenômenos podem ser abstraídos na forma de uma série temporal, o estudo destas tem relevância em diversas áreas, tais como finanças [5], processamento de sinais [6] e meteorologia [7].

Um dos focos do estudo em séries temporais é a previsão, que visa a gerar aproximações do comportamento futuro com o maior grau de acurácia possível, supondo que o seu comportamento passado contenha as informações necessárias para tal. A busca por uma maior acurácia e a ausência de um método que seja ótimo para todos os problemas têm motivado o desenvolvimento de diversos modelos, nos campos da estatística (e.g. ARIMA, Holt-Winters) e da Inteligência Computacional (e.g. Redes Neurais, Sistemas de Inferência Fuzzy (SIFs)). Essa dissertação abordará elementos dos dois campos, porém com ênfase no segundo, em especial SIFs. O uso de SIFs traz uma camada adicional de valor à previsão, na forma de extração de conhecimento via interpretação da base de regras, quando esta é gerada a partir dos dados. A redução da dinâmica que rege o comportamento da série temporal estudada a regras na forma “se-então” oferece um grau de interpretabilidade inexistente em outros métodos orientados a dados, que costumam gerar sistemas do tipo caixa-preta.

É importante notar que os fenômenos representados por séries temporais não existem isoladamente, mas inseridos em um contexto, influenciando e sendo influenciados por outros fenômenos ao seu redor. Por exemplo, pode-se tentar inferir o preço futuro da energia elétrica no Brasil apenas pelo seu histórico, mas, ao incluir na análise fatores como índice pluviométrico,

importante no funcionamento de hidrelétricas, e o custo de combustíveis fósseis, influenciando no custo de operação das termelétricas, tem-se o fenômeno estudado contextualizado, potencialmente aprimorando a acurácia da previsão realizada. Essa noção motivou o desenvolvimento de modelos estatísticos multivariados como o VAR [8], que visa a capturar as interdependências entre séries, com o intuito de realizar previsões simultâneas mais acuradas. Os ganhos e perdas decorrentes da incorporação da lógica vetorial de métodos multivariados ao SIF serão abordados em detalhe nos próximos capítulos.

1.1 Inteligência Computacional para Previsão

A partir da segunda metade do século XX, problemas de previsão foram tratados tradicionalmente por modelos estatísticos lineares como o ARIMA, que se popularizaram especialmente após a publicação do método Box & Jenkins [9]. Notou-se, porém, que modelos lineares não se mostraram suficientemente adequados para resolver alguns problemas reais, como dinâmicas populacionais de espécies e ciclos de atividade solar. Isto motivou desenvolvimento de modelos não lineares como o ARCH [10], embora ainda em sua infância quando comparados aos modelos lineares, apresentando, em consequência, incertezas quanto ao seu melhor desempenho [11] [12]. Paralelamente, em especial nas últimas duas décadas, modelos da Inteligência Computacional têm surgido como alternativas aos modelos não lineares clássicos, com resultados ocasionalmente superiores [13] [14]. Esses modelos oferecem a vantagem de serem orientados a dados, dependendo apenas dos históricos para o aprendizado da relação entre o estado passado e o desenvolvimento futuro das séries, eliminando a necessidade de uma análise detalhada para definir as componentes do modelo, o que muitas vezes requer conhecimentos específicos do problema.

Os primeiros exemplos de sucesso na área de Inteligência Computacional (IC) em problemas de previsão foram obtidos com o uso de Redes Neurais Artificiais (RNAs), que lograram obter resultados superiores aos proporcionados por métodos estatísticos como regressão linear [15] e Box & Jenkins [16]. Ainda na década de 80, foram publicados resultados similarmente positivos para o uso de RNAs na modelagem de problemas não lineares [17]. Além das redes neurais, foi documentado o bom desempenho de outras técnicas como, por exemplo, Árvores de Decisão [18] e SVM [19], demonstrando o potencial da área nesse tipo de problema.

A evolução do desempenho de modelos da Inteligência Computacional vem sendo acompanhada através de diversas competições de previsão existen-

tes, que oferecem dados de fontes e condições diversas, como, por exemplo, a NN3 para séries mensais e a NN5 para séries diárias. Nos resultados das últimas edições pôde-se observar o destaque de novas abordagens baseadas em métodos bem estabelecidos, como as Redes Neurais Echo State [20] e o uso de comitês de Redes Neurais [21].

Paralelamente aos demais modelos de IC, estabeleceu-se o uso de Sistemas de Inferência Fuzzy, baseados na Teoria de Conjuntos Fuzzy [22]. Essa teoria foi criada com o intuito de incorporar um grau de aproximação e incerteza à lógica clássica, fazendo uso de valores intermediários ao "verdadeiro" e "falso" admitidos tradicionalmente. A popularização do uso da Lógica Fuzzy para solução de problemas reais só viria, porém, décadas depois, com os SIFs propostos por Takagi & Sugeno [23] [24] e Mamdani[25]. De maneira similar a modelos clássicos, esses sistemas apresentam o problema de dependerem do conhecimento de um especialista, envolvendo um oneroso processo iterativo para a construção e sintonia da base de regras. De modo a contornar esse problema, surgiram os sistemas híbridos, que fazem uso de outras técnicas de IC, como as Redes Neurais e os Algoritmos Genéticos, para a definição da base de regras a partir apenas dos dados históricos e para a sintonia de parâmetros, oferecendo uma solução orientada a dados mais transparente.

Propõe-se, nesta dissertação, investigar o uso de um SIF com uma abordagem orientada a dados para a solução de problemas de previsão multivariado. Esse modelo é denominado AutoMFIS (Automatic Multivariate Fuzzy Inference System).

1.2 Objetivo

O objetivo principal da dissertação é a construção de um sistema de previsão de múltiplas séries temporais a partir da combinação de técnicas fuzzy e híbridas existentes na literatura com a lógica vetorial presente em métodos de previsão multivariada (e.g. VAR), investigando seu desempenho no âmbito de acurácia e interpretabilidade, quando comparado ao de métodos univariados e multivariados tradicionais. Os pontos a ser investigados incluem:

- Construção do dicionário de termos linguísticos, incluindo seleção de variáveis e definição dos conjuntos fuzzy
- Geração da base de regras a partir de dados
- Procedimentos de inferência e defuzzificação
- Interpretabilidade da base de regras resultante

1.3

Metodologia de experimentação

Para verificar o desempenho do modelo proposto, o mesmo passará por duas etapas de experimentação:

1. Avaliação do desempenho do modelo de previsão para três séries multivariadas construídas a partir da base de dados de uma competição de previsão, utilizando como benchmark todos os métodos participantes.
2. Análise do potencial de interpretabilidade do modelo para dois estudos de caso construídos com dados reais públicos. Como os ajustes realizados para aumentar a interpretabilidade podem ocasionar perda de acurácia, esta também será avaliada.

1.4

Contribuições

Primeiramente, esse trabalho oferece uma contribuição conceitual, na forma do modelo AutoMFIS, empregado em problemas de previsão multivariada utilizando uma abordagem híbrida, que, além das vantagens associadas a métodos orientados a dados, oferece alto grau de transparência sobre o seu funcionamento. Esta abordagem foi construída para ser aplicada sob duas condições: quando se têm séries temporais correlacionadas e quando se deseja entender a dinâmica entre elas.

Em segundo lugar, o próprio processo de ajuste do modelo AutoMFIS para os estudos de caso selecionados oferece uma visão dos impactos de seus diferentes aspectos sobre os resultados, incluindo a definição dos conjuntos fuzzy, geração e filtragem da base de regras a partir dos dados e métodos de agregação e defuzzificação. Em adição, é sugerida uma nova abordagem para esta última, com o intuito de injetar a lógica de interdependência entre séries na forma de uma defuzzificação acoplada.

Por último, é oferecida uma contribuição prática na forma de um conjunto de resultados aplicados a problemas reais e bases de dados de competição. A comparação engloba modelos lineares e não lineares, univariados e multivariados, da estatística clássica e da Inteligência Computacional. Esses resultados contribuirão para a contínua investigação da eficácia de modelos multivariados, em comparação com abordagens tradicionais, para problemas em que existe um alto grau de correlação entre as séries. Um segundo aspecto da contribuição prática é no quesito interpretabilidade, onde se avalia se esta forma de extração de conhecimento agrega valor nos casos selecionados ou se o modelo só é válido no aspecto da previsão em si, analogamente aos modelos caixa-preta.

1.5 Organização

Essa dissertação está organizada da seguinte forma:

- Capítulo 2: cobre os fundamentos de previsão de séries temporais e justifica a importância de modelos multivariados. São apresentados alguns modelos existentes na estatística clássica e conceitos relevantes, como o de correlação entre séries. Será a base teórica para a construção do problema tratado nesse trabalho.
- Capítulo 3: trata de Sistemas de Inferência Fuzzy e de sua aplicação no campo de previsão de séries temporais múltiplas. É discutida a dicotomia entre complexidade e interpretabilidade e apresentam-se as soluções empregadas para a formulação do modelo AutoMFIS.
- Capítulo 4: o desempenho do modelo AutoMFIS é inicialmente avaliado para benchmarks da competição M3, de modo a confirmar o seu potencial de previsão. Em seguida, é explorada a aplicação do modelo em estudos de caso com dados reais, analisando sua viabilidade no quesito interpretabilidade.
- Capítulo 5: conclui o trabalho, discutindo os resultados obtidos e sugerindo caminhos possíveis para a continuidade da pesquisa.

2

Previsão de Séries Temporais

Tanto os métodos estatísticos quanto de Inteligência Computacional para previsão partem de um mesmo princípio, o de que o comportamento passado de uma série contém informações sobre seu desenvolvimento futuro, ou seja, é válido realizar previsões com base nos dados históricos.

Formalmente, para casos univariados, uma função de previsão é representada da seguinte forma:

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

onde \hat{y}_t e y_t denotam, respectivamente, os valores previsto e real para a série no instante t , h indica a quantidade de períodos à frente que se quer prever, e $f(\cdot)$ representa a função de previsão. Anteriormente à previsão, busca-se a $f(\cdot)$ de maneira que $\hat{y}_t = f(y_{t-1}, y_{t-2}, \dots)$ se aproxime do valor observado y_t . A quantidade de defasagens considerada na função $f(\cdot)$ varia de acordo com o problema e é um dos pontos de investigação na sua modelagem.

Em problemas reais, não é incomum que o valor de uma série seja afetado também por variáveis extrínsecas a ela, como, por exemplo, demanda por um produto influenciando seu custo. Em tal situação, pode-se reescrever a equação 2-2 como:

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

onde y_2, \dots, y_S são as séries utilizadas na previsão da série y_1 , além de defasagens da própria. Em algumas situações busca-se realizar previsões das S séries relacionadas. Nesse caso, pode-se generalizar a equação 2-3 para qualquer uma delas, denotada por s :

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

Um conjunto qualquer de séries temporais $y_{s,t}$, $s = 1, \dots, S$; $t = 1, \dots, T$ é denominado série temporal múltipla ou multivariada. Os modelos de previsão multivariados procuram, de maneira análoga ao caso univariado, aproximar as S funções $f_s(\cdot)$ de previsão das S séries temporais. Em contrapartida ao uso de múltiplos modelos univariados, esta abordagem permite capturar relações de interdependência entre as séries, resultando, potencialmente, em uma maior acurácia. Nesse capítulo serão apresentados os conceitos de modelagem multi-

variada utilizados no modelo proposto nessa dissertação, através da formulação de modelos bem fundamentados como o VAR.

Vale salientar que, em alguns casos, a abordagem multivariada pode não apresentar ganhos de desempenho, ou até ser inferior ao de modelos univariados [26, 27, 28]. Com o intuito de minimizar perdas potenciais pela injeção de erro ou aumento desnecessário de complexidade causada por séries potencialmente não relacionadas, será testada a eficácia do uso de técnicas de cálculo de correlação entre séries para definição das séries temporais a serem modeladas conjuntamente.

2.1 Modelagem da Série Temporal

O processo de modelagem da função de previsão, e a sua subsequente aplicação, é idêntico para os casos univariados e multivariados, sendo descrito na figura 2.1.

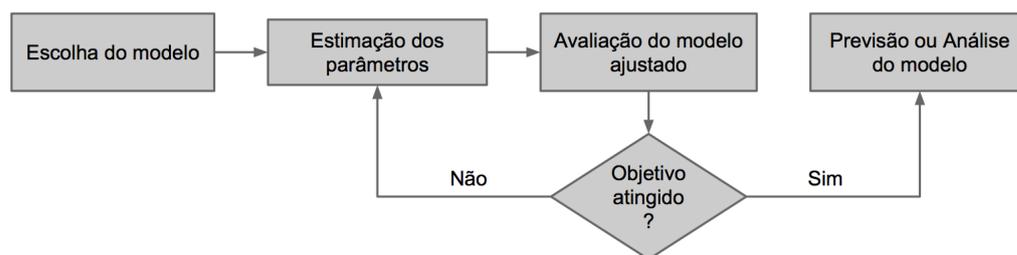


Figura 2.1: Fluxograma da modelagem de um conjunto de séries temporais.

O primeiro passo é a escolha do modelo mais apropriado, onde se consideram as características da(s) série(s), tais como linearidade e estacionaridade, além de fatores como o objetivo da análise, conhecimentos específicos sobre problema, volume de dados disponível, entre outros.

Existe uma distinção importante entre a escolha de modelos estatísticos e de Inteligência Computacional. Os modelos estatísticos já fixam a estrutura da função de previsão em sua definição, estipulando-se suas componentes. Um exemplo seria a inclusão de termos autorregressivos e sazonais, o que oferece um processo de ajuste mais direcionado, porém que ao mesmo tempo requer um conhecimento mais profundo da série modelada. Em contrapartida, os métodos de Inteligência Computacional podem aproximar uma infinidade de funções lineares e não-lineares com a mesma formulação, tendo sua estrutura definida a partir dos dados [29, 30], porém são mais dependentes do volume de informação disponível.

Selecionado um modelo, é necessário estimar os valores das variáveis e parâmetros que compõem a função de previsão modelada, a partir dos valores históricos da série. Para tal, cada modelo conta com um conjunto próprio de técnicas e metodologias.

Em métodos de ajuste iterativo da função, existe um passo adicional de verificação do modelo resultante da etapa de estimação. De acordo com a métrica escolhida, o modelo é avaliado quanto à acurácia com que representa a série temporal analisada, em relação a um limiar mínimo estipulado. Em caso negativo, retorna-se à etapa anterior, onde um novo ciclo de refinamento dos parâmetros é realizado, até que se obtenham resultados adequados segundo um critério de parada pré-determinado.

O modelo ajustado ser utilizado para as aplicações desejadas, que neste trabalho serão: previsão e análise do relacionamento entre as séries. Essa segunda aplicação será realizada por meio da interpretação da base de regras gerada para o SIF.

Será apresentado na próxima seção o modelo autoregressivo VAR, que serve como introdução dos conceitos de modelagem multivariada, que serão empregados no Sistema de Inferência Fuzzy aqui abordado.

2.2

Correlação entre Séries Temporais

Como explicitado anteriormente nas equações 2-2 e 2-3, uma função de previsão de uma série temporal pode ser enriquecida com a inclusão de outras séries relacionadas como variáveis de entrada. Como a inclusão destas informações exógenas afetará o comportamento da função ajustada, é necessário avaliar se as séries temporais incorporadas são verdadeiramente relacionadas com a série modelada, de modo a evitar uma perda de desempenho, tanto em termos de acurácia quanto de custo computacional, causada pela injeção de informações irrelevantes. Essa triagem pode ser feita por meio da análise da correlação entre as séries, que tem uma importância reconhecida em diversas áreas do conhecimento, como finanças [31, 32], meteorologia [33] e medicina [34], e qualquer outra em que é relevante analisar os fenômenos contextualizados pelos que estão à sua volta.

Um ponto a ser salientado é que a correlação é capaz de avaliar apenas o aspecto linear do relacionamento entre séries, o que entra em conflito com a capacidade de um SIF de extrair relações não-lineares. Por essa razão, as ferramentas exploradas nesta seção têm apenas um caráter de auxílio à definição dos problemas, ajudando a definir uma configuração inicial básica

para cada experimento.

A seguir serão expostas as abordagens para a análise de correlação entre séries utilizadas neste trabalho, assim como as aplicações e restrições de cada uma, com o intuito de construir o conjunto de ferramentas a ser utilizado nos experimentos apresentados no capítulo 4.

2.2.1 Correlação Cruzada

A correlação cruzada, aplicada ao contexto de processos estocásticos, é uma métrica de similaridade representada como uma função da defasagem entre eles. Sendo $Y_{i,t}$ e $Y_{j,t}$ dois processos estocásticos conjuntamente estacionários no sentido amplo¹ quaisquer, a correlação cruzada entre eles em função da quantidade de defasagens τ é dada por:

$$\rho_{y_i y_j}(\tau) = E[(Y_{i,t} - \mu_i)(Y_{j,t+\tau} - \mu_j)] / (\sigma_i \sigma_j) \quad (2-4)$$

onde μ_i e μ_j são as médias de $Y_{i,t}$ e $Y_{j,t}$, enquanto que σ_i e σ_j são os desvios padrões. Esses valores são constantes, devido à estacionaridade dos processos.

A métrica acima é, em suma, a normalização da covariância cruzada, assumindo valores entre -1 e 1. Pode-se notar também que para defasagem zero a função toma o valor do coeficiente de correlação. O fato de as correlações serem função das defasagens não só contribui para a validação da escolha das séries a serem analisadas conjuntamente, como também permite efetuar uma seleção mais bem embasada da quantidade de defasagens a serem consideradas pelo modelo de previsão – incluindo as defasagens da própria série modelada, por meio de sua autocorrelação cruzada – lembrando que essa seleção representa apenas um ponto de partida para o ajuste do modelo, sendo essa quantidade de defasagens variada durante o processo de busca da configuração ótima de cada problema.

A definição desta métrica, porém, traz duas questões. A primeira diz respeito à estacionaridade dos processos estocásticos estudados, que não é aplicável para uma grande quantidade de casos, enquanto a segunda se refere ao fato de as funções densidade de probabilidade que compõem os processos estocásticos não serem conhecidas, tendo-se apenas uma amostra dos mesmos, na forma das séries históricas analisadas.

A maneira de tratar a primeira questão acima pode variar de acordo com o problema, não existindo uma abordagem padrão. Assim, optou-se por uma

¹Processo estacionário no sentido amplo é aquele cujo primeiro momento e autocovariância não variam no tempo.

heurística comumente utilizada em economia [35], que consiste em realizar as análises com base nas taxas de variação, em vez de utilizar as séries puras:

$$Y_{dif} = (y_2 - y_1, y_3 - y_2, \dots, y_T - y_{T-1}) \quad (2-5)$$

onde Y_{dif} é a série de diferenças e y_t é o valor da série no instante t . Alternativamente, pode-se remover a tendência, estimada via métodos como a regressão linear.

O segundo problema é tratado por meio da utilização da correlação cruzada amostral [36]: a correlação cruzada é estimada a partir de uma amostra de pontos das séries temporais.

$$\hat{\rho}_{y_i y_j}(\tau) = \frac{\sum_{t=1}^{T-k} (y_{i,t} - \bar{\mu}_i)(y_{j,t+\tau} - \bar{\mu}_j)}{[\sum_{t=1}^T (y_{i,t} - \bar{\mu}_i)^2 \sum_{t=1}^T (y_{j,t} - \bar{\mu}_j)^2]^{1/2}} \quad (2-6)$$

onde $y_{i,t}$, $y_{j,t}$ são as duas séries a ser comparadas, $\bar{\mu}_i$, $\bar{\mu}_j$ as suas médias amostrais e τ a quantidade de defasagens entre as séries.

2.2.2

Clustering Hierárquico

Para analisar o grau de sincronia linear de um par de séries temporais, a correlação cruzada já atende às necessidades deste trabalho. Porém, quando há um grande número de séries, a identificação dos agrupamentos daquela que são relacionadas torna-se mais complexa, pois é necessário analisar uma quantidade de relações que cresce de maneira quadrática com o tamanho do conjunto. Para tratar deste problema, utiliza-se um método de agrupamento denominado Clustering Hierárquico [37]. Ao contrário das demais técnicas de agrupamento, que segmentam um conjunto em n clusters em um único passo, o clustering hierárquico oferece uma visualização de estágios incrementais de segmentação em função de uma métrica de similaridade arbitrária. Essa visualização leva o nome de dendrograma, apresentada na figura 2.2.

A vantagem deste método é o caráter exploratório que oferece, permitindo uma escolha mais bem embasada de limiar de similaridade para o agrupamento de séries.

O Clustering Hierárquico pode ser construído via duas abordagens, a Aglomerativa (bottom-up) e a Divisiva (top-down). Como os nomes sugerem, a primeira aglomera os elementos em grupos sucessivamente, chegando a um único cluster no final, enquanto a segunda traça o caminho contrário, partindo de um cluster único contendo todos os elementos e efetuando divisões sucessivas.

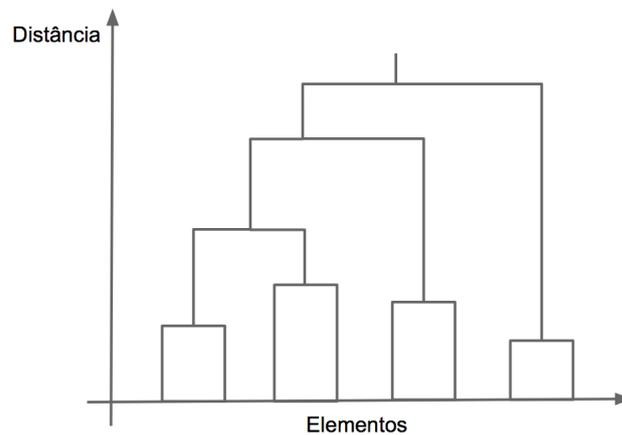


Figura 2.2: Exemplo de dendrograma de um Clustering Hierárquico.

De uma maneira geral, a abordagem divisiva requer maiores recursos computacionais, além de ter critérios de segmentação mais difíceis de estabelecer [37], fazendo com que o Clustering Aglomerativo seja mais utilizado na maioria das aplicações. Assim, este é o adotado neste trabalho. Para a aplicação deste método, algumas escolhas devem ser feitas:

1. **Métrica de distância entre séries:** pode ser qualquer função que respeite as quatro condições que definem uma métrica [38]; por exemplo, a distância Euclidiana.
2. **Métrica de distância entre clusters ou entre uma série e um cluster:** método de cálculo da distância entre clusters dadas as distâncias entre os elementos individuais; exemplo: média das distâncias entre elementos de dois clusters.
3. **Escolha do limiar de similaridade:** ponto de corte que definirá os clusters a utilizar.

Estes pontos são abordados a seguir.

Distância entre séries

De modo a manter a coesão entre as técnicas descritas neste capítulo, a métrica de distância utilizada será uma manipulação do coeficiente de correlação. Em Gu & Wang [39] foram analisadas duas métricas com base na correlação; emprega-se a que apresentou o melhor desempenho:

$$d_v(Y_{i,t}, Y_{j,t}) = \sqrt{\sum_{s=1}^S (M_{i,s} - M_{j,s})^2} \quad (2-7)$$

onde $Y_{i,t}$, $Y_{j,t}$ são um par de séries, e M_{ij} denota a correlação entre as séries $Y_{i,t}$ e $Y_{j,t}$, com $i, j \in \{1, \dots, S\}$ – sendo S a quantidade de séries a serem agrupadas.

Distância entre clusters

Dado o caráter gradual de agrupamento do método, é necessário definir a métrica para o cálculo da distância entre dois clusters ou uma série e um cluster. Algumas das chamadas métricas de ligação mais utilizadas são:

Métrica	Equação
Simple	$\min\{d(a, b) : a \in A, b \in B\}$
Completa	$\max\{d(a, b) : a \in A, b \in B\}$
Média	$\frac{1}{ A B } \sum_{a \in A} \sum_{b \in B} d(a, b)$

Tabela 2.1: Métricas de ligação.

Como o objetivo é agrupar elementos de forma que todos sejam satisfatoriamente similares, optou-se pelo uso da ligação completa, que usa a distância máxima. Assim, para um ponto de corte do agrupamento igual a d , a distância entre qualquer par de elementos dentro de um grupo será no máximo esse valor, tornando o processo de escolha mais intuitivo.

Escolha do limiar

Como neste trabalho não há interesse na hierarquia em si, mas sim em determinar uma partição de um conjunto de séries, é necessário definir um limiar de distância que servirá como o ponto de corte do dendrograma, estabelecendo os agrupamentos a utilizar.

Na seção 4.4.4 de [37] são discutidas algumas regras existentes na literatura. Nestas regras, de maneira geral, definem-se parâmetros de maneira arbitrária, tornando o processo ultimamente subjetivo. Observa-se, na comunidade científica, uma ausência de consenso quanto à empregabilidade de cada regra e é possível verificar resultados inconstantes de uma mesma regra para um mesmo conjunto de dados.

Neste trabalho, emprega-se uma metodologia mais informal, combinando a escolha de métricas mais interpretáveis, como descrito nas subseções acima, e a análise, caso a caso, dos dendrogramas gerados, utilizando como heurística básica que se evite formar grupos cujas dimensões não sejam comportadas pelos recursos disponíveis.

Dado que as séries temporais foram agrupadas com o auxílio dos métodos apresentados, é iniciado o processo de ajuste do modelo. Nas próximas seções será utilizado o modelo VAR para apresentar os conceitos comuns a todos os modelos de previsão multivariados, importantes para a construção do AutoMFIS.

2.3 Processo Autoregressivo Vetorial (VAR)

A maneira mais simples de aproximar a função de previsão de uma série temporal é representá-la como uma função linear de N observações passadas:

$$\hat{y}_{t+1} = v + a_1 y_t + a_2 y_{t-1} + \dots + a_N y_{t-L+1} \quad (2-8)$$

onde \hat{y}_{t+1} é a previsão um passo a frente, v é um intercepto, y_t são os valores reais da série até o instante t , sendo os a_i , $i = 1, \dots, N$, os coeficientes associados a cada defasagem.

Como \hat{y}_{t+1} potencialmente possui um erro em relação ao valor real y_{t+1} , pode-se definir esse último como:

$$\begin{aligned} y_{t+1} &= \hat{y}_{t+1} + u_{t+1} \\ &= v + a_1 y_t + a_2 y_{t-1} + \dots + a_N y_{t-L+1} + u_{t+1} \end{aligned} \quad (2-9)$$

onde o termo u_{t+1} se refere ao erro de previsão.

Com essa abordagem, tem-se um processo autoregressivo de ordem L , supondo que os erros u_t de períodos diferentes são decorrelatados, ou seja, não existem erros sistemáticos de previsão.

Expandindo a equação 2-8 com a lógica da série temporal múltipla explicitada em 2-3, tem-se a seguinte equação:

$$\begin{aligned} \hat{y}_{s,t+1} &= v + a_{s,1,1} y_{1,t} + a_{s,2,1} y_{2,t} + \dots + a_{s,S,1} y_{S,t} + \\ &\dots + a_{s,1,L} y_{1,t-L+1} + \dots + a_{s,S,L} y_{S,t-L+1} \end{aligned} \quad (2-10)$$

onde $s \in \{1, \dots, S\}$, com S sendo a dimensão da série temporal múltipla em questão. Adotando uma notação vetorial e matricial, o conjunto de S equações

de previsão pode ser escrito de maneira compacta:

$$y_t = (y_{1,t}, \dots, y_{s,t})' \quad (2-11)$$

$$\hat{y}_t = (\hat{y}_{1,t}, \dots, \hat{y}_{s,t})' \quad (2-12)$$

$$v = (v_1, \dots, v_s)' \quad (2-13)$$

$$A_l := \begin{bmatrix} a_{1,1,l} & \cdots & a_{1,s,l} \\ a_{2,1,l} & \cdots & a_{2,s,l} \\ \vdots & \ddots & \vdots \\ a_{s,1,l} & \cdots & a_{s,s,l} \end{bmatrix} \quad (2-14)$$

onde y_t e \hat{y}_t são, respectivamente, os vetores dos valores reais e estimados das S séries no instante t , v é o vetor de interceptos e A_l a matriz de coeficientes para a l -ésima defasagem.

O conjunto de equações acima pode ser representado na forma compacta similar a 2-8:

$$\hat{y}_{t+1} = v + A_1 y_t + A_2 y_{t-1} + \dots + A_L y_{t-L+1} \quad (2-15)$$

Observando o mesmo desenvolvimento da equação 2-8 para a 2-9, ter-se-á um processo autoregressivo vetorial, ou seja, um processo VAR de ordem L :

$$y_{t+1} = v + A_1 y_t + A_2 y_{t-1} + \dots + A_L y_{t-L+1} + u_{t+1} \quad (2-16)$$

2.3.1

Estimação e Verificação de um processo VAR

No ajuste de um processo VAR(L), na forma descrita na equação 2-16, os valores dos interceptos v , das matrizes de coeficientes A_l e da matriz de covariância Σ_u do erro u são inicialmente desconhecidos. Para estimá-los, utilizam-se os próprios valores históricos da série temporal múltipla a modelar. Estão disponíveis na literatura diversos estimadores; serão apresentados nas próximas subseções as versões multivariadas de dois deles (estimação por Mínimos Quadrados e por Máxima Verossimilhança), comumente utilizados em modelos univariados [8].

Estimação via Mínimos Quadrados Multivariados

Na formulação deste estimador para um processo VAR(L), segmenta-se o histórico de dados em conjuntos de amostra e pré-amostra, com o primeiro sendo y_1, \dots, y_T e o segundo constituído por pontos com dimensão igual à ordem do VAR, y_{-L+1}, \dots, y_0 , ambos para as S séries temporais consideradas. Faz-se

uso de uma notação simplificada:

$$Y := (y_1, \dots, y_T) \quad (S \times T) \quad (2-17)$$

$$B := (v, A_1, \dots, A_L) \quad (S \times (SL + 1)) \quad (2-18)$$

$$Z_t := \begin{bmatrix} 1 \\ y_t \\ \vdots \\ y_{t-L+1} \end{bmatrix} \quad ((SL + 1) \times 1) \quad (2-19)$$

$$Z := (Z_0, \dots, Z_{T-1}) \quad ((SL + 1) \times 1) \quad (2-20)$$

$$U := (u_1, \dots, u_T) \quad (S \times T) \quad (2-21)$$

$$y := \text{vec}(Y) \quad (ST \times 1) \quad (2-22)$$

$$\beta := \text{vec}(B) \quad ((S^2L + S) \times 1) \quad (2-23)$$

$$b := \text{vec}(B') \quad ((S^2L + S) \times 1) \quad (2-24)$$

$$u := \text{vec}(U) \quad (ST \times 1) \quad (2-25)$$

onde Y é a matriz que contém os T valores da amostra para as S séries temporais, B é a matriz formada pelo vetor de interceptos v e as matrizes de coeficientes A_l referentes às L defasagens consideradas. Utilizando essa notação, o modelo VAR(L) pode ser escrito como:

$$Y = BZ + U \quad (2-26)$$

ou

$$y = (Z' \otimes I_S)\beta + u \quad (2-27)$$

onde I_S é uma matriz identidade de dimensão S e \otimes é o produto de Kronecker.

Para encontrar estimador de Mínimos Quadrados de β , parte-se da definição de R como sendo o somatório do quadrado dos resíduos:

$$R(\beta) = y'(I_T \otimes \Sigma_u^{-1})y + \beta'(ZZ' \otimes \Sigma_u^{-1})\beta - 2\beta'(Z \otimes \Sigma_u^{-1})y \quad (2-28)$$

Derivando e igualando a zero, para encontrar o mínimo de R :

$$\frac{\partial R(\beta)}{\partial \beta} = 2(ZZ' \otimes \Sigma_u^{-1})\beta - 2(Z \otimes \Sigma_u^{-1})y = 0 \quad (2-29)$$

é obtido o estimador:

$$\hat{\beta} = ((ZZ')^{-1}Z \otimes I_S)y \quad (2-30)$$

Estimação via Máxima Verossimilhança

A estimação por Máxima Verossimilhança é aplicável quando se considera que a distribuição do processo é conhecida – em especial, supõe-se que o processo VAR é gaussiano, em outras palavras, o componente de erro u pode ser definido como:

$$u = \text{vec}(U) = \begin{bmatrix} u_1 \\ \vdots \\ u_T \end{bmatrix} \sim \mathcal{N}(0, I_S \otimes \Sigma_u) \quad (2-31)$$

Como se trata de uma distribuição normal, a função densidade de probabilidade de u é:

$$f_u(u) = \frac{1}{(2\pi)^{ST/2}} |I_T \otimes \Sigma_u|^{-1/2} \exp \left[-\frac{1}{2} u' (I_T \otimes \Sigma_u^{-1}) u \right] \quad (2-32)$$

onde S é a quantidade de séries temporais, T é o tamanho da amostra, I_T é uma matriz identidade de dimensão T , Σ_u é a matriz de covariância do erro u .

$$(y_t - \mu) = A_1(y_{t-1} - \mu) + \dots + A_p(y_{t-p} - \mu) + u_t \quad (2-33)$$

Utilizando a representação de um VAR(L) na *forma ajustada pela média* 2-33, obtém-se a função log-verossimilhança $\ln \ell$, como deduzido em [8, p. 88-89], a partir da qual as funções de estimação serão construídas:

$$\ln \ell(\mu, \alpha, \Sigma_u) = -\frac{ST}{2} \ln 2\pi - \frac{T}{2} \ln |\Sigma_u| - \frac{1}{2} \text{tr}[(Y^0 - AX)' \Sigma_u^{-1} (Y^0 - AX)] \quad (2-34)$$

onde

$$Y^0 := (y_1 - \mu, \dots, y_T - \mu) \quad (S \times T) \quad (2-35)$$

$$A := (A_1, \dots, A_p) \quad (S \times TL) \quad (2-36)$$

$$Y_t^0 := \begin{bmatrix} y_t - \mu \\ \vdots \\ y_{t-p+1} - \mu \end{bmatrix} \quad (KL \times 1) \quad (2-37)$$

$$X := (Y_0^0, \dots, Y_{T-1}^0) \quad (KL \times T) \quad (2-38)$$

$$y^0 := \text{vec}(Y^0) \quad (KT \times 1) \quad (2-39)$$

$$\alpha := \text{vec}(A) \quad (K^2L \times 1) \quad (2-40)$$

Aplicando derivadas parciais em relação a cada parâmetro desconhecido e igualando-as a zero, obtém-se um sistema de equações normais cuja solução gera os estimadores:

$$\frac{\partial \ln \ell}{\partial \mu} = [I_S - A(j \otimes I_S)]' \Sigma_u^{-1} \left[\sum_t (y_t - \mu - AY_{t-1}^0) \right] \quad (2-41)$$

$$\frac{\partial \ln \ell}{\partial \alpha} = (X \otimes \Sigma_u^{-1})(y - \mu^*) - (XX' \otimes \Sigma_u^{-1})\alpha \quad (2-42)$$

$$\frac{\partial \ln \ell}{\partial \Sigma_u} = -\frac{T}{2} \Sigma_u^{-1} + \frac{1}{2} \Sigma_u^{-1} (Y^0 - AX)(Y^0 - AX)' \Sigma_u^{-1} \quad (2-43)$$

Os dois tipos de estimadores operam de maneira análoga à do caso univariado, necessitando apenas que se parta da representação matricial e vetorial do modelo multivariado. Essa lógica de adaptação de técnicas univariadas bem fundamentadas para o caso multivariado será incorporada no Sistema Fuzzy proposto.

2.3.2 Previsão

Dado um processo VAR(L) ajustado para um conjunto de dados ω_t , onde:

$$\omega_T = \{y_t, t \leq T\} \quad (2-44)$$

$$y_t = (y_{1,t}, y_{2,t}, \dots, y_{S,t})' \quad (2-45)$$

pode-se desejar realizar previsões a partir do instante t , denominado origem, para h passos à frente, onde h é o "horizonte de previsão". Podem ser adotadas duas abordagens para tal: a primeira procura efetuar uma previsão do valor específico da série nos passos futuros, enquanto a segunda procura oferecer um intervalo de confiança para o valor a partir de suposições sobre o tipo de distribuição das componentes do modelo. Este trabalho concentra-se no primeiro tipo, que oferece um paralelo direto com o funcionamento dos métodos de Inteligência Computacional.

Para realizar uma previsão do primeiro tipo acima, faz-se necessário utilizar uma função de custo associada ao erro de previsão, que atinge seu valor mínimo quando a previsão é ótima. Como a tarefa de minimizar essa função é muitas vezes inviável na prática, costuma-se minimizar o valor esperado do custo. Vale salientar que a previsão varia de acordo com a função de custo, a qual é escolhida de acordo com as necessidades de quem está solucionando o problema. Em virtude do seu bom desempenho documentado na literatura, optou-se, neste trabalho, pelo Erro Médio Quadrático (Mean Square Error,

MSE) [26, 40].

Dado um VAR(L) S -dimensional $y_t = (y_{1,t}, \dots, y_{S,t})'$ e seu conjunto de dados histórico ω_t , a minimização do MSE para uma previsão h passos a frente será seu valor esperado condicional:

$$E_t(y_{t+h}) = E(y_{t+h}|\omega_t) = E(y_{t+h}|y_t|t \leq T) \quad (2-46)$$

Isto significa que se tem, para qualquer previsão $\bar{y}_t(h)$, de horizonte h e origem t :

$$MSE(\bar{y}_t(h)) \geq MSE(E_t(y_{t+h})) \quad (2-47)$$

$$E[(y_{t+h} - \bar{y}_t(h))(y_{t+h} - \bar{y}_t(h))'] \geq E[(y_{t+h} - E_t(y_{t+h}))(y_{t+h} - E_t(y_{t+h}))'] \quad (2-48)$$

Esta proposição é demonstrada em [8, p. 34]. A partir de 2-16, pode-se reescrever 2-46 como:

$$E_t(y_{t+h}) = v + A_1 E_t(y_{t+h-1}) + \dots + A_p E_t(y_{t+h-p}) \quad (2-49)$$

Vale notar que o termo $E_t(u_{t+h})$ foi ignorado pelo fato de que u_{t+h} é um ruído branco independente, ou seja, seu primeiro momento possui valor 0. A partir de ω_t , que contém os valores de y_t até o momento t , pode-se encontrar, de maneira recursiva, o valor da previsão h passos a frente a partir de $t+1$:

$$\begin{aligned} E_t(y_{t+1}) &= v + A_1 y_t + \dots + A_p y_{t-p+1} \\ E_t(y_{t+2}) &= v + A_1 E_t(y_{t+1}) + A_2 y_t + \dots + A_p y_{t-p+2} \\ &\vdots \end{aligned} \quad (2-50)$$

O processo iterativo acima é denominado previsão multistep, que significa uma previsão em múltiplos passos (figura 2.3). Este conceito também é utilizado por métodos da Inteligência Computacional, como Redes Neurais e SIFs.

A incorporação de outras séries temporais como variáveis explicativas na construção de um modelo de previsão é um conceito cujo uso é bem documentado em modelos univariados (e.g. ARIMA), porém sua aplicação é dificultada para uma previsão de múltiplos passos à frente (multistep), pois, naturalmente, só se têm os valores das variáveis explicativas até o momento atual. Esse tipo de problema exige a realização de estimativas das variáveis para

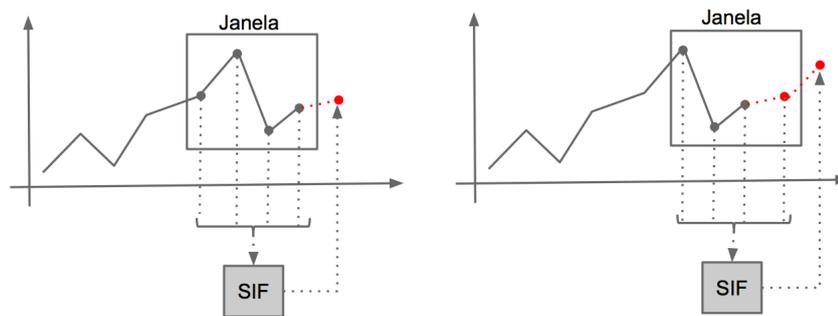


Figura 2.3: Multistep para um SIF.

alimentar o modelo nos passos subsequentes, o que por sua vez pode requerer o ajuste de novos modelos de previsão para cada uma das séries consideradas. Esta situação pode se tornar ainda mais complexa quando os modelos de previsão auxiliares utilizam também outras séries como variáveis explicativas. Nessa situação é útil aplicar o VAR – ou outro modelo multivariado qualquer – sobre o conjunto de séries temporais relevantes ao problema, agrupadas na estrutura denominada série temporal múltipla.

Um aspecto adicional de interesse é que o processo de ajuste do modelo multivariado definirá o impacto de cada série nas demais. Isto, por um lado, reduz a dependência de um especialista, mas, por outro, pode ocasionar o aumento do erro do modelo por meio de associações indevidas, que por qualquer motivo não tenham sido eliminadas pelo algoritmo de ajuste. No caso de se utilizar um método em que o modelo ajustado pode ser interpretado, é possível extrair conhecimento sobre a interação entre os fenômenos descritos pelas séries temporais. O modelo proposto nessa dissertação privilegia este aspecto, utilizando a acurácia da previsão do modelo apenas como um “termômetro” da sua validade como uma representação do comportamento das séries estudadas.

2.4 Outros Modelos Multivariados

Além do VAR - e modelos derivados como o VARMA - existem na literatura alguns outros exemplos de modelos de previsão adaptados para o caso multivariados, tanto no âmbito estatístico quanto no de Inteligência Computacional.

Os modelos autoregressivos tradicionais (como o VAR) partem do pressuposto que a variância ou matriz de covariância da distribuição é invariante no tempo, o que pode ser problemático em casos como o de séries com períodos de volatilidade exacerbada. Para lidar com esse tipo de problema, foi desenvolvida

uma família de modelos denominados ARCH (autoregressive conditional heteroskedasticity) [10]. Como os autoregressivos tradicionais, os modelos ARCH também podem ser multivariados, apresentando um desenvolvimento similar ao que ocorre entre um modelo AR e um VAR [8]. Estes modelos substituem a suposição de uma variância constante por uma variância condicional. Um modelo ARCH(q) pode ser representado pela seguinte equação:

$$\sigma_{t|t-1}^2 := \text{Var}(u_t|\Omega_{t-1}) = E(u_t^2|\Omega_{t-1}) \quad (2-51)$$

$$= \gamma_0 + \gamma_1 u_{t-1}^2 + \dots + \gamma_q u_{t-q}^2 \quad (2-52)$$

onde $\Omega_{t-1} = \{u_{t-1}, u_{t-2}, \dots\}$ contém os valores passados de u_t . Como ocorre no caso do VAR, pode-se fazer uso das equações univariadas como base para injetar uma notação vetorial e matricial. Pode-se observar que o ARCH(q) é um modelo AR(q) em σ^2 .

Para definir a versão multivariada do mesmo modelo, supõe-se primeiramente que $u_t = (u_{1,t}, \dots, u_{S,t})'$ é um processo S-dimensional de média zero, serialmente descorrelatado e resíduo de um modelo dinâmico. Pode ser representado como:

$$u_t = \Sigma_{t|t-1}^{1/2} \varepsilon_t \quad (2-53)$$

onde ε_t é um ruído branco S-dimensional $\varepsilon \sim i.i.d.(0, I_S)$, e $\Sigma_{t|t-1}$ é a matriz de covariância condicional de u_t , dado Ω_{t-1} . Com isso, a representação do ARCH(q) multivariado será:

$$\text{vech}(\Sigma_{t|t-1}^{1/2}) = \gamma_0 + \Gamma_1 \text{vech}(u_{t-1} u_{t-1}') + \dots + \Gamma_q \text{vech}(u_{t-q} u_{t-q}') \quad (2-54)$$

onde vech denota o operador que vetoriza uma matriz empilhando suas colunas da diagonal para baixo. Com γ_0 representando um vetor de interceptos e Γ_i sendo matrizes de coeficientes, o MGARCH(q) toma a forma de um modelo VAR(q).

De maneira análoga aos modelos estatísticos, os métodos da Inteligência Computacional utilizados para previsão também podem ser adaptados para o caso multivariado. Um exemplo que é o de uma rede neural MLP (Multi-layer Perceptron), conforme mostrado na figura 2.4.

Da mesma forma que nos demais modelos de previsão, as variáveis de entrada são as defasagens da série temporal estudada, cuja quantidade e seleção são definidas durante o processo de modelagem, sendo o valor previsto para a

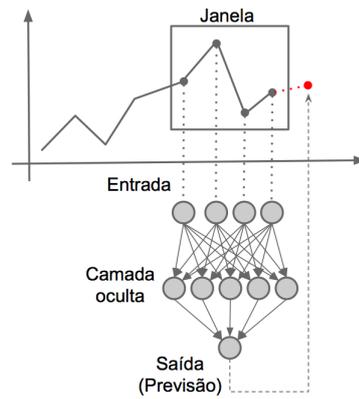


Figura 2.4: Esquema de uma rede MLP para previsão univariada.

série no passo seguinte a variável de saída. Com um conjunto de pares entrada-saída (conjunto de treinamento), faz-se o ajuste da rede, gerando a função de previsão.

A maneira mais direta de se adaptar o sistema acima para o caso multivariado consiste em simplesmente treinar múltiplas redes, onde as defasagens de todas as séries analisadas constituem as entradas, e os valores previstos das mesmas, as saídas. Cada rede contempla uma série temporal diferente na saída (figura 2.5).

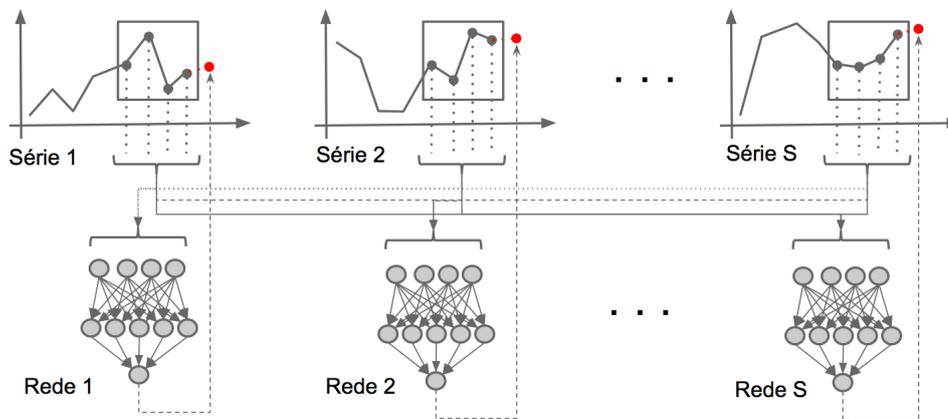


Figura 2.5: Esquema de redes MLP para previsão multivariada.

Este mesmo processo pode ser aplicado a um sistema de inferência fuzzy, ou seja, as defasagens de todas as séries analisadas constituem as variáveis de entrada dos sistemas, que compõem os antecedentes de regra. Ajusta-se um modelo para cada série, ou seja, constrói-se uma base de regras para cada. Esquemas do caso univariado e multivariado estão ilustrados nas figuras 2.6 e 2.7.

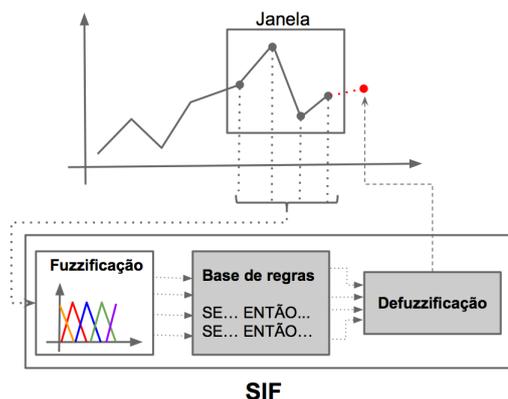


Figura 2.6: Esquema de um SIF para previsão univariada.

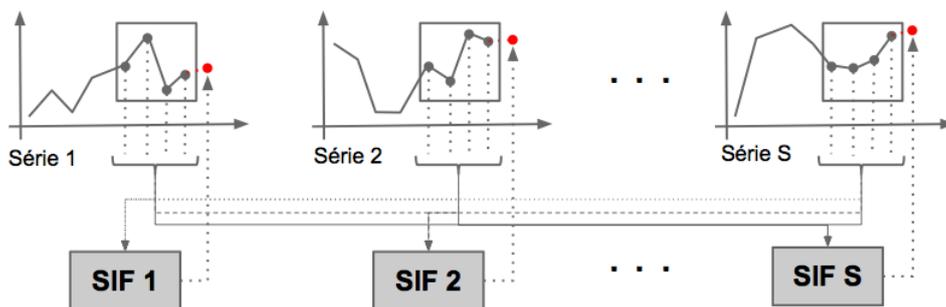


Figura 2.7: Esquema para previsão multivariada.

Pode-se observar que, em um nível superficial, a adaptação de um modelo univariado para o caso multivariado envolve uma abordagem comum de adaptação, que é análoga à transformação de um AR em um VAR. Dentro de cada método utilizado, porém, faz-se necessária a escolha das técnicas que mais bem irão suportar o modelo adaptado para a aplicação desejada.

Este trabalho, como já mencionado concentra-se na construção de um SIF multivariado, com ênfase na interpretação do modelo ajustado, conforme detalhamento a ser apresentado no capítulo 3.

Automatic Multivariate Fuzzy Inference System (AutoMFIS)

Os Sistemas de Inferência Fuzzy (SIFs) são a aplicação da teoria de conjuntos fuzzy mais difundida na indústria e academia, sendo, essencialmente, uma injeção dos conceitos fuzzy aos sistemas especialistas tradicionais, fato que oferece alguns atributos na solução de problemas. O primeiro atributo é que o sistema passa a conseguir representar conceitos linguísticos, inerentemente imprecisos, por meio de conjuntos fuzzy. O segundo atributo é sua natureza como aproximador universal [41], que permite que o sistema seja utilizado para modelar, teoricamente, qualquer problema aproximável por uma função real contínua, linear ou não-linear, características estas compartilhadas por outros métodos da inteligência computacional, como a rede neural MLP (Multi-layer Perceptron).

Como mencionado na introdução, um SIF pode ser construído com dois objetivos:

1. **Acurácia:** a acurácia matemática do modelo, ou seja, o grau de proximidade que o mesmo apresenta em relação ao comportamento do sistema modelado.
2. **Interpretabilidade:** ao invés de visar puramente à acurácia, pode-se desejar que o modelo ajustado seja interpretável, permitindo extração de conhecimento dos dados.

Quando se busca também interpretabilidade, o ajuste do modelo torna-se um processo multi-objetivo, pois é necessário que o seu desempenho, em termos de acurácia, seja razoavelmente próximo aos de outros métodos para que a base de regras gerada de fato descreva o fenômeno modelado. É importante salientar que estes dois objetivos podem ser conflitantes, de forma que, para se alcançar maior interpretabilidade, eventualmente será necessário sacrificar a acurácia.

O modelo AutoMFIS proposto nesse trabalho ajusta um SIF do tipo Mamdani considerando também a interpretabilidade. Ele é aplicado a problemas de previsão de séries temporais múltiplas, de maneira análoga aos modelos apresentados no capítulo anterior. O intuito é capturar o relacionamento entre as séries na forma de regras linguísticas, sendo de especial interesse as relações de natureza não-linear, que geralmente passam despercebidas com o uso de métricas como a correlação.

Nas próximas seções o modelo AutoMFIS será apresentado, iniciando-se pela introdução do funcionamento de um SIF para previsão de séries

temporais e a notação empregada. Em seguida, será abordado em profundidade o tema da modelagem linguística, incluindo a definição do dicionário de termos linguísticos, a geração e o refinamento da base de regras, e as etapas de agregação de regras e de defuzzificação.

3.1 Sistema de Inferência Fuzzy para Previsão

A figura 3.1 mostra, de maneira macroscópica, os componentes de um SIF do tipo Mamdani.

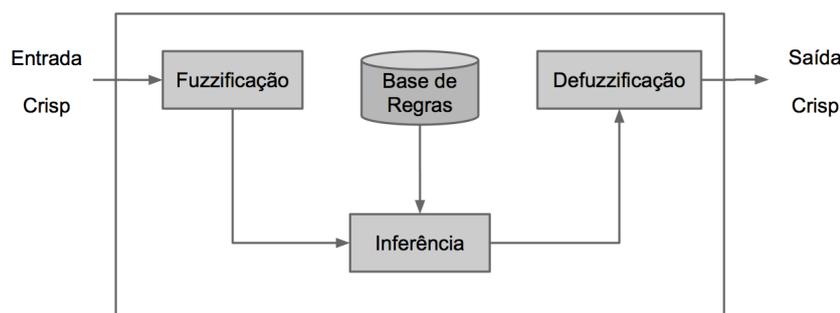


Figura 3.1: Diagrama de um SIF.

Em um SIF baseado em regras, cada variável tem o seu próprio dicionário de termos linguísticos, cada um deles representado por um conjunto fuzzy. Os graus de pertinência μ de valores do universo de discurso da variável aos conjuntos fuzzy são definidos por funções de pertinência. Estas mapeiam tais valores no intervalo $[0, 1]$, permitindo que se verifique a compatibilidade de cada um deles com os conjunto fuzzy, ou seja, com os conceitos expressos pelos termos linguísticos. A figura 3.2 ilustra um exemplo de conjuntos fuzzy, definidos por suas funções de pertinência, para uma variável y qualquer.

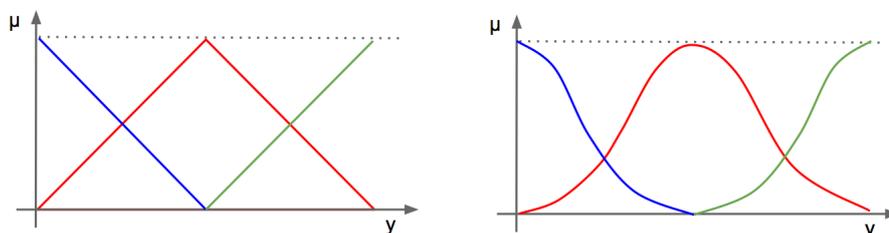


Figura 3.2: Exemplos de conjuntos fuzzy.

Os termos linguísticos, definidos na etapa de fuzzificação, são relacionados por operadores do tipo t-norma (conjunção) e t-conorma (disjunção). É mais comum utilizarem-se apenas t-normas nas premissas das regras. Para um operador ser considerado uma t-norma, devem ser satisfeitas as propriedades

listadas na tabela 3.1. Dos operadores disponíveis na literatura, serão utilizados neste trabalho os listados na tabela 3.2, incluindo os mais comumente utilizados mínimo e produto.

Propriedade	Descrição
Comutatividade	$T(a, b) = T(b, a)$
Monotonicidade	$T(a, b) \leq T(c, d)$ se $a \leq c$ e $b \leq d$
Associatividade	$T(a, T(b, c)) = T(T(a, b), c)$
Número 1 como identidade	$T(a, 1) = a$

Tabela 3.1: Propriedades de uma t-norma.

T-norma	Função
Mínimo	$T(a, b) = \min\{a, b\}$
Produto	$T(a, b) = a \cdot b$
Hamacher	$T(a, b) = \max\{0, a + b - 1\}$
Lukasiewicz	$T(a, b) = \begin{cases} 0, & \text{se } a + b = 0 \\ \frac{ab}{a+b-ab} c.c. & \end{cases}$

Tabela 3.2: T-normas utilizadas.

Ao se comporem, por meio de uma t-norma, os graus de pertinência aos termos de uma premissa, tem-se como resultado um valor denominado grau de ativação (da regra). Este servirá para se manipular o conjunto fuzzy referente ao termo linguístico do conseqüente de regra na etapa de defuzzificação.

Em alguns casos, uma determinada entrada de dados pode ativar diversas regras com o mesmo termo linguístico no conseqüente, tornando necessário, assim, que se efetue uma agregação. Alguns dos métodos de agregação são tão simples quanto utilizar a ativação máxima, mas pode-se empregar também um método de regressão (mínimos quadrados, por exemplo) para definir pesos para cada regra antes de agregá-las. Finalmente, na etapa de defuzzificação é realizada uma interpretação da informação resultante do processo de inferência, produzindo como resultado um valor preciso.

O uso de um SIF para previsão de séries temporais utiliza a mesma estrutura básica descrita acima, com peculiaridades quanto às variáveis de entrada e saída, além de sua rotina de funcionamento. Como descrito no capítulo 2, em um sistema de previsão, as variáveis de entrada são as defasagens da série a ser prevista e de outras séries relacionadas, enquanto que a saída é o valor previsto (figura 3.3).

No caso de previsão de uma única série temporal, as variáveis de entrada são definidas em um mesmo universo e os conjuntos fuzzy (funções de pertinência) são idênticos para todas elas. Designando x como sendo uma

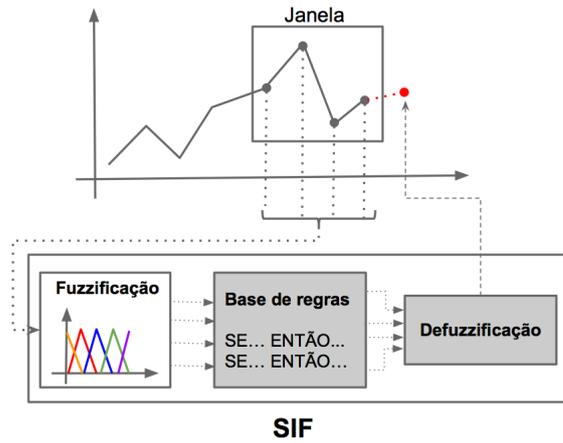


Figura 3.3: Esquema de um SIF para previsão univariada.

das variáveis de entrada (defasagens da série), $\mu_n(x)$ é o seu grau de pertinência ao conjunto fuzzy $n \in \{1, \dots, N\}$.

No caso multivariado, para S séries geram-se S SIFs, um para cada série considerada, porém com as defasagens de todas as séries sendo consideradas na entrada de cada SIF (figura 3.4). Neste caso, o grau de pertinência ao conjunto n da série $s \in \{1, \dots, S\}$ é denotado por $\mu_{s,n}(x_s)$.

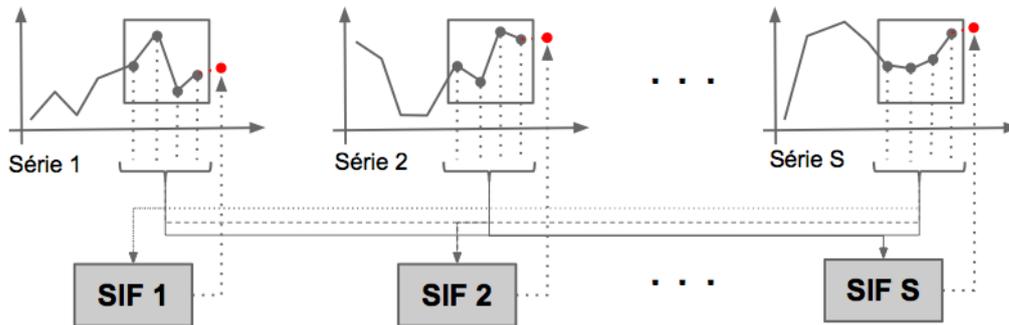


Figura 3.4: Esquema para previsão multivariada.

O processo de multistep funciona da mesma forma que em outros métodos: utiliza-se o valor previsto como entrada da função para prever os passos seguintes (figura 3.5). Como cada SIF utiliza todas as séries como entradas, deve-se realizar o passo de previsão para o instante t com todos os SIFs antes de efetuá-lo para o instante subseqüente $t + 1$.

Nas próximas seções, será formulado o modelo AutoMFIS, baseado nessa estrutura básica de um SIF para previsão multivariada.

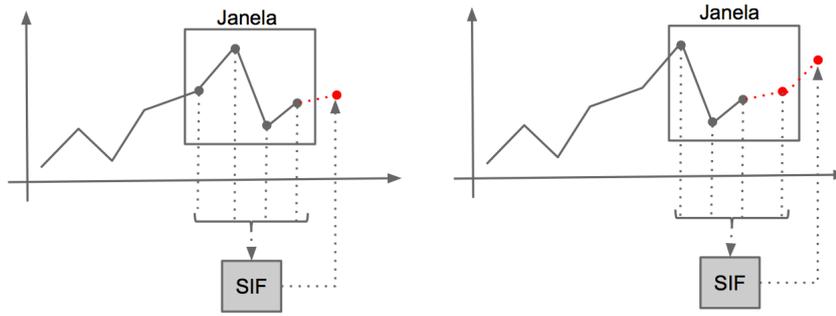


Figura 3.5: Multistep para um SIF.

3.2 Modelagem para Extração de Conhecimento

Como mencionado brevemente na introdução, existem duas variedades principais de SIFs: Mamdani e Takagi-Sugeno (TSK). Neste trabalho, optou-se pelo tipo Mamdani, que, ao contrário do modelo TSK, apresenta variáveis linguísticas tanto nos antecedentes como nos consequentes das regras, fazendo com que a interpretação seja mais intuitiva.

A interpretabilidade pode ser prejudicada no caso de um problema complexo como o de previsão multivariada, em que um número de variáveis envolvidas pode ser muito grande. De modo a garantir que o sistema ajustado seja interpretável, é necessário investigar medidas de combate à complexidade em todos os aspectos do SIF. Por outro lado, essas medidas podem gerar um sistema excessivamente simplificado, que não represente toda a informação contida nos fenômenos modelados. Essa dicotomia entre clareza e complexidade do sistema será levada em conta na construção do AutoMFIS.

Gacto et al.[42] sugere uma taxonomia de avaliação de interpretabilidade baseada nos aspectos (conflitantes) complexidade e interpretabilidade semântica, e levando em conta os elementos principais do SIF: o dicionário de termos linguísticos (variáveis e seus conjuntos fuzzy) e a base de regras (tabela 3.3).

	Base de Regras	Conjuntos Fuzzy
Complexidade	Núm. de regras e condições	Núm. de variáveis e conjuntos fuzzy
Semântica	Consistência e ativação simultânea	Cobertura, discernibilidade e complementaridade

Tabela 3.3: Parâmetros de interpretabilidade.

Um SIF pode ter sua base de regras gerada por especialistas ou extraída da base de dados. Como o AutoMFIS se destina à extração de conhecimento, este trabalho concentra-se no segundo caso. As regras representarão as relações

entre as séries temporais modeladas, que é exatamente a informação que se deseja obter. A figura 3.6 ilustra, em alto nível, o processo genérico de ajuste de um SIF a partir de dados.

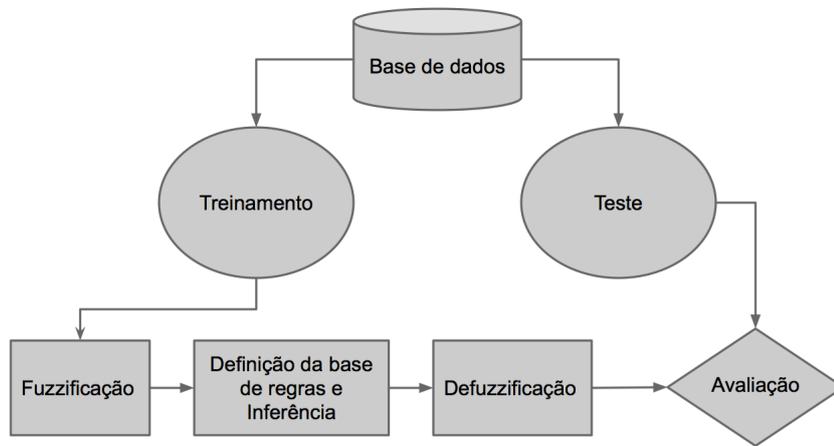


Figura 3.6: Ajuste de um SIF a partir de dados.

A base de dados – séries históricas disponíveis – é segmentada em dois conjuntos: treinamento e teste. Este é um conjunto de saídas esperadas formado pelos últimos H pontos da amostra das séries, onde H é o horizonte de previsão para o qual se quer ajustar o modelo. O conjunto de treinamento é construído pelos $T - H$ pontos restantes das séries, e toma a forma de uma série de entradas e saídas esperadas de dimensão $T - H - L$, onde os L primeiros pontos da amostra são utilizados para construir a primeira entrada de dados (tabela 3.4). Para o caso multivariado com S séries, cada par entrada-saída tem $S \cdot L$ entradas e S saídas, mantendo-se a mesma lógica da tabela.

Entradas				Saídas
$y_{1,1}$	$y_{1,2}$	\dots	$y_{1,L}$	$y_{1,L+1}$
$y_{1,2}$	$y_{1,3}$	\dots	$y_{1,L+1}$	$y_{1,L+2}$
		\vdots		\vdots
$y_{1,T-H-L}$	$y_{1,T-H-L+1}$	\dots	$y_{1,T-H-1}$	$y_{1,T-H}$

Tabela 3.4: Exemplo de conjunto de treinamento univariado

O conjunto de treinamento é utilizado no ajuste do modelo. O conjunto de teste serve para avaliar a resposta do modelo ajustado, segundo alguma métrica de erro.

Nas próximas seções será detalhado o modelo AutoMFIS, levando em consideração os pontos listados na tabela 3.3.

3.3

Modelo AutoMFIS

A figura 3.7 ilustra cada uma das etapas do ajuste do modelo AutoMFIS a partir de dados.

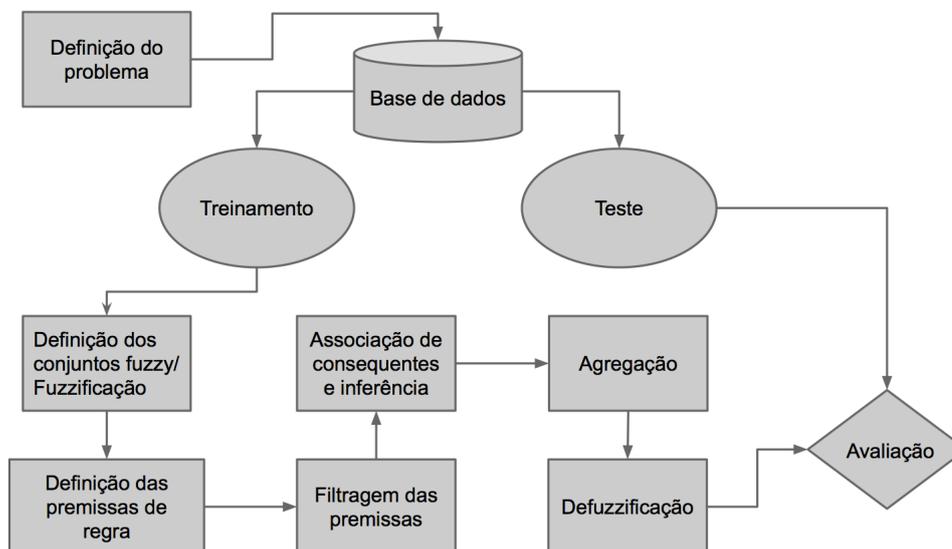


Figura 3.7: Resumo: Funcionamento do modelo AutoMFIS.

Essas etapas serão apresentadas neste trabalho divididas três em grupos, com cada um dos desses sendo dotado de significado análogo a um dos passos do processo genérico de ajuste de um SIF ilustrado na figura 3.6 (fuzzificação, geração da base de regras e defuzzificação). Esses grupos são:

1. **Definição do Dicionário Linguístico:** envolve definição das variáveis do problema e a quantidade e forma de seus conjuntos fuzzy.
2. **Construção da Base de Regras:** construção e otimização da base de regras a partir do dicionário linguístico.
3. **Geração dos Valores Crisp de Saída:** agregação das ativações dos consequentes e, em sequência, sua defuzzificação.

A segmentação das etapas dentro desses três grupos está ilustrada na figura 3.8. Nas próximas seções serão apresentados cada um dos grupos de etapas, finalizando com um resumo do modelo completo.

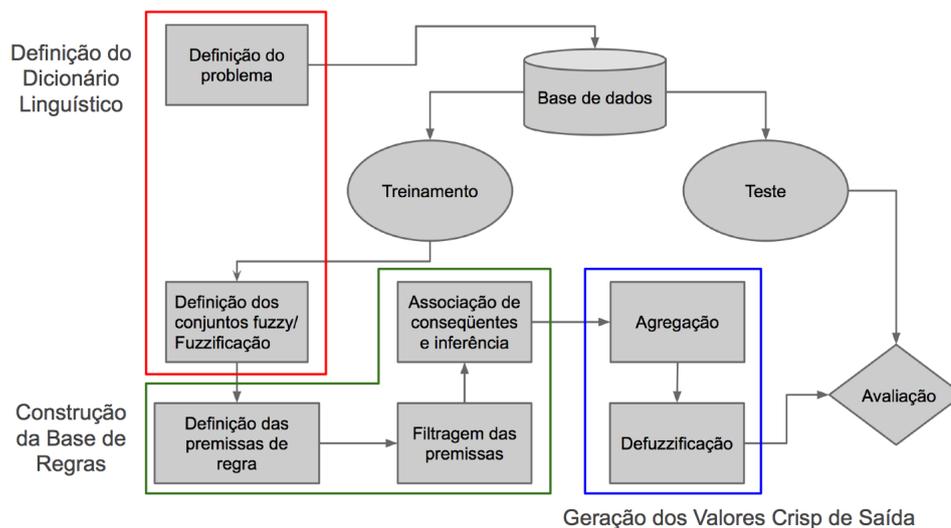


Figura 3.8: Resumo: Funcionamento do modelo AutoMFIS em 3 Etapas.

3.3.1 Definição do Dicionário Linguístico

O primeiro aspecto na modelagem do AutoMFIS é a definição do dicionário de termos linguísticos, tendo sempre em vista a questão da complexidade. Como listado na seção anterior, os pontos relevantes neste aspecto são:

1. **Número de variáveis:** a quantidade de variáveis consideradas aumenta diretamente o volume de informação tratado pelas regras e, conseqüentemente, o volume de informação a ser processado na interpretação o sistema ajustado.
2. **Número de termos linguísticos (conjuntos fuzzy):** similarmente, a quantidade de termos linguísticos, ou conjuntos fuzzy, define o quão extenso é o dicionário, influenciando diretamente na complexidade do sistema.

Em um problema de previsão de séries temporais, cada variável considerada pelo modelo não terá um significado conceitual próprio, já que as variáveis consideradas se referem à uma mesma série (ou conjunto de séries) em momentos distintos. Isso permite que se considere um número de variáveis maior do que o aceitável em outras aplicações. Todavia, por mais que a compreensibilidade frente a um número elevado de variáveis seja menos difícil neste caso, é de interesse que essa quantidade seja o mais limitada possível. Para tal, faz-se uso de uma seleção de variáveis. Na literatura é possível encontrar diversas metodologias, envolvendo algoritmos genéticos [43, 44] ou ordenação das variáveis segundo sua importância [45], em adição a métodos específicos para problemas

mais comumente abordados, como os de classificação [46, 47, 48]. No caso do AutoMFIS, a seleção de variáveis é mais delicada, pois a escolha das séries e defasagens a serem modeladas é alinhada ao próprio objetivo do modelo, que é a compreensão das relações entre séries. Eventuais remoções de variáveis podem significar uma oportunidade perdida de extração de conhecimento.

O primeiro aspecto da seleção de variáveis a ser tratado diz respeito à definição das séries temporais que irão compor a série temporal multivariada modelada. Tornar esse conjunto de séries o mais enxuto possível oferece ganhos em termos de complexidade e evita, potencialmente, que ruído seja injetado no sistema por séries não relacionadas – fato que pode ser verificado experimentalmente. Em alguns casos a escolha pode ser clara, como na modelagem da dinâmica populacional de um predador e sua presa. Por outro lado, pode-se ter em mãos um problema formado por uma elevada quantidade de séries, sem conhecimento prévio sobre quais delas se relacionam, a exemplo das séries de preços de ações que formam um portfólio de investimentos. Como, em um caso como este, pode-se ter dezenas de séries, é essencial que haja uma segmentação prévia. Isso é realizado por meio da metodologia de agrupamento descrita na seção 2.2.2. O agrupamento, porém, depende da definição de um ponto de corte de similaridade para definição dos grupos. Este ajuste é guiado pela visualização dos clusters de séries de maneira gráfica, e limitado pelo poder de processamento disponível, ou seja, quanto menores os recursos, menores os grupos deverão ser.

Uma outra forma de limitação da quantidade de variáveis com as quais o SIF terá que lidar é a escolha da quantidade de defasagens, embasada na análise das características das séries. Típicamente, o uso das correlações-cruzadas e auto-correlações (como descrito na seção 2.2.1) oferecem uma boa base para esta escolha, embora elas descrevam apenas a relação linear entre cada série e as defasagens, próprias e das outras séries relacionadas. De modo a contemplar a possibilidade de uma influência não-linear, as defasagens definidas por relações lineares serão apenas o ponto de partida, com sua quantidade sendo modificada iterativamente durante o processo de ajuste do SIF, ao mesmo tempo em que se analisam as alterações de desempenho do modelo. A quantidade de defasagens final será aquela que oferecer a melhor acurácia.

A quantidade de termos linguísticos (conjuntos fuzzy) também tem impacto direto na interpretabilidade. Quanto mais extenso o dicionário, mais difícil será visualizar a distinção entre cada termo. Uma diretriz que será seguida nesse trabalho é o do “número mágico” estabelecido por Miller [49], que indica como sendo 7 ± 2 a quantidade de informação que o ser humano

consegue processar simultaneamente. Seguindo essa diretriz, a quantidade de conjuntos fuzzy será de 5 a 9 por variável.

Quanto ao aspecto de interpretabilidade semântica do dicionário, alguns pontos devem ser contemplados na definição dos conjuntos fuzzy:

1. **Completeude ou cobertura:** o universo de discurso da variável deve ser coberto integralmente por conjuntos fuzzy, ou seja, todo valor do universo deve ter grau de pertinência diferente de zero a pelo menos um conjunto fuzzy.
2. **Discernibilidade:** os termos linguísticos representados pelos conjuntos fuzzy devem ser claros e facilmente distinguíveis um dos outros.
3. **Complementaridade:** para cada valor do universo de discurso de uma variável, a soma dos seus graus de pertinência deve ser próxima de um, de modo a garantir que todos os pontos do domínio possam ser descritos em sua totalidade pelos termos linguísticos.

A questão da completeude inicia-se na própria estimativa do universo de discurso da variável. Quando a faixa de valores possíveis não é clara, ela deve ser definida com base nos dados disponíveis, mas esta abordagem pode se mostrar ineficiente em alguns casos. Um exemplo de série com esse problema está ilustrado na figura 3.9. Devido a sua tendência, o domínio definido pela amostra que compõe o conjunto de treinamento não contém a faixa de valores de fora da amostra, fazendo com que estes não possam ser atingidos na previsão. Para um caso como esse, pode-se remover a tendência via regressão ou considerar, na modelagem, as diferenças das séries em questão.

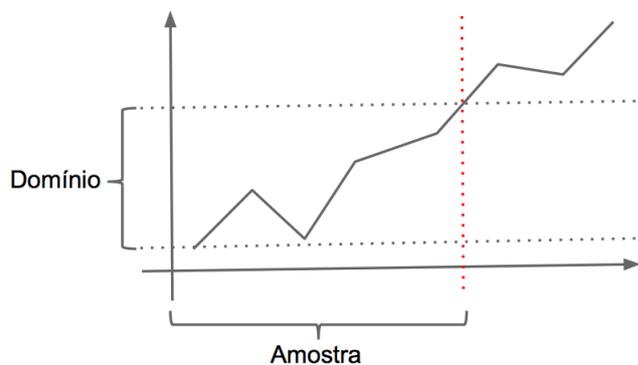


Figura 3.9: Série com tendência.

Mesmo para séries estacionárias há a possibilidade de se extrapolar o domínio definido no in-sample (figura 3.10). Supondo ignorância completa

a respeito da variável em questão, pode-se fazer uso de uma margem de segurança, como, por exemplo, uma folga de 25% em relação aos valores máximo e mínimo (figura 3.11). Esta folga afetará a construção da base de regras, pois “enfraquece” a ativação dos conjuntos fuzzy extremos no treinamento, tornando-os menos relevantes. Caso esse artifício seja utilizado, o tamanho da folga deve ser ajustado em conjunto com os outros parâmetros do modelo de modo a garantir um bom funcionamento do SIF.

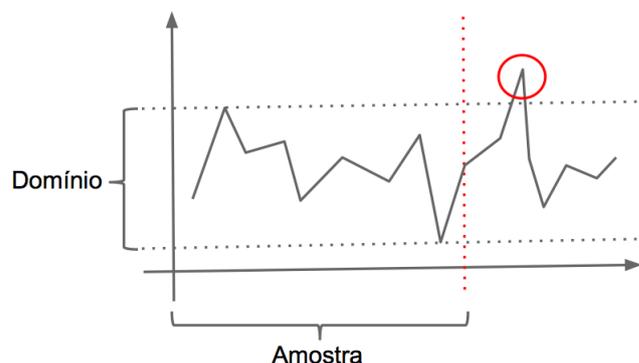


Figura 3.10: Série estacionária com out-of-sample extrapolando o domínio das funções de pertinência.

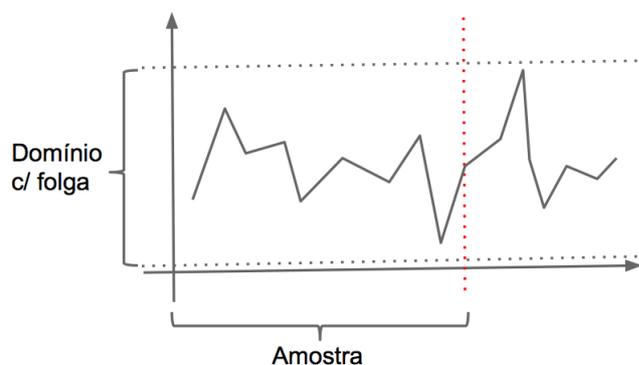


Figura 3.11: Folga na definição do domínio.

Em adição, pode-se ter conhecimento prévio dos limites das variáveis, sejam eles intrínsecos da grandeza representada (a população de uma espécie nunca é negativa) ou dentro do que se considera razoável (a temperatura média no Rio de Janeiro não passaria dos 50°C).

Uma outra questão importante é a disposição das funções de pertinência no universo. Quando um valor do domínio apresenta grau de pertinência não nulo a dois termos linguísticos, a interpretação fica possível, visualizando-se o valor como sendo compatível com os dois conceitos. Em contrapartida, se a sobreposição for mais complexa, esta avaliação é menos clara (figura 3.12).

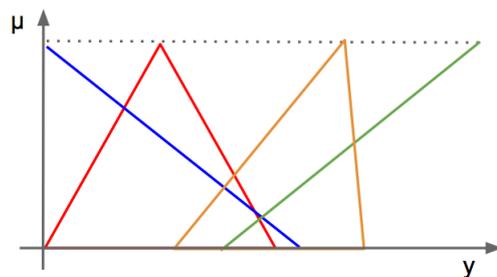


Figura 3.12: Dicionário de baixa interpretabilidade.

A estrutura ideal para se obter funções de pertinência normais, discerníveis e complementares é a da partição fuzzy forte, que é definida como aquela para a qual a soma dos graus de pertinência é sempre 1.0 para todos os valores do universo. A forma mais fácil de obter essa estrutura é fazer uso de funções de pertinência triangulares, com sobreposição de no máximo dois conjuntos fuzzy. Essa estrutura, por construção, garante a complementaridade e, em decorrência da transição linear entre conjuntos adjacentes, auxilia a discernibilidade.

O ajuste da quantidade e posição dos conjuntos fuzzy tem sido bastante estudada, como, por exemplo, na inclusão iterativa de conjuntos fuzzy nas faixas mais problemáticas do domínio [50] ou no ajuste por algoritmos genéticos [51]. Este tipo de ajuste numérico é a abordagem ideal para a obtenção de maior acurácia, porém será empregada neste trabalho uma simplificação para reduzir os custos computacionais.

A primeira forma de definição do posicionamento dos conjuntos fuzzy, mais trivial, é o espaçamento uniforme dos mesmos, com os centros dos conjuntos extremos posicionados nos limites inferior e superior do universo (figura 3.13).

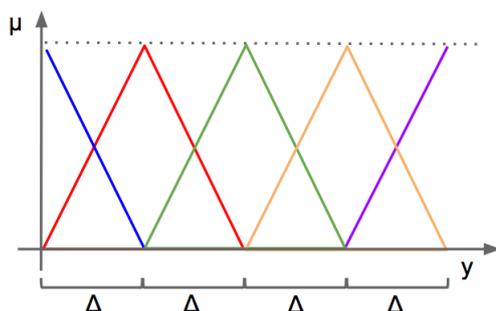


Figura 3.13: Segmentação uniforme do domínio.

No caso de séries cujos valores não estão bem distribuídos no domínio, pode-se espaçar os centros dos conjuntos igualmente em termos dos percentis

da distribuição de seus valores, fazendo com que conjuntos cobrindo áreas menos comuns tenham maior dimensão, equilibrando a ocorrência dos termos linguísticos associados, como observado na figura 3.14.

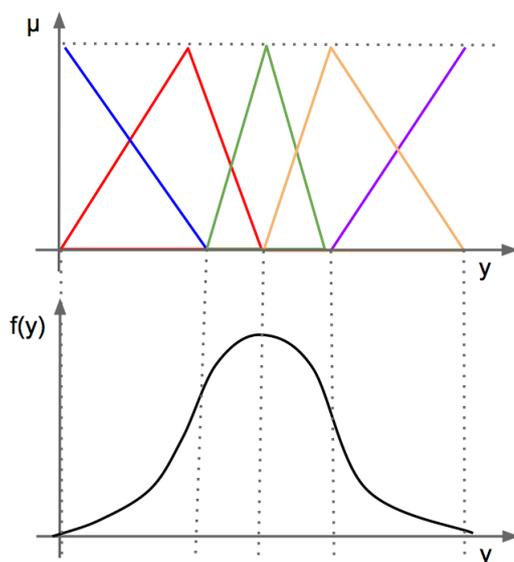


Figura 3.14: Segmentação pelos percentis da distribuição de valores.

3.3.2 Geração da Base de Regras

Definidos os termos linguísticos, o passo seguinte é a construção da base de regras. Esse passo é crítico, pois as regras serão o cerne da interpretação do modelo. Esse processo é realizado segundo as etapas ilustrados na figura 3.15.

De acordo com o diagrama, consideram-se duas etapas principais no processo de geração da base de regras a partir de dados: indução e otimização.

O passo de indução envolve a utilização do dicionário de termos linguísticos construído na etapa de fuzzificação, que são combinados para dar origem às regras. Nesta etapa estão incluídos o possível ajuste do dicionário e a definição dos consequentes mais apropriados a partir dos dados. Como exemplos de indução tem-se a geração de todas as regras possíveis [52], regras específicas para cada par entrada-saída da base de dados [53] e uso árvores de decisão [54, 55].

No passo de otimização, controla-se a complexidade da base de regras segundo alguma métrica pré-definida. Esta otimização inclui métodos de seleção de variáveis [43, 46, 48], não abordados aqui por motivos já descritos na seção anterior, e redução de regras, como por métodos estatísticos [56, 57].

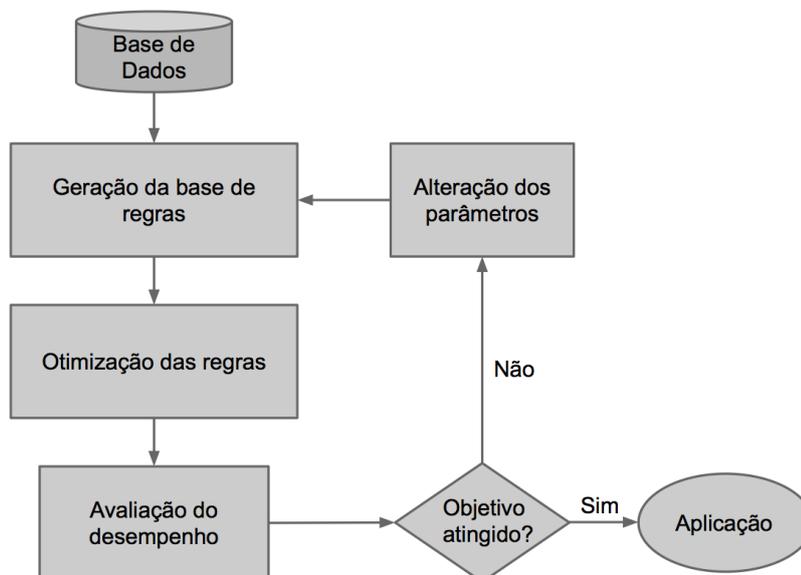


Figura 3.15: Geração orientada a dados.

No caso do AutoMFIS, os passos de indução e otimização ocorrem de maneira entrelaçada (figura 3.16). As premissas de regra são geradas pela combinação dos termos linguísticos por t-normas. As ativações das premissas são utilizadas para eliminar as que não forem consideradas importantes segundo algum critério. Os consequentes serão então associados às premissas por meio da comparação das ativações destas com os graus de pertinência aos possíveis consequentes. Na figura, é utilizada uma notação que será repetida ao longo do resto do capítulo, onde X designa o vetor de entrada de dados, e Y o de saída.

Como complemento ao processo descrito anteriormente, utilizar-se-á o método de Wang & Mendel para aprimorar a base de regras por meio da injeção de regras completas, mais específicas.

Da mesma forma que na definição do dicionário, a complexidade da base de regras deve ser controlada, durante seu processo de geração. Voltando à taxonomia de Gacto, os pontos referentes à complexidade da base de regras são:

1. **Número de regras:** especialmente no que tange à interpretabilidade, o melhor modelo é aquele mais simples, contendo menos regras. Um dos objetivos do ajuste do modelo é manter um bom desempenho de previsão com uma base de regras o mais enxuta possível.
2. **Número de elementos nas premissas:** uma base de regras não é complexa somente pela sua própria dimensão, mas também pela

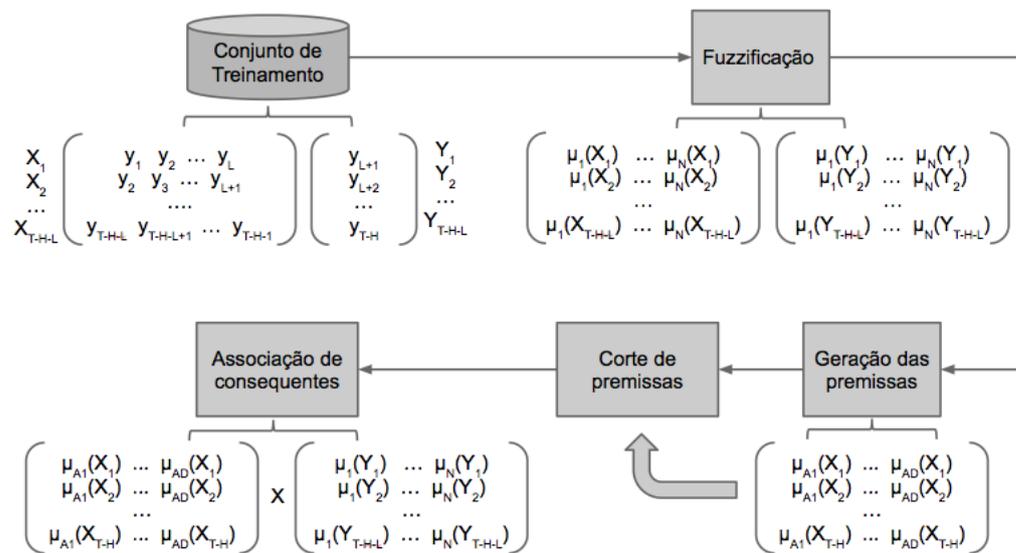


Figura 3.16: Geração de regras do AutoMFIS.

dimensão das regras que a compõem. Quanto mais termos linguísticos figurarem em uma premissa, mais difícil será a interpretação da regra.

A quantidade de regras geradas é proporcional à dimensão do dicionário de termos linguísticos, o que faz com que as medidas utilizadas para reduzir a complexidade do dicionário impactem diretamente na complexidade da base de regras. Em adição, o passo de otimização da base de regras também atuará diretamente nesse ponto.

Da mesma forma da quantidade de conjuntos fuzzy por variável, a quantidade de elementos nas premissas obedecerá o critério de Miller, ou seja, as premissas ficam limitadas ao tamanho 9, limite superior de compreensão. Com isto, permite-se que sejam geradas regras mais ricas e específicas, que são boas fontes de informação, sem perda de capacidade de compreensão. Essa limitação restringe também o uso do método de Wang & Mendel para os problemas contendo no máximo 9 variáveis, dado que sempre gera regras completas.

Em adição às questões de complexidade, há que considerar os seguintes pontos no que tange à interpretabilidade semântica:

1. **Consistência:** para uma base de regras ser consistente, regras com premissas similares devem ter consequentes similares.
2. **Ativação simultânea de regras:** quando um grande número de regras é ativado simultaneamente para uma mesma entrada de dados, o significado individual das regras não se preserva completamente.

A consistência da base de regras dependerá do funcionamento correto do passo de associação de consequentes. Isto será apresentado mais adiante.

A questão da ativação simultânea é complexa, e será tratada apenas indiretamente, por meio da redução da dimensão da base de regras. A idéia é que uma base de regras mais enxuta terá menos regras redundantes, diminuindo a probabilidade de muitas regras serem ativadas simultaneamente.

As próximas sub-seções detalham os processos de geração da base de regras.

Geração semi-exaustiva

A geração semi-exaustiva da base de regras é um processo iterativo baseado no princípio de indução exaustiva combinado a um passo de eliminação de regras via critérios de corte. Esta metodologia será a fonte principal de regras do AutoMFIS e se divide em duas etapas: geração das premissas e definição dos consequentes (associação).

O processo de geração das premissas é realizado de maneira iterativa, com cada ciclo de geração tratando de regras de um mesmo tamanho, ou seja, contendo a mesma quantidade de termos linguísticos na premissa, iniciando-se pelas regras mais simples, de tamanho 1, até chegar às de tamanho 9 (novamente segundo o critério de Miller). Para a geração das regras de tamanho n , compõem-se todas as regras de tamanho $n - 1$ com todo o dicionário de termos linguísticos por meio da t -norma escolhida. Ao final de um ciclo de geração, verifica-se o desempenho de cada premissa segundo alguma métrica. Todas as regras que estiverem abaixo do limiar estabelecido serão removidas da base de regras e, conseqüentemente, não serão consideradas no próximo ciclo de geração.

Partindo de um dicionário contendo os termos linguísticos a_1 , a_2 e a_3 , o diagrama representado na Figura 3.17 mostra o três ciclos de geração de regras, partindo das regra de tamanho 1 formadas pelos termos isolados.

A partir do segundo ciclo, as premissas geradas no ciclo anterior são combinadas com todos os termos do dicionário, gerando novas premissas, com tamanho $n + 1$. Como visto na figura, as regras duplicadas são eliminadas a cada ciclo, efetivamente impedindo o crescimento do galho. Em adição às deduplicadas, premissas também podem ser eliminadas segundo critérios de corte, novamente impedindo a geração de novas premissas a partir da mesma. Vale salientar que só são combinados termos linguísticos pertencentes a variáveis diferentes; caso contrário, podem ser geradas regras inconsistentes (e.g. se a temperatura está alta e a temperatura está baixa).

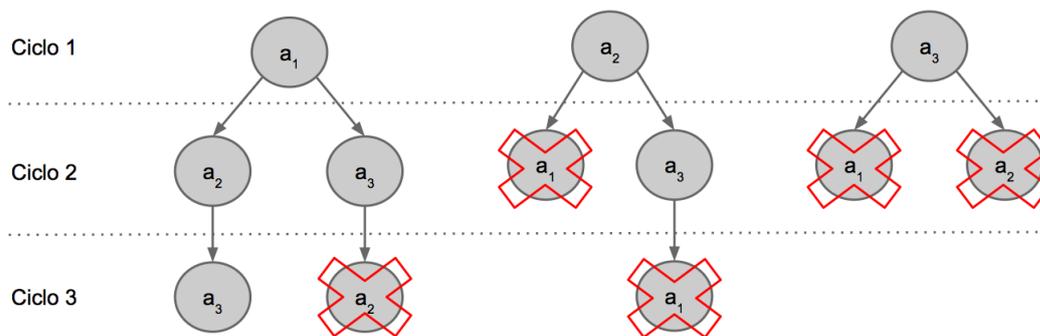


Figura 3.17: Exemplo de geração iterativa.

Como apresentado anteriormente na seção 3.2, o conjunto de treinamento será formado por $T - H - L$ amostras, cada uma representando um instante de tempo. Em cada instante de tempo, as premissas serão ativadas, em diferentes graus, pela amostra do conjunto de treinamento. Os graus de ativação são usados como parâmetros para os critérios de corte, regidos pelas seguintes métricas:

1. **Frequência:** frequência com que a premissa tem ativação acima de zero.
2. **Cardinalidade:** ativação média da premissa.
3. **Ativação:** híbrido das métricas 1 e 2, utiliza a ativação média das amostras com ativação tiver sido diferente de zero.

As métricas de frequência e cardinalidade funcionam com base no conceito de que a regra deve ter alguma representatividade no conjunto de treinamento para ser considerada válida. Esta abordagem, porém, pode apresentar algumas falhas, o que motivou a criação da terceira métrica.

Como abordado na seção sobre fuzzificação, os dados do conjunto de treinamento podem não estar uniformemente distribuídos nos domínios das variáveis. Caso esta questão não tenha sido tratado durante a definição do dicionário, os termos linguísticos cobrindo as regiões mais “raras” são pouco ativados no conjunto de treinamento, o que afeta as premissas que os contêm. Em adição, mesmo que, individualmente, os termos linguísticos tenham uma ativação equilibrada, algumas combinações deles podem ser raras, porém relevantes. Nestes dois casos, as métricas de frequência e cardinalidade podem ocasionar a eliminação indevida de regras. Outra falha possível reside no fato de que essas métricas permitem a sobrevivência de uma premissa com ativação frequente porém com grau baixo, o que pode indicar que ela é excessivamente genérica e que não descreve com precisão comportamento algum das séries.

Ao realizar o corte pela média das ativações não-nulas, a terceira métrica permite a sobrevivência de regras de baixa ocorrência mas que descrevem com alta precisão alguma situação específica, assim como eliminar regras frequentes porém com baixa ativação. Como cada problema e base de dados têm suas próprias características, as três métricas serão avaliadas durante os experimentos.

Geradas todas as premissas de regra, deve-se encontrar o conseqüente mais apropriado para cada uma delas. Cada premissa gerará S regras, uma para cada série temporal, deve-se encontrar o conjunto fuzzy de cada série que mais bem se relaciona com cada premissa. Esta etapa, denominada associação, será realizada de acordo com métodos disponíveis na literatura, em adição a alguns sugeridos em [58]. Esse processo pode ser sumarizado nos três passos seguintes:

1. Calculam-se as ativações de cada premissa e cada conseqüente possível para todas as amostras do conjunto de treinamento.
2. Aplica-se a métrica de associação para todos os pares premissa-conseqüente.
3. Para cada série, encontra-se o conjunto fuzzy que apresenta o maior valor segundo a métrica escolhida.

Sendo A_d a d -ésima premissa gerada, sua ativação é definida como $\mu_{A_d}(X)$, onde X é um vetor de entrada de dados. Para cada amostra do conjunto de treinamento – representada por um par entrada-saída (X, Y) , com $Y = \{y_1, \dots, y_S\}$ –, μ_{A_d} assume um determinado valor. Similarmente, têm-se para estas amostras as ativações $\mu_{s,n}(y_s)$, referentes a cada conjunto fuzzy n de cada uma das S séries temporais. Ao se calcularem essas ativações para todos as T amostras, obtém-se, para cada um destes elementos, um vetor de ativações de dimensão T . As métricas de associação a ser apresentadas permitem comparar o vetor de ativação das premissas com os vetores de ativação de cada conseqüente candidato, de modo a se encontrar o que apresenta maior compatibilidade. Para cada um dos pares premissa-conseqüente candidato, calcula-se o grau de compatibilidade e define-se como o conseqüente aquele que apresenta grau máximo.

Tanto $\mu_{A_d}(X)$ quanto $\mu_{s,n}(y_s)$ representam vetores no hipercubo $[0, 1]^{(T-H-L)}$ (figura 3.18). Desta forma, pode-se aplicar métricas de similaridade de vetores comprovadas, disponíveis na literatura, como as métricas de compatibilidade para definição dos conseqüentes.

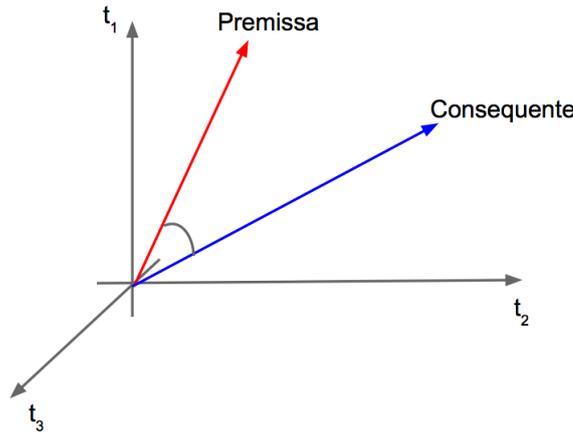


Figura 3.18: Representação dos vetores da premissa e consequente no hiper-cubo $[0, 1]^3$.

A métrica de similaridade mais difundida, e comumente utilizada em problemas de mineração de dados e agrupamento [59, 60], é o cosseno de similaridade, denominado Grau de Confiança Fuzzy:

$$FCD_{s,n} = \frac{\langle \mu_{A_d}(X), \mu_{s,n}(y_s) \rangle}{(\|\mu_{A_d}(X)\|_2 \|\mu_{s,n}(y_s)\|_2)} \quad (3-1)$$

onde $\langle ., . \rangle$ refere-se ao produto interno e $\|\cdot\|_2$ a norma euclidiana. Caso a premissa e o consequente sejam ortogonais, a compatibilidade será zero; se forem linearmente dependentes, a métrica atinge valor máximo 1.

Alternativamente, uma outra métrica comumente utilizada é a similaridade de Jaccard generalizada, representada pela razão dos mínimos pelos máximos entre os dois vetores em cada posição (equação 3-2).

$$GJS_{s,n} = \sum_{t=1}^T \frac{\min(\mu_{A_d}(X_t), \mu_{s,n}(y_{s,t}))}{\max(\mu_{A_d}(X_t), \mu_{s,n}(y_{s,t}))} \quad (3-2)$$

onde X_t é a entrada da amostra t e $y_{s,t}$ é sua saída esperada referente à série $s \in \{1, \dots, S\}$.

É possível construir métricas de similaridade a partir de métricas de distância comumente utilizadas, como a distância L_1 (também conhecida como Taxicab ou Manhattan). De modo a obedecer as propriedades de uma métrica de similaridade, deve-se realizar algumas manipulações na distância, nominalmente normalizá-la e utilizar o seu complemento (equação 3-3).

$$SD_{s,n} = 1 - \frac{\sum_{t=1}^T (\|\mu_{A_d}(X_t) - \mu_{s,n}(y_{s,t})\|_1)}{T} \quad (3-3)$$

A métrica WM avalia apenas a frequência de ativação simultânea entre premissa e consequente, independentemente do grau de ativação das partes, sendo basicamente uma contagem da quantidade de vezes em que a premissa e o consequente candidato figuram com ativações diferentes de zero simultaneamente em uma mesma amostra (equação 3-4).

$$WM = \sum_{t=1}^T (\sigma(t)) \quad (3-4)$$

$$\sigma(t) = \begin{cases} 1, & \text{se } \mu_{A_d}(X_t) \cdot \mu_{s,n}(y_{s,t}) > 0. \\ 0, & \text{caso contrário.} \end{cases}$$

Além do uso de métricas de similaridade de vetores, pode-se utilizar a Máxima Credibilidade. Nesse método, estimam-se pesos para cada consequente candidato por Mínimos Quadrados Restritos e resolve-se o seguinte problema de otimização:

$$\begin{aligned} \min : \quad & |||U(\mu_{A_d})\beta - \Omega(\mu_n)|||_2^2 \\ \text{s.a. : } & \beta \mathbf{1} = 1 \\ & \beta \geq 0 \end{aligned} \quad (3-5)$$

onde $U(\mu_{A_d}) = [\mu_{A_1}(X), \mu_{A_2}(X), \dots, \mu_{A_D}(X)]$, ou seja, a matriz cujas colunas se referem às ativações de cada regra A_d para as amostras do conjunto de treinamento. Ainda na função de objetivo, o termo $\Omega(\mu_n) = [\mu_1(Y), \mu_2(Y), \dots, \mu_N(Y)]$ é a matriz que contém as ativações dos consequentes, novamente dispostas nas colunas. β é a matriz de pesos $[b_{d1} \ b_{d2} \ \dots \ b_{dN}]$ a ser encontrada, com D linhas e N colunas, ou seja, cada linha se refere a uma premissa, enquanto as colunas contêm os pesos de cada consequente candidato. Finalmente, o operador $|||\cdot|||_2$ utilizado na função objetivo se refere a norma de Frobenius.

As restrições impostas à função a ser minimizada são: os pesos devem ser sempre maiores ou iguais a zero e a soma dos pesos de todas as regras associadas a um determinado consequente candidato deve ser igual a 1. O vetor N -dimensional $\mathbf{1} = [1, 1, \dots, 1]^T$ é apenas um operador utilizado para representar a segunda restrição.

Para calcular os pesos contidos na matriz beta, é necessário resolver o problema de minimização (convexo com restrições lineares) acima, que possui uma solução ótima única [61]. Contudo, em virtude da restrição de desigualdade $beta \geq 0$, este problema não tem solução analítica, devendo,

portanto, ser resolvido por um método numérico. Exista uma gama de métodos para tal, como, por exemplo, os métodos de ponto interior ou de gradiente. No AutoMFIS será utilizado o método do conjunto ativo [62], com representação do problema por multiplicadores de Lagrange, como deduzido [58].

Cada par premissa-consequente terá um peso b_{an} , que indica a “credibilidade” da premissa na explicação o consequente. Essa maneira é distinta, conceitualmente, da abordagem via métrica de similaridade pois, enquanto esta busca o melhor consequente para cada premissa de maneira isolada, o método da máxima credibilidade contextualiza a credibilidade do par premissa-consequente em relação aos demais, fazendo com que o consequente escolhido seja aquele mais compatível com a premissa.

No caso de problemas mais simples, pode-se utilizar a geração de regras pelo método de Wang & Mendel para complementar a base de regras, como detalhado na próxima sub-seção.

Wang & Mendel

Uma das formas de geração de regras mais difundidas é o método de Wang & Mendel [53]. Essa metodologia foi elaborada com o intuito de complementar uma base de regras construída via conhecimento de especialistas, costumeiramente formada por regras incompletas mais genéricas, com regras completas mais específicas. Em problemas de baixa dimensão, com menos de 10 variáveis, pode-se utilizar esse método para complementar as regras geradas por meio da geração semi-exaustiva, pois nesse caso as regras completas ainda têm uma dimensão que permite interpretação.

Diferentemente da geração exaustiva, que avalia cada regra para o conjunto de treinamento inteiro, o método de Wang & Mendel visa a gerar a regra que, individualmente, mais bem representa cada ponto do conjunto de treinamento, sem avaliá-las globalmente. Dado que se tem S séries temporais, e são utilizadas L defasagens como variáveis de entrada, tem-se o conjunto de treinamento formados por pares entrada-saída da seguinte forma:

ENTRADAS

$y_{1,t-L}, y_{1,t-L+1}, \dots, y_{1,t-1}, y_{2,t-L}, \dots, y_{S,t-1}$

SAÍDAS

$y_{1,t}, \dots, y_{S,t}$

Em cada ponto do conjunto de treinamento, cada variável pode ser

descrita como um conjunto de graus de pertinência aos seus conjuntos fuzzy:

ENTRADAS

$$\begin{aligned}
 y_{1,t-L} &= (\mu_{1,1}(y_{1,t-L}), \mu_{1,2}(y_{1,t-L}), \dots, \mu_{1,N}(y_{1,t-L})) \\
 y_{1,t-L+1} &= (\mu_{1,1}(y_{1,t-L+1}), \mu_{1,2}(y_{1,t-L+1}), \dots, \mu_{1,N}(y_{1,t-L+1})) \\
 &\vdots \\
 y_{1,t-1} &= (\mu_{1,1}(y_{1,t-1}), \mu_{1,2}(y_{1,t-1}), \dots, \mu_{1,N}(y_{1,t-1})) \\
 y_{2,t-L} &= (\mu_{2,1}(y_{2,t-L}), \mu_{2,2}(y_{2,t-L}), \dots, \mu_{2,N}(y_{2,t-L})) \\
 &\vdots \\
 y_{S,t} &= (\mu_{S,1}(y_{S,t}), \mu_{S,2}(y_{S,t}), \dots, \mu_{S,N}(y_{S,t}))
 \end{aligned}$$

SAÍDAS

$$\begin{aligned}
 y_{1,t} &= (\mu_1(y_{1,t}), \mu_2(y_{1,t}), \dots, \mu_N(y_{1,t})) \\
 &\vdots \\
 y_{S,t} &= (\mu_S(y_{S,t}), \mu_2(y_{S,t}), \dots, \mu_N(y_{S,t}))
 \end{aligned}$$

Como se quer encontrar a regra completa que melhor descreve um determinado ponto do conjunto de treinamento, obtém-se qual é o termo linguístico que apresenta o maior grau de pertinência para cada variável de entrada, gerando um premissa. Da mesma forma que na geração semi-exaustiva, para cada premissa geram-se S regras, uma para cada série temporal contemplada, onde o termo linguístico que figura como conseqüente é também o de grau de pertinência máximo para aquele ponto.

Dado um conjunto de treinamento de tamanho T , poderão ser geradas até $T \cdot S$ regras completas, ou seja, para cada uma das T amostras, são geradas S regras, uma para cada série modelada. Nesse processo, porém, é possível que dois ou mais pontos gerem regras com premissas idênticas, porém com conseqüentes distintos, o que gera um conflito no funcionamento do SIF. Tradicionalmente, como o método foi elaborado para trabalhar junto com especialistas, tem-se a possibilidade de utilizá-los para resolver os conflitos de acordo com seu conhecimento. No caso do AutoMFIS, aplica-se uma heurística baseada apenas nos dados: calcula-se a ativação da regra (via composição dos termos por t-norma) para os pontos conflitantes e se conserva aquela gerada pelo ponto onde a ativação atingiu o maior valor, pois nesse ponto tem-se que a regra foi gerada com maior confiança.

Além dessa redundância “interna” de regras geradas via Wang & Mendel,

há também a possibilidade de que haja conflitos com as regras já existentes na base a qual se quer complementar. Nesse caso, será dada preferência às regras criadas pelo método semi-exaustivo, tendo em vista que os consequentes das mesmas são definido levando em consideração todo conjunto de treinamento, sendo assim mais gerais.

3.3.3

Geração dos Valores Crisp de Saída

A obtenção dos valores crisp de saída envolve dois passos: agregação e defuzzificação.

Agregação

No exemplo abaixo, assume-se duas regras arbitrárias A_1 e A_2 contendo o mesmo termo linguístico no consequente, $B_{1,1}$.

$$\begin{aligned} \text{regra } A_1 &: a_{1,1} \cdot a_{2,1} \rightarrow B_{1,1} \\ \text{regra } A_2 &: a_{1,1} \cdot a_{2,2} \rightarrow B_{1,1} \\ \mu_{B_{1,1}} &= \begin{cases} \mu_{a_{1,1}} \cdot \mu_{a_{2,1}} & \text{ou} \\ \mu_{a_{1,1}} \cdot \mu_{a_{2,2}} & ? \end{cases} \end{aligned} \quad (3-6)$$

Como as premissas das regras A_1 e A_2 são distintas, serão geradas, tipicamente, duas ativações também distintas para para uma entrada qualquer X^* , gerando uma ambiguidade na ativação do consequente $B_{1,1}$. Assim, é necessário efetuar uma agregação das ativações das regras de mesmo consequente, de modo a gerar uma ativação única para o mesmo.

Definindo $A_d^{(s,n)}$ como sendo a d -ésima regra associada ao consequente $B_{s,n}$, uma métrica de agregação é uma função da forma:

$$\mu_{B_{s,n}} = g(\mu_{A_1^{(s,n)}}(X^*), \mu_{A_2^{(s,n)}}(X^*), \dots) \quad (3-7)$$

onde $g : [0, 1]^{D^{(s,n)}} \rightarrow [0, 1]$; $D^{(s,n)}$: quantidade de regras associadas ao consequente $B_{s,n}$.

A agregação pelo maximum, representada abaixo, é a mais corriqueiramente empregada em SIFS do tipo Mamdani.

$$\mu_{B_{s,n}} = \max(\mu_{A_1^{(s,n)}}(X^*), \mu_{A_2^{(s,n)}}(X^*), \dots) \quad (3-8)$$

No caso do AutoMFIS, as regras mais simples, de menor dimensão, tendem a

ter ativação mais alta que regras mais completas. Como o método de geração o cria regras partindo das de tamanho 1, o uso do maximum fará com que o sistema tenha seu comportamento dominado por estas regras mais simples. Isto é especialmente problemático para séries em que regras mais curtas não contêm a quantidade mínima de informação para descrever com alguma propriedade o comportamento futuro de uma série como a exemplificada na figura 3.19.

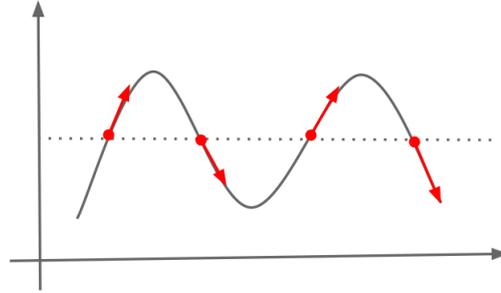


Figura 3.19: Série não explicável por regras de tamanho 1.

Na modelagem desta série, o conhecimento de apenas um ponto (regras de tamanho 1) não é suficiente para retratar o seu comportamento. Uma premissa ambígua tenderá a ter um conseqüente que não descreve com precisão as situações futuras possíveis, enquanto que o processo de associação escolherá um conseqüente intermediário que maximize a confiança, resultando em um baixo desempenho do sistema. Para contornar este problema, faz-se uso de um método de agregação ponderado.

De modo similar à associação por máxima credibilidade, é possível definir os pesos de cada regra por regressão via mínimos quadrados restritos, e então aplicar uma métrica sobre suas ativações ponderadas:

$$\mu_{B_{s,n}} = g(\omega_1^{(s,n)} \mu_{A_1^{(s,n)}}(X^*), \omega_2^{(s,n)} \mu_{A_2^{(s,n)}}(X^*), \dots) \quad (3-9)$$

O problema de minimização do erro quadrático, que representará o ajuste de pesos, é:

$$\begin{aligned} \min \sum_{t=1}^T (\mu_{B_{s,n}}(y_{s,t}) - \sum_{d^{(s,n)}=1}^{D^{(s,n)}} w_d^{(s,n)} \mu_{A_{d^{(s,n)}}}(X_t))^2 \\ \text{s.a. : } \sum_{d^{(s,n)}=1}^{D^{(s,n)}} w_d^{(s,n)} = 1 \\ w_d^{(s,n)} \geq 0 \end{aligned} \quad (3-10)$$

onde $\mu_{B_{s,n}}(y_{s,t})$ se refere ao vetor de ativações do conseqüente $B_{s,n}$ para as

saídas esperadas $y_{s,t}$, e $\mu_{A_{d^{(s,n)}}}$ ao vetor de ativações da premissa $A_{d^{(s,n)}}$ para as entradas X_t , ambos definidos pelas T amostras do conjunto de treinamento. $w_d^{(s,n)}$ é o peso da d -ésima regra referente ao consequente n da série s . Com os pesos definidos, a função de agregação será:

$$\mu_{B_{s,n}} = \sum_{d^{(s,n)}=1}^{D^{(s,n)}} w_d^{(s,n)} \mu_{A_{d^{(s,n)}}}(X^*) \quad (3-11)$$

Por mais que os pesos tenham sido ajustados especificamente eles podem ser utilizados como uma medida de “confiança” e explorados em outras métricas, como a própria agregação pelo máximo:

$$\mu_{B_{s,n}} = \max(w_1^{(s,n)} \mu_{A_1^{(s,n)}}(X^*), w_2^{(s,n)} \mu_{A_2^{(s,n)}}(X^*), \dots) \quad (3-12)$$

Esta abordagem ponderada é especialmente interessante por representar um passo adicional de filtragem da base de regras, pois esse processo de otimização associa peso zero (ou aproximadamente zero) às regras menos valiosas, aumentando a acurácia e diminuindo a complexidade do sistema.

Na etapa de fuzzificação, a seguir, são levadas em consideração todas as regras ativadas, como tradicionalmente ocorre em SIFs.

Defuzzificação

Neste trabalho, consideram-se os métodos de defuzzificação do Centróide e da Altura, que proporcionam resultados similares, com o segundo sendo de execução mais rápida. Ambos levam em consideração a informação contida na totalidade do conjunto fuzzy, em oposição a métodos como o do Máximo ou da Média dos Máximos, que privilegiam as regras de maior ativação.

Tradicionalmente, o processo de defuzzificação é realizado para cada um dos S sistemas, ou seja, efetivamente criando um desacoplamento das previsões das séries em um mesmo instante t . Com o objetivo de levar em conta o acoplamento entre séries, o valor defuzzificado para cada série deve levar em conta os conjuntos fuzzy de saída para todas as séries. Este método de defuzzificação será denominado VAR.

Para que a defuzzificação VAR seja possível, o dicionário de conjuntos fuzzy de cada variável de saída deverá ser preparado para receber uma representação dos conjuntos fuzzy das demais variáveis. Assim, a defuzzificação VAR consiste em duas etapas: ajuste do dicionário de conjuntos fuzzy e defuzzificação.

O ajuste do dicionário é feito por meio da definição de pesos que multiplicarão os centros de cada conjunto fuzzy – efetivamente movendo-os

ao longo do universo.

Após se aplicar o passo de agregação para todas as amostras do conjunto de treinamento, tem-se a seguinte matriz de ativações:

$$CP = \begin{bmatrix} \mu_{B_{1,1}}(X_1) & \mu_{B_{1,2}}(X_1) & \dots & \mu_{B_{1,N}}(X_1) & \mu_{B_{2,1}}(X_1) & \dots & \mu_{B_{S,N}}(X_1) \\ \mu_{B_{1,1}}(X_2) & \mu_{B_{1,2}}(X_2) & \dots & \mu_{B_{1,N}}(X_2) & \mu_{B_{2,1}}(X_2) & \dots & \mu_{B_{S,N}}(X_2) \\ \vdots & & & \ddots & & & \vdots \\ \mu_{B_{1,1}}(X_T) & \mu_{B_{1,2}}(X_T) & \dots & \mu_{B_{1,N}}(X_T) & \mu_{B_{2,1}}(X_T) & \dots & \mu_{B_{S,N}}(X_T) \end{bmatrix} \quad (3-13)$$

Essas ativações são normalizadas de modo que $\sum_{s=1}^S \sum_{n=1}^N \mu_{B_{s,n}}(X_t) = 1$, $t \in \{1, \dots, T\}$. Com as ativações normalizadas, faz-se o produto de cada uma delas pelo centro de seu conjunto fuzzy associado, gerando a matriz CP. Os pesos multiplicarão cada centro dos conjuntos fuzzy podem ser encontrados por:

$$pesos = (CP' * CP)^{-1} * (CP' * y_s) \quad (3-14)$$

onde y_s é a série de saídas esperadas no conjunto de treinamento associado à série s .

Este procedimento é similar ao do método da altura, com a diferença de que faz uso das ativações de todos os conjuntos fuzzy de todas as variáveis de saída e dos seus centros modificados, em vez dos reais definidos nos dicionários.

É importante salientar que o deslocamento dos conjuntos - provocado pela aplicação dos pesos - pode dificultar a interpretação do problema, em virtude da eventual perda de significado semântico ou por causa de uma sobreposição acentuada dos conjuntos. Para problemas complexos envolvendo muitas séries temporais, com dezenas de conjuntos fuzzy para cada variável de saída, haverá um colapso da compreensão das regras. Assim, no contexto de modelos interpretáveis, a aplicação desse método de defuzzificação fica restrito apenas a problemas mais simples, como o de um par de séries, ou a casos em que o seu uso proporcione ganho de desempenho significativo, compensando a dificuldade de interpretação.

3.3.4

Conclusão

Nas seções anteriores foram apresentados e discutidos todos os aspectos do funcionamento do modelo AutoMFIS, que pode ser resumido nos passos

descritos no diagrama da figura 3.20.

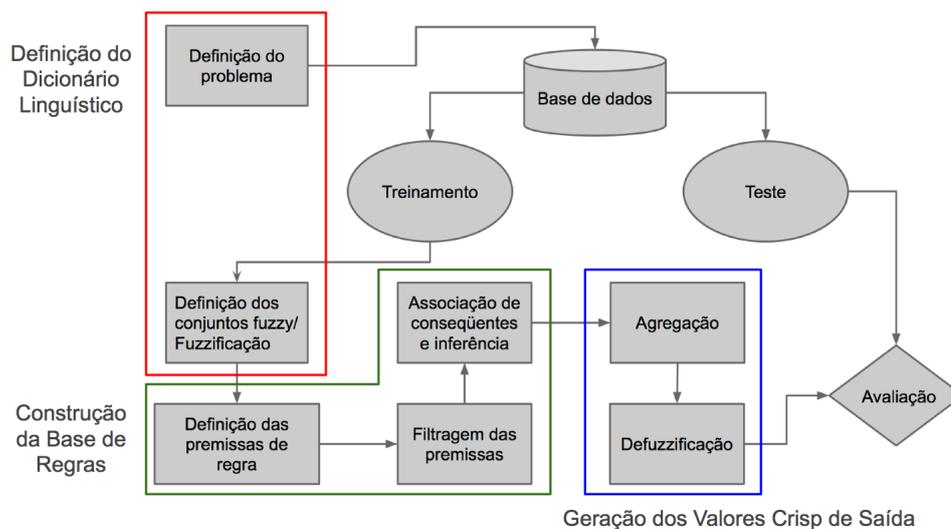


Figura 3.20: Resumo: Funcionamento do modelo AutoMFIS.

A primeira etapa diz respeito à construção do problema multivariado. Para tal foram apresentadas técnicas da estatística para guiar a escolha das séries e a quantidade de defasagens a serem contempladas pelo SIF gerado. Este passo é importante pois define quais variáveis apresentarão relação.

Com as variáveis selecionadas, organizam-se os dados conjuntos de treinamento e teste e constrói-se, então, o SIF. A primeira parte desse processo é a definição do dicionário fuzzy. Foram apresentadas algumas formas para definição dos domínios das funções de pertinência a partir dos conjuntos de treinamento e teste e suas consequentes segmentações em conjuntos fuzzy triangulares. Procura-se manter o equilíbrio entre a resolução do sistema e a compreensão da distinção entre os termos linguísticos.

Na geração da base de regras do SIF, estabelecem-se, primeiramente, as premissas das regras, que são eliminadas de acordo com a sua relevância no conjunto de treinamento, segundo uma das métricas escolhidas. A seguir, na etapa de associação, define-se o consequente de regra mais apropriado para cada premissa do conjunto pós-eliminação. Formadas as regras, são agregadas aquelas que apresentam o mesmo consequente.

A etapa final é a de defuzzificação. Como, no caso de previsão multivariada, há múltiplas variáveis de saída, cada qual com seus próprios conjuntos fuzzy, foram apresentadas duas abordagens para a defuzzificação: a tradicional, que utiliza apenas os conjuntos fuzzy referentes à própria variável para a geração do valor preciso, e a multivariada, que combina os conjuntos de todas as variáveis, fazendo com que a previsão de uma série afete as demais.

Quanto à questão da complexidade segundo a taxonomia de Gacto (tabela 3.3), cada ponto foi tratado da seguinte maneira:

- **Número de variáveis:** ditado pelos objetivos e por técnicas de seleção de variáveis, como o clustering hierárquico.
- **Número de conjuntos fuzzy:** limitado pelo critério de Miller (7 ± 2).
- **Cobertura, discernibilidade e complementaridade:** uso de uma partição fuzzy forte, com conjuntos fuzzy triangulares.
- **Número de regras:** redução do número de variáveis e conjuntos fuzzy, corte de premissas, eliminação de regras por agregação ponderada.
- **Número de condições:** limitado pelo critério de Miller (7 ± 2).
- **Consistência:** garantida pelos métodos de associação de consequentes – associação de consequentes similares à premissas com ativações similares.
- **Ativação simultânea:** reduzida pela diminuição da base de regras.

4

Experimentos

Este capítulo descreve uma sequência de experimentos destinados à avaliação de cada aspecto do modelo AutoMFIS, cobrindo algumas variedades de problemas de previsão multivariada. Os experimentos, realizados utilizando rotinas criadas no software Matlab [63], podem ser divididos em duas etapas no que tange ao foco da avaliação:

1. **Acurácia da previsão:** antes de analisar o potencial de extração de conhecimento do modelo, deve-se ter certeza de que ele é capaz de ajustar um função de previsão razoável, ou seja, apresentar acurácia competitiva com as de outros métodos. Para isto, foram construídas algumas séries multivariadas, para as quais o desempenho do modelo será comparado com os de benchmarks.
2. **Extração de conhecimento:** comprovado o bom desempenho do AutoMFIS, avalia-se a interpretabilidade dos SIFs por ele ajustados. Serão apresentados dois estudos de caso com dados reais, representando problemas de previsão multivariada.

Nas próximas seções serão apresentados as metodologias e dados empregados em cada experimento.

4.1

Comparação com métodos da competição M3

O modelo foi inicialmente avaliado sob o ponto de vista exclusivo da acurácia da previsão, para depois se verificar a questão de extração de conhecimento (interpretabilidade). Para tal, faz-se uso de uma competição de previsão.

As competições são estudos empíricos de larga escala que têm como objetivo avaliar metodologias de diferentes áreas, na certeza de que os participantes não pouparam esforços para atingir os melhores resultados possíveis. As competições de previsão mais notórias foram as organizadas por Spyros Makridakis, conhecidas como Competições M. Estas competições nasceram da contestação, por membros da comunidade científica, dos resultados obtidos por Makridakis e Hibon [64], na avaliação de desempenho de diversos modelos para um conjunto de 111 séries temporais cobrindo uma grande gama de situações reais, como dados da indústria e economia. De modo a eliminar qualquer dúvida e

possíveis vieses de metodologia, esse estudo empírico tomou a forma das Competições M [65], abrindo espaço para que os críticos pudessem realizar suas próprias simulações.

A competição mais recente foi a M3 [66], que contou com um total de 24 métodos participantes. Estes métodos, utilizados como benchmarks para a avaliação do AutoMFIS, são mostrados na tabela 4.1.

Método	Competidores	Descrição
<i>Ingênuo / simples</i>		
1. Naïve2	M. Hibon	Ingênuo Desazonalizado (Passeio Aleatório)
2. Single	M. Hibon	Alisamento Exponencial Simples
3. Holt	M. Hibon	Alisamento Exponencial Linear de Holt Automático (modelo de dois parâmetros)
4. Robust-Trend	N. Meade	Versão não-paramétrica do modelo linear de Holt com tendência estimada pela mediana
5. Winter	M. Hibon	Alisamento Exponencial Linear e Sazonal de Holt-Winter (modelo de dois ou três parâmetros)
6. Dampen	M. Hibon	Alisamento Exponencial com Tendência Amortecida
7. PP-autocast	H. Levenbach	Alisamento Exponencial com Tendência Amortecida
8. Theta-sm	V. Assimakopoulos	Alisamento sucessivo em adição de um conjunto de regras para amortecimento da tendência
9. Comb S-H-D	M. Hibon	Combinação de três métodos: Single, Holt e Dampen
<i>Decomposição</i>		
10. Theta	V. Assimakopoulos	Técnica específica de decomposição, seguida de projeção e combinação dos componentes
<i>ARIMA/ARARMA</i>		
11. B-J automatic	M. Hibon	Metodologia de Box-Jenkins
<i>Sistema Especialista</i>		
12. Autobox1	D. Reilly	ARIMA univariado robusto Box-Jenkins com ou sem detecção de intervenção
13. Autobox2		
14. Autobox3		
15. AAM1	G. Melard, J.M. Pasteels	Modelagem ARIMA automática com ou sem análise de intervenção
16. AAM2		

Método	Competidores	Descrição
17. ARARMA	N. Meade	Metodologia de Parzen automatizada com filtro auto regressivo
18. ForecastPro	R. Goodrich, E. Stellwagen	Seleciona entre vários métodos: Alisamento Exponencial, Box-Jenkins, Poisson e modelos binomiais negativos, método de Croston, Média Móvel Simples
19. SmartFes	C. Smart	Sistema Especialista Automática de Previsão, que conduz um torneio de previsão entre quatro métodos de alisamento exponencial e dois de média móvel
20. RBF	M. Adya, S. Armstrong, F. Collopy, M. Kennedy	Previsão baseada em regras: utiliza três métodos - Passeio Aleatório, Regressão Linear e Holt - para estimar nível e tendência, envolvendo correções e simplificações, identificação automática de características e re-calibração
21. Flores/Pearce1	B. Flores, S. Pearce	Sistema Especialista que escolhe entre quatro métodos baseados nas características dos dados
22. Flores/Pearce2	J. Galt	
23. ForecastX	J. Galt	Executa testes para sazonalidade e outliers, selecionando entre diversos métodos: Alisamento Exponencial, Box-Jenkins e Croston
Redes Neurais		
24. Automat ANN	K. Ord, S. Balkin	Redes Neurais Artificiais automatizadas para previsão

Tabela 4.1: Métodos da Competição M3, divididos em categorias.

O conjunto de séries temporais foi ampliado a cada competição, chegando a 3003 séries para a M3. Esse conjunto de séries pode ser segmentado quanto ao seu contexto semântico (“tipo”), e o intervalo entre as amostras (tabela 4.2).

Intervalo	Micro	Indústria	Macro	Finanças	Demog.	Outros
Anual	146	102	83	58	245	11
Trimestral	204	83	336	76	57	0
Mensal	474	334	312	145	111	52
Outro	4	0	0	29	0	141

Tabela 4.2: Séries temporais por tipo e intervalo.

Devido à limitação de tempo e recursos disponíveis, o modelo AutoMFIS foi aplicado a três séries multivariadas construídas a partir do conjunto

completo de séries. Os pré-requisitos para que séries possam ser agrupadas numa multivariada é que elas devem ter o mesmo intervalo, tamanho e origem (data inicial).

Como visto na tabela, existem séries mensais, trimestrais, mensais e algumas categorizadas como “outros”. Como a informação sobre o intervalo das séries desta última categoria não está disponível, é impossível agrupá-las em séries temporais múltiplas, já que todas séries agrupadas devem ter o mesmo intervalo. Dos três grupos restantes, selecionou-se o de séries mensais, que contém um maior volume de dados – essencial para métodos orientados a dados –, ao mesmo tempo em que exige um horizonte de previsão também maior – prova mais rígida de desempenho.

Para compensar a ausência de informações específicas sobre o significado de cada série, os experimentos concentraram-se no grupo de séries de finanças, área na qual é comum o uso de modelos multivariados devido à natureza interconectada dos fenômenos. Neste grupo, o maior subconjunto de séries apresentando os dois outros pré-requisitos – origem e tamanho – é o das que se iniciam em janeiro de 1983 e possuem tamanho 134. Estas 62 séries temporais podem ser observadas na figura 4.1.

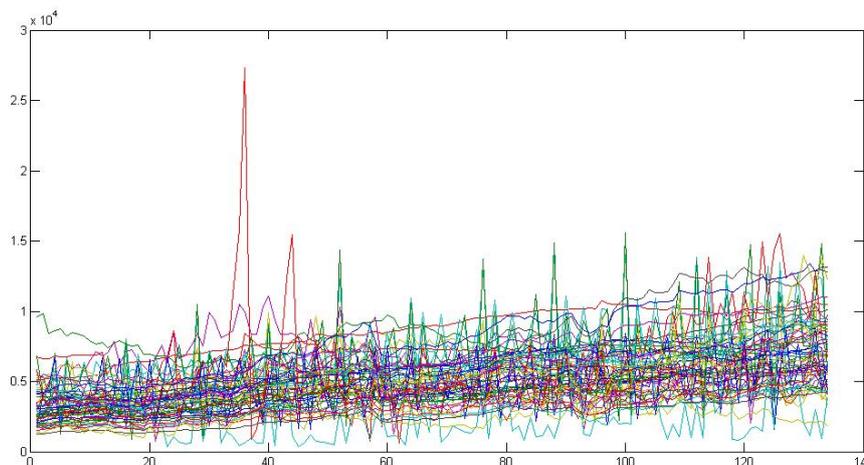


Figura 4.1: 62 séries temporais mensais de finanças - Competição M3.

Uma seleção de variáveis por clustering hierárquico foi efetuada nesse conjunto de séries; os clusters resultantes constituíram as séries multivariadas a ser modeladas. Utilizando as métricas e métodos apresentados no capítulo 2, obteve-se o dendrograma ilustrado na figura 4.2.

Como discutido anteriormente, não existe um ponto de corte absoluto para o clustering hierárquico: é necessário realizar um julgamento específico para cada problema e objetivo. Neste caso, o procedimento para definir o

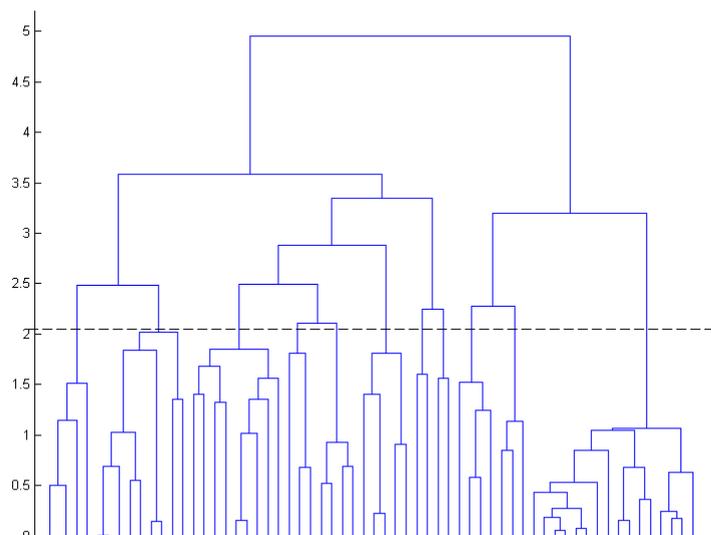


Figura 4.2: Dendrograma do clustering hierárquico.

ponto de corte levou em conta, primeiramente, as dimensões dos clusters formados. A quantidade de séries utilizadas para formar uma série multivariada está sujeita ao limite de recursos computacionais disponíveis. Verificou-se, empiricamente, que o limite de cinco séries temporais permite uma maior liberdade na escolha de configurações. Em segundo lugar, considerando que se deseja apenas três séries multivariadas com dimensão próxima a cinco séries, deve ser definido o quão “conservadora” deve ser a escolha do ponto de corte. A menor distância que gera três clusters com dimensão de aproximadamente 5 elementos é o ponto de corte mais conservador, com menor risco de se agrupar séries não relacionadas. Todavia, optou-se por uma abordagem mais “arriscada”: verificação da capacidade do AutoMFIS de lidar com associações menos ideais. Definiu-se como ponto de corte a distância 2,1, para a qual são gerados clusters de diferentes tamanhos, desde aqueles formados por uma única série até aqueles formados por mais de dez; entre eles encontram-se três clusters da dimensão desejada, ilustrados nas figuras 4.3, 4.4 e 4.5.

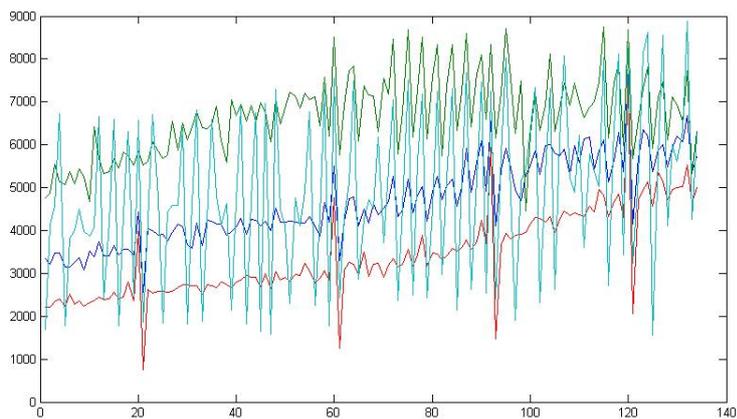


Figura 4.3: Séries do cluster 1.

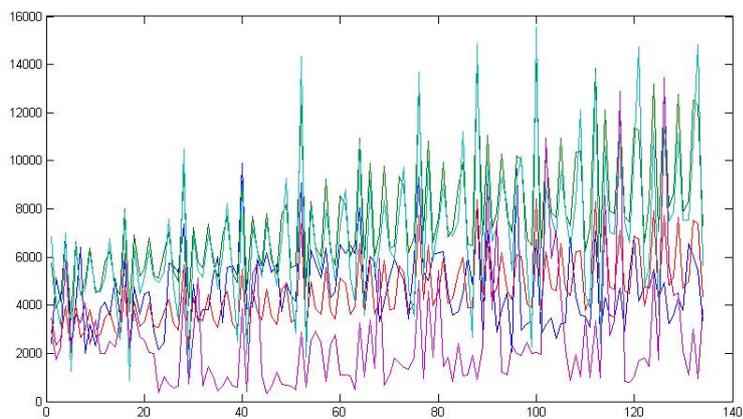


Figura 4.4: Séries do cluster 2.

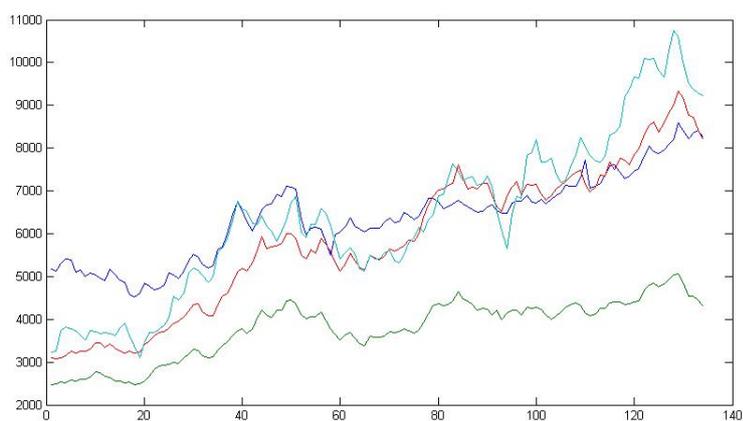


Figura 4.5: Séries do cluster 3.

A competição M3 estipula um horizonte de previsão de tamanho 18 para as séries selecionadas. Por essa razão, o conjunto de teste será formado pelos últimos 18 pontos, com o restante compondo o conjunto de treinamento.

As três séries multivariadas resultantes apresentam tendência de crescimento. Conforme discutido na seção 3.3, sobre a construção do dicionário linguístico, a presença de uma tendência impede que o sistema fuzzy opere corretamente. Para tratar dessa questão, avaliaram-se os dois métodos apresentados: eliminação da tendência estimada via regressão linear ou uso das diferenças. Como nesta etapa a preocupação é com a acurácia, a escolha do pré-processamento não possui tanta importância, mas, como será discutido no último estudo de caso, este tipo de manipulação altera o caráter da informação extraída pelo SIF.

Devido à grande variedade de opções para a configuração do modelo, é difícil avaliar todas as combinações de maneira exaustiva. Assim, em uma primeira etapa, foram avaliadas combinações de métodos e métricas, mantendo-se fixas algumas opções de ajuste fino, tais como limiar de corte – selecionados de modo que todos os critérios gerem a mesma quantidade de regras, dentro dos limites de recursos computacionais –, t-norma (produto) e número de conjuntos fuzzy (cinco). As opções testadas nessa etapa foram:

1. **Pré-processamento:** Eliminação de tendência ou diferenciação.
2. **Fuzzificação:** funções de pertinência triangulares, com espaçamentos uniforme ou por percentis da distribuição de valores.
3. **Corte de regras:** todas as métricas (frequência, cardinalidade e ativação).
4. **Particionamento:** todas as métricas de máxima confiança, além do método de máxima credibilidade.
5. **Agregação:** todas as métricas (máximo, máximo ponderado e média ponderada).
6. **Defuzzificação:** método da altura, representando a abordagem clássica, e método VAR, representando a defuzzificação acoplada.

O desempenho das combinações de todas essas opções foi avaliado segundo a métrica de erro SMAPE (symmetric mean absolute percentage error) – métrica principal utilizada na competição. Destes resultados, foram armazenados os dez melhores para utilização no passo seguinte, que consistiu no ajuste fino dos aspectos restantes do modelo:

1. **Fuzzificação:** testes com 7 e 9 conjuntos, com espaçamentos uniforme ou por percentis
2. **T-normas:** mínimo, Hamacher e Lukasiewicz
3. **Corte de regras:** incremento do limiar até regra geração de base de regras vazia.
4. **Defuzzificação (adendo):** Centróide, para as configurações com defuzzificação por Altura.

Essa sequência é repetida para diferentes quantidades de defasagens consideradas, tomando-se um valor base, obtido via análise de correlações, e pequenas variações em seu entorno.

Com esse processo, obteve-se uma aproximação das melhores configurações para cada série multivariada usando 3 defasagens, cujo desempenho pode ser visto comparado ao dos 24 benchmarks da competição M3 na tabela 4.3.

Pode-se observar que o modelo AutoMFIS supera todos os métodos univariados competidores para os clusters 1 e 3, em especial para esse último, apresentando um desempenho muito acima da mediana. Para o cluster 2, porém, o modelo apresentou mais dificuldades. Observando o gráfico para este cluster, verifica-se um comportamento sazonal, na forma de picos nos seus valores a cada 12 meses. Para captar esta sazonalidade, ter-se-ia que considerar um grande número de defasagens, o que inviabilizaria o ajuste do modelo devido complexidade gerada pela grande quantidade de variáveis de entrada. Alternativamente, optou-se por uma forma menos automatizada: enriquecer o modelo com a inclusão, apenas, da 12ª defasagem. Esta informação adicional permite ao modelo identificar a ocasião em que o passo a ser previsto fecha o ciclo de 12 meses.

Aplicando novamente as duas etapas de avaliação para a nova configuração do AutoMFIS para o cluster 2, obteve-se um resultado muito mais acurado, como visto na tabela 4.4.

O aumento da quantidade de defasagens utilizadas pelo sistema para o cluster 2 aumentou consideravelmente o custo computacional, o que impediu que fossem avaliadas as configurações com 7 e 9 conjuntos fuzzy sem que o limiar de corte fosse também aumentado, eventualmente eliminando regras valiosas. Em virtude desta limitação, existe um potencial não explorado de melhoria de acurácia. Mesmo assim, o modelo apresentou um dos melhores resultados do conjunto de métodos, com desempenho superior à mediana.

Método	Cluster 1	Cluster 2	Cluster 3
Naïve2	36,24%	24,03%	9,54%
Single	17,02%	22,72%	9,54%
Holt	15,53%	20,42%	6,43%
Dampen	17,54%	21,48%	9,64%
Winter	15,96%	43,35%	6,43%
Comb S-H-D	15,64%	21,36%	8,44%
B-J auto	17,03%	17,40%	9,74%
AutoBox1	15,66%	25,03%	6,39%
AutoBox2	16,07%	17,02%	8,39%
AutoBox3	15,50%	37,77%	8,41%
Robust-Trend	21,19%	27,33%	6,17%
ARARMA	19,14%	19,12%	5,75%
Auto-ANN	17,68%	18,47%	10,14%
Flors-Pearc1	16,14%	22,87%	9,18%
Flors-Pearc2	15,62%	25,56%	8,55%
PP-Autocast	20,12%	23,61%	9,64%
ForecastPro	16,62%	19,96%	9,73%
SmartFcs	16,20%	23,05%	9,58%
Theta-sm	27,53%	22,05%	10,24%
Theta	16,99%	20,91%	7,97%
RBF	17,45%	22,21%	5,96%
ForcX	15,56%	18,51%	9,70%
AAM1	17,81%	17,65%	5,67%
AAM2	17,81%	17,65%	5,67%
Desempenho Mediano	17,01%	21,77%	8,49%
Melhor Desempenho	15,50%	17,02%	5,67%
AutoMFIS	14,60%	27,66%	4,59%

Tabela 4.3: Comparativo dos resultados: Benchmarks M3 x AutoMFIS.

O caso do cluster 2 mostra como uma abordagem completamente orientada a dados, sem análise das séries modeladas, pode resultar em uma perda de desempenho. O uso de métricas como a correlação cruzada para definição das defasagens utilizadas na entrada do sistema falha em contemplar componentes como sazonalidade e ciclos, que exigem tratamento específico, assim como a própria componente de tendência. Todavia, com um tratamento correto, o modelo AutoMFIS tem um alto potencial de previsão, como evidenciado pelos resultados. Essa conclusão obtida a partir dos SMAPEs acumulados pode ser corroborada aplicando testes estatísticos sobre as amostras resultante de erros de previsão.

Como o horizonte de previsão é 18 para todos os casos, é possível construir para cada um dos métodos listados uma amostra dos erros SMAPE acumulados em cada instante – de $t + 1$ a $t + 18$ –, totalizando 18 elementos. Como essas

Método	Cluster 1	Cluster 2	Cluster 3
Desempenho Mediano	17,01%	21,77%	8,49%
Melhor Desempenho	15,50%	17,02%	5,67%
AutoMFIS	14,60%	19,77%	4,59%

Tabela 4.4: Comparativo dos resultados: Benchmarks M3 x AutoMFIS modificado.

amostras foram consideradas pequenas demais para permitir que suposições sejam feitas sobre suas distribuições (≤ 30 elementos), foram utilizados apenas testes não-paramétricos. Para compensar o seu menor poder estatístico, fez-se uso de dois tipos: teste dos sinais e o teste de Wilcoxon. Ambos almejam testar a hipótese nula de que a diferença entre as medianas de duas amostras é zero com nível de significância $p = 0.05$, que, neste caso, significa que os métodos comparadas demonstraram acurácia equivalente. Os resultados desses testes são mostrados nas Tabelas 4.5, 4.6 e 4.7, onde H_0 é a hipótese nula (0 significa que a mesma não pode ser rejeitada, enquanto 1 significa o oposto) e IC é o intervalo de confiança para a diferença das medianas no nível de confiança de 95% – onde um valor negativo significa que a mediana da amostra dos erros do AutoMFIS é menor do que a obtida pelo método ao qual o AutoMFIS está sendo comparado.

Para o primeiro cluster, os testes mostram que o AutoMFIS superou de 7 a 10 métodos ($H_0 = 1$ com IC negativo), e teve desempenho equivalente aos demais 14 a 17 métodos ($H_0 = 0$). Para aqueles em que hipótese nula não é rejeitada, os intervalos de confiança possuem um viés claro para o espectro negativo, indicando uma possível superioridade do AutoMFIS, porém não confirmada para o grau de confiança definido. Para o cluster 3, os resultados são ainda melhores, com 14 a 15 métodos sendo derrotados, e 9 a 10 sendo similares, com o mesmo viés aparecendo nos intervalos de confiança. Para o cluster 2, o modelo superou de 7 a 8 métodos, resultado comparável ao obtido no primeiro cluster. Todavia, os intervalos de confiança simétricos e centrados em zero não permitem confirmar a aparente superioridade, sugerida pelo erro SMAPE acumulado (Tabela 4.3), dos métodos B-J auto, AutoBox2, ARARMA, Auto-ANN, AAM1 e AAM2 em relação ao AutoMFIS.

A acurácia competitiva das previsões indica que o AutoMFIS foi bem sucedido em modelar as séries temporais e, conseqüentemente, que o conhecimento extraído das bases de regras descreve bem o comportamento das séries. Porém, este bom desempenho não é necessariamente um resultado da modelagem multivariada, pois o processo de filtragem da base de regras admite que as regras que relacionam as séries possam ser eliminadas. Por essa razão, mesmo

Tabela 4.5: Testes estatísticos: Cluster 1

Método	Teste dos sinais		Teste de Wilcoxon	
	H_0	IC	H_0	IC
Naive2	1	[-0,32;-0,19]	1	[-0,31;-0,17]
Single	0	[-0,14;0,00]	0	[-0,11;0,01]
Holt	0	[-0,09;0,01]	0	[-0,10;0,03]
Dampen	1	[-0,14;-0,00]	0	[-0,11;0,00]
Winter	0	[-0,13;0,01]	0	[-0,11;0,01]
Comb S-H-D	0	[-0,12;0,01]	0	[-0,10;0,02]
B-J auto	0	[-0,12;0,00]	0	[-0,11;0,00]
AutoBox1	0	[-0,11;0,02]	0	[-0,10;0,02]
AutoBox2	0	[-0,10;0,01]	0	[-0,10;0,02]
AutoBox3	0	[-0,11;0,02]	0	[-0,10;0,02]
Robust-Trend	1	[-0,17;-0,04]	1	[-0,17;-0,04]
ARARMA	1	[-0,15;-0,01]	1	[-0,13;-0,00]
Auto-ANN	1	[-0,13;-0,00]	1	[-0,12;-0,00]
Flors-Pearc1	0	[-0,10;0,01]	0	[-0,10;0,02]
Flors-Pearc2	0	[-0,10;0,01]	0	[-0,10;0,02]
PP-Autocast	1	[-0,17;-0,03]	1	[-0,16;-0,03]
ForecastPro	0	[-0,11;0,01]	0	[-0,11;0,01]
SmartFcs	0	[-0,13;0,01]	0	[-0,11;0,02]
Theta-sm	1	[-0,24;-0,10]	1	[-0,22;-0,10]
Theta	0	[-0,14;0,00]	0	[-0,12;0,01]
RBF	1	[-0,14;-0,00]	0	[-0,12;0,00]
ForcX	0	[-0,11;0,01]	0	[-0,10;0,01]
AAM1	1	[-0,12;-0,00]	0	[-0,11;0,00]
AAM2	1	[-0,12;-0,00]	0	[-0,11;0,00]

que neste estudo de caso não se permita interpretação devido à opacidade de significado das séries da competição M3, é importante observar as bases de regra geradas. As tabelas 4.8, 4.9 e 4.10 mostram alguns exemplos de regras para os três clusters de séries modelados, com $y_i(t)$ sendo o valor da i -ésima série do cluster no instante t , e o dicionário fuzzy sendo formado por pelos termos "Muito Baixo" (MB), "Baixo" (B), "Médio" (M), "Alto" (A) e "Muito Alto" (MA).

As regras geradas são simples – com um baixo número de termos em suas premissas – e de fato estabelecem uma conexão entre as diferentes séries temporais, confirmando que o objetivo de extração do relacionamento entre as séries foi atingido. Como mostrado na tabela 4.11, as bases de regra também são compactas – em alguns dos casos pode ser vantajoso desconsiderar regras com peso insignificante –, o que permite um bom nível de interpretabilidade.

Estabelecido que o AutoMFIS é competitivo em relação aos modelos da competição M3 em termos de acurácia, e que também é capaz de detectar o

Tabela 4.6: Testes estatísticos: Cluster 2

Método	Teste dos sinais		Teste de Wilcoxon	
	H_0	IC	H_0	IC
Naive2	0	[-0,12;0,00]	1	[-0,11;-0,05]
Single	0	[-0,13;0,01]	0	[-0,10;0,00]
Holt	0	[-0,12;0,00]	0	[-0,09;0,01]
Dampen	0	[-0,13;0,00]	0	[-0,09;0,01]
Winter	1	[-0,22;-0,06]	1	[-0,22;-0,08]
Comb S-H-D	0	[-0,12;0,01]	0	[-0,09;0,01]
B-J auto	0	[-0,07;0,06]	0	[-0,06;0,04]
AutoBox1	1	[-0,11;-0,00]	1	[-0,11;-0,02]
AutoBox2	0	[-0,06;0,05]	0	[-0,05;0,05]
AutoBox3	1	[-0,28;-0,14]	1	[-0,26;-0,15]
Robust-Trend	1	[-0,18;-0,05]	1	[-0,16;-0,06]
ARARMA	0	[-0,06;0,05]	0	[-0,06;0,03]
Auto-ANN	0	[-0,05;0,06]	0	[-0,05;0,04]
Flors-Pearc1	1	[-0,15;-0,03]	1	[-0,11;-0,01]
Flors-Pearc2	1	[-0,18;-0,06]	1	[-0,14;-0,04]
PP-Autocast	0	[-0,10;0,02]	0	[-0,09;0,00]
ForecastPro	0	[-0,08;0,04]	0	[-0,07;0,02]
SmartFcs	1	[-0,13;-0,03]	1	[-0,12;-0,01]
Theta-sm	0	[-0,11;0,02]	0	[-0,09;0,01]
Theta	0	[-0,12;0,01]	0	[-0,09;0,01]
RBF	0	[-0,13;0,00]	0	[-0,10;0,00]
ForcX	0	[-0,09;0,04]	0	[-0,07;0,03]
AAM1	0	[-0,07;0,06]	0	[-0,06;0,04]
AAM2	0	[-0,07;0,06]	0	[-0,06;0,04]

relacionamento entre as séries temporais, foram avaliados em seguida mais dois estudos de caso, utilizando as mesmas duas etapas de busca pelas configurações ótimas, mas enfocando a questão da interpretabilidade, segundo a taxonomia apresentada no capítulo 3. Isto permitirá analisar em quais casos pode ser válido sacrificar acurácia por uma maior interpretabilidade.

4.2

Avaliação da interpretabilidade

Com o potencial do modelo AutoMFIS para previsão comprovado pelos benchmarks da competição M3, é explorado o seu segundo objetivo, a interpretabilidade do SIF ajustado. O intuito da interpretação de um SIF para previsão multivariada é compreender o relacionamento entre as séries temporais modeladas, porém para isso, é necessário primeiro entender os tipos de relacionamentos possíveis, ilustrados na figura 4.6. Para uma quantidade qualquer de séries, pode-se ter uma infinidade de combinações, porém baseadas em

Tabela 4.7: Testes estatísticos: Cluster 3

Método	Teste dos sinais		Teste de Wilcoxon	
	H_0	IC	H_0	IC
Naive2	1	[-0,05;-0,00]	1	[-0,04;-0,01]
Single	1	[-0,05;-0,00]	1	[-0,04;-0,01]
Holt	0	[-0,02;0,00]	0	[-0,02;0,00]
Dampen	1	[-0,05;-0,00]	1	[-0,04;-0,01]
Winter	0	[-0,02;0,00]	0	[-0,02;0,00]
Comb S-H-D	1	[-0,04;-0,00]	1	[-0,03;-0,00]
B-J auto	1	[-0,05;-0,00]	1	[-0,04;-0,01]
AutoBox1	0	[-0,01;0,02]	0	[-0,01;0,01]
AutoBox2	0	[-0,04;0,00]	1	[-0,03;-0,00]
AutoBox3	0	[-0,04;0,01]	0	[-0,03;0,00]
Robust-Trend	0	[-0,01;0,00]	0	[-0,02;0,00]
ARARMA	0	[-0,01;0,00]	0	[-0,01;0,00]
Auto-ANN	1	[-0,05;-0,00]	1	[-0,04;-0,01]
Flors-Pearc1	1	[-0,04;-0,00]	1	[-0,03;-0,01]
Flors-Pearc2	1	[-0,04;-0,01]	1	[-0,03;-0,01]
PP-Autocast	1	[-0,05;-0,00]	1	[-0,04;-0,01]
ForecastPro	1	[-0,05;-0,00]	1	[-0,04;-0,01]
SmartFcs	1	[-0,05;-0,00]	1	[-0,04;-0,01]
Theta-sm	1	[-0,05;-0,00]	1	[-0,04;-0,01]
Theta	1	[-0,03;-0,00]	1	[-0,03;-0,00]
RBF	0	[-0,01;0,00]	0	[-0,01;0,00]
ForcX	1	[-0,05;-0,00]	1	[-0,04;-0,01]
AAM1	0	[-0,01;0,00]	0	[-0,01;0,00]
AAM2	0	[-0,01;0,00]	0	[-0,01;0,00]

duas possibilidades básicas.

A primeira possibilidade é que um par qualquer de séries que compõem uma série multivariada não tenha qualquer relação. Isto pode ser resultado de um equívoco na modelagem ou fruto de um processo de exploração, onde não se sabe quais séries se relacionam a priori. Devido à forma de operação do AutoMFIS, serão geradas regras relacionando essas duas séries desconexas, resultando em possíveis regras "ruído", caso os passos de filtragem da base de regras não as elimine corretamente. A sensibilidade do SIF gerado pelo AutoMFIS à existência de séries não relacionadas será um dos pontos avaliados nos experimentos. A expectativa é que, em um caso como este, as séries dependam apenas de suas próprias defasagens, não sendo geradas regras que as relacionem.

Na existência de um relacionamento, deverão ser geradas regras descrevendo a interação entre as séries. A dimensão das regras e, conseqüentemente, sua interpretabilidade, dependerá da quantidade de séries. Uma regra que re-

Regra	Peso
SE $y_1(t-2)$ é MA & $y_3(t-3)$ é M ENTÃO $y_1(t)$ é MA	1,0
SE $y_2(t-1)$ é MA & $y_4(t-3)$ é MA ENTÃO $y_1(t)$ é MB	0,7573
SE $y_3(t-3)$ é M & $y_4(t-2)$ é MB ENTÃO $y_2(t)$ é MB	0,3955
SE $y_1(t-2)$ é MA & $y_3(t-3)$ é M ENTÃO $y_3(t)$ é MA	0,6828
SE $y_1(t-3)$ é M & $y_4(t-1)$ é M ENTÃO $y_4(t)$ é MB	0,3062

Tabela 4.8: Base de Regras - M3 Cluster 1.

Regra	Peso
SE $y_5(t-1)$ é B ENTÃO $y_1(t)$ é A	0,2927
SE $y_2(t-2)$ é MA & $y_3(t-2)$ é MA & $y_4(t-12)$ é A & $y_5(t-3)$ é M ENTÃO $y_2(t)$ é MA	0,2199
SE $y_2(t-12)$ é A ENTÃO $y_3(t)$ é A	0,4579
SE $y_5(t-2)$ é MB ENTÃO $y_4(t)$ é MB	1,0
SE $y_1(t-3)$ é A & $y_5(t-12)$ é A ENTÃO $y_5(t)$ é A	0,3993

Tabela 4.9: Base de Regras - M3 Cluster 2.

Regra	Peso
SE $y_2(t-1)$ é B ENTÃO $y_1(t)$ é M	0,7255
SE $y_1(t-1)$ é M ENTÃO $y_2(t)$ é A	1,0
SE $y_4(t-3)$ é M ENTÃO $y_3(t)$ é M	0,9487
SE $y_3(t-1)$ é A ENTÃO $y_4(t)$ é M	1,0

Tabela 4.10: Base de Regras - M3 Cluster 3.

Cluster	Todas as regras	Peso > 0.05	Peso > 0.1
1	53,75	24,50	13,75
2	62,40	33,80	11,60
3	25,75	25,50	24,75

Tabela 4.11: Tamanho média da base de regras por série.

lacione uma grande quantidade de séries diferentes é de difícil compreensão, o que faz com que o uso de pequenos conjuntos de 2 e 3 séries seja o mais indicado para a extração de conhecimento.

Com esses pontos em mente, foram construídos dois estudos de caso. No primeiro, abordando um problema do tipo Predador-Presa, foi estudado um relacionamento de causalidade com uma referência para a extração de conhecimento na forma do modelo de Lotka-Volterra, o mais difundido para a aplicação. Em adição, foi analisado o impacto da inclusão de uma terceira série não relacionada, por meio da injeção de uma série aleatória. O segundo caso é um problema artificial, criado a partir de uma série de dados reais de consumo de energia elétrica e de uma cópia defasada. Para esse caso, existe um

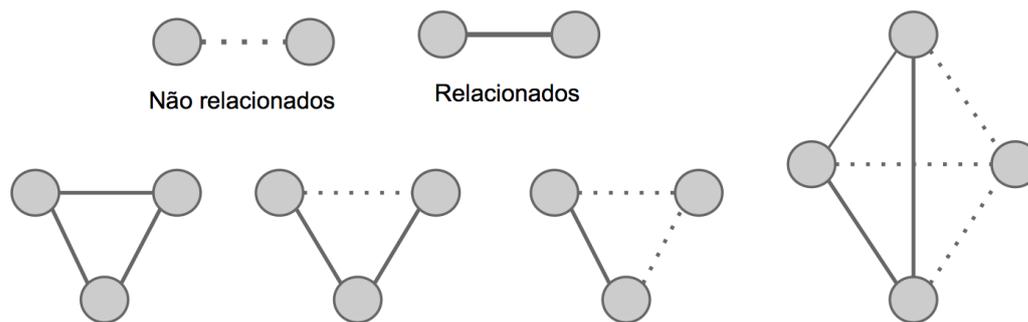


Figura 4.6: Exemplos de relacionamentos entre séries.

expectativa mais específica da resposta que o modelo deve oferecer, permitindo algumas verificações não possíveis para o caso Predador-Presa.

O objetivo é obter uma base de regras o mais compacta possível, em tamanho e quantidade de regras, cuja descrição do comportamento seja compatível com a referência estipulada previamente.

4.2.1

Caso 1: Predador-Presa

Neste estudo de caso, a série multivariada é formada por duas séries temporais com um relacionamento causal entre elas. Um problema de previsão com esta dimensão tem a vantagem de ser mais facilmente construído, além de contar com baixa complexidade, o que é de interesse em termos de interpretabilidade.

No caso das séries de um predador e sua presa, tem-se conhecimento prévio, em função da natureza do problema, de que existe uma relação de causalidade entre os comportamentos de cada uma das séries. Além disso, a dinâmica de tal problema pode ser compreendida em alto nível – i.e. sem conhecimento específico do problema –, o que facilita a validação das regras geradas pelo AutoMFIS. Um dos modelos disponíveis na literatura para tal problema é o descrito pelo sistema de equações de Lotka-Volterra [67]:

$$\frac{dx}{dt} = \alpha x - \beta xy \quad (4-1)$$

$$\frac{dy}{dt} = \delta xy - \gamma y \quad (4-2)$$

onde x é a população da presa e y a do predador; t representa o tempo e α , β , γ , δ são parâmetros reais que descrevem a interação entre as espécies.

A primeira equação mostra que a população da presa cresce proporcionalmente ao seu próprio valor, ou seja, na ausência do predador, crescerá indis-

crimindamente (supondo suprimento de comida ilimitado). O segundo termo, negativo, imprime a ação da população do predador sobre a da presa. Esse termo depende de ambas as populações e é proporcional à taxa com que estas se encontram. Quando qualquer uma delas atinge valor nulo, a predação cessa. De maneira oposta, a população do predador, descrita na segunda equação, cresce de acordo com a sua taxa de predação, semelhantemente à redução da população da presa. Observe-se que estes dois termos são multiplicados por constantes diferentes, pois a taxa com que a presa é consumida não é necessariamente a mesma com a qual o predador se multiplica. Em adição, este tem um decaimento natural proporcional a sua população atual, representando morte natural ou emigração, levando à extinção do predador na ausência de presa. Resumindo, a presa apresenta um crescimento natural subtraído da taxa com que sofre predação, enquanto o predador cresce de acordo com a disponibilidade de alimento subtraído de sua morte natural. Esse comportamento esperado foi o guia na avaliação da interpretabilidade.

Na validação das regras geradas pelo AutoMFIS, não haverá um paralelo direto com as equações de Lotka-Volterra. Nestas, a dinâmica do problema é descrita pelas taxas de variação das populações em função de seus valores absolutos, enquanto o SIF relacionará valores absolutos, tanto na entrada quanto na saída do sistema.

Um par predador-presa bem difundido na literatura é o do lince-do-canadá e sua presa, a lebre americana. Essa série multivariada tem origem na publicação "Fundamentals of Ecology" [68], onde aparece na forma de um gráfico, reproduzido na figura 4.7.

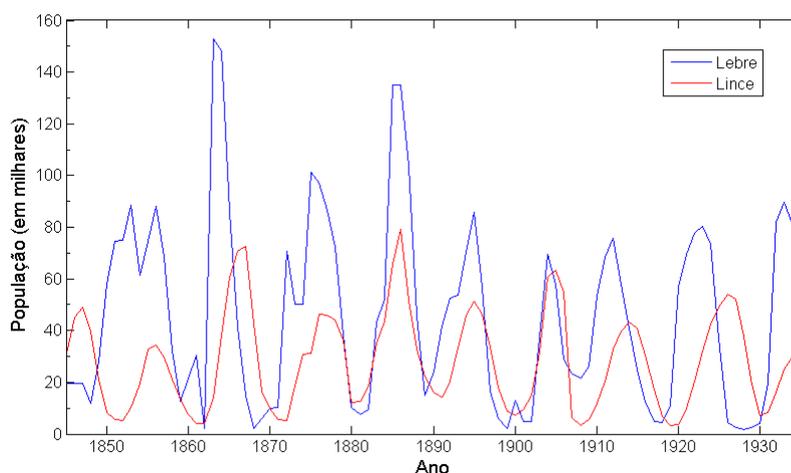


Figura 4.7: Gráfico das séries populacionais Lince-Lebre.

Esta base de dados é amplamente utilizada, mas existem algumas ressal-

vas a fazer. As séries não representam contagens precisas das populações das espécies nos períodos indicados, e sim registros históricos de Hudson's Bay a respeito da quantidade de peles comercializadas, o que gera apenas uma estimativa das populações reais. Além disso, alguns autores apontam que há dados faltantes na população de lince. Apesar destas ressalvas, o uso abrangente desta série ratifica a sua validade [69, 70, 71].

O primeiro passo neste estudo de caso é o estabelecimento de uma solução ótima em termos de acurácia, que pode ser realizado repetindo-se o processo de ajuste da seção anterior para as séries da competição M3. Observe-se, no entanto, que, como visto na figura 4.7, as séries são estacionárias, não havendo necessidade de pré-processamento. Quanto ao número de defasagens a considerar, parte-se da análise da correlação cruzada entre as séries e de suas autocorrelações, representadas nas figuras 4.8, 4.9 e 4.10.

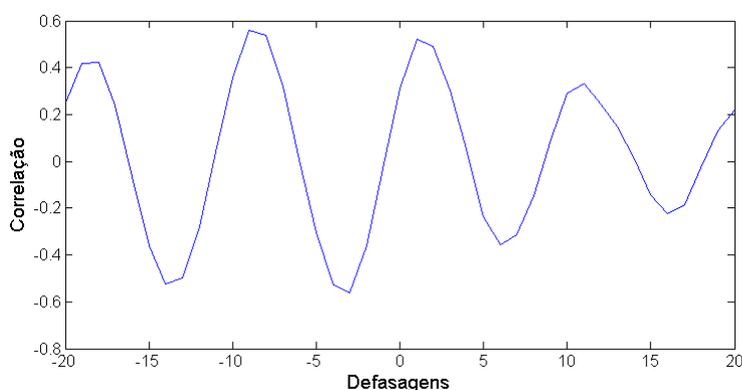


Figura 4.8: Correlação cruzada Lince X Lebre.

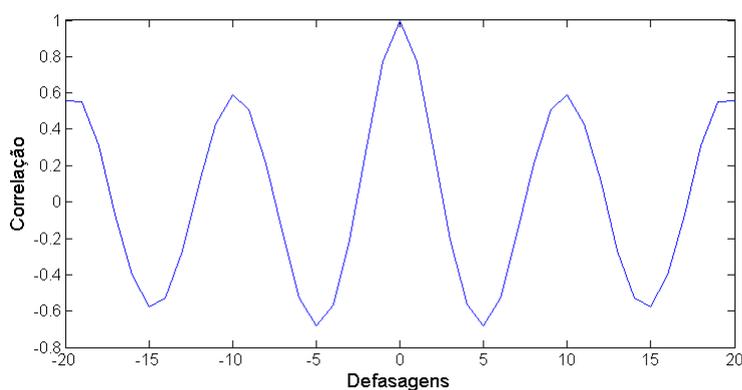


Figura 4.9: Autocorrelação Lince.

Essa série multivariada tem um aspecto cíclico, refletido na correlação, oscilando entre valores altos e baixos, em função da defasagem. Assim, não há sentido em utilizar uma grande quantidade de defasagens, bastando considerar

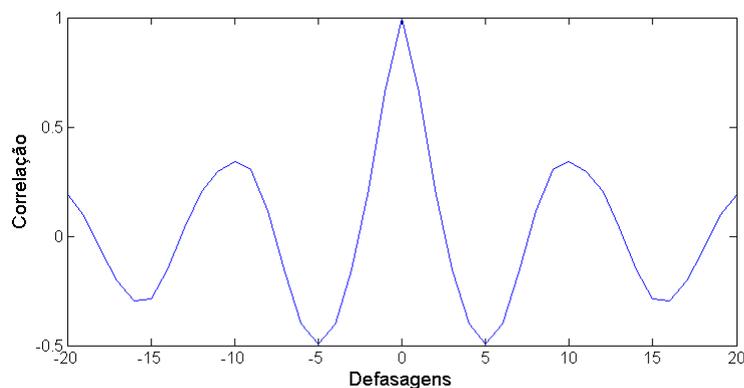


Figura 4.10: Autocorrelação Lebre.

apenas o primeiro período. É possível observar nos gráficos de correlação cruzada e autocorrelação que as correlações atingem valor mínimo por volta de 5 defasagens – voltando a crescer subsequentemente –, de forma que foi este o valor inicial utilizado na modelagem.

Como este é um problema construído a partir de dados reais, não há uma diretriz para o horizonte de previsão, que depende do objetivo da análise. Por essa razão, optou-se por seguir o padrão de 6 passos a frente da competição M3 para séries anuais. O tamanho do conjunto de teste será, conseqüentemente, também 6, com o conjunto de treinamento sendo formado pelo restante dos dados, analogamente ao caso anterior.

Como as duas abordagens para defuzzificação tendem a oferecer propostas de interpretabilidade distintas, buscou-se encontrar os melhores modelos para cada uma delas separadamente. Foi realizada a busca da melhor configuração seguindo a mesma metodologia da etapa de benchmarking, para as duas abordagens, para em seguida serem avaliados os impactos da variação dos seguintes parâmetros no modelo ótimo:

1. Critério de corte de regras
2. Limiar de corte de regras
3. Agregação ponderada X não-ponderada
4. Quantidade de conjuntos fuzzy

Após esses testes, foi também investigado o impacto da inclusão de uma série artificial aleatória no desempenho do modelo, e, finalmente, avaliou-se o aspecto da interpretabilidade do modelo ótimo ajustado.

Comparação das abordagens de defuzzificação

Observando-se apenas o SMAPE (tabela 4.12), a defuzzificação acoplada VAR apresentou desempenho superior à tradicional. Por outro lado, como o intuito é verificar se o comportamento do modelo segue a lógica das séries modeladas, vale observar também os gráficos das previsões para tentar entender a origem da discrepância de desempenho.

Defuzzificação	Lebre	Lince
Acoplada	60,61%	11,58%
Tradicional	57,29%	29,89%

Tabela 4.12: Erro SMAPE das melhores configurações para Predador-Presa.

Analisando-se visualmente os gráficos 4.11 e 4.12, a superioridade da abordagem acoplada em relação à tradicional não é tão clara.

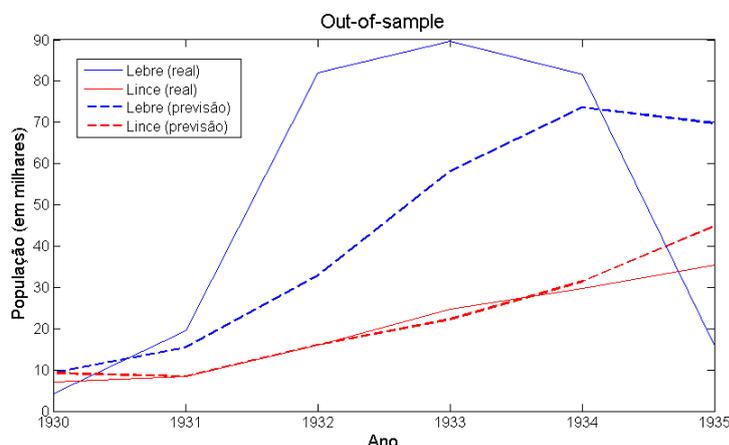


Figura 4.11: Previsão fora da amostra pela abordagem acoplada de defuzzificação.

Primeiramente, observa-se que a população de lince apresenta um crescimento sutil. Na defuzzificação pela Altura, o primeiro passo de previsão está um pouco afastado da referência, mas, após, apresenta o crescimento esperado, finalizando num valor próximo ao real. O VAR tem um comportamento muito próximo em todos os pontos do horizonte de previsão, com um leve desvio no último. O menor SMAPE deste método é resultado do melhor desempenho para a série do Lince.

Em segundo lugar, nenhum dos dois métodos pode prever a velocidade com que a população de lebres cresceu, para em seguida decair no último período. Em ambos os casos, porém, a população cresce até um valor máximo e, em seguida, inicia seu decaimento, representando um alinhamento conceitual, mesmo que numericamente não tão acurado quanto o desejado. Em con-

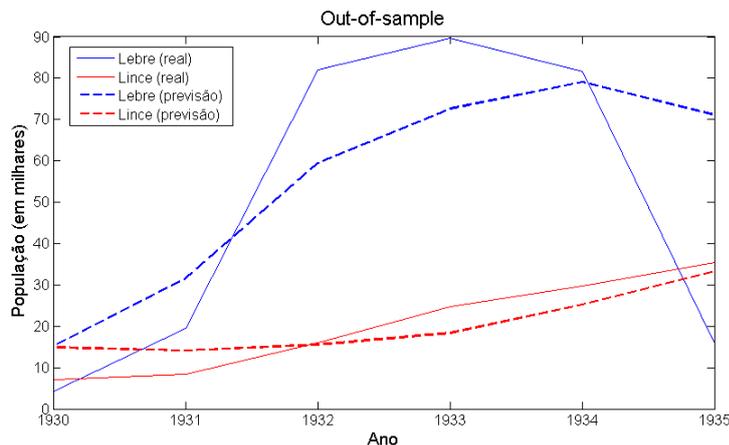


Figura 4.12: Previsão fora da amostra pela abordagem tradicional de defuzzificação.

trapartida ao que aconteceu para o caso do Lince, o método de defuzzificação tradicional ofereceu uma previsão mais próxima do comportamento real da série populacional da lebre.

Neste caso, dá-se preferência ao método tradicional por duas razões. Primeiramente, na configuração com a defuzzificação tradicional, foram geradas, em média, 24 regras, contra 52 no método acoplado (tabela 4.13). Em adição, no método acoplado, os termos linguísticos das variáveis de saída são embaralhados, com seus centros realocados de acordo com a metodologia apresentada no capítulo 3. Em segundo lugar, – mais importante para interpretabilidade –, na defuzzificação acoplada, há modificações nos dicionários das variáveis de saída. Como visto no capítulo 3, o método acoplado é uma variação da defuzzificação por altura, onde se consideram os centros de todos os conjuntos fuzzy de todas as variáveis, não só os da variável a ser defuzzificada. Para incorporar as demais variáveis na defuzzificação de cada uma, os centros são ajustados, conforme mostrado na tabela 4.14.

Defuzzificação	Lebre	Lince
Acoplada	48	56
Tradicional	25	23

Tabela 4.13: Quantidade de regras das melhores configurações para Predador-Presa.

Observa-se que a interpretabilidade é comprometida, com alguns conjuntos fuzzy inutilizados (centro nulo) e outros posicionados fora de seus respectivos universos de discurso. Mesmo em casos onde não ocorrem esses pontos, o dicionário ainda assim se tornará bem mais vago, com alto índice de so-

Originais Lebre	Originais Lince	Ajustados Lebre	Ajustados Lince
1,80		-14,45	16,31
20,66		1,85	20,58
39,51		5,25	28,59
58,37		44,74	26,00
77,23		43,37	20,69
96,08		53,62	36,43
114,94		111,25	1,61
133,79		-0,00	0,00
152,65		156,54	46,43
	3,19	13,36	-26,60
	12,71	21,96	-10,91
	22,23	8,34	-0,82
	31,75	27,42	2,76
	41,27	20,86	19,25
	50,79	33,06	17,38
	60,31	9,13	43,11
	69,83	-0,00	0,00
	79,35	6,83	53,08

Tabela 4.14: Centros dos conjuntos Fuzzy: Originais X Ajustados.

breposição de conjuntos fuzzy, fato que se agrava com séries multivariadas de maior ordem.

Critérios de corte de regra

Um outro ponto a ser avaliado é o dos critérios de corte de regras utilizados pela geração semi-exaustiva. A melhor configuração foi obtida para o critério de ativação com limiar 0,4, ou seja, quando a premissa é ativada, ela apresenta uma ativação média acima de 0,4. As duas outras formas de eliminar premissas são por cardinalidade e frequência. Alterar apenas o critério de corte na melhor configuração resulta em uma grande perda de desempenho, com erros SMAPE de 57,68% e 68,41%. Estes valores não são, necessariamente, indicativos do desempenho inerente a estes critérios, pois os resultados do modelo são uma composição dos diferentes parâmetros de configuração. Assim este resultado negativo meramente indica que os dois critérios não funcionam bem para essa configuração específica. De toda maneira, esse experimento oferece uma perspectiva interessante quanto à distribuição de regras por consequente resultante do uso de cada um dos critérios. Essas distribuições podem ser visualizadas na tabela 4.15.

Dado que os limiares dos critérios alternativos foram ajustados de modo que a base de regras tivesse a mesma dimensão, há uma grande discrepância

Série	Conjunto Fuzzy	Ativação	Cardinalidade	Frequência
Lebre	A	49,09%	24,70%	0,35%
	B	29,40%	22,91%	18,57%
	C	14,88%	48,43%	79,95%
	D	1,85%	1,14%	0,18%
	E	4,79%	2,82%	0,95%
Lince	A	4,31%	25,59%	4,31%
	B	13,70%	29,63%	18,33%
	C	12,64%	39,85%	76,94%
	D	21,55%	4,26%	0,36%
	E	47,80%	0,67%	0,06%

Tabela 4.15: Quantidade de regras por conjunto fuzzy das variáveis de saída (% do total).

na cobertura oferecida aos conjuntos fuzzy das variáveis de saída. Devido ao caráter cíclico, entradas de dados que resultam em valores centrais de saída são muito mais comuns do que os extremos, ou seja, as regras que representam estas entradas serão menos ativadas no treinamento – fator determinante para ambos os critérios alternativos. Na tabela está ilustrado o viés dos critérios de cardinalidade e frequência pelos termos linguísticos centrais, sendo especialmente intenso para este último, em que o conjunto fuzzy central é o consequente para 75% a 80% das regras. Este tipo de comportamento é minimizado quando os conjuntos fuzzy são triangulares e espaçados de acordo com a distribuição de valores, equilibrando suas ocorrências na base de dados. De qualquer forma, uma métrica menos sensível à definição dos conjuntos fuzzy é preferível, tornando o modelo menos instável a ajustes de parâmetros.

No entanto, o critério de corte da ativação não é isento de crítica. Analisando a sua distribuição, fica claro que dá preferência aos extremos em detrimento do conjunto central, apresentando um viés oposto aos outros dois critérios. Observando a distribuição para as duas séries, é possível notar, porém, que este critério não beneficia os conjuntos extremos indiscriminadamente, sendo, no caso de ambas as séries, apenas um conjunto extremo a demonstrar maior representatividade na base de regras. Esse comportamento pode vir a ser o ideal para este problema, como também pode ser inerente ao critério. Esse ponto será analisado no próximo estudo de caso.

Variação do limiar de corte de regras

Ainda no contexto dos critérios de corte, é útil analisar os impactos do ajuste do limiar na base de regras. Este parâmetro é um ajuste direto entre complexidade e interpretabilidade. À medida que se reduz o limiar, a

base de regras aumenta de dimensão, dificultando a interpretabilidade. Este crescimento da base de regras, porém, não necessariamente resulta em uma maior acurácia, podendo saturar o sistema com regras de baixa confiança. O resultado do estudo do impacto da variação do limiar está ilustrado na tabela 4.16.

Limiar	SMAPE Médio	# médio de regras/série
0,3	53,98%	34
0,35	60,09%	26
0,4	43,59%	24
0,45	48,76%	17,5
0,5	71,83%	21,5
0,55	50,00%	21,5

Tabela 4.16: Erro e quantidade de regras por limiar de corte.

Os resultados desse experimento negam a noção prévia de que o aumento do limiar sempre resulta em uma menor complexidade. Isso é apenas verdade quando se considera apenas a etapa de geração semi-exaustiva, porém, devido ao uso de um método de agregação ponderado – que por si só representa um passo de filtragem da base de regras –, os resultados acabam se tornando mais imprevisíveis. Faz-se necessário, assim, avaliar uma grande faixa de valores para o limiar, maximizando a probabilidade de se encontrar a configuração mais adequada aos objetivos.

Agregação ponderada X não-ponderada

A imprevisibilidade da resposta do modelo a alterações de parâmetros é uma clara desvantagem na busca pela melhor configuração. Por esta razão, o uso dos métodos ponderados devem que oferecer uma clara vantagem em relação a métodos não-ponderados, como o do maximum. O primeiro ponto é verificar a importância desta segunda camada de filtragem da base de regras. A quantidade de regras por método de agregação é apresentada na tabela 4.17.

Método	# médio de regras/série
Não-ponderado	14186
Ponderado	24

Tabela 4.17: Quantidade de regras por método de agregação.

A diferença é extrema, sendo possível argumentar que o modelo não é interpretável sem o uso do método ponderado. Um outro ponto a ser avaliado é a hipótese levantada no capítulo 3, de que o uso de um método não-ponderado tornaria o sistema insensível, com algumas poucas regras de pequeno porte e

baixa confiança ditando o seu comportamento, fazendo com que as previsões ficassem flutuando em torno do mediana do universo de discurso das variáveis. Comparando as previsões dentro da amostra das duas abordagens, a suspeita aparenta ter se confirmado, como visto nas figuras 4.13 e 4.14.

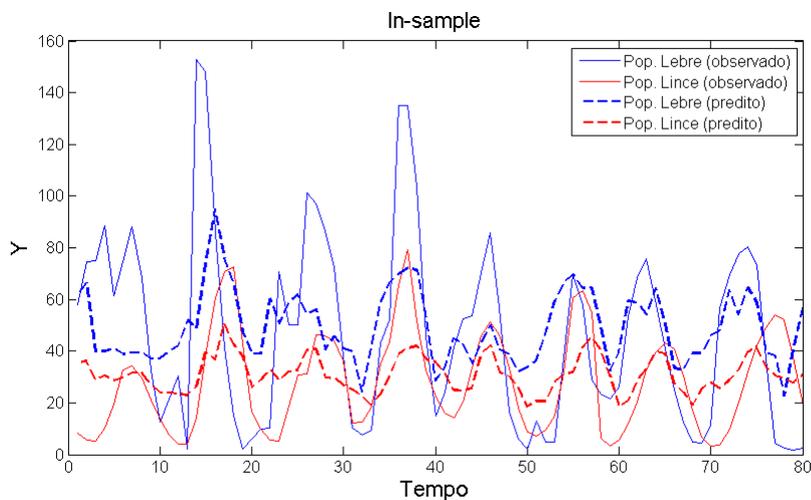


Figura 4.13: Previsões dentro da amostra com agregação não-ponderada.

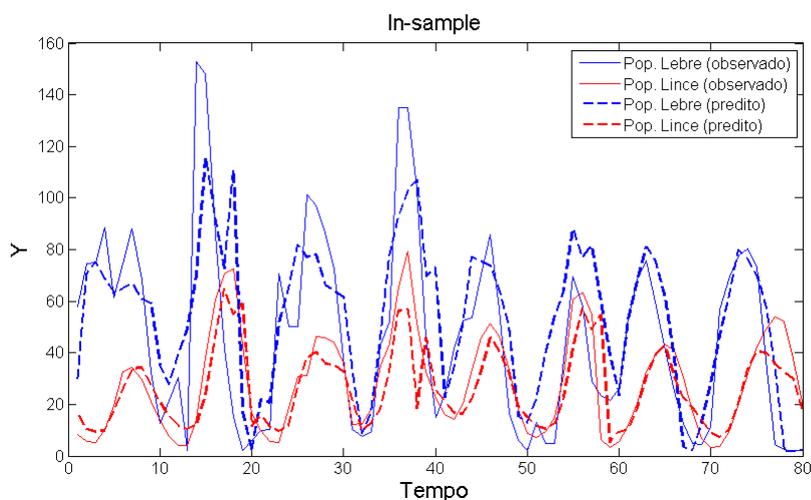


Figura 4.14: Previsões dentro da amostra com agregação ponderada.

Fica clara a dificuldade do modelo de prever valores próximos aos extremos. Quando se observa também a figura 4.15, a agregação não-ponderada se confirma como ineficaz também no âmbito de acurácia de previsão.

Varição da quantidade de conjuntos fuzzy

Finalmente, o desempenho do sistema é avaliado para 5, 7 e 9 conjuntos fuzzy por variável. O aumento da quantidade de conjuntos fuzzy aumenta também a especificidade de cada regra, que passa a tratar de uma faixa mais

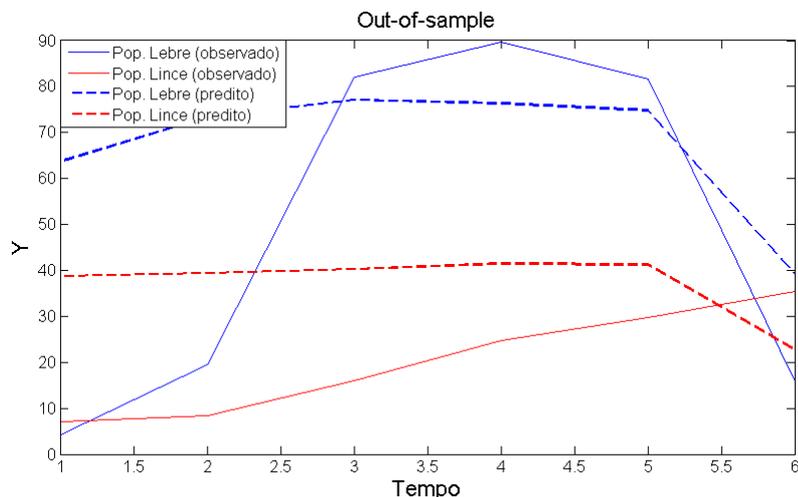


Figura 4.15: Previsão fora da amostra com agregação não-ponderada.

restrita de valores. Com o aumento do dicionário, é aumentada também a dimensão da base de regras, o que dificulta a interpretabilidade e torna ainda maiores os custos computacionais. O valor de 5 conjuntos fuzzy da melhor configuração já apresenta uma base significativa para previsão, estando dentro da faixa de limite de compreensão de 7 ± 2 conceitos linguísticos, o que torna quaisquer ganhos em especificidade menos valiosos. As figuras 4.16 e 4.17 exibem as previsões da melhor configuração quando a quantidade de conjuntos é aumentada.

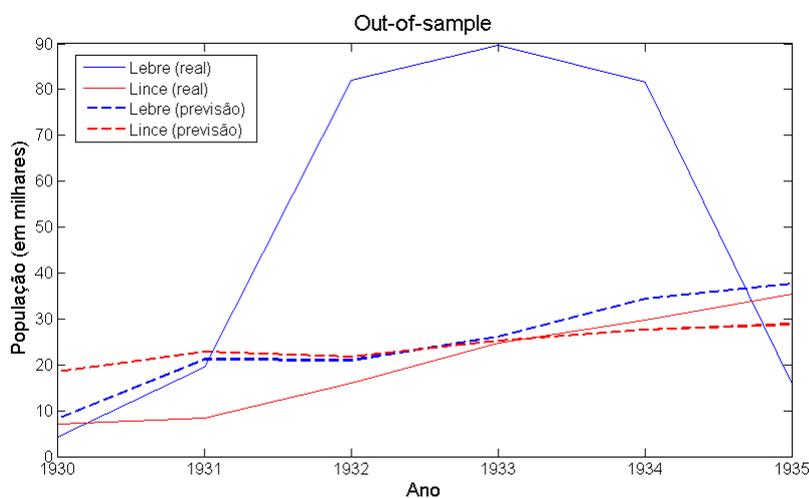


Figura 4.16: Previsão fora da amostra com 7 conjuntos fuzzy.

Neste caso, especificamente, o aumento da quantidade de conjuntos fuzzy ocasiona perdas significativas de acurácia, ao mesmo tempo que crescem o custo computacional e a complexidade na interpretação. A escolha inicial de 5 conjuntos se mostrou a mais eficaz.

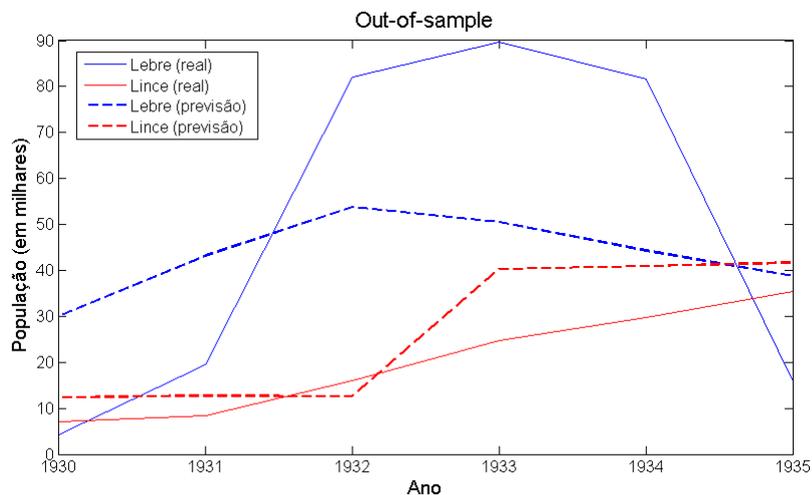


Figura 4.17: Previsão fora da amostra com com 9 conjuntos fuzzy.

Com o modelo que oferece o melhor equilíbrio entre acurácia e interpretabilidade definido, avalia-se a sua sensibilidade do à injeção de ruído na série temporal múltipla, na forma de uma terceira série não relacionada, gerada de maneira aleatória. Essa avaliação é de grande importância para entender se o AutoMFIS pode ser utilizado de maneira puramente orientada a dados, supondo ignorância total acerca destes, ou se são exigidos cuidados na construção do problema de modo a se obterem resultados significativos.

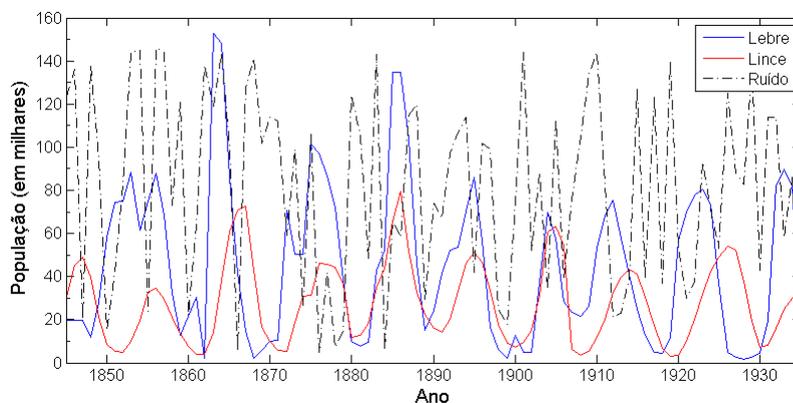


Figura 4.18: Séries populacionais Lince-Lebre adicionadas de uma série aleatória.

Avaliando a correlação cruzada entre cada uma das duas séries de população e a nova série aleatória de ruído, confirma-se que esta tem baixa correlação com as séries reais, conforme ilustrado nas figuras 4.19 e 4.20.

A inclusão da série adicional faz com que não seja possível avaliar a configuração ótima do problema. Para contornar esse problema, aumentou-se o limiar de corte para 0,55. Mesmo com um limiar mais agressivo e a

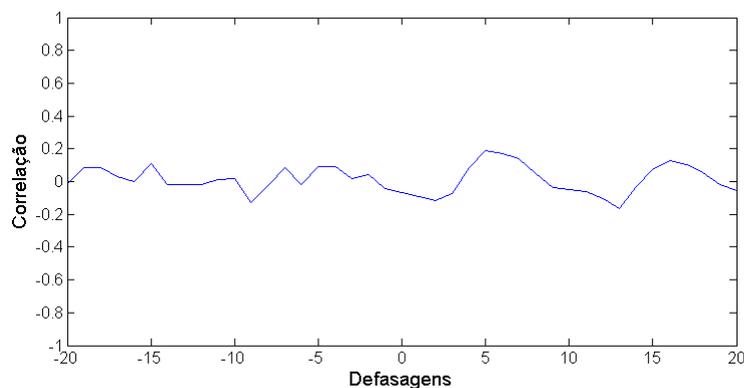


Figura 4.19: Correlação cruzada Lince X Ruído.

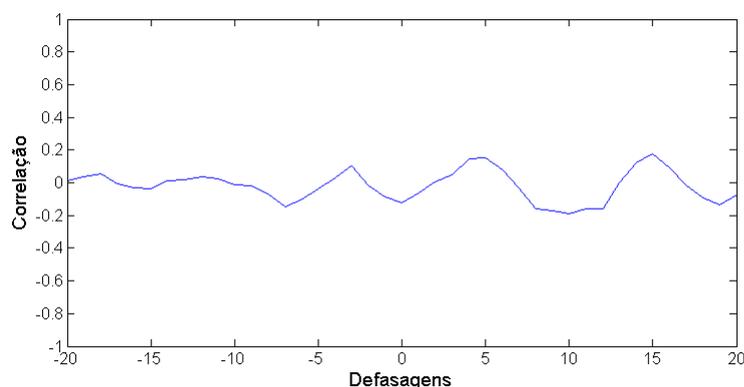


Figura 4.20: Correlação cruzada Lebre X Ruído.

utilização da agregação ponderada, percebe-se que a inclusão da série falsa causou uma perturbação no comportamento do modelo, evidenciada pelos novos erro SMAPE e tamanho da base de regras (tabela 4.18).

Séries	SMAPE Médio	# médio de regras/série
Lince-Lebre	50,00%	30
Lince-Lebre + Ruído	54,95%	25,5

Tabela 4.18: Erro e quantidade de regras com e sem série ruído.

Mesmo a adição de uma série com baixíssima correlação causa alterações no SIF gerado, o que corrobora a idéia de que uma abordagem puramente orientada a dados, supondo ignorância total, não é a ideal, como já se havia observado no caso do cluster 2 dos benchmarks M3.

Extração de conhecimento

Observa-se, agora, a base de regras para avaliação do conhecimento gerado pelo ajuste do modelo. Os 5 termos linguísticos são: ‘muito baixo’ (MB),

‘baixo’ (B), ‘médio’ (M), ‘alto’ (A) e ‘muito alto’ (MA). As principais regras geradas para a série da Lebre estão mostradas na tabela 4.19.

Regra	Peso
SE lince(t-3) é MB & lebre(t-1) é MB ENTÃO lebre é MB	0,6721
SE lince(t-4) é M & lebre(t-3) é M ENTÃO lebre é MB	0,0437
SE lince(t-5) é B ENTÃO lebre é B	0,0666
SE lince(t-1) é MB & lebre(t-2) é MB ENTÃO lebre é B	0,3863
SE lince(t-1) é A & lebre(t-4) é MB ENTÃO lebre é B	0,1303
SE lince(t-2) é B ENTÃO lebre é M	0,1396
SE lince(t-3) é MB ENTÃO lebre é M	0,3179
SE lince(t-4) é M & lebre(t-3) é MB ENTÃO lebre é M	0,1222
SE lince(t-4) é M & lebre(t-1) é MA ENTÃO lebre é A	0,6380
SE lince(t-5) é MB & lebre(t-5) é MB ENTÃO lebre é A	0,1896
SE lebre(t-4) é B & lince(t-1) é MA & lince(t-4) é M ENTÃO lebre é A	0,0320
SE lince(t-2) é M & lebre(t-3) é MA ENTÃO lebre é MA	0,0588
SE lince(t-4) é MB & lebre(t-1) é MA ENTÃO lebre é MA	0,0408
SE lince(t-1) é M & lince(t-2) é M & lince(t-3) é B & lebre(t-4) é MB ENTÃO lebre é MA	0,1001

Tabela 4.19: Base de Regras - Lebre.

De uma maneira geral, é possível observar que, mesmo fixando-se um limite superior de 9 termos nas premissas, a base de regras é formada por regras de baixa dimensão. Isto é condizente com o modelo (de referência) Lotka-Volterra, que faz uso de pouca informação para realizar suas estimativas.

Um tema recorrente na série da Lebre é regras que envolvem maiores defasagens descreverem as relações esperadas de maneira indireta. Um exemplo disso é a regra “**SE** lince(t-5) é B **ENTÃO** lebre é B”, que é um descritor do comportamento cíclico do sistema: onde se o Lince tem um população baixa agora, daqui a alguns períodos será a vez da Lebre. Esta regra já sugere que as duas séries competem, mesmo de maneira pouco clara. Por outro lado, há regras como “**SE** lince(t-4) é M & lebre(t-1) é MA **ENTÃO** lebre é A” para de difícil. O significado da população do lince estar “média” há 4 instantes não diz muito, pois pode ter crescido ou diminuído até o instante (t-1). O fato de a população da Lebre estar “muita alta” parece ser um indicativo de que na verdade a do Lince decaiu, voltando a aumentar em t-1 por causa da abundância de comida (Lebre MA) – causando o decaimento da população de Lebres para “alta”. É possível traçar este tipo de linha de raciocínio quando se tem uma base de comparação, mas, na extração de conhecimento em um problema desconhecido, uma regra como esta teria valor baixo.

Foram geradas, também, regras mais imediatas, com menores defasagens na premissa, como, por exemplo: “**SE** *lince*($t-1$) é MB & *lebre*($t-2$) é MB **ENTÃO** *lebre* é B”. Esta regra mostra que, quando as duas populações estão no extremo mais baixo, a da Lebre consegue crescer de “muito baixa” para “baixa”. As defasagens de menor ordem permitem um paralelo mais direto e fácil com a referência, e, na ausência de uma referência, são de melhor compreensão.

É possível observar uma situação análoga à acima na base de regras referente ao Lince (tabela 4.20), com regras mais claras, utilizando menores defasagens (“**SE** *lince*($t-2$) é M & *lebre*($t-1$) é MA **ENTÃO** *lince* é MA”), e outras nem tanto, fazendo uso de maiores defasagens (“*lince*($t-5$) é MB & *lebre*($t-2$) é M **ENTÃO** *lince* é M”).

Regra	Peso
SE <i>lince</i> ($t-5$) é M & <i>lebre</i> ($t-1$) é MB ENTÃO <i>lince</i> é MB	0,1579
SE <i>lebre</i> ($t-5$) é M ENTÃO <i>lince</i> é B	0,1468
SE <i>lince</i> ($t-1$) é M & <i>lebre</i> ($t-1$) é M ENTÃO <i>lince</i> é M	0,3591
SE <i>lince</i> ($t-5$) é MB & <i>lebre</i> ($t-2$) é M ENTÃO <i>lince</i> é M	0,1526
SE <i>lince</i> ($t-5$) é MB & <i>lebre</i> ($t-5$) é MB ENTÃO <i>lince</i> é A	0,2602
SE <i>lince</i> ($t-5$) é MB & <i>lince</i> ($t-4$) é MB & <i>lince</i> ($t-1$) é M & <i>lebre</i> ($t-3$) é MB ENTÃO <i>lince</i> é A	0,4336
SE <i>lince</i> ($t-2$) é M & <i>lebre</i> ($t-1$) é MA ENTÃO <i>lince</i> é MA	0,3949
SE <i>lince</i> ($t-5$) é MA & <i>lebre</i> ($t-3$) é MA ENTÃO <i>lince</i> é MA	0,3677
SE <i>lince</i> ($t-3$) é B & <i>lince</i> ($t-2$) é M & <i>lebre</i> ($t-4$) é MB ENTÃO <i>lince</i> é MA	0,2374

Tabela 4.20: Base de Regras - Lince.

Como o AutoMFIS já garante regras de menor porte, e o uso de muitas defasagens torna a interpretação menos direta, realiza-se um teste adicional: avaliar a hipótese do uso de apenas uma defasagem (como no Lotka-Volterra), ignorando o que a análise de correlação cruzada indica sobre os termos autoregressivos do sistema. Repetindo o processo de busca da solução ótima, foram obtidos os resultados expostos nas tabelas 4.21 e 4.22.

Defasagens	Lebre	Lince
5	57,29%	29,89%
1	33,92%	35,07%

Tabela 4.21: Erro SMAPE das melhores configurações para Predador-Presa, por quantidade de defasagens.

Surpreendentemente, o desempenho do modelo com apenas uma defasagem foi, na média, superior ao de 5 defasagens: um pouco inferior para a

Defasagens	Lebre	Lince
5	25	23
1	15	15

Tabela 4.22: Quantidade de regras das melhores configurações para Predador-Presa, por quantidade de defasagens.

série do Lince mas consideravelmente superior na da Lebre, apresentando um resultado mais equilibrado. Em relação à configuração com a defuzzificação acoplada, agora o resultado médio mostrou-se levemente superior, e sem o ônus de perda de interpretabilidade. A menor quantidade de variáveis desta configuração resultou em uma base de regras menor, com uma média de 15 regras, o que contribuiu para um menor custo computacional. Como se consideram menos defasagens, as regras também são mais compactas, apresentando premissas menores, representando um ganho duplo no quesito complexidade. A melhor previsão realizada pode ser vista na figura 4.21.

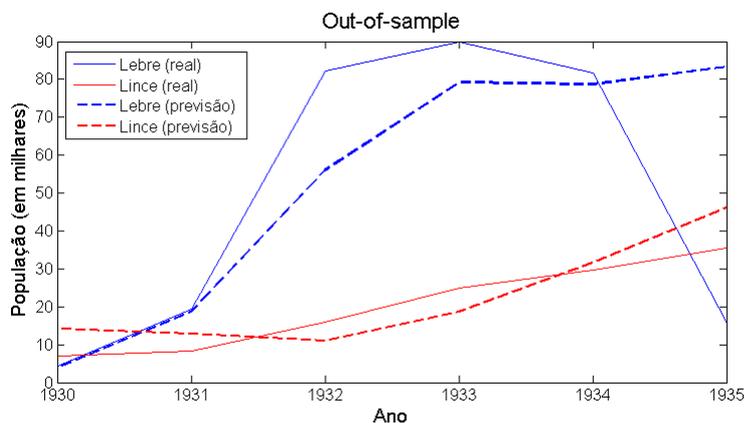


Figura 4.21: Previsão fora da amostra para o modelo com 1 defasagem.

A maior falha ocorreu no último passo de previsão, devida à queda repentina da população da Lebre, mas nenhuma das outras configurações foi precisa em descrever esse comportamento.

A melhor configuração para uma defasagem utilizou 7 conjuntos fuzzy – com a adição dos termos “muito muito baixo” (MMB) e “muito muito alto” (MMA) ao dicionário. As bases de regras resultantes podem ser vistas nas tabelas 4.23 e 4.24.

No caso da série da Lebre, apenas os três conjuntos fuzzy superiores (A, MA e MMA) foram contemplados com regras relacionando as duas séries, mas estas estão consistentes. Quando a população de lince e lebre está na metade mais baixa do domínio, a população de lebre tenderá a crescer, dada a baixa taxa de predação. Os demais conjuntos fuzzy da Lebre receberam regras

Regra	Peso
SE lince(t-1) é M & lebre(t-1) é MB ENTÃO lebre é A	0,0518
SE lince(t-1) é MB & lebre(t-1) é B ENTÃO lebre é A	0,0219
SE lince(t-1) é B & lebre(t-1) é B ENTÃO lebre é A	0,3802
SE lince(t-1) é B & lebre(t-1) é M ENTÃO lebre é A	0,3802
SE lince(t-1) é M & lebre(t-1) é B ENTÃO lebre é MA	0,5496
SE lince(t-1) é M & lebre(t-1) é M ENTÃO lebre é MA	0,3087
SE lince(t-1) é MMB & lebre(t-1) é MMB ENTÃO lebre é MMA	0,5890
SE lince(t-1) é MMB & lebre(t-1) é MB ENTÃO lebre é MMA	0,3438

Tabela 4.23: Base de Regras - Lebre (1 defasagem).

simples de tamanho 1, dependendo apenas de sua própria defasagem, o que significa que o AutoMFIS não conseguiu detectar o impacto da população de lince no decaimento da população de lebres. Vale salientar que, como apenas os conjuntos superiores aparecem nos consequentes, ao mesmo tempo que as premissas contêm apenas conjuntos da parte inferior do universo (MMB, MB, B e M), este SIF tende rapidamente a não apresentar resposta em uma previsão multi-step, por não ter regras contemplando as situações a que o sistema é levado. Caso se deseje uma base de regras mais completa, as configurações podem ser modificadas de maneira similar ao que já foi apresentado nesta seção.

Regra	Peso
SE lebre(t-1) é MB ENTÃO lince é MMB	0,2065
SE lebre(t-1) é MMB & lince(t-1) é MB ENTÃO lince é MMB	0,2157
SE lebre(t-1) é MMB ENTÃO lince é MB	0,4971
SE lebre(t-1) é B & lince(t-1) é MB ENTÃO lince é B	0,5563
SE lebre(t-1) é M & lince(t-1) é B ENTÃO lince é B	0,4437
SE lebre(t-1) é M ENTÃO lince é MA	0,1211
SE lebre(t-1) é A ENTÃO lince é MMA	0,7189

Tabela 4.24: Base de Regras - Lince (1 defasagem).

As primeiras regras abordam uma situação similar à da série das lebres: um cenário de baixa predação. Neste caso, como esperado, o predador mantém ou diminui sua população baixa. Em adição, as duas últimas regras mostram o cenário de abundância de alimento, quando a predação atinge altos valores.

A despeito de alguns lacunas, a opção com uma única defasagem possibilitou uma extração significativa de informação, condizente com a expectativa para um problema do tipo predador-presa, além de oferecer o melhor desempenho em acurácia.

Este caso corrobora a noção de que o uso “às cegas” do modelo não resultará em resultados tão bons quanto os que se obteriam arquitetando o AutoMFIS especificamente para cada problema (fato que já havia sido observado no cluster 2 da fase de benchmarking). O modelo não é capaz de compensar uma definição menos rigorosa do problema, sendo extremamente suscetível a variações na configuração e nos dados.

4.2.2

Caso 2: Consumo de energia (indústria)

O uso de um problema real é interessante pois simula uma situação real de uso de um modelo como o AutoMFIS, comprovando seu valor prático.

No caso Predador-Presa, utilizou-se como base de comparação um modelo matemático, por trás do qual há uma lógica. Tal comparação, porém, não é exata, pois não se tem em mente exatamente as regras que o AutoMFIS deve gerar; conhece-se apenas um comportamento de alto nível que deve ser seguido de modo a avaliar a consistência da base de regras.

Uma alternativa interessante para avaliar a capacidade do modelo de extrair relações de causalidade é construir um problema artificial a partir de séries reais. Seleciona-se uma série real qualquer – dotada de um caráter ruidoso e de maior dificuldade de modelagem quando comparada a séries geradas por funções – e duplicá-la, defasando a cópia. Isso cria um efeito de causalidade da série original na atrasada, na exata quantidade de defasagens definida, permitindo a criação de um conjunto de regras esperadas. Por exemplo, se a série duplicada for atrasada em 3 instantes em relação à original, deve-se esperar regras do tipo “*SE* $y_1(t - 3)$ *É ‘alto’ ENTÃO* $y_2(t)$ *É ‘alto’*”, onde $y_1(t)$ e $y_2(t)$ dizem respeito às séries original e atrasada, respectivamente.

Escolheu-se a série de consumo de energia pela indústria no sudeste do Brasil, durante toda a década de 90 [72]. O segmento e região foram escolhidas por oferecerem um período de crescimento sem mudanças bruscas de nível e sem apresentar caráter sazonal, como no caso do consumo residencial. As figuras 4.22 e 4.23 representam, respectivamente, a série original (com o período utilizado marcado entre as linhas vermelhas) e as séries duplicada e defasada.

Selecionou-se uma defasagem de 2 passos, de modo a poder considerar uma faixa de defasagens contínua, sem recorrer a “saltos” nas defasagens utilizadas pelo modelo, como observado no cluster 2 do experimento de benchmarking.

Pode-se notar que a série apresenta uma tendência de crescimento. Conforme discutido anteriormente, essa característica deve tratada por eliminação

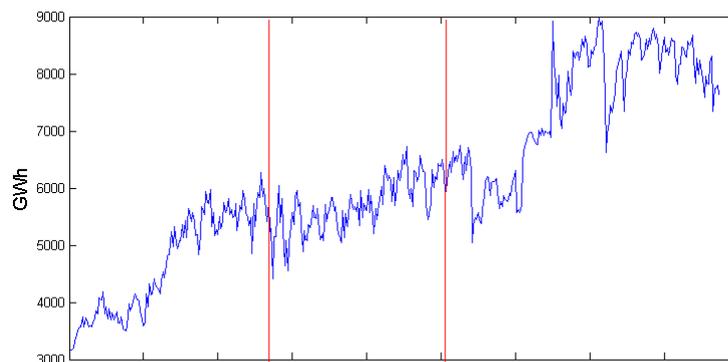


Figura 4.22: Série de consumo de energia (Indústria, Sudeste).

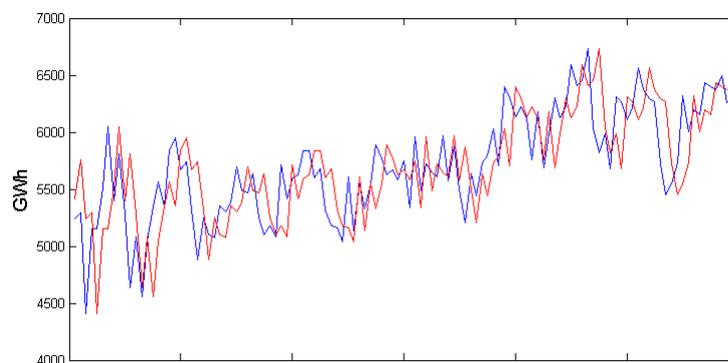


Figura 4.23: Série de consumo de energia multivariada.

da tendência ou diferenciação. Há que ter cautela neste quesito devido à questão da interpretabilidade.

O uso das diferenças é intuitivo, já que se consideram as taxas de variação. Os conceitos linguísticos "aumentou muito" e "diminuiu pouco" são bem compreendidos. Em contrapartida, a eliminação de tendência cria certas dificuldades em termos de interpretação, pois informação é retirada das séries, ao mesmo tempo em que se cria um referencial mais vago para a série resultante – pode haver uma situação em que um valor negativo ainda possa representar um aumento em relação ao instante anterior. Como ambos os métodos são comparáveis em termo de acurácia, aliado ao fato de que o objetivo é oferecer a melhor interpretabilidade, opta-se pela diferenciação.

A partir das séries de diferenças, são calculadas a autocorrelação e a correlação cruzada, de modo a estabelecer a quantidade de defasagens considerada pelo modelo. Observando-se a autocorrelação da série, há uma correlação forte apenas na primeira defasagem (figura 4.24), indicando uma correlação negativa da série com o seu valor um passo atrás. Como esperado,

a correlação cruzada (figura 4.25) apresentou um pico de valor máximo na segunda defasagem.

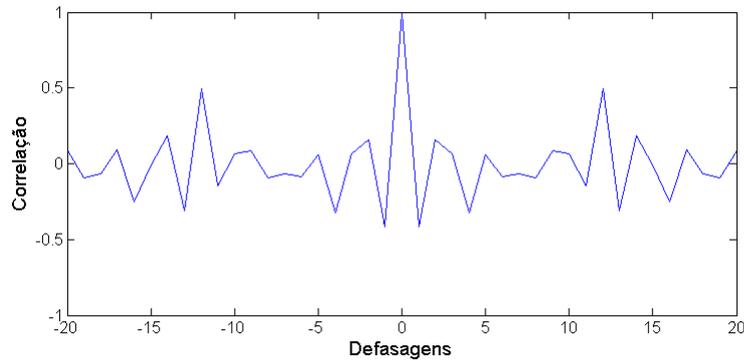


Figura 4.24: Autocorrelação do consumo de energia.

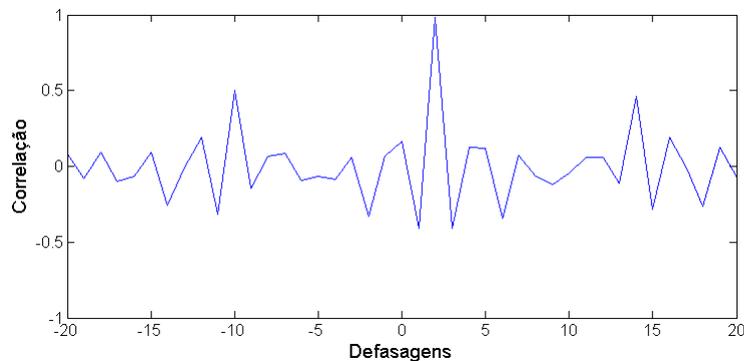


Figura 4.25: Correlação Cruzada do consumo de energia com sua cópia defasada.

Manteve-se, neste caso, o processo de busca da configuração ótima utilizado nos outros experimentos. Como se trata de uma série mensal, considerou-se um horizonte de 18 passos – o que também define o tamanho dos conjuntos de teste e treinamento, como nos dois primeiros casos –, utilizando, novamente, o padrão da competição M3 para séries com esse intervalo. Além disso, muitas das descobertas realizadas no caso Predador-Presa foram aproveitadas para esse segundo caso. Foram desconsideradas, por exemplo, a defuzzificação acoplada e a agregação não-ponderada. O ajuste concentrou-se no parâmetro que demonstrou ser o mais crucial nos outros experimentos: escolha das defasagens. O processo de busca da configuração ótima foi realizada separadamente para diferentes quantidades de defasagens consideradas, desde 2, valor mínimo necessário para captar a causalidade construída, até 4, número que vai além do que parece razoável pela análise da autocorrelação e correlação cruzada. Os resultados estão mostrados nas tabelas 4.25, em termos de erro SMAPE de previsão out-of-sample.

Defasagens	Consumo	Consumo Atrasado
2	3,86%	3,61%
3	3,48%	3,30%
4	3,91%	3,75%

Tabela 4.25: Erro SMAPE das melhores configurações para Consumo de Energia.

As três opções de defasagens apresentaram erro SMAPE similar. Notam-se diferenças nas bases de regras, cujas dimensões podem ser vistas na tabela 4.26. Em teoria, todas as regras necessárias para a previsão da série atrasada são geradas em todos os três casos, podendo a diferença estar na base de regras para a série original. Em todo caso, avaliar-se-á se as regras esperadas foram de fato captadas pelo AutoMFIS em todos os casos.

Defasagens	Consumo	Consumo Atrasado
2	13	7
3	30	10
4	18	11

Tabela 4.26: Quantidade de regras das melhores configurações para Consumo de Energia.

A composição da base de regras do Consumo atrasado (excluindo as de peso insignificante) pode ser observada nas tabelas 4.27, 4.28 e 4.29, para 2, 3 e 4 defasagens, respectivamente.

Regra	Peso
SE consumo(t-2) é B ENTÃO consumo atrasado é B	0,9973
SE consumo(t-2) é M ENTÃO consumo atrasado é M	0,9991
SE consumo(t-2) é A ENTÃO consumo atrasado é A	0,9921

Tabela 4.27: Base de Regras - Consumo Atrasado (2 defasagens).

Regra	Peso
SE consumo(t-2) é MB ENTÃO consumo atrasado é MB	0,9945
SE consumo(t-2) é B ENTÃO consumo atrasado é B	0,9975
SE consumo(t-2) é M ENTÃO consumo atrasado é M	1,0
SE consumo(t-2) é A ENTÃO consumo atrasado é A	0,9962
SE consumo(t-2) é MA ENTÃO consumo atrasado é MA	0,9871

Tabela 4.28: Base de Regras - Consumo Atrasado (3 defasagens).

As regras desnecessárias receberam peso insignificante, tendo impacto quase nulo no comportamento do SIF. As únicas regras com impacto relevante

Regra	Peso
SE consumo(t-2) é B ENTÃO consumo atrasado é B	0,9989
SE consumo(t-2) é M ENTÃO consumo atrasado é M	0,9994
SE consumo(t-2) é A ENTÃO consumo atrasado é A	0,9985

Tabela 4.29: Base de Regras - Consumo Atrasado (4 defasagens).

são as que se esperava que o AutoMFIS gerasse. Todavia, no caso das 2 e 4 defasagens, há uma omissão clara dos dois conjuntos fuzzy extremos, MB e MA. As regras, quando geradas devidamente, naturalmente recebem peso próximo a 1,0, o que indica que nos casos de 2 e 4 defasagens elas não estão chegando à etapa de ponderação. Isso pode indicar dois problemas:

- As premissas estão sendo eliminadas durante a geração semi-exaustiva
- Não estão sendo associados os consequentes adequados às premissas

A melhor configuração para o problema com 2 defasagens utilizou o critério de corte de frequência, que é mais penalizante para as regras contemplando os extremos, mais raros, como já havia sido observado no estudo de critérios de corte do caso Predador-Presa. Como é um caso simples, foi investigada a retirada completa de corte de regras, mas as regras esperadas continuaram a não ser geradas. Os conjuntos MB e MA, receberam, neste caso, as regras mostradas na tabela 4.30, que possuem a informação correta do uso da segunda defasagem da série de Consumo na premissa, mas combinada com uma informação desnecessária. Isso indica que o problema de fato está na segunda opção, a associação de consequentes.

Regra	Peso
SE consumo(t-2) é MB & consumo(t-1) é MA ENTÃO consumo atrasado é MB	1,0
SE consumo(t-2) é MA & consumo atrasado(t-2) é M ENTÃO consumo atrasado é MA	1,0

Tabela 4.30: Regras adicionadas pela eliminação do critério de corte (2 defasagens).

Foi observado que o método de associação de consequentes para 2 e 4 defasagens era o mesmo (WM), porém diferia daquele utilizado para 3 defasagens (FCD). Mantendo o limiar de corte zerado, avaliou-se a presença das regras esperadas para cada método de associação, como visto na tabela 4.31. Apenas dois métodos não efetuaram as associações de maneira completamente correta: o WM, como visto anteriormente, e o SD, que omitiu apenas uma das

regras (referente ao conjunto fuzzy M) mas gerou duas outras regras não ideais, listadas na tabela 4.32. Estas duas regras são similares ao que era esperado, mas incluem uma informação desnecessária, que é a própria defasagem da série atrasada. Isto não é o ideal mas ainda assim permite a extração do conhecimento.

Método de associação	# das regras esperadas
FCD	5
SD	4
GJS	5
WM	3
Máxima Credibilidade	5

Tabela 4.31: Presença de regras esperados por método de associação.

Regra	Peso
SE consumo(t-2) é M & consumo atrasado(t-1) é B ENTÃO consumo atrasado é M	0,7119
SE consumo(t-2) é M & consumo atrasado(t-1) é M ENTÃO consumo atrasado é M	0,2881

Tabela 4.32: Regras não ideais para conjunto fuzzy M (SD)

Este parâmetro não havia sido problemático para os outros casos, pois em nenhuma das melhores configurações o critério WM fora utilizado, mas os resultados indicam que o seu uso não é recomendado. Retirado esse método de associação, todos geram as regras esperadas (ou próximo disso no caso do SD), com alto peso. Assim, verifica-se que uma relação de causalidade clara é captada mesmo quando se inserem mais variáveis desnecessárias no sistema.

Nos dois estudos de casos, os impactos dos ajustes do AutoMFIS na extração de conhecimento permitem estabelecer algumas conclusões:

- **Construção do problema:** recomenda-se considerar um único par de séries por vez quando se almeja interpretabilidade, pois mesmo em casos mais simples pode-se chegar à uma base de regras de difícil interpretação. A informação fornecida pelas correlações cruzadas e autocorrelações para definição das defasagens não é absoluta; e pode ser benéfico avaliar modelos mais simplificados, com menos defasagens.
- **Pré-processamento:** é recomendável empregar diferenciação das séries; outros tipos de manipulação diluem o significado das regras, reduzindo ou até eliminando seu valor como extração de conhecimento.

- **Fuzzificação:** de uma maneira geral, o uso 5 conjuntos fuzzy costuma ser satisfatórios em termos de acurácia; a complexidade é baixa, favorecendo a interpretabilidade.
- **Crítérios de corte de regras:** o uso dos critérios de frequência e cardinalidade só é recomendado em casos onde o limiar de corte é mais relaxado, pois estes métodos podem eliminar regras raras e relevantes. No caso do Predador-Presa com 5 defasagens, o critério de ativação foi nitidamente superior. Em casos mais simples, não houve superioridade nítida de um método nas melhores configurações, talvez porque o corte de regras fosse desnecessário.
- **Limiar de corte:** não há diretrizes específicas; deve ser verificado empiricamente para cada caso, de acordo com as limitações de recursos computacionais.
- **Associação de consequentes:** a única ressalva recai sobre o uso da métrica WM, que, no segundo estudo de caso, não proporcionou associações óbvias de consequente.
- **Agregação:** os métodos ponderados são os mais indicados. A ponderação é crucial para compactação da base de regras, além de apresentar melhor desempenho em termos de acurácia.
- **Defuzzificação:** a abordagem acoplada é inviável em termos de interpretabilidade, embora possa proporcionar maior acurácia.

5

Conclusão

Como estipulado na introdução, o objetivo desta dissertação foi a construção de um sistema fuzzy para previsão de séries temporais multivariadas (AutoMFIS), observando os quesitos de acurácia de previsão e interpretabilidade do modelo ajustado.

O modelo AutoMFIS foi testado, no que tange à acurácia, para séries temporais selecionadas da base da competição M3, oferecendo previsões competitivas e, em alguns casos, superiores às dos métodos participantes.

Nos dois estudos de caso destinados fundamentalmente à análise da interpretabilidade, as bases de regras resultantes mostraram-se legíveis e consistentes com o comportamento esperado, confirmando o potencial do modelo de extrair conhecimento das relações de interdependência entre séries temporais. Os dois estudos de caso foram selecionados por apresentarem um comportamento mais previsível, permitindo uma avaliação mais imediata da capacidade do AutoMFIS de extrair conhecimento.

O modelo mostrou-se sensível ao aumento de complexidade, com seu custo computacional crescendo exponencialmente com a quantidade de séries e defasagens consideradas. Esta é uma característica inerente ao método semi-exaustivo de geração de regras, causada por seu caráter combinatorial, mas é também um problema comum a outros métodos multivariados como o VAR. Este aumento de complexidade tem reflexos na interpretabilidade do sistema, que se reduz à medida que se consideram mais séries e defasagens. No primeiro estudo de caso, foi observado o efeito do aumento da complexidade. Inicialmente, considerando uma quantidade considerável de defasagens, obteve-se um sistema de bom desempenho e consistente em significado, porém um pouco vago e de avaliação onerosa. A simplificação proposta em seguida foi tão acurada quanto, porém oferecendo regras muito mais simples.

Tanto nos experimentos de benchmarking como nos de análise de interpretabilidade, ficou claro que não se pode adotar uma abordagem completamente orientada a dados. Agrupar séries não-relacionadas em uma série multivariada e utilizar uma quantidade excessiva de defasagens pode impactar negativamente nos resultados, pois o modelo não necessariamente consegue eliminar as informações irrelevantes. É importante que o analista construa o problema mais simples e direto para a aplicação desejada, evitando eventuais perdas de acurácia e aumento da complexidade do modelo ajustado. A escolha da quantidade de defasagens é crítica e não pode ser baseada completamente

em métodos como a análise da correlação cruzada – como visto nos experimentos – devendo o valor extraído destas análises ser utilizado apenas como um ponto de partida.

Em um trabalho futuro, o processo de busca da configuração ótima para um determinado problema poderia ser aprimorado através da utilização de uma meta-heurística como o algoritmo genético. Isto ofereceria uma varredura mais completa e rigorosa do espaço de busca das configurações do AutoMFIS, aumentando a probabilidade de o melhor modelo encontrado ser de fato o ótimo.

A distinção entre causalidade bidirecional e unidirecional não foi abordada por se constituir em um tema altamente complexo. Mesmo em se tratando de um mesmo problema, é possível que a direção da causalidade se inverta como resultado de alterações, por exemplo, no período de tempo considerado ou na granularidade dos dados. Um exemplo seria o estudo da causalidade entre o PIB de um país e o seu consumo de energia. De acordo com o levantamento realizado por Lee [73], há muita divergência para este problema: os trabalhos de Masih&Masih [74] e Asafu-Adjaye [75] mostram resultados distintos para Indonésia, assim como os de Soyta&Sari [76] e Oh&Lee [77], para a Coreia do Sul. Neste sentido, poder-se-ia avaliar o potencial do AutoMFIS como um teste de causalidade, a exemplo da utilização do modelo VAR como base para o teste de causalidade de Granger [78].

Referências bibliográficas

- [1] KHAN, J.; WEI, J. S.; RINGNER, M.; SAAL, L. H.; LADANYI, M.; WESTERMANN, F.; BERTHOLD, F.; SCHWAB, M.; ANTONESCU, C. R.; PETERSON, C. ; OTHERS. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nature Medicine*, v.7, n.6, p. 673–679, 2001.
- [2] LEE, C.-C. Fuzzy logic in control systems: fuzzy logic controller. II. *Systems, Man and Cybernetics, IEEE Transactions on*, v.20, n.2, p. 419–435, 1990.
- [3] DEAVEN, D.; HO, K. Molecular geometry optimization with a genetic algorithm. *Physical Review Letters*, v.75, n.2, p. 288, 1995.
- [4] SAPANKEVYCH, N. I.; SANKAR, R. Time series prediction using support vector machines: a survey. *Computational Intelligence Magazine, IEEE*, v.4, n.2, p. 24–38, 2009.
- [5] TAYLOR, S. J. **Modelling Financial Time Series**. World Scientific Publishing, 2007.
- [6] SCHARF, L. L. **Statistical Signal Processing**, volume 98. Addison-Wesley Reading, MA, 1991.
- [7] CHEONG, B.; PALMER, R. ; XUE, M. A time series weather radar simulator based on high-resolution atmospheric models. *Journal of Atmospheric and Oceanic Technology*, v.25, n.2, p. 230–243, 2008.
- [8] LUTKEPOHL, H. **New Introduction to Multiple Time Series Analysis**. Springer, 2007.
- [9] GEORGE, B. **Time Series Analysis: Forecasting & Control**.
- [10] ENGLE, R. F. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, p. 987–1007, 1982.
- [11] DE GOOIJER, J. G.; KUMAR, K. Some recent developments in non-linear time series modelling, testing, and forecasting. *International Journal of Forecasting*, v.8, n.2, p. 135–156, 1992.

- [12] CLEMENTS, M. P.; FRANSES, P. H. ; SWANSON, N. R. Forecasting economic and financial time-series with non-linear models. **International Journal of Forecasting**, v.20, n.2, p. 169–183, 2004.
- [13] GURESEN, E.; KAYAKUTLU, G. ; DAIM, T. U. Using artificial neural network models in stock market index prediction. **Expert Systems with Applications**, v.38, n.8, p. 10389–10397, 2011.
- [14] DHAMIJA, A.; BHALLA, V. Financial time series forecasting: comparison of neural networks and ARCH models. **International Research Journal of Finance and Economics**, v.49, p. 185–202, 2010.
- [15] WERBOS, P. Beyond regression: New tools for prediction and analysis in the behavioral sciences. 1974.
- [16] WERBOS, P. J. Generalization of backpropagation with application to a recurrent gas market model. **Neural Networks**, v.1, n.4, p. 339–356, 1988.
- [17] LAPEDES, A.; FARBER, R. **Nonlinear signal processing using neural networks: Prediction and system modelling**. Technical report, 1987.
- [18] WANG, J.-L.; CHAN, S.-H. Stock market trading rule discovery using two-layer bias decision tree. **Expert Systems with Applications**, v.30, n.4, p. 605–611, 2006.
- [19] CAO, L. Support vector machines experts for time series forecasting. **Neurocomputing**, v.51, p. 321–339, 2003.
- [20] JAEGER, H. The “echo state” approach to analysing and training recurrent neural networks-with an erratum note. **Bonn, Germany: German National Research Center for Information Technology GMD Technical Report**, v.148, p. 34, 2001.
- [21] ZHOU, Z.-H.; WU, J. ; TANG, W. Ensembling neural networks: many could be better than all. **Artificial Intelligence**, v.137, n.1, p. 239–263, 2002.
- [22] ZADEH, L. A. Fuzzy sets. **Information and Control**, v.8, n.3, p. 338–353, 1965.
- [23] TAKAGI, T.; SUGENO, M. Fuzzy identification of systems and its applications to modeling and control. **Systems, Man and Cybernetics, IEEE Transactions on**, , n.1, p. 116–132, 1985.

- [24] SUGENO, M.; KANG, G. Structure identification of fuzzy model. **Fuzzy Sets and Systems**, v.28, n.1, p. 15–33, 1988.
- [25] MAMDANI, E. H.; ASSILIAN, S. An experiment in linguistic synthesis with a fuzzy logic controller. **International Journal of Man-Machine Studies**, v.7, n.1, p. 1–13, 1975.
- [26] GRANGER, C. W. J.; NEWBOLD, P. **Forecasting Economic Time Series**. Academic Press, 2014.
- [27] SWANSON, N. R.; WHITE, H. A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks. **Review of Economics and Statistics**, v.79, n.4, p. 540–550, 1997.
- [28] SWANSON, N. R.; WHITE, H. A model-selection approach to assessing the information in the term structure using linear models and artificial neural networks. **Journal of Business & Economic Statistics**, v.13, n.3, p. 265–275, 1995.
- [29] CYBENKO, G. Approximation by superpositions of a sigmoidal function. **Mathematics of Control, Signals and Systems**, v.2, n.4, p. 303–314, 1989.
- [30] WANG, L.-X.; MENDEL, J. M. Fuzzy basis functions, universal approximation, and orthogonal least-squares learning. **Neural Networks, IEEE Transactions on**, v.3, n.5, p. 807–814, 1992.
- [31] MARKOWITZ, H. Portfolio selection. **The Journal of Finance**, v.7, n.1, p. 77–91, 1952.
- [32] ELTON, E. J.; GRUBER, M. J.; BROWN, S. J. ; GOETZMANN, W. N. **Modern Portfolio Theory and Investment Analysis**. John Wiley & Sons, 2009.
- [33] WIGLEY, T. M.; BRIFFA, K. R. ; JONES, P. D. On the average value of correlated time series, with applications in dendroclimatology and hydrometeorology. **Journal of Climate and Applied Meteorology**, v.23, n.2, p. 201–213, 1984.
- [34] FRISTON, K. J.; JEZZARD, P. ; TURNER, R. Analysis of functional MRI time-series. **Human Brain Mapping**, v.1, n.2, p. 153–171, 1994.

- [35] MANTEGNA, R. N.; STANLEY, H. E. **Introduction to Econophysics: Correlations and Complexity in Finance**. Cambridge University Press, 1999.
- [36] WEI, W. W.-S. **Time Series Analysis**. Addison-Wesley, 1994.
- [37] EVERITT, B. S.; LANDAU, S.; LEESE, M. ; STAHL, D. **Cluster Analysis**. 5th. ed., John Wiley & Sons, Ltd, 2011.
- [38] KELLEY, J. L. **General Topology**. Springer Science & Business Media, 1975.
- [39] GU, Y.; WANG, C. **A study of hierarchical correlation clustering for scientific volume data**. In: Boyle, R.; Parvin, B.; Koracin, D.; Chung, R.; Hammoud; Hussain, M.; Tan, K.-H.; Crawfis, R.; Thalmann, D.; Kao, D. ; Avila, L., editors, **ADVANCES IN VISUAL COMPUTING**, p. 437–446. Springer, 2010.
- [40] GRANGER, C. W. Prediction with a generalized cost of error function. **OR**, p. 199–207, 1969.
- [41] KOSKO, B. Fuzzy systems as universal approximators. **Computers, IEEE Transactions on**, v.43, n.11, p. 1329–1333, 1994.
- [42] GACTO, M. J.; ALCALÁ, R. ; HERRERA, F. Interpretability of linguistic fuzzy rule-based systems: An overview of interpretability measures. **Information Sciences**, v.181, n.20, p. 4340–4360, 2011.
- [43] RUSSO, M. FuGeNeSys - A fuzzy genetic neural system for fuzzy modeling. **Fuzzy Systems, IEEE Transactions on**, v.6, n.3, p. 373–388, 1998.
- [44] CORDÓN, O.; HERRERA, F.; MAGDALENA, L. ; VILLAR, P. A genetic learning process for the scaling factors, granularity and contexts of the fuzzy rule-based system data base. **Information Sciences**, v.136, n.1, p. 85–107, 2001.
- [45] TIKK, D.; GEDEON, T. D. ; WONG, K. W. **A feature ranking algorithm for fuzzy modelling problems**. In: Casillas, J.; Cordón, O.; Herrera Triguero, F. ; Magdalena, L., editors, **INTERPRETABILITY ISSUES IN FUZZY MODELING**, p. 176–192. Springer, 2003.
- [46] HONG, T.-P.; CHEN, J.-B. Finding relevant attributes and membership functions. **Fuzzy Sets and Systems**, v.103, n.3, p. 389–404, 1999.

- [47] VANHOUCKE, V.; SILIPO, R. **Interpretability in multidimensional classification**. In: INTERPRETABILITY ISSUES IN FUZZY MODELING, p. 193–217. Springer, 2003.
- [48] PAL, N. R. Soft computing for feature analysis. **Fuzzy Sets and Systems**, v.103, n.2, p. 201–221, 1999.
- [49] MILLER, G. A. The magical number seven, plus or minus two: some limits on our capacity for processing information. **Psychological Review**, v.63, n.2, p. 81, 1956.
- [50] BORTOLET, P. **Modelisation et commande multivariables floues: application a la commande d'un moteur thermique**. 1998. Tese de Doutorado - Toulouse, INSA.
- [51] PULKKINEN, P.; KOIVISTO, H. A dynamically constrained multiobjective genetic fuzzy system for regression problems. **Fuzzy Systems, IEEE Transactions on**, v.18, n.1, p. 161–177, 2010.
- [52] ISHIBUCHI, H.; NOZAKI, K.; TANAKA, H.; HOSAKA, Y. ; MATSUDA, M. Empirical study on learning in fuzzy systems by rice taste analysis. **Fuzzy Sets and Systems**, v.64, n.2, p. 129–144, 1994.
- [53] WANG, L.-X.; MENDEL, J. M. Generating fuzzy rules by learning from examples. **Systems, Man and Cybernetics, IEEE Transactions on**, v.22, n.6, p. 1414–1427, 1992.
- [54] QUINLAN, J. R. Induction of decision trees. **Machine Learning**, v.1, n.1, p. 81–106, 1986.
- [55] ICHIHASHI, H.; SHIRAI, T.; NAGASAKA, K. ; MIYOSHI, T. Neuro-fuzzy id3: a method of inducing fuzzy decision trees with linear programming for maximizing entropy and an algebraic method for incremental learning. **Fuzzy Sets and Systems**, v.81, n.1, p. 157–167, 1996.
- [56] CHEN, S.; COWAN, C. F. ; GRANT, P. M. Orthogonal least squares learning algorithm for radial basis function networks. **Neural Networks, IEEE Transactions on**, v.2, n.2, p. 302–309, 1991.
- [57] CHEN, S.; BILLINGS, S. A. ; LUO, W. Orthogonal least squares methods and their application to non-linear system identification. **International Journal of Control**, v.50, n.5, p. 1873–1896, 1989.

- [58] KOSHIYAMA, A. **GPFIS: Um sistema fuzzy-genético genérico baseado em programação genética**. 2013. Dissertação de Mestrado - PUC-Rio.
- [59] HAN, J.; KAMBER, M. ; PEI, J. **Data Mining: Concepts and Techniques**. Elsevier, 2011.
- [60] DHILLON, I. S.; MODHA, D. S. Concept decompositions for large sparse text data using clustering. **Machine Learning**, v.42, n.1-2, p. 143–175, 2001.
- [61] BOYD, S.; VANDENBERGHE, L. **Convex Optimization**. Cambridge University Press, 2004.
- [62] MURTY, K. G.; YU, F.-T. **Linear Complementarity, Linear and Nonlinear Programming**. Citeseer, 1988.
- [63] MATLAB. **version 7.10.0 (R2010a)**. Natick, Massachusetts: The MathWorks Inc., 2010.
- [64] MAKRIDAKIS, S.; HIBON, M. ; MOSER, C. Accuracy of forecasting: An empirical investigation. **Journal of the Royal Statistical Society. Series A (General)**, p. 97–145, 1979.
- [65] MAKRIDAKIS, S.; ANDERSEN, A.; CARBONE, R.; FILDES, R.; HIBON, M.; LEWANDOWSKI, R.; NEWTON, J.; PARZEN, E. ; WINKLER, R. The accuracy of extrapolation (time series) methods: Results of a forecasting competition. **Journal of Forecasting**, v.1, n.2, p. 111–153, 1982.
- [66] MAKRIDAKIS, S.; HIBON, M. The M3-Competition: results, conclusions and implications. **International Journal of Forecasting**, v.16, n.4, p. 451–476, 2000.
- [67] MURRAY, J. D. **Mathematical Biology I: An Introduction**, volume 17 de **Interdisciplinary Applied Mathematics**. Springer, New York, NY, USA,, 2002.
- [68] ODUM, E. P.; ODUM, H. T. ; ANDREWS, J. **Fundamentals of Ecology**, volume 3. Saunders Philadelphia, 1971.
- [69] AKCAKAYA, H. R. Population cycles of mammals: evidence for a ratio-dependent predation hypothesis. **Ecological Monographs**, p. 119–142, 1992.

- [70] BLASIUS, B.; HUPPERT, A. ; STONE, L. Complex dynamics and phase synchronization in spatially extended ecological systems. *Nature*, v.399, n.6734, p. 354–359, 1999.
- [71] STENSETH, N. C.; FALCK, W.; BJØRNSTAD, O. N. ; KREBS, C. J. Population regulation in snowshoe hare and canadian lynx: asymmetric food web configurations between hare and lynx. *Proceedings of the National Academy of Sciences*, v.94, n.10, p. 5147–5152, 1997.
- [72] BANCO CENTRAL DO BRASIL. **Consumo de energia elétrica - sudeste (industrial)**.
- [73] LEE, C.-C. Energy consumption and GDP in developing countries: a cointegrated panel analysis. *Energy Economics*, v.27, n.3, p. 415–427, 2005.
- [74] MASIH, A.; MASIH, R. Electricity consumption, real income and temporal causality: results from a multi-country study based on cointegration and error correction modeling techniques. *Energy Economics*, v.18, p. 165—183, 1996.
- [75] ASAFU-ADJAYE, J. The relationship between energy consumption, energy prices and economic growth: time series evidence from Asian developing countries. *Energy Economics*, v.22, n.6, p. 615–625, 2000.
- [76] SOYTAS, U.; SARI, R. Energy consumption and GDP: causality relationship in G-7 countries and emerging markets. *Energy Economics*, v.25, n.1, p. 33–37, 2003.
- [77] OH, W.; LEE, K. Causal relationship between energy consumption and GDP revisited: the case of Korea 1970–1999. *Energy Economics*, v.26, n.1, p. 51–59, 2004.
- [78] GRANGER, C. W. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, p. 424–438, 1969.
- [79] MIŚKIEWICZ, J. Analysis of time series correlation. the choice of distance metrics and network structure. *Acta Physica Polonica A*, v.121, n.2B, 2012.
- [80] GUILLAUME, S. Designing fuzzy inference systems from data: an interpretability-oriented review. *Fuzzy Systems, IEEE Transactions on*, v.9, n.3, p. 426–443, 2001.